Estadística para las ciencias sociales

SEGUNDA EDICIÓN

Ferris J. Ritchey
Department of Sociology
University of Alabama at Birmingham

Revisión técnica

Cecilia Balbás
Universidad Anáhuac Norte

McGraw Hill

MÉXICO • BOGOTÁ • BUENOS AIRES • CARACAS • GUATEMALA • LISBOA
MADRID • NUEVA YORK • SAN JUAN • SANTIAGO • AUCKLAND
LONDRES • MILÁN • MONTREAL • NUEVA DELHI • SAN FRANCISCO • SÃO PAULO
SINGAPUR • SAN LUIS • SIDNEY • TORONTO

Director Higher Education: Miguel Ángel Toledo Castellanos
Director editorial: Ricardo A. del Bosque Alayón
Editor sponsor: Noé Islas López
Supervisor de producción: Zeferino García García
Traducción: Jorge Humberto Romo Muñoz
Jorge Yescas Milanes
Javier León Cárdenas
Jorge Alberto Velázquez Arellano

ESTADÍSTICA PARA LAS CIENCIAS SOCIALES
Segunda edición

McGraw-Hill Interamericana

A Subsidiary of ***The McGraw-Hill*** *Companies, Inc.*
Corporativo Punta Santa Fe
Prolongación Paseo de la Reforma 1015, Torre A
Piso 17, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN-10: 970-10-6699-5
ISBN-13: 978-970-10-6699-7
(ISBN: 970-10-3141-5 edición anterior)

Translated from the 2nd edition of THE STATISTICAL IMAGINATION: ELEMENTARY STATISTICS FOR THE SOCIAL SCIENCES

0-07-294304-1
CMC 05/15

1402356789 2346789015

Impreso en México *Printed in Mexico*

Impreso por CM CUADRADO, S.A DE C.V. Printed by CM CUADRADO, S.A DE C.V.

The ***McGraw-Hill*** *Companies*

A Wanda, Daniel, Holly, Alan, Sarah, Kitty, Dorrance y Agnes por su amor y su inspiración.

A Daniel O. Price y P. Neal Ritchey por su generoso apoyo.

Al amoroso recuerdo de Phillip Ritchey.

CONTENIDO BREVE

CONTENIDO

PREFACIO

Todos utilizamos el pensamiento estadístico, el cálculo de probabilidades, en nuestra vida diaria. La simple decisión sobre si llevar paraguas implica la estimación de la probabilidad de lluvia. Las probabilidades tienen relevancia cuando se toman decisiones importantes en la vida, por ejemplo casarse, aceptar un empleo, invertir en acciones o cambiar de carriles en el tráfico. Incluso una cantidad moderada de especialización estadística en el trabajo ofrece una ventaja competitiva a un empleado. Para los estudiantes en áreas científicas, el pensamiento estadístico resulta un ingrediente esencial para la comprensión clara del mundo natural, el orden social y el comportamiento humano. En un plano menos formal, el pensamiento estadístico constituye la base de los juegos de azar; así como jugar y apostar son divertidos, la estadística también lo es.

Por desgracia, los estudiantes no siempre aprecian qué tan divertido puede ser un curso sobre estadística. Los alumnos que estudian especialidades en ciencias sociales por lo común tienen antecedentes limitados en matemáticas y se quejan por sentirse obligados a tomar estos cursos. Algunos textos de estadística desatienden este hecho y presentan fórmulas complejas, lo cual causa ansiedad matemática innecesaria. Otros textos son simplificados en exceso para reducir la ansiedad matemática, pero a menudo sacrifican principios estadísticos básicos. El presente texto intenta enseñar los conceptos difíciles de la estadística sin sacrificar conceptos matemáticos y sus cálculos. No obstante, está diseñado para convencer a los estudiantes de que las matemáticas son sólo una herramienta, no la esencia, para aprender estadística.

Aprendí que la estadística puede enseñarse integralmente sin enfatizar demasiado en las matemáticas cuando tuve la gran fortuna de trabajar como asistente con Daniel O. Price, a quien dedico este texto. Su entusiasmo por la materia, junto con sus claras explicaciones de procesos lógicos, logró enamorarme de la materia. Al igual que Dan, me he esforzado por más de 25 años en enseñar estadística para desarrollar técnicas para compartir tal entusiasmo con los estudiantes. En particular, he puesto en la mira varias barreras conceptuales con la idea de ayudar a los estudiantes a librarlas. El diseño del presente texto sigue cuatro principios básicos:

- La estadística no es parte de las matemáticas. En cambio, es una forma sabia de pensar acerca de las cosas.
- Desde el comienzo deben diseñarse tareas para fortalecer la confianza de los estudiantes.
- El dominio de los elementos básicos del razonamiento estadístico facilita el dominio de elementos más complejos; por consiguiente, el proceso de aprendizaje es acumulativo.
- La estadística se aprende con práctica. Habrá muchas labores, pero la materia es inherentemente interesante y agradable. Diviértase y tenga confianza en que el trabajo le recompensará.

Permítame describir estos principios con un poco más de detalle. El primero consiste en que la estadística no es acerca de las matemáticas *per se*, sino sobre el pensamiento proporcional: la visualización de una parte en un todo. A esta idea acerca de la realidad la llamo *imaginación estadística,* cuyo concepto se relaciona con la idea de C. Wright Mills de la imaginación sociológica, que define la relación del individuo con la sociedad en su conjunto. De igual

forma, la imaginación estadística requiere ver datos en varios contextos mayores. Primero, las observaciones del comportamiento individual son vistas dentro del contexto de la estructura social mayor. Segundo, las conclusiones sobre una población grande de sujetos con base en una muestra de esos sujetos son vistas como sólo uno de muchos conjuntos de conclusiones, debido a que una segunda muestra producirá resultados ligeramente diferentes. Tercero, las interpretaciones de datos estadísticos deben tomar en cuenta las circunstancias prácticas y las realidades culturales que proporcionan el significado esencial de los números.

El segundo principio de este texto radica en que el diseño del curso debe permitir a los estudiantes tener éxito inicial para así crear confianza y aliviar el miedo al fracaso. En las primeras páginas del texto se presentan cálculos estadísticos simples, pero esenciales, de fracciones, proporciones y porcentajes. Dichos cálculos se presentan como formas de cuantificar el pensamiento proporcional, reforzando la idea de que las matemáticas son sólo una herramienta, no la esencia, para aprender estadística. Es más, a las abstracciones estadísticas que aparecen como barreras para numerosos estudiantes (por ejemplo, la desviación estándar, puntuaciones estandarizadas, error de muestreo y distribuciones muestrales) se les presta mucha atención. El tema del control del error se enfatiza para mostrar la importancia de la diligencia en el trabajo estadístico y animar a los estudiantes a desarrollar un sentido de competencia.

El tercer principio consiste en que para que los estudiantes comprendan con éxito la lógica de la estadística inferencial y la comprobación de hipótesis, deben dominar bien los elementos básicos del procedimiento de comprobación. Parte de este objetivo se logra a través del diseño del texto. Se da mucha cobertura para trabajar con áreas bajo una curva gráfica en forma de campana, llamada curva de distribución normal. Se asigna un tiempo considerable para generar distribuciones muestrales: descripciones de los resultados que ocurren cuando, por decir, se lanzan al aire 10 monedas repetidamente, o también de modo repentino se extraen frijoles de una caja. A través del muestreo repetido, los estudiantes aprenden que los estadísticos de cualquier muestra única constituyen sólo un conjunto de muchas posibles estimaciones para el grupo mayor del cual se tomó la muestra. Este método desmitificará este concepto que está en el centro del razonamiento estadístico. Los estudiantes aprenden que las distribuciones muestrales son conjeturas reales, no abstractas. El texto también presenta la lógica de la comprobación de hipótesis en seis pasos, uno de los cuales requiere dibujar curvas de distribución muestral. La atención en el detalle facilita e impulsa el pensamiento proporcional. La otra parte del objetivo de dominar las ideas básicas es que los estudiantes deben aplicarse y mantenerse al ritmo del material y las tareas del curso.

El cuarto principio se relaciona con el anterior: la estadística se aprende trabajando. Cada capítulo incluye preguntas y ejercicios que fomentan el pensamiento proporcional que es la base del análisis estadístico. Para grupos donde se usan computadoras, ofrece el programa *SPSS for Windows.* El sitio web presenta ejercicios para los capítulos, ilustraciones detalladas sobre la interpretación de los resultados y una variedad de conjuntos de datos elegidos para estimular el interés, así como ejemplos para exponer a los estudiantes a la investigación del mundo real.

AL PROFESOR

Este curso está diseñado para cubrir los elementos básicos de la comprobación de hipótesis, de tal manera que cuando se introduce la estadística inferencial (capítulos 9 en adelante), los conceptos abstractos se adquieren con facilidad. El Manual del Profesor, el Banco de Pruebas y el Manual de Soluciones (disponibles en inglés en el sitio web), ofrecen detalles de los indicadores citados líneas adelante, junto con sugerencias de exposiciones y ejemplos de problemas para presentar en clase, tareas de diversas formas para asignar a los estudiantes, pruebas, exámenes breves y consejos de cómo calificar. Los ejercicios de los capítulos se ofrecen en cuatro conjuntos.

Si bien la mayoría de profesores ha desarrollado sus propias técnicas efectivas, he encontrado que el siguiente régimen pedagógico aumenta al máximo el éxito de los estudiantes. Tal régimen se ha probado en clase más de 25 veces, y se basa en la idea de que las tareas y los cuestionarios constituyen ensayos para los exámenes mayores. En mi experiencia, los exámenes mayores deben aplicarse "a libro cerrado" (salvo fórmulas y tablas estadísticas). Los exámenes a libro abierto crean hábitos de estudio negativos, pero intento aliviar las presiones de un examen a libro cerrado al ofrecer a los estudiantes muy buena oportunidad para aprender de errores cometidos en tareas.

- Solicite tareas semanales cuya fecha de entrega sea para la sesión de clase inmediata posterior a completar la exposición del material de un capítulo.
- Devuelva en la siguiente clase las tareas calificadas y deje en el laboratorio, o en algún otro lugar, las guías y claves de respuestas de las tareas (del banco de pruebas y el manual de soluciones). Puesto que se pueden ver las claves de respuesta, calificar las tareas no requiere de muchas "marcas con rojo". (He encontrado que la disponibilidad de las claves de las tareas no representa un riesgo para el trabajo del siguiente curso. Como se mencionó antes, los ejercicios están diseñados con un formato de pares e impares para alternarlos entre los cursos.)
- En la siguiente clase o en el laboratorio haga preguntas a los estudiantes sobre el material del capítulo. Recoja los exámenes cinco minutos después que el primer alumno haya terminado. Distribuya copias limpias del examen y presente o permita a los estudiantes que presenten las respuestas de inmediato.
- Realice dos o tres exámenes parciales, así como uno final (todos a libro cerrado, salvo fórmulas y tablas estadísticas).

En mi experiencia, varios temas del curso deben prestarse con suficiente atención porque, de no ser así, se perderá mucho tiempo al intentar rellenar lagunas en los contenidos.

- Para eliminar la ansiedad matemática, haga que los estudiantes disfruten el éxito inicial con tareas sobre proporciones, distribuciones de frecuencia y gráficas. Es más, una revisión completa de proporciones y porcentajes facilita la instrucción de la teoría de probabilidad, distribución muestral, valores *p*, errores tipo I y II, etcétera.
- Para fomentar el pensamiento lineal y capacidad de pensamiento proporcional, tome el tiempo suficiente para explicar la desviación estándar y los resultados estandarizados, y haga que los estudiantes resuelvan muchos problemas acerca de la partición de áreas bajo la curva normal.
- Genere al menos dos distribuciones muestrales en clase. Después de ello, cuando se analice el concepto, los estudiantes entenderán perfectamente qué es una distribución muestral.
- Exija a los estudiantes que desarrollen las etapas de los seis pasos de la inferencia estadística, especialmente dibujar la curva de la distribución muestral en el paso 2, en cada prueba de hipótesis de las tareas y exámenes. La repetición de este procedimiento mejora la comprensión en todos los estudiantes. Algunos dominarán rápido los detalles (capítulo 9). Para el capítulo 11, todos los estudiantes que de verdad trabajen duro habrán captado la lógica. Después de esto, se podrá avanzar con un ritmo más sereno porque los aspectos pedagógicos de los seis pasos resultarán familiares para los estudiantes. Así, en capítulos posteriores acerca del análisis divariado, el profesor podrá concentrarse en problemas conceptuales relacionados con la comprobación de hipótesis y en las ideas de investigación.

CARACTERÍSTICAS ESPECIALES DEL LIBRO

- **Legibilidad.** El material se ha probado muchas veces en clase.
- **Temas conceptuales para despertar el interés.** El texto está diseñado alrededor de varios temas conceptuales que hacen de la estadística una materia agradable. Primero, nos habla del pensamiento proporcional, y los cálculos matemáticos sólo son herramientas para ayudar en este proceso. Segundo, cuando se usa imaginación estadística, las estimaciones se interpretan respecto a contextos mayores no sólo de una población de sujetos, sino también de una "población" de ideas, valores, fuerzas normativas, circunstancias prácticas y teorías. Se realizan distinciones entre la significación estadística y la significación práctica/teórica. Tercero, el libro destaca la importancia de la precisión, la diligencia y el profesionalismo en la conducción de una investigación.
- **Dirigir los resultados al público apropiado.** Se incluyen análisis sobre cómo presentar los resultados a audiencias científicas y al público en general, junto con ejemplos variados de la presentación de tablas.
- **Superar barreras conceptuales.** Se identifican barreras conceptuales y se emplean muchas estrategias aprendidas por el autor a través de largos años de enseñanza, para que los estudiantes las superen. Tales estrategias incluyen una delineación completa de la desviación estándar, una amplia cobertura de puntuaciones estandarizadas y distribuciones muestrales, así como una clara explicación de los grados de libertad.
- **Un capítulo separado acerca de las distribuciones muestrales.** Las distribuciones muestrales se presentan e ilustran para dar el ingrediente esencial del pensamiento proporcional.
- **Seis pasos de inferencia estadística.** Los procedimientos lógicos de la prueba de hipótesis se presentan en forma consistente como "los seis pasos de la inferencia estadística". Cada prueba estadística se ilustra dentro de tales lineamientos. Las ilustraciones se ponen de relieve.
- **Los cuatro aspectos de una relación.** Las interpretaciones de las pruebas estadísticas divariadas contienen los cuatro aspectos de una relación: existencia, dirección, fuerza y aplicaciones prácticas.
- **Ejemplos completos de cada procedimiento estadístico.** Al ceñirse a los seis pasos de la inferencia estadística y los cuatro aspectos de una relación, los ejemplos completos mantienen informados a los estudiantes sobre lo que se espera en las tareas y en los exámenes. Las distinciones entre "especificaciones" y "cálculos" facilitan la solución del problema.
- **Pautas para elegir la prueba estadística apropiada.** Cada prueba de hipótesis viene precedida por un cuadro que describe cuándo usar una prueba (ejemplo, número de muestras, nivel de medición de las variables, tamaño de la muestra). Un diagrama de árbol de decisiones acumulativas al final de cada capítulo que implica la comprobación de hipótesis refuerza el proceso de selección de la prueba.

- **Se resaltan los términos y fórmulas importantes.** Los conceptos y fórmulas se encuadran para permitir una revisión fácil, y cada capítulo incluye un resumen de fórmulas. El índice es detallado y se presentan símbolos y fórmulas en la tercera de forros.
- **Diagramas conceptuales.** Para enseñar a los estudiantes a pensar proporcionalmente, todas las pruebas de hipótesis se presentan con diagramas conceptuales que distinguen entre poblaciones y parámetros a partir de muestras y estadísticos.
- **Variados ejercicios de capítulo.** Los ejercicios presentan una buena mezcla de problemas cotidianos prácticos y de problemas científicos sobre diversas ciencias sociales y sobre publicaciones de la salud. Los ejercicios están ordenados de simples a complejos. En el apéndice C se dan las respuestas a algunos ejercicios seleccionados.
- **Aplicaciones opcionales en computadora.** Ya sea que en clase se usen o no computadoras, en todo el texto se describe su utilidad. El disco compacto contiene el *SPSS for Windows* y conjuntos variados de datos, como el estudio social general, un conjunto de datos ecológicos extraído de datos del Censo de Estados Unidos, así como estudios sobre personas sin hogar y temores médicos de litigios por negligencia.
- **Insensatez y falacias estadísticas.** Consistente con el tema del control del error, cada capítulo presenta equívocos comunes (y a menudo hasta graciosos) de interpretaciones erróneas de la estadística en la vida cotidiana, y por los medios de comunicación masiva e investigados.
- **El sitio web.** Además de los conjuntos de datos y extensos ejercicios de aplicación en computadora, el sitio web del texto en www.mhle.com/ritchey3 da recursos en inglés tanto al estudiante como al profesor. Para estudiantes, hay compendios de capítulos, cuestionarios de autoevaluación, tarjetas de terminología rápida, transparencias PowerPoint, glosarios y enlaces a fuentes de datos y sitios estadísticos. En el seguro rincón de profesores, hay archivos PDF que se pueden imprimir y descargar para un Banco de Pruebas, un Manual del Instructor, un Manual de Soluciones con claves completas para los cuatro conjuntos de ejercicios del capítulo, transparencias PowerPoint y enlaces de recursos de enseñanza.

MEJORAS A LA SEGUNDA EDICIÓN

- La adición del apéndice D, *Guía para el SPSS for Windows*, está organizada por capítulo. Este apéndice presenta diagramas de flujo de secuencias de comando de punto y *clic* para cada uno de los procedimientos, así como una breve descripción de la salida de computadora.
- La inclusión del apéndice E, Potencia estadística, donde se realiza el análisis de este tema tan importante, desarrollado *ex profeso* para la segunda edición en español.
- Una sección titulada "Extensiones del capítulo sobre el sitio web *The Statistical Imagination*" aparece al final de cada capítulo de este libro, dirigiendo a los lectores al sitio web. Estas extensiones están en archivos PDF que con toda facilidad se pueden descargar e imprimir. Las extensiones del capítulo comprenden: (1) temas de la primera edición que se tratan con poca frecuencia, por ejemplo el gamma (capítulo 16 de la primera edición); (2) recién agregados, conceptos y procedimientos ligeramente avanzados, por ejemplo cálculos de potencia estadística; (3) introducciones a correlación parcial y múltiple y ANOVA *N*-Way con términos de interacción; y (4) problemas metodológicos, como es la validez y confiabilidad de escalas de medición. Estas adiciones facilitan el uso de este libro de texto en un curso de graduados de primer nivel. En términos generales, cualquier material que se haya retirado de la primera edición, por ejemplo las variables ordinales semejantes a intervalos, está retenido en el sitio web.

- Aparecen resúmenes de capítulos cerca del final de cada capítulo.
- Se duplica el número de ejercicios para resolver manualmente. Cuatro conjuntos de ejercicios paralelos se adaptan a profesores que impartan numerosas secciones por año.
- Se actualizaron ilustraciones de procedimientos e innumerables tablas de datos.
- Para cada procedimiento estadístico, en rectángulos aparece un ejemplo completo sobre "Cómo" y "Solución" para dar guías a estudiantes para trabajar ejercicios del capítulo.
- Más adelante en el capítulo 10 se introduce la distribución t aproximadamente normal, en coordinación con la prueba pequeña de medias de muestra individual (prueba t).
- La prueba grande de medias de muestra individual para $n > 121$ que usa la estadística de prueba Z, y la tabla de curva normal, se introduce en el capítulo 9 para facilitar la instrucción sobre la lógica de la prueba de hipótesis.
- En los seis pasos de inferencia estadística, varios cambios de formateo simplifican y abrevian la presentación: (1) *Hipótesis estadística* (Stat. *H*) se denomina ahora consistentemente *hipótesis nula* (H_0), y una sección del capítulo 9 explica de manera completa lo que significa *nula*. (2) En la Preparación de prueba y Seis pasos, se eliminan redundancias.
- El sitio web incluye un Manual del Instructor que contiene un método de "llavero" para una instrucción eficiente. Las técnicas de instrucción están basadas en el taller "Successfully Teaching Statistics Without Catering Down", que se ha presentado en numerosas reuniones profesionales por el autor y por Thomas A. Petee de la Universidad Auburn.

RECONOCIMIENTOS

Muchos miembros de mi familia me fueron de especial ayuda para escribir este libro. Gracias a Wanda por su amor, ayuda y paciencia. Sarah y Kitty fueron especialmente serviciales revisando los primeros proyectos, y Daniel colaboró en formulaciones matemáticas y en las aplicaciones de computadora. Gail proporcionó consejos en los gráficos. Lynn Harper Ritchey, una socióloga colega, me dio estímulo y asistencia.

Estoy especialmente agradecido por la ayuda de dos personas. La primera es Daniel O. Price, quien fue mi mentor cuando fui estudiante en la Universidad de Texas en Austin. Dan fue coautor de un texto de estadística con Margare Hagood en los años cincuenta y le dio clases a Hubert M. Blalock, cuyo texto *Social Statistics* (McGraw-Hill) fue un apoyo fundamental para muchos estudiantes en los años setenta y ochenta. Muchas de las ideas y estrategias pedagógicas de este texto, el énfasis en las distribuciones muestrales, los seis pasos de la inferencia estadística, los cuatro aspectos de una relación, los aprendí de Dan. De hecho, inicialmente me sugirió que escribiera un texto como su coautor. El tiempo pasó y se retiró antes que el proyecto avanzara, pero ha sido de mucha ayuda desde entonces. Segundo, un agradecimiento especial a P. Neal Ritchey, mi hermano y colega sociólogo de la Universidad de Cincinnati. Cuando encontré desafíos conceptuales, él siempre estaba ahí con su consejo y respuestas correctas. Mientras este texto pasaba por borradores, él aportó espacios dentro de su ocupado horario para leer, criticar y editar para mí; también me ayudó a reunir conjuntos de datos para las aplicaciones en computadora. De verdad aprecio las ventajas de tener un hermano mayor en el mismo campo. Mi amor y agradecimiento para Neal, quien me ha guiado en muchas obras durante años.

Extiendo un cordial agradecimiento a quienes fueron generosos al brindarme su tiempo y ayuda: Jeffrey E. Hall, Brian P. Hinote, Jason Wasserman, Catherine Moran y Mercy Mwaria, cuyos efectos son visibles en el sitio web; Jackie Skeen, Charlotte Edwards y LaShundra Wormsley-Dooley por su ayuda en el ensamble. Deseo dar gracias a Sean-Shong Hwang por ayudarme a hallar formas de presentar los temas. También doy gracias a las siguientes personas por sus constructivas sugerencias: Julie Locher, Akilah Dulin, Cullen Clark, Thomas Petee, Darlene Wright, Jennifer Moren-Cross y Lynn Gerald. Y gracias a William C. Cockerham, Mark LaGory, Patricia Drentea, Jeffrey Clair, Mike Wilson, Becky Trigg, Shelia Cotten, Ken Wilson, Kevin Fitzpatrick, Abdullah Khatri, Harry Hamilton, Gregory Sheinfeld y Tennat S. McWilliams por su apoyo, estímulo y consejo.

Deseo dar gracias a los siguientes y varios revisores anónimos por sus sugerencias completas y constructivas:

Neil W. Henry, Virginia Commonwealth University,
Therese Seibert, Keene State College,
Surendar Yadava, University of Northern Iowa,
Jay Alperson, Palomar College,
Christopher Bradley, Indiana-Purdue University Fort Wayne,
Furjen Deng, Sam Houston State University,
Lisa Pellerin, Ball State University,
Robin Perrin, Pepperdine University, y
William Wells, Southern Illinois University.

Por último, agradezco mucho la guía y cooperación de las personas maravillosas de McGraw-Hill. Mi editora, Sherith Pankratz y editoras anteriores Sally Constable, Carolyn Meier y Hill Gordon, y también a Julie Abodeely, Kathy Shackleford y Gina Boedecker. Gracias a los editores de desarrollo Robin Reed y Beth Baugh, y a la editora de producción Valerie Heffernan, de Carlisle Publishing Services; Jill Rietema de SPSS; y otras personas ayudaron a que este proyecto se realizara.

CAPÍTULO 1

La imaginación estadística

RESUMEN DEL CAPÍTULO

Introducción

Un día, cuando un pollo estaba rascando entre las hojas, una bellota cayó del árbol y lo golpeó en la cola. "¡Oh", dijo el pollo, "¡el cielo se está cayendo! Voy a avisarle al rey".

El pollo hizo algo que todos hacemos de cuando en cuando: poner las cosas fuera de proporción. Aun cuando ésta es una reacción normal para un libro de cuentos de animales y muchos seres humanos, los peritos en estadística no deben reaccionar demasiado rápido ni emocionalmente ante dichas situaciones. Un experto en estadística debe retroceder y observar desapasionadamente para mantener un sentido claro de equilibrio y proporción.

El **campo de la estadística** es *un conjunto de procedimientos para reunir, medir, clasificar, codificar, computar, analizar y resumir información numérica adquirida sistemáticamente.* Un curso de estadística suele ser percibido como aquel que involucra muchas fórmulas y cálculos. De hecho, intervienen algunas operaciones matemáticas, pero no constituyen el catalizador de la estadística y por lo general las computadoras se encargan de esta parte. En realidad, la estadística implica aprender una nueva manera de ver las cosas, adquirir una visión de la realidad basada en el análisis cuidadoso de hechos, en lugar de reacciones emocionales ante experiencias aisladas.

> **El campo de la estadística** Conjunto de procedimientos para reunir, medir, clasificar, codificar, computar, analizar y resumir información numérica adquirida sistemáticamente.

No todas las búsquedas requieren representaciones exactas y objetivas de la realidad.[1] Los medios populares de comunicación para el entretenimiento: películas, televisión, novelas románticas, etc., son, por definición, ficción y fantasía, con personajes y eventos imaginarios. Están diseñados para emocionar, alegrar, entristecer o inspirar. Del mismo modo, la publicidad llega al mundo entre la realidad y la fantasía, apelando no sólo a la razón sino también a las emociones, para convencer de que con una compra uno se *sentirá* bien. Las campañas políticas apelan a las emociones de orgullo, patriotismo, temor y odio. Si bien la mayoría de los candidatos son servidores públicos especializados, no todos los políticos se apegan a los hechos, ni se les exige hacerlo; muchos contratan "expertos" para lograr una imagen.

El foro político mantiene un fuerte contraste con la ciencia, pues ésta conlleva un esfuerzo específicamente diseñado para generar un entendimiento más claro de la naturaleza. La ciencia se practica, en la mayor medida posible, de manera independiente de la influencia política o ideológica. El análisis estadístico es una parte vital del método científico. Existe una gran diferencia entre las estadísticas objetivas de las encuestas científicas independientes y las opiniones tendenciosas de encuestadores contratados por políticos ambiciosos. Mientras la meta del personal de las campañas políticas es reforzar la confianza del electorado, empresas independientes intentan estimar la opinión pública. Por ejemplo, para mostrar que un candidato al Congreso va adelante en las preferencias electorales, el personal de la campaña puede contratar una empresa encuestadora que esté dispuesta a hacer preguntas capciosas y preguntar sólo a votantes que han donado dinero. Por supuesto, semejante sondeo revelará un fuerte apoyo y quizá el personal olvide mencionar a los medios de comunicación que la muestra no era representativa de todos los electores. Tal manipulación de información numérica hace recordar el dicho de Mark Twain: "Hay mentirosos, mentirosos detestables y estadistas."

Si un estadista profesional dirigiera la misma encuesta, el estudio no ocultaría datos ni incluiría preguntas capciosas. En cambio, un estadista sigue procedimientos cuidadosamente controlados y realiza un muestreo a partir de la población entera de votantes. Los resultados se presentan con un rango de error y un grado de confianza conocidos, por ejemplo, más menos 3 puntos porcentuales con 95% de confianza. En la estadística se trata de lograr una perspectiva equilibrada y una alta precisión en la recolección y presentación de información.

El principal objetivo de este libro es mostrar que el campo de la estadística versa sobre la obtención de un sentido de proporción exacto con respecto a la realidad; esto significa ver objetivamente las cosas, hacer apreciaciones justas sobre eventos y conductas, dar la cantidad de atención correcta a las cosas que en verdad importan y no distraerse con eventos irrelevantes. Un sentido de proporción ayuda a moderar los sentimientos subjetivos, los sesgos y prejuicios que distorsionan la percepción de la realidad. Aprender a ubicar las cosas en la

perspectiva apropiada requiere imaginación, y en ello descansa el potencial de ver el análisis estadístico como un esfuerzo interesando y que se puede disfrutar.

La imaginación estadística

Como ya mencionamos, el objetivo de este texto es proporcionar una nueva visión de la realidad basada en un análisis estadístico. Llamaremos a esta visión imaginación estadística. El científico social C. Wright Mills (1959) definió la **imaginación sociológica** como *un conocimiento de la relación del individuo con la sociedad y con la historia.* La imaginación *sociológica* es el reconocimiento de que el comportamiento individual se rige en función de estructuras sociales más grandes; que la mayoría de las acciones individuales involucra apegarse a las reglas de la *sociedad* y no a la iniciativa personal; y que, bien o mal, tales reglas se definen dentro de un contexto cultural. La imaginación sociológica implica ver un detalle aislado (una parte) con respecto a una representación más amplia (el todo); ver el bosque, así como los árboles.

Del mismo modo, la imaginación estadística consiste en percibir una parte en relación con el todo. La **imaginación estadística** es *una apreciación de cómo un evento usual o inusual, circunstancia o conducta está en relación con un conjunto mayor de eventos similares, y una apreciación de las causas y consecuencias de un evento.*

> **La imaginación estadística** Una apreciación de qué tan usual o inusual es un evento, circunstancia o comportamiento, en relación con un conjunto mayor de eventos similares y una apreciación de las causas y consecuencias del mismo.

Poseer la imaginación es entender que la mayoría de los eventos son predecibles (es decir, éstos tienen una probabilidad de ocurrencia basada en tendencias y circunstancias a largo plazo).[2] La imaginación estadística es la habilidad para pensar a través de un problema y mantener un sentido de proporción o equilibrio cuando se pondera la evidencia contra nociones preconcebidas; es reconocer eventos muy raros por lo que son y no por la reacción ante ellos.

Ser estadísticamente *falto* de imaginación es poner las cosas fuera de proporción, para pensar de manera reaccionaria en lugar de proporcional. Por ejemplo, en 1991 muchas personas se perturbaron con noticias sobre una persona que se había inclinado al canibalismo, como en el caso célebre del asesino en serie Jeffrey Dahmer. Mientras este evento suscitó furia, temor y disgusto, muchos lo vieron como un símbolo del declive moral en Estados Unidos. Semejante noción es reaccionaria. ¡El canibalismo es tan raro ahora como siempre lo ha sido! La imaginación estadística dice: mire esto a la larga, ¿está pasando con frecuencia? ¿Se involucran muchas personas en esta conducta? ¿Es probable que me convierta en el almuerzo de alguien? De hecho, el incidente de Jeffrey Dahmer fue un caso aislado que involucró a sólo una de entre 250 millones de personas. Ver este evento en su proporción adecuada da la razón a los argumentos sobre la representación más grande de la estabilidad cultural.

Adquirir la imaginación estadística es abrir los ojos a una representación más amplia de la realidad y superar los malentendidos, prejuicios y la estrechez de pensamiento. Por ejemplo, funcionarios de salud pública informan que más de 40 000 personas mueren cada año en accidentes automovilísticos. Confunden el hecho de que los estadounidenses no ven esta importante causa de muerte como un problema de salud pública, uno relacionado con la seguridad de los caminos y el diseño de los automóviles y, por consiguiente, un problema que debería ser resuelto mediante políticas gubernamentales; en cambio, el público ve los fallecimientos en accidentes automovilísticos como infortunios o fallas individuales. Suponemos que las muertes en el tráfico son resultado de la mala suerte (la víctima se cruzó en el camino de un conductor descuidado), tontería, imprudencia o descuido (la víctima manejaba a exceso de velocidad o se quedó dormida), mezquindad (demasiado tacaño para comprar nuevos neumáticos) o inmoralidad (la víctima no debió estar bebiendo). ¿Por qué el público no apela a explicaciones individuales pasadas? Una razón es que las muertes y lesiones de tráfico no golpean con frecuencia a una familia en particular y, por consiguiente, parece que ocurren a "otra persona". Mientras estemos convencidos de que la víctima se lo buscó, nos sentiremos tranquilos de que no nos pasará a nosotros. Por supuesto, nunca beberíamos al mismo tiempo que manejamos y sólo aceleraríamos donde fuese seguro hacerlo.

La imaginación estadística, sin embargo, nos permite reconocer el efecto a gran escala de este medio de transporte. Miramos la representación amplia de cómo los accidentes de tráfico afectan a la población en contraposición a los individuos. Calculamos las muertes totales y las tasas de mortalidad por millones de millas recorridas usando datos que abarcan muchos años. Determinamos qué condiciones inseguras del camino resultan en pérdida de vidas cuando los individuos *son* descuidados. Por ejemplo, se sabe bien que ocurren más muertes en caminos de dos carriles que en carreteras interestatales de cuatro carriles. De hecho, tomando en cuenta el incremento de automóviles y conductores (traducido en millones de millas recorridas), las tasas de mortalidad en el tráfico han disminuido de manera significativa desde que el sistema de carreteras interestatales se construyó en los años cincuenta y sesenta. Enfocándonos en el grupo y examinando las circunstancias, además de los individuos, colocamos las muertes en el tráfico en el amplio contexto de salud pública. Sólo entonces empezamos a considerar el valor de la seguridad de otros medios de transporte, como autobuses y trenes subterráneos.

Enlace de la imaginación estadística con la imaginación sociológica

Normas estadísticas y normas sociales

Una visión equilibrada requiere más que un cálculo matemático cuidadoso. Por ejemplo, aun cuando se tenga conciencia del número de muertes anuales por accidentes automovilísticos, existen "prejuicios" a favor de este medio privado de transporte. Nos resistimos ante los esfuerzos para su sustitución por los sistemas de tránsito masivo, porque los automóviles encarnan el valor social de libertad individual, fuertemente arraigado entre los norteamericanos. Estamos dispuestos a enfrentar lesión o muerte por libertad y comodidad.

Cuando los seres humanos usan su ilustre cerebro para calcular proporciones, porcentajes y otras estadísticas, están simplemente esforzándose por obtener una medida de la realidad. Una estadística, sin embargo, no significa mucho por sí sola. Un principio importante de la imaginación estadística es que al hacer interpretaciones estadísticas se deben tener en cuenta las circunstancias de un fenómeno, incluso los valores de la sociedad o algún grupo dentro de ella. Los valores sociales pueden llevar a limitar, o quizá incrementar, la respuesta humana a una estadística. En este sentido, cualquier estadística está sujeta a la cultura, es decir, es **normativa**: *su interpretación depende del lugar, tiempo y cultura donde se observa.* Una **norma social** es *una idea compartida de la conducta que es apropiada o inapropiada en una situación determinada y en una cultura dada.* En una palabra, una norma es una regla y las normas son peculiares a una sociedad particular, a un periodo de la historia y a la situación específica en que la acción ocurre. Lo que se considera correcto o incorrecto, mucho

o poco, depende del lugar y tiempo. Por ejemplo, estar desnudo en la ducha es normal; de hecho, sería peculiar bañarse con la ropa puesta. Estar desnudo en el salón de clases, sin embargo, es un comportamiento desviado (o anormal).

¿Cuándo poco es mucho? Cualquier estadística carece de sentido si no se establece alguna base de comparación, una norma estadística. Una **norma estadística** es *una tasa promedio de ocurrencia de un fenómeno.* Semejante promedio puede diferir de una sociedad a otra o de un grupo a otro porque cualquier norma estadística es influenciada por normas sociales. Para ilustrar las normas estadísticas y su relación con las normas sociales, comparemos algunas tasas de mortalidad infantil nacional (TMI), el número de niños que mueren en el primer año de vida por cada 1 000 nacidos vivos. La tabla 1-1 presenta la TMI de países seleccionados durante 2003. En Estados Unidos, la TMI fue aproximadamente de nueve muertes por cada 1 000 nacidos vivos.

¿Esta tasa fue alta o baja comparada con la norma estadística? Fue baja respecto de la norma estadística mundial de 51.1, pero los estadounidenses no deben sentirse satisfechos por ello. La TMI está estrechamente ligada al desarrollo económico; por consiguiente, la tasa de Estados Unidos se compara en forma más apropiada con las normas estadísticas de culturas y economías parecidas, como la de los países industrializados (Japón y los países de Europa Occidental). Resulta que la TMI es bastante alta comparada con la TMI de estas naciones y los funcionarios de salud pública de Estados Unidos están preocupados por ello. Tomado en este contexto, poco es mucho. Cualquier muerte infantil es significativa para la familia de la víctima; pero esto no inquieta a los funcionarios de salud pública en un país pobre con una alta TMI (así como los estadounidenses no ven las muertes en el tráfico con alarma). Primero, los altos niveles de TMI han perdurado por siglos, haciendo que la norma estadística del país parezca estable. Segundo, las circunstancias culturales, la higiene y el cuidado médico escasos y la falta de recursos económicos pueden desafiar enormemente los esfuerzos para reducir esta tasa. Tercero, otras causas de muerte, como el síndrome de inmunodeficiencia adquirida (SIDA), son tan grandes que hacen parecer la TMI bastante baja o simplemente parte de un problema más grande. La forma en que los funcionarios públicos o una sociedad como un todo interpreten una estadística, depende de las circunstancias en un momento dado. Como sugiere este análisis rudimentario de la TMI en la tabla 1-1, la situación cultural influye en la interpretación de los hallazgos estadísticos.

Para algunas mediciones, como aquellas sobre el desempeño cognoscitivo o conductual, el estado de salud y el logro académico, las normas estadísticas son necesarias incluso para dar sentido a una puntuación. Por ejemplo, con pruebas de coeficiente intelectual (pruebas de CI), las puntuaciones son normadas contra el *juicio* informado de la comunidad de investigación psicológica sobre lo que constituye la *inteligencia promedio.* Así, las pruebas de CI a menudo son específicamente diseñadas con una norma estadística de 100, número con el que estamos familiarizados y nos sentimos cómodos. Una persona con inteligencia presumiblemente promedio puntúa 100, mientras aquellas que obtienen una puntuación mayor tienen un CI arriba del promedio, y quienes presentan una puntuación menor poseen un CI inferior al promedio.

Ideales estadísticos y valores sociales

Una discusión sobre tasas de mortalidad infantil trae a la mente otra distinción que une a la estadística con la realidad social: aquella entre las normas y los ideales. Mientras una norma estadística es un promedio existente, un **ideal estadístico** es *una tasa de ocurrencia so-*

TABLA 1-1 | Tasas de mortalidad infantil en países seleccionados en 2003

País	Tasa de mortalidad infantil (muertes de menores de 1 año por 1 000 nacidos vivos)
Ya industrializados	
Japón	3.3
Islandia	3.5
Suiza	2.8
Alemania	4.2
Canadá	4.9
Inglaterra	5.3
Estados Unidos	6.8
Por industrializarse	
México	22.5
China	25.3
India	59.6
Haití	76.0
Etiopía	103.2
Afganistán	142.5
En el mundo	51.1

Fuente: U.S. Bureau of the Census, International Date Base. http://www.census.gov/ipc/www/idbnew.html.

cialmente deseada de un fenómeno, una tasa óptima que se quiere alcanzar. Los ideales estadísticos reflejan a menudo **valores sociales**, es decir, *ideas compartidas entre los miembros de una sociedad con respecto de cómo deberían ser las cosas.* Los valores son las nociones comunes de una sociedad sobre lo que una sociedad realmente buena tendría en alta estima. En Estados Unidos, por ejemplo, libertad, igualdad, logro, bienestar material, eficacia y nacionalismo son muy valorados (Williams, 1970: 452-500). Estos valores sociales son sólo ideas y nunca se realizan en un sentido puro. Por ejemplo, mientras la libertad individual es altamente valorada, la libertad pura, es decir, que cada individuo establezca sus propias reglas, representa la anarquía. Los valores funcionan como los faros en las playas rocosas. Estas luces sirven como guías, pero alcanzarlas completamente sería arriesgado.

En respuesta a los valores sociales, los ideales estadísticos (tasas de ocurrencia óptimas) a menudo se sustituyen por normas estadísticas. Por ejemplo, la tasa de muerte infantil en Estados Unidos (6.8 muertes por 1000 nacidos vivos) es un dato estadístico y dicha tasa es más alta que la norma para la mayoría de los países *ya industrializados*, como Japón y Alemania. De esta manera, los funcionarios de salud pública de Estados Unidos quizá señalen las tasas de estos países como un ideal estadístico, una tasa a alcanzar. El público, sin embargo, tal vez no esté dispuesto a aceptar los cambios necesarios para lograr eso, como impuestos más altos y mayor injerencia gubernamental en el cuidado de la salud. Los debates sobre ideales estadísticos a menudo revelan conflictos subyacentes y opiniones sobre valores sociales. Tales ideales, entonces, son únicamente eso y son fuertemente influenciados y restringidos por los valores sociales.

El significado de cualquier estadística a veces depende tan sólo de circunstancias prácticas. Por ejemplo, las normas e ideales estadísticos basados en la biofísica abundan en deportes competitivos, reflejando las limitaciones prácticas de la física. Por ejemplo, ¿cuatro minutos es un tiempo largo? Es un tiempo demasiado largo para completar la vuelta en una carrera automovilística, pero notablemente corto para correr a pie una milla. Un jugador profesional de baloncesto con un porcentaje de tiros libres de 50% está arriesgándose a perder un contrato multimillonario. Sin embargo, un jugador de béisbol de ligas profesionales necesita dar de hit a la pelota aproximadamente 33% de las oportunidades (un promedio de bateo de 0.333) para ganar el título de bateo en la liga y conseguir un aumento multimillonario. A veces poco es mucho. La importancia de una estadística depende de las normas estadísticas (promedios), los ideales estadísticos (tasas señaladas como óptimas) y las circunstancias prácticas. La imaginación estadística se emplea para escoger las normas y los ideales estadísticos apropiados con los que se comparan estadísticas y observaciones.

La imaginación estadística, con su conocimiento de la relación entre medidas estadísticas y datos sociales, requiere un grado de escepticismo, actitud crítica y suspicaz. Así como un estadista se muestra escéptico ante lo que se informe como un hecho por aquellos con intereses políticos y económicos, debe aplicarse el escepticismo a la labor de un estadista, especialmente en el trabajo científico.

Estadísticas y ciencia: herramientas para el pensamiento proporcional

Como hemos mencionado, la estadística trata de observar y organizar información numérica sistemáticamente adquirida. *La información sistemáticamente adquirida que se organiza siguiendo los procedimientos de la ciencia y la estadística* se llama **dato** o **datos**.

Las estadísticas y la recolección de datos no son actividades informales, pero son empresas que requieren un esfuerzo máximo. El análisis estadístico implica precisión, es decir, se refiere a seguir procedimientos, y realizar mediciones precisas y predicciones exactas sobre cómo ocurrirán los eventos en el mundo. Cuando el análisis estadístico se hace de manera apropiada, el analista conoce las limitaciones del razonamiento y de los procesos matemáticos, y sabe cuándo las predicciones sobre eventos o conductas son menos que precisas; además, puede expresar el grado de confianza que tiene al hacer una conclusión. En cuanto a esto, el objetivo de la estadística consiste en controlar el error. Los errores estadísticos no son equivocaciones. El **error estadístico** se refiere al *grado conocido de imprecisión en los procedimientos utilizados para reunir y procesar información.* Controlar el error significa ser tan preciso como sea necesario para reforzar la confianza en las conclusiones derivadas de los hallazgos estadísticos.

Error estadístico Grado conocido de imprecisión en los procedimientos utilizados para reunir y procesar información.

La imaginación estadística no sólo requiere un sentido de proporción acerca de la realidad, sino también la diligencia para mantenerse al tanto de los detalles para minimizar el error.

Estadística descriptiva e inferencial

Los datos se reúnen para diferentes propósitos estadísticos. Un propósito del análisis estadístico consiste en tomar muchos datos sobre una categoría de personas u objetos, y resumir esta información en pocas cifras matemáticas exactas, tablas o gráficas. Este primer paso en la estadística se llama estadística descriptiva.

La **estadística descriptiva** informa *cuántas observaciones fueron registradas y qué tan frecuentemente ocurrió en los datos cada puntuación o categoría de observaciones.* Por ejemplo, datos tomados de 291 encuestados muestran que 40 por ciento son varones y tienen una edad promedio de 21 años, siendo el más joven de 19 años y el más viejo de 51. La estadística descriptiva es empleada tanto por científicos como por encuestadores, analistas de mercadotecnia, proyectistas urbanos y muchos otros profesionales. Estos cálculos informan al público sobre qué productos adquirir, a qué políticos creer, qué acciones comprar, qué automóviles son los más confiables, a qué edad se recomiendan las revisiones físicas anuales y asuntos por el estilo. La estadística descriptiva también es utilizada por científicos como un primer paso en el análisis de hipótesis de investigación científica, que es una tarea de la estadística inferencial.

Un segundo propósito del análisis estadístico es extraer conclusiones sobre las relaciones matemáticas entre las características de un grupo de personas u objetos. Por ejemplo, podríamos investigar si los estadounidenses con un nivel educativo mayor tienden a creer menos que los de nivel educativo más bajo que el diablo existe. Este tipo de análisis corresponde a la **estadística inferencial** y se calcula *para mostrar relaciones de causa-efecto, así como para probar hipótesis y teorías científicas.* (Inferir quiere decir *sacar conclusiones sobre algo.*) La mayor parte de este texto trata sobre estadística inferencial. Entender los principios básicos de la ciencia es imprescindible para comprender la estadística inferencial; por tanto, repasaremos estos principios.

¿Qué es la ciencia?

La **ciencia** es *un método sistemático para la explicación de los fenómenos empíricos.* **Empírico** significa *observable y medible.* Los fenómenos (forma plural derivada de la palabra latina *phanomenon*) son hechos, situaciones, eventos, circunstancias, o bien, simplemente "cosas que existen naturalmente". Los fenómenos empíricos, entonces, son cosas que pueden observarse y medirse, como condiciones naturales, procesos, eventos, situaciones, objetos, grupos de personas, conductas, pensamientos, creencias, conocimientos, opiniones, emociones y sentimientos.

No todo es medible y observable. Por ejemplo, si existe vida después de la vida no se observa fácilmente, aunque cerca del 70 por ciento de los adultos estadounidenses afirma creer en ello. Además, muchas cosas intangibles, como las emociones, sentimientos y creencias deben medirse de forma indirecta. En las ciencias sociales tales mediciones indirectas incluyen encuestas que sondean opiniones, conocimiento, actitudes e incluso conducta. Los científicos físicos también usan mediciones indirectas; por ejemplo, los físicos indirectamente observan neutrinos, partículas subatómicas tan diminutas y rápidas que por lo general atraviesan la Tierra sin golpear con algo. (Millones de ellas están atravesándola ahora mismo.) En ocasiones, un neutrino desplaza una molécula de agua, liberando energía observable, y este efecto puede medirse. Un aspecto importante de la expansión de la ciencia está en encontrar nuevas formas de medir con exactitud los fenómenos que no son visibles al ojo humano. Microscopios, máquinas computarizadas de rayos X y sismómetros, así como instrumentos de investigación son las herramientas que los científicos emplean para extender el alcance de sus capacidades de medición.

El propósito de la investigación científica El objetivo principal de la ciencia es explicar los fenómenos. Una explicación científica está basada en procedimientos estrictos y se llama teoría. Una **teoría científica** es *un conjunto de aseveraciones interrelacionadas y lógicamente organizadas que explican un fenómeno de especial interés y que han sido corroboradas a través de la observación y el análisis.* Las teorías describen situaciones y cómo funcionan éstas. La colección de ideas que constituyen una teoría se prueba contra los hechos observados. Una teoría se "corrobora" cuando sus ideas predicen con éxito estos hechos observables. Una teoría no es un hecho en sí misma; es una explicación bien organizada de hechos. Cuando un fenómeno se entiende mejor, una teoría se modifica y se refina para aumentar su poder de predicción. Así, su desarrollo es un proceso acumulativo que ocurre durante un largo periodo.

Una teoría científica adecuada logra dos cosas. Primero, proporciona un sentido de comprensión sobre un fenómeno: cómo, cuándo, por qué y bajo qué condiciones sucede; descrito de manera sencilla, da sentido a las cosas. Segundo, una teoría nos permite realizar predicciones empíricas, respondiendo la pregunta de bajo qué condiciones y en qué grado un fenómeno ocurrirá. Tales predicciones son posibles porque los cambios en un fenómeno se relacionan con cambios en otros fenómenos. Por ejemplo, predecimos una mayor oportunidad de lluvia cuando la humedad atmosférica aumenta, o un incremento en la tasa de delitos en una comunidad cuando hay crisis económica.

Escepticismo científico e imaginación estadística

La ciencia *requiere* que sus ideas resistan la prueba de predecir observaciones. Los científicos especializados son escépticos, mantienen una actitud crítica y suspicaz y están dispuestos a tolerar la incertidumbre, por lo que no son demasiado rápidos para obtener conclusiones. Un escéptico vacila en creer algo simplemente porque sus amigos de confianza, los medios de comunicación masiva o personas en posiciones de autoridad, como líderes gubernamentales o incluso sus padres, lo consideran verdadero. Un vistazo a la cultura popular, sobre todo a las ideas lanzadas en los medios de comunicación masiva, sugiere que la mayoría de las personas son altamente crédulas (inclinadas a creer) incluso en ausencia de evidencia o en presencia de evidencia contradictoria. El renombrado científico Carl Sagan, un portavoz del valor de la ciencia, observó que muchas personas son muy rápidas para "cesar la *incredulidad*". Por ejemplo, él notó cómo las personas crédulas estaban cayendo en el engaño de los "círculos de la cosecha", al creer, durante 15 años, que los enormes y elegantes pictogramas descubiertos en campos de cultivo ingleses habían sido dejados ahí por extraterrestres. (Un par de bribones llamados Bower y Chorley finalmente confesaron ser los autores del engaño.) Sagan argumentaba que no somos lo suficiente escépticos ante mucho de lo que se presenta como un hecho sólo por alegato. Él exhortó a tomar la ciencia más en serio, porque el proceso científico se diseña especialmente para separar el hecho de la ficción (Sagan, 1995a, 1995b). Además, afirmó que una sociedad que fomenta el aprendizaje de la ciencia producirá ciudadanos mejor informados. Sagan señaló lo siguiente:

> En la universidad ... empecé a aprender un poco sobre cómo funciona la ciencia ... qué tan rigurosas deben ser las normas de evidencia ... cómo nuestros prejuicios pueden afectar nuestra interpretación de la evidencia, cómo los sistemas de creencias ampliamente ... apoyados por las jerarquías políticas, religiosas y académicas a menudo resultan no sólo ligeramente errados sino grotescamente equivocados...
>
> Los principios del escepticismo no exigen un grado avanzado de dominio... Lo que la ciencia pide es que empleemos los mismos niveles de escepticismo que usamos para comprar un automóvil usado...
>
> (Sagan, 1995b: 10-13)*

*De Carl Sagan, *The Demon-Haunted World*. Copyright © 1995 por Random House. Reimpreso con permiso del editor.

Las explicaciones científicas basadas en la observación, los procedimientos estrictos y el escrutinio colectivo de la comunidad científica a menudo contradicen el sentido común, así como las ideas propuestas por líderes políticos. Esto no significa que la ciencia abandone el sentido común. La ciencia utiliza el **sentido común informado**, *el que es evaluado y doblemente verificado contra datos cuidadosamente recogidos.* ¡El sentido común desinformado es demasiado común! El escepticismo científico requiere aprender habilidades de procedimiento y desarrollar una actitud de cuestionamiento. De igual manera, poseer la imaginación estadística involucra aprender habilidades (por ejemplo, cómo calcular probabilidades y pensar proporcionalmente) y estar listo para preguntar si un fenómeno observado es razonable.

Al mismo tiempo, sin embargo, la ciencia tiene limitaciones. Primero, se restringe a exámenes de fenómenos empíricos, observables y medibles. La fe, no la ciencia, debe resolver, por ejemplo, la pregunta de si Dios y la salvación del alma existen. Segundo, muchos sondeos, objetivamente basados en argumentos científicos, carecen de apoyo político o de los contribuyentes. Por ejemplo, la investigación revela que la pobreza en Estados Unidos se reduciría si se extienden los programas gubernamentales de ayuda familiar, como capacitación para el trabajo y servicios de cuidado infantil. Estos "programas de asistencia", sin embargo, son costosos y a menudo carecen de apoyo del contribuyente; la reciente legislación da simplemente un tiempo limitado a los destinatarios de la asistencia pública para resolver estos problemas por sí mismos. Una tercera limitación de la ciencia es que provoca dilemas éticos y resistencia ante su aplicación. Por ejemplo, un economista podría elaborar un argumento convincente de que la eutanasia o "muerte por piedad" ahorraría miles de millones de dólares en gastos médicos destinados a enfermos terminales. Obviamente, muchos cuestionarían tal argumento no con base en la cuestión monetaria, sino en la moral. Hay más qué explicar acerca de la existencia humana que de los costos.

La ciencia no tiene todas las respuestas y los científicos deben mantenerse escépticos de las respuestas que tengan. Cuando se explica la realidad empírica, sin embargo, el método científico constituye el mejor método. Un rasgo importante del método científico es el análisis estadístico.

Concepción de los datos

Variables y constantes *Los fenómenos medibles que varían (cambian) a través del tiempo o que difieren de un lugar a otro o de un individuo a otro* se denominan **variables**. Las variables son características de los sujetos (estudiantes, personas sin hogar, habitantes de St. Louis, ratas de laboratorio) u objetos (edificios, árboles, inundaciones, bacterias, delitos) bajo estudio. (De aquí en adelante, se empleará el término *sujeto* para designar tanto a personas como a objetos.) Por ejemplo, al estudiar a los individuos se notarían diferencias en las variables de edad, peso, estatura, rasgos de personalidad, raza y nivel socioeconómico.

> **Variable** Fenómeno medible que varía (cambia) a través del tiempo, o que difiere de un lugar a otro o de un individuo a otro.

Se utiliza el término **variación** para referirse a *cuánto difieren las mediciones de una variable entre los sujetos en estudio.* Se comparan las diferencias en la variación entre grupos. Existe mucha variación, por ejemplo, en las edades de los estudiantes universitarios en las grandes ciudades, quizá van de 17 a 70 años. En contraste, esta variación en universidades en pequeños "pueblos aledaños a un campus" es comúnmente más pequeña, de 17 a 25 años.

Algunas variables muestran poca o ninguna variación dentro de un grupo, como las edades de los alumnos de primer grado. Las *características de los sujetos en estudio que no varían* se llaman **constantes**. A veces, de manera intencional "mantenemos constantes las variables". Por ejemplo, en un experimento sobre los efectos de las bebidas alcohólicas en la conducta del automovilista, usaríamos sujetos de más o menos el mismo peso, porque se sabe que las personas de menor peso se embriagan con mayor rapidez que las más pesadas. De esta manera, la reducción del tiempo de reacción al manejar se atribuiría a la cantidad de alcohol consumida, en lugar de las diferencias en peso. "Al mantener constante el peso" eliminamos sus efectos en la conducta del automovilista; puesto que el peso no variaba, una variación en el peso no podría explicar los resultados del experimento. Si se mantiene constante el peso y cualquier otra variable que afecte la embriaguez, somos capaces de aislar los efectos del consumo de alcohol en la conducta de quien maneja.

La variable dependiente y las variables independientes que la explican Por lo común, al recolectar datos, nuestro propósito consiste en investigar una sola variable que es de especial interés para nosotros. Queremos saber qué provoca un incremento o disminución en la cantidad de esta variable. ¿Qué causa dicha "variación"? ¿Cuáles son sus puntuaciones dependientes? Esta variable de principal interés se denomina **variable dependiente**, *la variable cuya variación queremos explicar.* Por ejemplo, los años sesenta se caracterizaron por la violencia urbana, con disturbios en más de 40 ciudades durante un periodo de tres años. En un esfuerzo por entender y prevenir los disturbios, la Comisión Nacional de Asesoría sobre Desórdenes Civiles (1968) se formó para dirigir un estudio científico. La incidencia de conducta tumultuaria fue la variable dependiente. La comisión quiso explicar por qué los disturbios ocurrieron en algunas ciudades, pero no en otras.

Las variables que se sospechaba estaban relacionadas con un incremento o disminución en la conducta tumultuaria también se midieron. Tales variables incluyeron la tasa de pobreza en las comunidades, el número de quejas por brutalidad policiaca, las perturbaciones raciales en las semanas previas a un disturbio y el número de "simpatizantes comunistas" conocidos en una ciudad. Estas *variables de predicción, que están relacionadas o que predicen la variación en la variable dependiente,* se conocen como **variables independientes**.[3] La tabla 1-2 distingue las características de las variables independientes y de las variables dependientes.

Un enunciado que predice algo acerca de la relación entre variables se denomina hipótesis. Específicamente, una **hipótesis** es *una predicción sobre la relación entre dos variables; en ella se afirma que los cambios en la medida de una variable independiente corresponderán a cambios en la medida de una variable dependiente.*

> **Hipótesis** Predicción sobre la relación entre dos variables; en ella se afirma que los cambios en la medida de una variable independiente corresponderán a cambios en la medida de una variable dependiente.

La Comisión sobre Desórdenes Civiles examinó ciudades como sus sujetos de investigación. El equipo de investigación de la comisión encontró que la incidencia de los disturbios (la variable dependiente) estaba relacionada con las variables independientes como el porcentaje de familias que vivían en la pobreza, la suficiencia de programas de asistencia social, el grado de participación gubernamental de las minorías, la ocurrencia de "incidentes de elevada tensión" y, especialmente persistentes, informes de brutalidad policiaca. Para mostrar que existe relación entre la incidencia de brutalidad policiaca y la incidencia de disturbios, se dice que las ciudades con elevada brutalidad policiaca también tendieron a sufrir muchos disturbios. Esta afirmación y otras similares involucran otras variables independientes que conforman una *teoría de protesta de la conducta tumultuaria.* Esta teoría propuso el argumento de que las personas provocan disturbios en respuesta a las acciones policiacas opresivas y debido a la frustración por escasos servicios del gobierno. Las ideas y datos estimulados por esta teoría finalmente auspiciaron cambios en políticas gubernamentales locales y una reducción en los desórdenes civiles (Johnson, 1973: 376).

Los hallazgos de la comisión refutaron algunos **mitos**, *creencias ampliamente sostenidas que son falsas y que estaban muy arraigadas.* Específicamente, se desacreditó la *teoría de la conspiración comunista*, un argumento político que sostenía que los disturbios eran parte de una revolución organizada para derrocar al gobierno de Estados Unidos (Johnson, 1973: 376). ¿Por qué de manera tan inmediata las personas creyeron que los disturbios representaban una conspiración comunista? Los mitos a menudo surgen de explicaciones de sentido común reforzadas por eventos aislados o esporádicos y por la retórica política que aviva los miedos del electorado. La violencia urbana de los años sesenta ocurrió durante un periodo de cambio social rápido e incierto. En el frente nacional interno hubo un movimiento de derechos civiles organizado por minorías raciales, en especial de afroamericanos, que exigían eliminar la discriminación en la contratación para empleos, en las escuelas y en el uso de instalaciones públicas. En la escena mundial había, al mismo tiempo, una "guerra fría" entre los países capitalistas de Occidente y los países comunistas; de estos últimos, sobre todo, la antigua Unión Soviética y la República Popular China, cuyos gobiernos realizaron llamados abiertos a las armas y buscaron infiltrar espías en Estados Unidos, los cuales exaltarían a los pobres y a las "minorías reprimidas" a sublevarse. En esta atmósfera, la presencia de disturbios en barrios habitados por minorías pobres por todo Estados Unidos parecía, a muchas personas, un resultado verosímil de la conspiración comunista.

Como ocurrió, los hechos demostraron que era sumamente difícil encontrar a los comunistas entre los participantes del disturbio. Además, no había ninguna diferencia en el número de simpatizantes comunistas en ciudades donde los disturbios ocurrieron y en las ciudades donde no los hubo. El argumento de la conspiración comunista fue refutado, no fue apoyado por datos y no resistió el escrutinio del análisis estadístico.

TABLA 1-2 | Posibles relaciones entre las variables independientes y dependientes

Variable independiente		Variable dependiente
Causa	→	Efecto
Predictor	→	Resultado
Estímulo	→	Respuesta
Intervención (acción tomada)	→	Resultado
Correlación: cambio en una variable	→	Cambio asociado en otra variable

Una teoría científica es un argumento organizado que debe ser corroborado por la evidencia empírica. Una teoría se "corrobora" cuando sus ideas predicen con éxito mediciones observables.[4] Cuantos más datos se adquieran, más se modificarán y refinarán las teorías para mejorar su poder de predicción y su sentido de comprensión.

El proceso de investigación

El proceso de investigación implica organizar ideas en una teoría, realizar predicciones empíricas que apoyen la teoría y después reunir datos probatorios de tales predicciones. El proceso de investigación es acumulativo, es decir, un proceso continuo de acumulación de conocimiento. El proceso de investigación científica comprende siete pasos que serán enseñados en diferentes etapas: del 1 al 3 son los principales temas en los cursos de teoría en ciencias sociales, los pasos 4 y 5 se cubren en cursos de metodología y los pasos 6 y 7 se enseñan en cursos de estadística. Los siete pasos son los siguientes:

1. *Especifique la pregunta de investigación.* Planteamos una pregunta e identificamos la variable dependiente. Por ejemplo, podemos preguntar ¿por qué están ocurriendo disturbios en algunas ciudades?
2. *Revise la literatura científica.* Hacemos esto para asegurarnos de que no se desperdicien tiempo y dinero recolectando datos que ya existen. Buscamos la "frontera del conocimiento", los límites exteriores de lo que ya ha sido aprendido, por ejemplo, sobre los disturbios. La investigación bien informada y publicable extiende el conocimiento más allá de las fronteras.
3. *Proponga una teoría y formule una hipótesis.* La teoría involucra la organización de ideas en una forma lógica que pueda explicar la variación en la variable dependiente. Al desarrollar una teoría, identificamos las variables independientes y hacemos declaraciones predecibles respecto de cómo pensamos que afectan a la variable dependiente, asumiendo que la teoría es comprensible.

Las hipótesis se generan o "motivan" por la teoría, ideas probadas que han sido encontradas en la literatura científica, con modificaciones innovadoras del investigador. La teoría nos lleva a esperar ciertos resultados observados de los datos. Si estos resultados se presentan, la teoría se corrobora. Por ejemplo, la teoría de la protesta en la conducta de disturbios motiva la siguiente hipótesis:

H_1: Las ciudades con alta incidencia de brutalidad policiaca (variable independiente) están sujetas a tener una elevada incidencia de desórdenes civiles (variable dependiente).

En contraste, la teoría de la conspiración comunista en la conducta de disturbios da origen a la siguiente hipótesis:

H_2: Las ciudades con un gran número de comunistas (variable independiente) están sujetas a tener una elevada incidencia de desórdenes civiles (variable dependiente).

La teoría, basada en la revisión de la literatura, también guía en la selección de variables de "control". Por ejemplo, al medir desórdenes civiles, debemos controlar la tasa de delitos. Esto nos asegura que las ciudades con una alta incidencia de disturbios no son simplemente ciudades con alta criminalidad, en las cuales la tasa de casos de delitos, no sólo la brutalidad policiaca, explicaría parte de la incidencia de desórdenes civiles.

También debemos notar que no todos los estudios científicos emplean teoría. Mucha *investigación se lleva a cabo para resolver problemas prácticos inmediatos o para explorar nuevos fenómenos sobre los que se conoce tan poco que formular una teoría sería imposible.* Tales estudios se llaman **estudios exploratorios**. Por ejemplo, alguien que explora cuestiones privadas en internet empezaría con ideas y preguntas vagamente organizadas.

4. *Seleccione un diseño de investigación.* En el diseño de investigación se detalla cómo se medirán, muestrearán y reunirán los datos. Los métodos comunes de la ciencia social incluyen la observación directa del comportamiento, el experimento de laboratorio, la encuesta, el análisis de contenido en los medios de comunicación y el análisis de datos existentes o "secundarios" (como informes policiacos y censos de población).
5. *Recolecte datos.* Ésta es normalmente la parte más costosa de la investigación. Se trata de "entrar en el campo" para informar a las personas sobre el estudio y recolectar datos utilizando el plan desarrollado en el paso 4. También es una de las partes más agradables de la investigación, pues permite al investigador salir de la oficina y conocer nuevas y, a menudo, interesantes personas.
6. *Analice los datos y saque conclusiones.* Es aquí donde entra el análisis estadístico, tema principal de este libro. Las hipótesis se prueban mediante la comparación de observaciones con predicciones teóricas. En el ejemplo de los disturbios, los datos recolectados por la Comisión sobre Desórdenes Civiles apoyaron la hipótesis 1 y refutaron la 2, otorgando mayor credibilidad a la teoría de protesta.
7. *Difunda los resultados.* Difundir significa diseminar ampliamente y compartir. Los hallazgos científicos se comparten con dos tipos de "audiencias": el público en general y la comunidad científica.

Las audiencias públicas incluyen no sólo a los ciudadanos, sino también a grupos políticos y empresariales, religiosos, de caridad y educativos. Los investigadores pueden exponer en foros públicos a manera de conferencias de prensa, entrevistas, reuniones oficiales en la ciudad, reuniones comunitarias y clases de bachillerato. Tales charlas deben ser sencillas conceptual y estadísticamente.

Para la audiencia científica, la difusión de descubrimientos de investigación consiste en presentar los descubrimientos en conferencias científicas y libros que serán publicados o, más comúnmente, en artículos cortos en revistas especializadas. La publicación de la investigación es un arduo proceso de revisión entre pares (un sistema de comprobaciones y acuerdos) que idealmente aumenta al máximo la probabilidad de que un trabajo publicado sea preciso e imparcial. Un manuscrito científico sigue un procedimiento estricto. Cuando se completa, se somete al juicio del editor de una revista especializada en la materia, quien, a su vez, envía copias sin identificación del autor a otros científicos competentes en la materia. Esta revisión "ciega" minimiza el prejuicio personal; obliga a los revisores, sin embargo, a que sean altamente escépticos con el manuscrito. Ellos escrutan cada detalle buscando fallas lógicas, interpretaciones tendenciosas, muestras sin sondeo, medición deficiente o análisis y conclusiones equivocados. Si varios revisores están de acuerdo en que la investigación es sólida y que contribuirá al avance del conocimiento, el editor podrá aceptar la publicación si el espacio de impresión está disponible. Las principales revistas, en la mayoría de los campos, son sumamente selectivas, pues publican sólo una de cada 10 solicitudes. Este proceso asegura que la investigación seleccionada para su publicación alcanza estándares profesionales. Los investigadores que publican regularmente son practicantes de la ciencia altamente calificados.

Pensamiento proporcional: cálculo de proporciones, porcentajes y tasas

El término *proporción* es un concepto matemático relacionado con fracciones y porcentajes. Un buen sentido de proporción sobre un fenómeno requiere más que tener una buena percepción respecto de lo que trata el fenómeno. Entender las proporciones requiere **pensamiento proporcional**: *sopesar la parte contra el todo y calcular la probabilidad de que a la larga ocurra el fenómeno.* Tener un sentido de proporción y calcular las proporciones matemáticas son esencialmente lo mismo. Estas últimas son simplemente expresiones precisas de nuestras intuiciones sobre la importancia de ciertos hechos. Calcular una proporción es una manera de medir y evaluar un sentido de probabilidad y significación respecto de las observaciones realizadas.

Para empezar de manera adecuada, repasaremos brevemente los cálculos básicos de fracciones, proporciones y porcentajes. (Un repaso adicional se proporciona en el apéndice A.) Cada aspecto del trabajo estadístico, desde la medición y la presentación gráfica hasta el cálculo de probabilidades estadísticas, implica trabajar con proporciones matemáticas; por consiguiente, esta revisión ofrece una buena orientación a los cálculos estadísticos.

Las **proporciones matemáticas** son simplemente *problemas de división que comparan una parte (el numerador) contra un todo (el denominador).* Para calcular una proporción, empezamos con una **fracción**, *una forma de expresar qué parte del todo (o número total) constituye una categoría de observaciones.*

$$\text{Fracción} = \frac{(\text{numerador})}{(\text{denominador})} = \frac{(\text{parte})}{(\text{todo})}$$

Cálculo de una fracción

$$\text{Fracción} = \frac{\#\text{ en una categoría}}{(\#\text{ en un grupo total})}$$

donde # se lee "número" o "número de".

Por ejemplo, en un estudio de presos que ocupan la cárcel del condado de Washington, se determina que entre la población total de la cárcel de 149 presos, 112 fueron acusados de delitos relacionados con las drogas (DRD) como la posesión o venta de una sustancia ilegal. ¿Es ésta una gran parte de la población de la cárcel? Si es así, ¿qué dice esto sobre la naturaleza del delito y la aplicación de la ley en el condado de Washington? Para tener un buen sentido de proporción, las dos cifras, 112 y 149, deben integrarse en una fracción:

$$\begin{array}{c}\text{Fracción de presos en la}\\ \text{cárcel del condado de}\\ \text{Washington acusados de DRD}\end{array} = \frac{\#\text{ de acusados de DRD}}{\text{total de la población de presos}} = \frac{112}{149}$$

Una interpretación más fácil de esta fracción se lograría transformándola en una gran proporción. **Proporción** significa *parte de un todo, o parte de la cantidad total o número de observaciones, expresada en forma decimal.* Las fracciones se reducen a proporciones (o expresiones decimales) dividiendo el numerador de una fracción entre su denominador para obtener un cociente. (Un cociente es *la respuesta a un problema de división.*) Así

$$p\ \begin{bmatrix}\text{de presos en la cárcel}\\ \text{del condado de Washington}\\ \text{acusados de DRD}\end{bmatrix} = \frac{\#\text{ de acusados de DRD}}{\text{población total de presos}} = \frac{112}{149} = 0.7517$$

donde p simboliza la proporción, la información entre paréntesis describe el total de la población señalada (el denominador) seguido por la característica señalada (el numerador), y el símbolo (#) se lee como "número".

> **Proporción** Parte de la cantidad total o número de observaciones, expresada en forma decimal.

Cálculo de una proporción

$$p\,[\text{del grupo total en una categoría}] = \frac{\#\text{ en una categoría}}{(\#\text{ en grupo total})} = \text{cociente}$$

donde p = la proporción, y el cociente se redondea a cuatro lugares decimales (ejemplo, la diezmilésima más cercana). El cociente siempre tendrá un valor entre 0 y 1.

Esta proporción para el condado de Washington es correcta, pero torpemente expresada como "punto siete cinco uno siete" o "siete mil quinientos diecisiete milésimas". Para el público en general, entonces, damos un paso más allá y transformamos esta proporción en la expresión más reconocible: un *porcentaje.* **Porcentaje** significa "por cien" y es igual a *una proporción multiplicada por 100.* El porcentaje nos dice cuántos de cada 100 presos están acusados de DRD. Así,

$$\%\ \begin{bmatrix}\text{de presos en la cárcel}\\ \text{del condado de Washington}\\ \text{acusados de DRD}\end{bmatrix} = p\,(100) = (0.7517)(100) = 75.17\%$$

Cálculo de un porcentaje

$$\%\,[\text{del grupo total en una categoría}] = p\,(100)$$

donde p = proporción del grupo total en una categoría. El cociente siempre tendrá un valor entre 0 y 100 por ciento.

En este momento deberíamos tener la sensación de que el abuso de sustancias prohibidas es un serio problema para la aplicación de la ley en el condado de Washington. De hecho, más de 75 de 100 presos son encarcelados bajo cargos de DRD. Evidentemente, *es muy probable y común* para una persona encarcelada haber tenido problemas con drogas. El sistema de justicia en este condado está seriamente agobiado por estos casos.

Las proporciones y porcentajes son medios preferidos para expresar "la parte del todo". Las proporciones siempre tendrán respuestas entre 0 (ninguno) y 1 (todos). De manera semejante, los porcentajes siempre oscilan entre el 0 y el 100 por ciento. Aparte de su simplicidad comparada con la forma fraccionaria, las proporciones y porcentajes son útiles para producir rápidamente comunes denominadores para dos o más fracciones. Las proporciones proveen el común denominador 1.00, mientras que los porcentajes dan el común denominador 100. Por ejemplo, suponga que comparamos los casos de DRD en las cárceles de los condados de Jefferson y de Washington. Jefferson tiene 42 casos en una población total de 45 presos. ¿Qué fracción es más grande: 112 de 149 o 42 de 45? Obtenemos un común denominador calculando las proporciones y porcentajes. Para el condado de Jefferson, entonces:

$$p\ \begin{bmatrix}\text{de presos en el}\\ \text{condado de Jefferson}\\ \text{acusados de DRD}\end{bmatrix} = \frac{\#\text{ de acusados de DRD}}{\text{población total de presos}} = \frac{42}{45} = 0.9333$$

$$\%\ \begin{bmatrix}\text{de presos en el}\\ \text{condado de Jefferson}\\ \text{acusados de DRD}\end{bmatrix} = p\,(100) = (0.9333)\,(100) = 93.33\%$$

Los porcentajes permiten ver que, de hecho, la población de la cárcel del condado de Jefferson está más densamente poblada por delincuentes relacionados con drogas que la del condado de Washington (93.33 por ciento contra 75.17 por ciento, respectivamente), aun cuando hay más casos de DRD en el condado de Washington.

Por medio de estos cálculos podemos observar que, para cambiar una fracción a una proporción, dividimos el numerador entre el denominador para obtener el cociente "en forma decimal". Para cambiar una proporción a un porcentaje, multiplicamos la proporción por 100 moviendo el punto decimal dos lugares a la derecha. Para transformar un porcentaje en una proporción, movemos el punto decimal dos lugares a la izquierda, lo cual es simplemente una cuestión de dividir entre 100. Para expresar una proporción como una fracción, debemos tener buen dominio sobre los lugares decimales. Si es necesario, repase las posiciones de lugares decimales en el apéndice A. Finalmente, como regla general (con pocas excepciones), redondeamos las proporciones a cuatro lugares decimales a la derecha del punto decimal, y los porcentajes a dos lugares decimales.

Un porcentaje es una manera muy común de estandarizar estadísticas de grupos diferentes. A veces, sin embargo, los porcentajes no transmiten un sentido significativo de proporción. Por ejemplo, ¿cuál es la probabilidad de morir a causa de un relámpago? En 2000, se encontró que la población de Estados Unidos era de 281 421 906 (U.S. Bureau of the Census, 2000). Gracias a un meteorólogo se determinó que 51 personas perdieron la vida al ser alcanzados por relámpagos durante ese año (National Oceanic and Atmospheric Administration, 2003). La proporción y el porcentaje de la población muerta por relámpagos se calcularía así:

$$p \text{ [de la población de Estados Unidos en 2000 muerta por relámpago]} = \frac{\text{\# muertos por relámpagos}}{\text{tamaño total de la población}} = \frac{51}{281\,421\,906} = 0.00000018$$

$$\% \text{ [de la población de Estados Unidos en 2000 muerta por relámpagos] } (p) = (100) = 0.000018\%$$

Así, suponiendo que 2000 es un año típico, la probabilidad de perder la vida por relámpagos es de 18 cienmilésimas de uno por ciento. Esto es difícil de concebir incluso por un individuo matemáticamente astuto. Un denominador de 100 es confuso cuando al menos una persona de cada 100 está en riesgo. La imaginación estadística nos llama poderosamente a encontrar otra manera de interpretar este riesgo.

Otra forma de estandarizar es calculando una **tasa**, *la frecuencia de que ocurra un fenómeno en relación con el número "base" especificado de sujetos en una población.* El número base se coloca en el denominador para que la tasa pueda representar los casos por mil, por diez mil, por cien mil, por un millón, y así sucesivamente. Un número base útil es aquel que con claridad especifica "la población de riesgo" para un fenómeno. Con un grupo grande como la población de Estados Unidos, se necesita un número base mayor en lugar del "por 100" usado con los porcentajes. Recuerda que cuando transformamos una proporción en un porcentaje, multiplicamos por 100. De manera semejante, podemos multiplicar una proporción por otros múltiplos de 10 para obtener tasas con denominadores mayores.

Cálculo de una tasa

Tasa de ocurrencia = (p) (un número base útil)

donde p = proporción del grupo total en una categoría y el número base útil es un múltiplo de 10.

Un número base útil para una tasa es aquel que considera la dimensión del conteo de un fenómeno. En este ejemplo contamos a personas fallecidas. Nuestra tasa, entonces, debe presentarse en números enteros con dígitos a la izquierda del punto decimal. Al observar nuestra proporción de 0.00000018, para obtener un conteo de personas, debemos mover el punto decimal siete lugares a la derecha. El repaso de posiciones de lugares decimales en el apéndice A muestra que esto es equivalente a multiplicar por 10000000. Así,

$$\begin{aligned}\text{Tasa de muertes por relámpagos de una población de diez millones} &= (p)\,(10\,000\,000)\\ &= (0.00000018)(10\,000\,000) = 1.8 \text{ fallecimientos por relámpago}\\ &\quad \text{por cada } 10\,000\,000 \text{ de personas}\end{aligned}$$

Este cálculo es explícito y útil. Manifiesta que sólo dos de cada 10 millones de personas mueren por relámpagos cada año. Podemos estimar en 10 millones la población de una ciudad grande (como Nueva York). Al imaginar una ciudad y pensar proporcionalmente, obtenemos la noción de que el riesgo de muerte por relámpagos es muy pequeño. Sólo cerca de dos personas en una gran ciudad tienen probabilidad de morir así cada año. (De hecho, si estuviéramos en una ciudad en el desierto, donde rara vez llueve, se reduciría esta cifra.)

Otro cálculo rápido nos permite colocar el numerador de esta tasa a una persona en lugar de 1.8. Esto nos da el número de personas de la población por cada muerte por relámpago. Una razón de 1.8 a 10000000 es igual a una razón de 1 a 5555556. Esto se obtiene al dividir 10000000 entre 1.8:

$$\frac{1.8 \text{ muertes por relámpagos}}{10\,000\,000 \text{ de personas}} = \frac{1 \text{ muerte por relámpagos}}{X \text{ personas}}$$

$$X = \frac{10\,000\,000}{1.8} = 5\,555\,556 \text{ personas}$$

donde X es el número de personas de la población por cada persona muerta por relámpago. Entonces, la probabilidad de morir por relámpagos en un año es, aproximadamente, de 1 en 5 ½ millones, la población de la zona metropolitana de Boston, Massachussets, por ejemplo. En consecuencia, nuestras probabilidades de morir por un relámpago son mínimas.

Comparación de dos o más grupos de diferente tamaño

Estandarice la fracción usando un común denominador:

Las proporciones tienen un común denominador de 1.

Los porcentajes tienen un común denominador de 100.

Las tasas tienen un común denominador útil seleccionado en múltiplos de 10.

No hemos llegado muy lejos en nuestra discusión introductoria de estadística y ya hemos identificado la importancia de la comunicación precisa. Las fórmulas matemáticas son bastante estrictas en su forma. Todas las que aquí presentamos tendrán los siguientes elementos:

Presentación de respuestas de forma que estimulen el pensamiento proporcional

Símbolo = fórmula = contenidos de la fórmula = respuesta

Observe estos elementos en nuestros cálculos de presos por DRD.

Estos cálculos básicos se introdujeron al principio de este libro porque tener sentido de comprender la realidad y entender la matemática de proporciones van de la mano. Medidas de "parte del todo" son comúnmente los primeros cálculos realizados en cualquier análisis estadístico. El pensamiento proporcional es una característica básica de la imaginación estadística.

Cómo tener éxito en este curso y disfrutarlo

En mis años de enseñar estadística he visto que el estudiante debe estar dispuesto a trabajar y seguir con este curso. La atención y el éxito con las primeras tareas hacen que las tareas más abstractas que sigan serán más fáciles de entender. Tener éxito en un curso de estadística es muy semejante a conseguir que un avión despegue: se emplea mucha energía para alcanzar altitud (capítulos 1 al 9), pero luego el avión puede volar todo el resto de la ruta (capítulos 10 al 15). Este texto está diseñado para un éxito temprano y disipar temores y dejar ver lo agradable e interesante que es esta materia. Incluso el estudiante promedio que esté dispuesto a poner tiempo y esfuerzo puede obtener una A en este curso y además divertirse, pero este curso requiere de tareas prácticas. Aprender estadística es como aprender a tocar un instrumento musical; puedes estudiar teoría musical todo el día, pero hasta que practiques en un instrumento no aprenderás a tocarlo. La clave del éxito para tocar el instrumento o aprender estadística es una "práctica" bien organizada.

Si temes que este curso te condene al fracaso por tu debilidad percibida en matemáticas, deja esos temores de lado. El curso empieza con cálculos sencillos y se basa en ellos. Si trabajas duro y sigues en el curso, las matemáticas no serán un problema. Empieza por hacer un repaso de los procedimientos básicos de matemáticas en el apéndice A. He aquí algunos consejos de estudio:

- Organiza tus apuntes de estudio, tareas, papeles devueltos y otras cosas por el estilo en un cuaderno de argollas. Esto te permite insertar materiales corregidos y papeles devueltos en su lugar y hace más eficiente la preparación de un examen.
- Utiliza una técnica apropiada de lectura, es decir, échale un vistazo al capítulo unos 20 o 30 minutos antes de leerlo en detalle. Lee los capítulos antes que los presenten en clase.
- Nunca te pierdas una sesión de clase o laboratorio. El material de este curso es acumulativo. Todo lo que se aprende al principio se aplica en capítulos posteriores. Cada uno de los capítulos es un enlace en cadena y una cadena es tan fuerte como lo es su eslabón más débil. Continúa y verás que este curso es divertido; si te atrasas, se hace innecesariamente difícil.
- En este curso, no temas devolver lo que está en el libro que presenta ejercicios completos de muestra para todos los procedimientos, y hay un resumen de fórmulas al final de cada capítulo. Los ejercicios y tablas distinguen entre "datos" (información dada para un problema de investigación) y "cálculos" (lo que debe hacerse para completar el problema). Sigue la forma de estos ejercicios y "presenta el procedimiento" así como la respuesta. De hecho, las respuestas a algunos de los problemas se dan en el apéndice C, de modo que puedas verificar tu avance en casa. Una computadora inerme también puede generar números. La correcta interpretación de la respuesta es lo que es importante y el trabajo detallado es necesario para aprender la lógica que hay detrás de un procedimiento.
- Entrega el trabajo a tiempo. Revisa las tareas devueltas y corrígelas de inmediato.
- Pide ayuda cuando la necesites. No hay preguntas tontas en este curso, pero no preguntar sí que es tonto.
- Acepta el hecho de que este curso es agradable. Un esfuerzo concentrado será recompensado no sólo en términos de obtener una calificación, sino también en términos de aprender valiosos conocimientos en el trabajo.

TABLA 1-3 | Cambio porcentual del número de muertes por SIDA reportadas en la sección 11 de Salud Pública de Alabama por género

Género	Número de muertos por SIDA Datos de 1995	Número de muertos por SIDA Datos de 1996	Cambio en porcentaje (%) de 1995 a 1996
Hombres	43	44	2
Mujeres	6	10	67
Total	49	54	10

Fuente: Datos del Centro de Estadísticas de Salud de Alabama.

Insensatez y falacias estadísticas: el problema de los denominadores pequeños

Debes tener cuidado al interpretar proporciones y porcentajes basados en grupos sumamente pequeños; los números pequeños en la línea base en reportes de cambio de porcentaje son una fuente particular de confusión. La tabla 1-3 presenta un ejemplo de la epidemia del SIDA (Alabama Center for Health Statistics, 2004).

El cambio porcentual se calcula como sigue:

$$\text{Cambio porcentual} = \frac{\#\text{ al tiempo 2} - \#\text{ al tiempo 1}}{\#\text{ al tiempo 1}}(100)$$

La tabla muestra que el *incremento porcentual* de la incidencia de muertes por SIDA fue mucho mayor para mujeres que para hombres entre los dos años. Este tipo de estadísticas se publicaron a menudo como evidencia de que la epidemia estaba extendiéndose en forma mucho más rápida entre mujeres que entre hombres, sugiriendo que el SIDA de pronto se había vuelto una enfermedad "femenina". De hecho, en 1996, sólo 10 nuevos casos de muerte ocurrieron entre mujeres en comparación con 44 entre hombres. El aparentemente "femenino" fenómeno se debió al problema de un denominador pequeño. En tal caso, un buen estadístico simplemente informaría que hubo pocos casos de mujeres para que las comparaciones del cambio en porcentaje fueran significativas.

RESUMEN

1. Las estadísticas son una forma divertida, imaginativa e informativa de ver el mundo empírico. No es sólo un ejercicio de matemáticas. Comprende una cuidadosa observación, medición y análisis, así como poner en forma creativa los resultados para tener un uso práctico y científico.
2. En la imaginación estadística interviene una forma equilibrada de observar el mundo, la capacidad de pensar un problema y mantener un sentido de proporción cuando se comparen evidencias contra nociones preconcebidas. Observa la imagen en lo amplio y requiere de una vista crítica.
3. La estadística para consumidores, diseñada para audiencias públicas, recibe el nombre de estadística descriptiva. El análisis estadístico que comprende probar una teoría científica (prueba de hipótesis) se denomina estadística inferencial.

4. La finalidad de una investigación científica es explicar el mundo empírico. Estas explicaciones toman la forma de una teoría, es decir, ideas organizadas que dan un sentido de comprensión y la capacidad de hacer predicciones. Los científicos están capacitados para ser escépticos y aceptan resultados de un análisis estadístico sólo después de un cuidadoso escrutinio y crítica.
5. La ciencia tiene limitaciones, pues sólo comprende la investigación de fenómenos empíricos. Numerosos argumentos científicos lógicos y basados en datos carecen de apoyo político o del público. Con frecuencia surgen dilemas éticos causados por la investigación científica y crean resistencia a su aplicación.
6. Para la investigación científica, el análisis estadístico comprende la recolección de datos y prueba de hipótesis acerca de las relaciones entre variables independientes y dependientes.
7. Una parte importante del análisis estadístico consiste en controlar el error estadístico.
8. Los siete pasos del proceso de investigación son: especificar la pregunta de investigación, revisar la literatura científica, proponer una teoría y formular las hipótesis, seleccionar un diseño de investigación, recolectar datos, analizar los datos y sacar conclusiones (la etapa cubierta por este curso), así como diseminar los resultados.
9. Fracciones, proporciones, porcentajes y tasas son simplemente formas de medir un sentido de proporción, lo que gana un sentido de equilibrio al comparar una parte contra un todo.
10. Al contar con un denominador común, proporciones, porcentajes y tasas dan una forma de estandarizar una clasificación de observaciones de varios grupos de diferentes tamaños.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

El material de texto que se encuentra en el sitio web *The Statistical Imagination* de Extensiones del capítulo 1, en www.mhhe.com/ritchey2, tiene material adicional para cada capítulo que expone temas del capítulo o contiene técnicas avanzadas. Estos materiales se actualizan periódicamente.

FÓRMULAS PARA EL CAPÍTULO 1

Muestra del trabajo cuando se realizan cálculos:

Símbolo = fórmula = contenido de fórmula = respuesta

Cálculo de una fracción:

$$\text{Fracción} = \frac{\text{numerador}}{\text{denominador}} = \frac{\text{parte}}{\text{todo}}$$

Cálculo de una proporción:

$$\text{Proporción} = p \text{ [del grupo total en una categoría]} = \frac{\text{\# en una categoría}}{\text{\# en un grupo total}} = \text{cociente}$$

Cálculo de un porcentaje:

$$\text{Porcentaje} = \% \text{ [de grupo total en una categoría]} = p\,(100)$$

Cálculo de una tasa:

$$\text{Tasa de ocurrencia} = (p)\ \text{(un número base útil)}$$

Cálculo de cambio de porcentaje:

$$\text{Cambio de porcentaje} = \left(\frac{\text{\# en tiempo 2} - \text{\# en tiempo}}{\text{\# en tiempo 1}}\right)(100)$$

PREGUNTAS PARA EL CAPÍTULO 1

1. Un entrevistador en un estudio entendió mal a un encuestado y anotó incorrectamente la edad. ¿Fue ésta una equivocación o un error estadístico? Explique.
2. Mary Jones se ha preocupado por los desastres naturales, al notar que en un solo año hubo inundaciones en el medio oeste, sequía en el sur y grandes terremotos en el oeste. Ella cree que dichos eventos constituyen la prueba de que el fin del mundo está cerca. Considerando que Mary tiene una imaginación vívida, explique por qué carece de imaginación estadística.
3. En un estudio sobre los *estudiantes de último año* en una prestigiada universidad, se registra el área de especialización de los alumnos (psicología, sociología, química, inglés, arte, etc.) y su año de escolaridad (primero, segundo, tercero, último año). En dicho estudio, ¿cuál de estas mediciones representa una variable y cuál una constante?
4. En un estudio sobre los estudiantes universitarios del último año, se mide su promedio académico y su consumo de alcohol durante el mes anterior. Formula una hipótesis para estas dos variables e indica cuál es la variable independiente y cuál es la variable dependiente. ¿Podrían generalizarse los resultados de *estudiantes del último año* al *conjunto de estudiantes* de la universidad? ¿Por qué sí o por qué no?
5. Para una muestra de personas sin hogar, estás interesado en la relación entre género y tipos de lugares para dormir (donde el sujeto pasó la noche anterior). ¿Cuál es la variable independiente y cuál la dependiente?
6. Bob posee una tienda de libros y computadoras llamada InfoManiacs. Él calculó la proporción de sus ganancias que resultan de vender programas de computadora y obtuvo una respuesta de 2.49. ¿Esto podría ser correcto? Explica.
7. ¿Cuál es la característica esencial de la ciencia que se distingue de otras formas de indagar sobre la naturaleza?
8. Identifica una creencia común que sospechas se trata de un mito. Sugiere qué tipos de datos se recolectarían para descubrir el mito. ¿Cómo se podría aplicar el pensamiento proporcional para desafiar esta falsedad ampliamente sostenida?
9. Para darse una idea de qué tan organizados están los procedimientos científicos, consulta una hemeroteca y hojea varias publicaciones científicas, como *American Sociological Review, American Journal of Sociology, Journal of the American Psychological Association, Journal of Health and Social Behavior, Administrative*

Science Quarterly, Criminology, Social Services Review, American Journal of Psychology, American Political Science Review, Review of Public Administration y *Political Science Quarterly.* Observa la abundancia de tablas estadísticas en estos artículos. Nota también que todos los artículos en los diversos volúmenes tienen encabezados de sección semejantes.

a) Haz una lista de los encabezados de al menos cinco artículos de al menos tres revistas diferentes.

b) Compara estas listas con las siete etapas del proceso de investigación y comenta.

10. Supongamos que, en Estados Unidos, un estado tuvo una tasa de mortalidad infantil de 8.6 muertes por 1 000 nacidos vivos en 1998. En el año 2000, el departamento de salud pública estatal organiza una conferencia llamada Metas 2010, en la que políticos y funcionarios del gobierno buscan mejorar la salud pública en el nuevo milenio. Ellos establecieron una tasa óptima de 6.0 muertes infantiles por 1 000 nacidos vivos para el año 2010. Esta tasa óptima es un ______.

11. En la mayoría de los estados, el límite de velocidad interestatal es de 70 millas por hora. Se toman muestras aleatorias de la velocidad de vehículos, con un radar detector de velocidad y se determina que la velocidad promedio es de 74 millas por hora. El límite de velocidad establecido es un________estadístico, mientras que 74 millas por hora es un________estadístico.

12. Resuelve este viejo acertijo:

Cuando yo iba a St. Ives, conocí a un hombre con siete esposas.

Cada esposa tenía siete bolsas, cada bolsa tenía siete gatas.

Cada gata tenía siete gatitos.

Gatitos, gatas, bolsas y esposas; ¿cuántos iban a St. Ives?

EJERCICIOS PARA EL CAPÍTULO 1

Conjunto de problemas 1A

1A-1. Completa los espacios en blanco de la siguiente tabla (ver apéndice A como repaso).

	Fracción	Proporción	Porcentaje
a)	$\frac{27}{198}$	0.1364	______
b)	$\frac{1}{598}$	______	______
c)	$\frac{36}{12000}$	______	______
d)	______	0.2321	______
e)	______	______	44.63
f)	______	______	91.35

1A-2. Como estudiante de introducción a la estadística, tu profesor te pidió calcularas la precisión de goles de campo pateados durante los juegos de la temporada pasada que se jugaron en el Estadio Universitario. Los pateadores del equipo local anotaron 16 goles de campo en 21 intentos y los de los equipos visitantes anotaron 17 goles de campo en 24 intentos. ¿Cuál equipo tuvo los mejores pateadores de goles de campo, el local o los visitantes? ¿Por qué?

1A-3. Según la U.S. Federal Bureau of Prisons (2003), 19.7 por ciento de internos en prisiones federales están en instalaciones de seguridad mínima, 38.7 están en instalaciones de seguridad baja, 24.7 por ciento están en instalaciones de seguridad media y 10.8 por ciento están en instalaciones de alta seguridad. La población total de internos en prisiones federales en 2003 fue de 145 290. ¿Cuántos internos están en cada una de estas categorías de seguridad?

1A-4. Tú estás interesado en realizar un proyecto de investigación que comprende niveles de logros educativos en Alaska. La U.S. Bureau of the Census (2000) indica que la población en Alaska, de 25 años de edad y mayores, es de 379 556. Completa la tabla siguiente al insertar la proporción (*p*) de esta población que haya completado diversos niveles de educación. Muestra la fórmula general y cálculos para personas con menos del noveno grado de educación.

Menos de noveno grado	*n*	*p*
Noveno a 12o. grado, sin diploma	15 663	
Graduado de preparatoria (o equivalente)	28 619	
Universidad, sin título	105 812	
Graduado adjunto	108 442	
Título de adjunto	27 213	
Con licenciatura	61 196	
Graduado o profesional	32 611	
Totales		

1A-5. La North Atlantic Treaty Organization (NATO) está formada por 19 países miembros. De estas naciones, sólo dos, Estados Unidos y Canadá, están en Norteamérica. ¿Qué proporción de naciones miembros de la NATO *no están* en Norteamérica? ¿Qué porcentaje?

1A-6. Cinco departamentos académicos han sido seleccionados para enviar 50 estudiantes graduados, y 150 pasantes para representar a la universidad en una conferencia nacional sobre liderazgo. No obstante, la afiliación al departamento y el nivel de grado deben estar representados de manera proporcional. Por ejemplo, una proporción de 0.1945 de pasantes provienen del Departamento de Sociología, de modo que .1945 de los representantes vendrán de este departamento. La tabla siguiente muestra números de estudiantes por departamento y nivel de grado. Completa la proporción

(p) y número (#) de estudiantes asistentes en cada celda vacía. Presenta la fórmula general y cálculos para al menos un cálculo.

Departamento	Graduado	p	# asistentes	Pasante	p	# asistentes
Sociología	58			135		
Psicología	69			189		
Historia	50			122		
Antropología	44			118		
Ciencias políticas	48			130		
Totales			50			150

1A-7. Tú estás interesado en examinar el estado civil en el estado de California. La población con 15 años de edad o mayores en California es 26076163 (U.S. Bureau of the Census 2000). Usando los datos siguientes, calcula las tasas por población de 100000 para cada una de las categorías que se muestran en la tabla. Presenta la fórmula general y un ejemplo de cálculos para la categoría de los que nunca han estado casados.

Estado civil	n	Tasa por 100000 hab (>15 años)
Nunca han estado casados	7843907	
Nunca casados, excepto separados	13657201	
Separados	642670	
Viudas	1457818	
Divorciados	2474567	
Totales	26076163	

1A-7. Cockerham, Snead y DeWaal (2002) examinaron el impacto de la ideología socialista y conductas negativas de salud en Rusia. Entre una muestra de 8701 residentes rusos, 4437 indicaron consumo normal de alcohol y 3704 dijeron que fumaban. Además, 3292 de la muestra dijeron que tenían orientación prosocialista y 4868 estaban casados.

a) ¿Qué proporción de la muestra reportó un consumo normal de alcohol?
b) ¿Qué porcentaje de la muestra dijo que *no* fumaba?
c) ¿Qué porcentaje de quienes respondieron dijeron no tener orientación prosocialista?
d) ¿Qué proporción de quienes respondieron dijeron que estaban casados?

Conjunto de problemas 1B

1B-1. Completa los espacios en blanco de la siguiente tabla (ver apéndice A como repaso).

	Fracción	Proporción	Porcentaje (%)
a)	$\frac{51}{207}$	0.2464	____
b)	$\frac{24}{503}$	____	____
c)	$\frac{663}{13200}$	____	____
d)	____	0.0784	____
e)	____	____	38.35

1B-2. Durante la temporada regular de 2003, en el equipo de béisbol de la universidad jugaron dos receptores, David "Plate Guarder" Feinberg y Byron "Face Mask" Taylor. El manejador Smith debe decidir cuál de ellos iniciará en el próximo campeonato de invitación contra un equipo que tiene fama de "robar" bases. Su selección se apoyará en el éxito en la temporada regular para poner fuera a jugadores contrarios que hayan tratado de "robar" bases. David sacó a 17 que trataron de robar base en 48 intentos y Byron sacó a 14 de 32. ¿Cuál receptor empezará? ¿Por qué?

1B-3. El U.S. Federal Bureau of Investigation (FBI) (2002a, 2002b) compila periódicamente estadísticas del Uniform Crime Report para todo tipo de actividad delictiva. Para 2002, las estadísticas del FBI revelaron que 9721 casos de delitos se cometen con violencia y predisposición. De éstos, 44.9 por ciento fueron por odio racial, 21.6 por ciento por odio a etnias, 18.8 por motivos religiosos y 14.3 por estar contra orientación sexual. ¿Cuántos de estos delitos con violencia ocurrieron por cada categoría?

1B-4. Como parte de una investigación de convenios de unidades habitacionales en el estado de Louisiana, debes examinar las edades de propietarios de casas en todo el estado. Hay 1656053 unidades ocupadas en el estado (U.S. Bureau of the Census, 2000). Completa la tabla siguiente para la proporción (p) de propietarios de casas en cada una de las categorías de edades. Demuestra la fórmula general y cálculos para propietarios de casas de edades de 15 a 24 años.

Edades de propietarios de casa	n	p
15 a 24 años	102760	
25 a 34 años	282345	
35 a 44 años	367556	
45 a 54 años	335157	
55 a 64 años	228754	
65 años y más	339481	
Totales	1656053	

1B-5. Un fabricante internacional de automóviles y de maquinaria pesada opera 112 plantas en todo el mundo. Sólo 42 de estas plantas están situadas en el hemisferio oriental. ¿Qué proporción de estas plantas están ubicadas en el hemisferio occidental? ¿Qué porcentaje?

1B-6. Una asociación deportiva va a seleccionar a 100 miembros para asistir a las Olimpiadas de Invierno, mitad hombres y mitad mujeres. Los grupos de edad serán representados de acuerdo con la proporción de la membresía. Por ejemplo, la proporción de hombres en el grupo de edades de 21-30 es 0.0992, de modo que la proporción de hombres que harán el viaje caerá en esa categoría de edades. La tabla siguiente muestra un desglose de la membresía por edad y género. Calcula y anota la proporción (p) y número (# a asistir) para cada grupo de edades para cada género. Demuestra la fórmula general y el cálculo para los hombres del grupo de edades de 21 a 30 años.

Grupo de edad	Hombres	p	# a asistir	Mujeres	p	# a asistir
21-30	49			80		
31-40	170			217		
41-50	169			176		
51-60	84			91		
61+	22			48		
Totales			50			50

1B-7. Alrededor del 20% de la población en Estados Unidos se ve afectada por enfermedades mentales cada año (U.S. Dept. of Health and Human Services, 1999). La población total, según el censo decenal de 2000, es de 281 421 906 (U.S. Bureau of the Census, 2000). La tabla siguiente es una lista de varias enfermedades mentales y su frecuencia estimada en la población general. Calcula los números (n) y tasas de estas enfermedades por 100 000 habitantes en Estados Unidos, y llena las celdas apropiadas de la tabla. Demuestra las fórmulas generales y cálculos para fobia simple.

Enfermedad	Frecuencia estimada (%)	n	Tasa por 100 000 habitantes
Fobia simple	8.3		
Esquizofrenia	1.3		
Mal humor	7.1		
Estrés postraumático	3.6		
Anorexia nerviosa	0.1		

1B-8. En su análisis de estilo de vida de rusos en su salud, Cockerham, Snead y DeWaal (2002) publicaron la siguiente distribución de logros educativos entre una muestra

de 8 657 personas que respondieron. Completa las columnas de proporción y porcentaje de la tabla siguiente. Demuestra la fórmula general y los cálculos para quienes no tuvieron cursos profesionales.

Educación	n	p	(%)
Sin cursos profesionales	2 113		
Cursos profesionales	1 037		
Capacitación profesional sin educación secundaria	713		
Capacitación profesional con educación secundaria	1 154		
Escuela técnica	1 854		
Universidad	1 700		
Escuela de graduados	86		
Totales	8 657		

Conjunto de problemas 1C

1C-1. Llena los espacios en blanco de la tabla siguiente (ver apéndice A como repaso).

	Fracción	Proporción	Porcentaje (%)
a)	____	____	60.46
b)	____	0.2736	____
c)	____	____	94.32
d)	$\frac{1922}{8998}$	____	____
e)	$\frac{163}{7231}$	____	____

1C-2. La Asociación de Estudiantes Graduados (GSA) patrocinó una función nocturna de cine de estreno y de boliche. Ambos eventos recibieron publicidad, el cine en medios impresos y el de boliche en la radio del campus. La GSA proyectó una asistencia de 200 estudiantes para el cine y de 150 para el boliche, pero se presentaron 92 estudiantes para el cine y sólo 84 para el boliche. Con base en la proporción de asistencia proyectada, ¿qué medio publicitario pareció ser una forma más eficaz de anunciar eventos de la GSA?

1C-3. Según la Federal Bureau of Investigation (FBI) (2002b) hubo un total de 11 451 delitos con violencia publicados en 2001. De éstos, 67.8 por ciento fueron delitos contra personas y 31.5 por ciento fueron delitos contra propiedades. De los delitos indicados contra personas, 55.9 por ciento fueron actos de intimidación. De los delitos contra propiedades, 83.7 por ciento se clasificaron como destrucción, daños o vandalismo.

a) ¿Cuántos delitos con violencia contra personas se cometieron?

b) ¿Cuántos delitos con violencia contra propiedades se cometieron?

c) ¿Cuántos delitos con violencia contra personas se consideraron actos de intimidación?

d) ¿Cuántos delitos con violencia contra propiedades se consideraron actos de destrucción/daños/vandalismo?

1C-4. Scott, Sam y Sid, tres amigos que empacan comestibles en un mercado local, decidieron juntar sus propinas una tarde para comprar un regalo para su amiga Cindy, quien convalecía en un hospital. Scott cooperó con $15, Sam con $12 y Sid con $10. ¿Con qué proporción de dinero contribuyó cada uno para el regalo?

1C-5. Te interesa el fenómeno de las placas de matrícula personalizadas, en las que el dueño lleva su nombre o una frase. En una muestra aleatoria de 341 placas, encuentras que 73 son personalizadas. ¿Qué proporción de placas *no* son personalizadas? ¿Qué porcentaje?

1C-6. El alumnado de una gran universidad ha de elegir representantes ante la Asociación de Estudiantes Graduados (GSA) y la Asociación de Estudiantes Pasantes (USA). Los asientos disponibles se asignan en proporción a la inscripción de estudiantes en cada departamento. Por ejemplo, si 10 por ciento de los pasantes están inscritos en el departamento de biología, entonces 10 por ciento de los representantes de la USA vendrá de estudiantes de biología. La tabla siguiente presenta la inscripción de estudiantes para departamentos académicos seleccionados, que juntos llenarán 22 asientos de representantes de la GSA y 62 de la USA. Completa la tabla para mostrar la proporción (p) y el número (#) de representantes de cada organización estudiantil para cada departamento. Demuestra la fórmula general y cálculos para la proporción y número de representantes ante la GSA para biología.

Departamento	Graduado	p	# a la GSA	Pasante	p	# a la USA
Biología	43			119		
Química	33			98		
Ciencias de la computación	45			122		
Matemáticas	29			88		
Física	28			76		
Totales			22			62

1C-7. Completa la tabla siguiente al calcular la tasa de internamiento (en prisiones y hospitales de salud mental) por 100 000 habitantes. Demuestra la fórmula general y el cálculo para Anderson, Indiana.

Ciudad	Población	Número de personas internadas	Tasa por 100 000 habitantes
Anderson, Indiana	130669	3981	
Bellingham, Washington	127780	1602	
Duluth, Minnesota	239971	4610	
Modesto, California	370522	4456	

1C-8. Turner (1995) investigó los efectos del desempleo. Se comunicó con 5 612 personas, a quienes consideró elegibles para el estudio porque habían estado desempleadas por lo menos una vez, desde que se incorporaron a la fuerza laboral. De estas personas elegibles, realmente entrevistó a 3 617, entre las cuales 1 252 se integraron en un estudio a largo plazo. En el grupo de estudio a largo plazo, 154 estuvieron "recientemente desempleadas", pues perdieron sus trabajos en los últimos tres años; de ellas, 45 seguían desempleadas.

a) ¿Qué proporción de sujetos elegibles fue entrevistada?

b) ¿Qué porcentaje de los que fueron entrevistados *no* participó en el estudio a largo plazo?

c) De los que sí participaron, ¿qué porcentaje quedó desempleado en los últimos tres años?

d) ¿Qué porcentaje de los recientemente desempleados regresó a trabajar?

Conjunto de problemas 1D

1D-1. Llena los espacios en blanco para la tabla siguiente (ver apéndice A como repaso).

	Fracción	Proporción	Porcentaje (%)
a)	______	______	29.67
b)	______	0.7243	______
c)	______	______	87.63
d)	$\frac{2485}{6773}$	______	______
e)	$\frac{9228}{11621}$	______	______
f)	______	0.6827	______

1D-2. Según la U.S. Bureau of the Census (2000), la población de Alabama es 4 447 100, la población de Oregon es 3 421 399, y la población de Texas es 20 851 820. El estado con la proporción más elevada de personas de 65 años de edad o más recibirá fondos federales para apoyar programas para adultos mayores. El número de ciudadanos de más de 65 años de edad es 579 798 en Alabama, 438 177 en Oregon y 2 072 532 en Texas. ¿Cuál estado recibirá los fondos federales?

1D-3. Según la U.S. Federal Bureau of Prisons (2003), en la primera mitad de 2003, un total de 153 205 internos fueron sentenciados a prisión. De éstos, 55.6 por ciento fueron sentenciados por delitos por drogas; 11.3 por ciento por armas, explosivos o incendio premeditado; 10.7 por ciento por violaciones de inmigración; 6.7 por ciento por robo, y 15.7 por ciento por otros delitos. ¿Cuántos internos fueron sentenciados por cada uno de estos delitos?

1D-4. En una conversación informal después de clase, Jimena y Ana descubren que tienen un hábito común. Para aliviar el estrés, no pierden de vista la frecuencia con la que pueden meter un papel en la lata de basura cuando lo tiran. Jimena se jacta de que, de 250 tiros, acertó 128 veces en la lata. En 265 tiros, Ana hizo 157 "canastas". ¿Quién es mejor lanzadora? ¿Por qué?

1D-5. Para cumplir con las directrices de la Environmental Protection Agency (EPA), la cantidad de partículas de materia en el aire de ciudades puede rebasar las 69 partes por millón sólo 15 por ciento de los días del año, sin incurrir en contingencias ambientales. ¿Cuántos días es esto?

1D-6. Una empresa con oficinas en Los Ángeles y Nueva York capacitará a sus asistentes administrativos en un nuevo programa de software. Cada sesión de capacitación admite hasta 40 personas de cada ciudad. Para que los empleados de cada departamento reciban capacitación inmediata, se seleccionan participantes con base en cuotas sobre el número de empleados de un departamento. Por ejemplo, si 10 por ciento de los asistentes administrativos están en un departamento particular en Los Ángeles, entonces 10 por ciento de los 40 alumnos provendrá de ese departamento. La tabla siguiente da un desglose de membresía por departamento y ciudad. Escribe la proporción (p) y el número (# de asistentes) por cada departamento y ciudad. Demuestra la fórmula y cálculo para el departamento de personal de Los Ángeles.

Departamento	Los Ángeles	p	# de asistentes	Nueva York	p	# de asistentes
Personal	36			43		
Marketing	81			93		
Embarques	65			78		
Administración	24			31		
Contabilidad	25			38		
Totales			40			40

1D-7. Completa la tabla siguiente, calculando la tasa de internamiento (en prisiones y hospitales mentales) por 100 000 habitantes. Demuestra la fórmula general y el cálculo para Bakersfield, California.

Ciudad	Población	Número de personas internadas	Tasa por 100 000 habitantes
Bakersfield, California	543 477	10 808	
Burlington, Carolina del Norte	108 213	1 158	
Great Fall, Montana	77 691	787	
Poughkeepsie, Nueva York	259 462	11 082	

1D-8. Turner (1995) investigó los efectos del desempleo. Él se comunicó con 5 612 personas, a quienes consideró elegibles para el estudio porque habían estado desempleadas por lo menos una vez desde que se incorporaron a la fuerza laboral. De estas personas elegibles, realmente se entrevistó a 3 617, entre las cuales 1 252 se identificaron para el estudio a largo plazo. Imagina que de las 1 252 personas del grupo de estudio a largo plazo, 732 eran hombres y 520 eran mujeres. De las 154 recientemente desempleadas, 80 eran hombres y 74 mujeres. Entre las 45 desempleadas, 25 eran hombres y 20 eran mujeres.

a) En este grupo de estudio a largo plazo, ¿qué género tuvo mayor proporción de desempleo reciente?

b) Entre los recientemente desempleados, ¿fueron hombres o mujeres los más afortunados al regresar a la fuerza de trabajo?

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 1

Si en tu grupo el profesor asigna ejercicios en computadora, el texto viene con un disco compacto que contiene el programa de cómputo SPSS o Paquete Estadístico para las Ciencias Sociales para Windows, Versión del Estudiante (SPSS por sus siglas en inglés). El programa está restringido en tiempo. Funcionará 13 meses a partir de la fecha en que lo cargues en tu computadora. Este programa incluye material didáctico y buenos menús de ayuda que facilitan su operación.

El apéndice D de este texto, "Guía del SPSS para Windows", contiene un repaso conciso de las operaciones básicas del paquete de software, así como instrucciones de capítulo por capítulo. Además, el sitio web *The Statistical Imagination* en www.mhhe.com/ritchey2 contiene ejercicios específicos de cada capítulo y más instrucciones sobre procedimientos para ejecutar e interpretar los resultados. Usa el apéndice D para iniciarte.

El Paquete Estadístico para las Ciencias Sociales para Windows, *Versión del estudiante*, es más adecuado para aprender estadísticas básicas, pero si quieres dirigir tu propia investigación, quizá desees tener acceso a todo el sistema base completo de la versión regular (no del estudiante) de *SPSS para Windows*, o el *SPSS for Windows Graduate Pack*, que existe en universidades con licencia de versión completa. Estas versiones del *SPSS* tienen varias ventajas, incluyendo una ventana "Editor de Sintaxis" que pega y guarda comandos con el ratón para su uso posterior. Es más, en la versión regular del *SPSS* no hay limitaciones en el número de variables o el tamaño de las muestras de los archivos de datos.

Todos los ejercicios de aplicaciones en computadora están en el sitio web *The Statistical Imagination.* El sitio contiene ejercicios de computadora para cada capítulo, que se pueden descargar fácilmente, además de instrucciones detalladas sobre los procedimientos estadísticos empleados en un capítulo, conjuntos de datos con los cuales se realizan ejercicios del capítulo, así como los códigos que describen las variables en cada uno de los conjuntos de datos.

Se han hecho pequeñas modificaciones a conjuntos de datos del sitio web *The Statistical Imagination* para facilitar la instrucción y, por tanto, no son suficientes para verdaderos fines de investigación. El lector podrá solicitar conjuntos de datos originales no modificados a las fuentes que se detallan en las guías de codificación del sitio web. Hay otros conjuntos de datos de algunas dependencias gubernamentales, con acceso a internet, así como fundaciones de investigación como el Inter University Consortium for Political and Social Research (ICPSR) y el National Opinion Research Center (NORC). El sitio web contiene enlaces a estas fuentes y a otras, así como para sitios relacionados con la estadística que revelan las numerosas aplicaciones del trabajo de estadística. Estos sitios incluyen reportes de población de la U.S. Bureau of the Census, datos de delincuencia del U.S. Department of Justice, datos estadísticos abstractos, sitios de estadísticas de deportes, interesantes encuestas de marketing y juegos de cuestionarios.

Para iniciar con el SPSS

Empieza por ver el apéndice D y leer las instrucciones del capítulo 1. Para descargar el *SPSS para Windows, Versión del Estudiante* en tu computadora, inserta el disco compacto del *SPSS* en la unidad de disco compacto. Haz clic en "Install *SPSS for Windows Student Version*" y sigue las instrucciones. El apéndice D contiene información adicional para guardar un icono del *SPSS for Windows* en el escritorio de tu computadora y abrir el programa para usarlo. Una vez que abras el *SPSS* puedes seguir el material didáctico para familiarizarte con ventanas básicas, iconos y menús.

Ejercicios del capítulo 1

Entra al sitio web *The Statistical Imagination* www.mhhe.com/ritchey2 para tener acceso a ejercicios de aplicaciones en computadora del capítulo 1. Este primer ejercicio comprende una orientación al software estadístico de *SPSS for Windows* con instrucciones sobre cómo descargar y recuperar archivos de datos.

NOTAS

1. Cuando alguien hace una aseveración objetiva acerca de un objeto (o persona o situación), la frase describe una característica que es verdaderamente parte del objeto, por ejemplo, la frase de "la luz del semáforo es roja". Cuando alguien hace una aseveración subjetiva, ésta en realidad describe una característica del observador "sujeto" más que del objeto. Las aseveraciones subjetivas, por tanto, son puntos de vista personales u opiniones que reflejan las inclinaciones, distorsiones, opiniones personales o prejuicios de la persona que hace la aseveración. Por ejemplo, alguien ciego a los colores podría decir "la luz del semáforo es gris". Lo "gris" no es parte del semáforo, sólo es la percepción del observador sujeto.
2. La abreviatura latina *i.e.* significa "es decir" (*id est*); *e.g.* significa "por ejemplo" (*exempli gratia*).

3. El término *independiente* proviene de ciencias de laboratorio, donde las variables pronosticadoras se manipulan independientemente de los resultados. Por ejemplo, en un estudio de los efectos de una droga en ratas, la droga se administra a algunas ratas (el grupo experimental), en tanto que un placebo (o droga falsa) se da a un grupo comparado de ratas (el grupo de control). La elección de cuáles ratas se asignan a cada grupo se hace independientemente de medir cuáles ratas mejoran.
4. El término *teoría* se usa con frecuencia para representar una idea no corroborada, por ejemplo: "eso es sólo una teoría". Éste es un uso no científico del término. Las teorías científicas están basadas *no* en conjeturas u opiniones, sino en un análisis objetivo de datos reunidos con todo cuidado.

CAPÍTULO

2

Organización de los datos para reducir al mínimo el error estadístico

RESUMEN DEL CAPÍTULO

Introducción

Así sea realizada para la investigación científica, la mercadotecnia de un producto, un pronóstico meteorológico o una simple apuesta, la predicción del futuro es un pasatiempo común. Los científicos realizan predicciones empíricas para probar la exactitud de sus ideas. Por ejemplo, ¿cuál es la probabilidad de que seas víctima de un delito en tu área de trabajo? Madriz (1996) encontró tres factores de predicción basados en la idea de que el riesgo de ser víctima de un delito puede reducirse por medio del estudio cuidadoso de actividades rutinarias. Un primer factor de riesgo es la exposición, o vulnerabilidad circunstancial, como

trabajar solo por la noche en una tienda. Un segundo factor es la proximidad a delincuentes potenciales, como trabajar en una tienda ubicada en una zona con un alto índice delictivo. Un tercero es el atractivo del objetivo, es decir, desear la propiedad de una víctima, por ejemplo: tener grandes cantidades de dinero disponibles.

Si el dueño de una tienda pusiera a sus empleados en riesgo innecesario, un robo o asesinato no sería un suceso aleatorio o una equivocación; sería un error. En el capítulo 1 notamos que los **errores** son *grados conocidos de imprecisión*. Conocer la relación entre las circunstancias y la probabilidad de un robo permite realizar mediciones preventivas que reduzcan las oportunidades calculadas para que los "errores" ocurran. Las mediciones para la reducción del riesgo podrían incluir tener al menos dos empleados presentes, cerrar a las 11:00 p.m., ubicarse en un lugar de tránsito denso, manejar pequeñas cantidades de dinero disponible e instalar sistemas de alarma. La reducción del error depende de la comprensión de las relaciones que predicen entre variables.

Como brevemente anotamos en el capítulo 1, la estadística trata sobre la comprensión exacta y el **control del error estadístico**, los *grados conocidos de imprecisión en los procedimientos utilizados para reunir y procesar información.* Los errores no son equivocaciones. Los errores son cantidades conocidas de imprecisión que pueden calcularse y reducirse por una selección cuidadosa e informada de diseños de muestreo, instrumentos de medición y fórmulas estadísticas.

> **Error estadístico** Grado conocido de imprecisión en los procedimientos utilizados para reunir y procesar información.

El análisis estadístico comúnmente implica un muestreo: analizar sólo una pequeña parte del grupo que se estudia. Por ejemplo, para aprender acerca de todas las tiendas pequeñas, podríamos estudiar una muestra de 20 tiendas. ¿Pueden los datos de la muestra de 20 tiendas revelar con precisión cómo funcionan todas ellas? La investigación también comprende la observación y la medición. ¿Podemos suponer que nuestras mediciones son completamente exactas? El muestreo y la medición son dos fuentes potenciales de error al obtener conclusiones en la investigación. El **error de muestreo** representa la *inexactitud en las predicciones sobre una población que resulta del hecho de que no observamos a todos los sujetos de la población.* El **error de medición** es *la inexactitud en la investigación que se deriva de instrumentos de medición imprecisos, de las dificultades en la clasificación de las observaciones y de la necesidad de redondear los números.* Después de estudiar cada uno de estos tipos de error, mostraremos cómo están relacionados.

Control del error de muestreo

Analizar significa escoger algo y examinarlo con detalle de manera organizada. Al realizar trabajo estadístico, analizamos grupos de personas, objetos o acontecimientos y medimos variables para obtener promedios, tendencias o porcentajes. La *medición de una sola persona*, por ejemplo, registrar como 19 años de edad de María López, no proporciona un estadístico; simplemente es una **observación**. Sin embargo, determinar que la edad promedio de un grupo de 30 estudiantes es de 19.5 años es calcular un estadístico con base en un conjunto de observaciones. El campo de la estadística implica el **resumen de cálculos** de numerosas observaciones, es decir, *la adición de un grupo de mediciones.* Nuestros intereses se enfocan en observar muchos casos, recabar información precisa sobre ellos y hacer declaraciones concisas sobre el grupo, no sobre los individuos.

El grupo de sujetos que observamos a menudo es bastante pequeño. Nuestro propósito es estudiar el número pequeño de sujetos para obtener conclusiones sobre la población más grande a la cual esos sujetos pertenecen. Estudiar cada caso de un fenómeno es impráctico, costoso e innecesario. Por ejemplo, no tenemos que encuestar a cada votante probable para determinar el apoyo al candidato A. En cambio, podemos encuestar una muestra representativa de votantes probables, quizá 500. Este grupo más pequeño se llama muestra, mientras el grupo más grande, completo, al que pertenece se denomina población o universo.

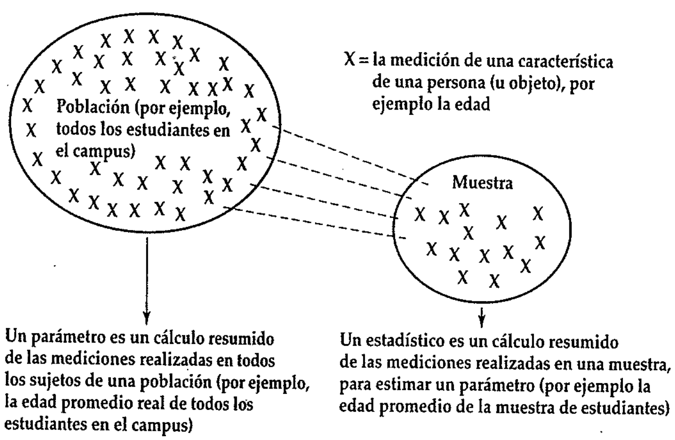
La figura 2-1 ejemplifica la noción del muestreo. La **población** (o universo) es *un grupo grande de personas de interés particular que deseamos estudiar y entender.* Con frecuencia las poblaciones estudiadas incluyen a las personas de un país, estado o comunidad; los presos en las instalaciones correccionales de un estado; los estudiantes actualmente inscritos en una universidad; las familias con hijos en edad escolar; los pacientes de un hospital; los jefes de cocina en restaurantes de la ciudad de Nueva York, y los ejecutivos de corporaciones. Una **muestra** es *un subgrupo pequeño de la población; la muestra se observa y se mide y después se utiliza para obtener conclusiones sobre la población.*

Población (o universo) Grupo grande de personas de interés particular que deseamos estudiar y entender.

Muestra Subgrupo pequeño de la población; la muestra se observa y se mide y después se utiliza para obtener conclusiones sobre la población.

El muestreo es algo que hacemos todo el tiempo. Probamos una cucharada (una muestra) para decidir si agregamos más picante en polvo a la olla (la población). Para explorar una carrera académica, por ejemplo sociología, tomaríamos uno o dos cursos (una muestra) para determinar si el universo de ideas y actividades de la sociología nos agrada. Una primera cita con alguien es un muestreo de la personalidad del individuo, una primera exposición al universo de sus tendencias de conducta y actitudes. El muestreo es una conducta humana común y eficaz.

FIGURA 2-1
Relación de una población (universo) de mediciones con una muestra de mediciones



Nuestro interés, sin embargo, no está en la muestra por sí misma. En cambio, queremos aprender sobre la población entera. Para adquirir información completamente correcta respecto de una población entera, mediríamos *todos* sus miembros y resumiríamos los resultados en términos matemáticos, reportando porcentajes, tasas y promedios. *Al cálculo resumido de mediciones realizadas en todos los sujetos en una población* se le llama **parámetro**. Por ejemplo, el promedio de edad de presos en la prisión Sharpwire es un parámetro. El porcentaje de ejecutivos mujeres en la Menrule Plastics Corporation es un parámetro. Por desgracia, la mayoría de las poblaciones son tan grandes que no podemos invertir el tiempo y los recursos necesarios para medir a todos los miembros. Por ejemplo, sería absurdo medir las estaturas de todos los adultos en un país. A causa de los altos costos para medir a cada sujeto en una población, los verdaderos valores de los parámetros comúnmente son desconocidos.

Por fortuna, el muestreo nos permite *estimar* parámetros con precisión. Con las muestras calculamos estadísticos en vez de parámetros. **Un estadístico** es *un cálculo resumido de mediciones realizadas en una muestra para estimar un parámetro poblacional.* Por ejemplo, en una muestra de 800 republicanos registrados en Nueva Jersey, encontraríamos que 74 por ciento apoyan al gobernador. Este porcentaje constituye un estadístico: sólo una estimación del verdadero apoyo al gobernador. Una muestra y las estadísticas calculadas acerca de ésta son simples herramientas para obtener *conclusiones sobre una población en general,* la población no como un todo. Estas conclusiones, si se realizan siguiendo procedimientos estadísticos adecuados, se llaman generalizaciones estadísticas.

Parámetro Cálculo resumido de mediciones realizadas en todos los sujetos de una población.

Estadístico Cálculo resumido de mediciones realizadas en una muestra para estimar un parámetro poblacional.

Nunca debemos perder de vista el hecho de que la población es lo que nos preocupa. Por ejemplo, una muestra de votantes en una "encuesta de salida" (tomada cuando las personas salen de las casillas) sugeriría que el candidato A es el ganador. Ésta, sin embargo, es una estimación, una aproximación del nivel de apoyo real. El verdadero ganador sólo se conocerá después que se cuenten todos los votos, es decir, cuando la población entera de votantes haya sido medida.

Una manera de recordar que una muestra sólo proporciona estimaciones es comparar los resultados de varias muestras de la misma población. Si un profesor de estadística mandara a cada uno de los 30 miembros del grupo a reunir una muestra de 10 compañeros estudiantes y estimara el promedio de edad de los estudiantes, cada miembro del grupo obtendría un resultado ligeramente diferente. (Si no estás convencido, consigue tú mismo dos muestras.) Esta variabilidad en los resultados de las muestras sólo refleja el hecho de que el estadístico, en una muestra única, es sólo una estimación del verdadero parámetro de la población.

Entonces, ¿cómo confiaremos en los resultados de una sola muestra? La respuesta a esta pregunta implica una noticia buena y una mala; la mala es que el estadista debe reconocer que las conclusiones de una muestra no son totalmente correctas, dado que estos estadísticos son sólo estimaciones de parámetros; la buena noticia es que los procedimientos estadísticos y la lógica de la teoría de probabilidad permiten a los estadistas especificar un grado de error conocido en las predicciones y, por consiguiente, estipular el grado de confianza que tendríamos en una conclusión basada en estadísticos. En pocas palabras, aun cuando las estimaciones estadísticas no son perfectas, sabemos qué tan cerca están de la perfección.

Estimación estadística cuidadosa contra adivinación o estimación apresurada

La imaginación estadística enfatiza el entendimiento de un detalle en su contexto apropiado, teniendo cuidado de no emitir conclusiones simplistas o fantásticas. La estimación estadística es diferente del sentido común de la "adivinación o estimación apresurada", que a menudo es tendencioso. Una **estimación estadística** es *el informe de una medida de resumen basada en el muestreo sistemático y en mediciones precisas e informadas, con grados conocidos de error y confianza.* Una adivinación o **estimación apresurada** es *un informe de una medición de resumen basada en las experiencias personales limitadas y comúnmente subjetivas, evidencia anecdótica u observaciones informales apresuradas.*

La adivinación podría ocurrir cuando un reportero de noticias elige al candidato A como el seguro ganador porque el reporte de las encuestas de salida lo apoya con 52 por ciento de probables votantes. En contraste, tomando en cuenta el tamaño de la muestra, un estadístico sería más cauto y destacaría el hecho de que 52 por ciento significa 52 más y menos 5 puntos porcentuales; por consiguiente, el apoyo se encuentra entre 47 y 57 por ciento. La victoria del candidato A no está asegurada porque el apoyo podría ser *de sólo* 47 por ciento. Además, el estadista mantiene un grado de confianza para la estimación de 95 por ciento. (No podemos exigir 100 por ciento de confianza hasta que todos los votos se contabilicen.) La estimación estadística cuidadosa es diferente incluso de una buena suposición. El estadista difiere de otros "pronosticadores" en dos maneras importantes: el estadista (1) controla y maneja el grado de error en las estadísticas reportadas y (2) señala de forma precisa la confianza en sus conclusiones.

Un tipo particularmente insidioso de la estimación apresurada es un **estereotipo** prejuicioso, es decir, *una generalización falsa que implica que todos los individuos de una categoría comparten ciertas características, normalmente indeseables.* Existe un estereotipo racista, por ejemplo, en creer que los afroamericanos son ignorantes, perezosos o inmorales para mantener a sus familias y que ésta es la causa de pobreza en Estados Unidos. De hecho, casi 7 de cada 10 estadounidenses pobres son blancos y la mayoría de la gente pobre tiene empleo. Las estimaciones apresuradas a menudo se guían por sentimientos que refuerzan estereotipos y sentimientos como odio, temor y superioridad. En contraste, las generalizaciones estadísticas se interpretan con cautela y dentro del contexto más grande de comprobación científica con sus resguardos contra la subjetividad. La tabla 2-1 compara las estimaciones apresuradas con las estimaciones estadísticas.

TABLA 2-1 | "Estimación apresurada" del sentido común contra estimación estadística

Estimación apresurada del sentido común	Estimación estadística
La idea se basa en experiencias personales limitadas y comúnmente subjetivas, evidencia anecdótica u observaciones apresuradas.	La idea se basa en muestreo sistemático y en medición.
Produce conjeturas y conclusiones equivocadas.	Produce estimaciones confiables con grados conocidos de error y confianza.
Genera y refuerza estereotipos.	Produce generalizaciones estadísticas.
Usualmente es un asunto de opinión.	Usualmente es un asunto de hecho.

Error de muestreo y su manejo con la teoría de la probabilidad

Como la única manera de conocer un parámetro verdadero es mediante el sondeo de la población entera, cada estadístico calculado de una muestra es una estimación. Por casualidad, los estadísticos de algunas muestras están más cerca del valor del parámetro verdadero que otros. La **teoría de la probabilidad** (capítulo 6) consiste en *el análisis y la comprensión de las probabilidades de los acontecimientos.* Nos brinda un conjunto de reglas para determinar la exactitud de los estadísticos de la muestra y calcular los grados de confianza que tenemos en las conclusiones sobre una población.

Para manejar exitosamente el error de muestreo debemos concentrarnos en sus fuentes específicas: el tamaño y la representatividad de la muestra. El **tamaño de la muestra** se refiere al *número de casos u observaciones que constituyen una muestra: el número de personas u objetos observados.* De manera general, cuanto mayor sea la muestra, menor será el rango del error. Suponga que un investigador envía a dos asistentes para determinar la edad promedio de todo el alumnado. Uno les preguntó sus edades a 3 estudiantes, mientras que el segundo les preguntó a 1 000. La intuición nos lleva a tener mayor confianza en los resultados de la muestra mayor, porque la muestra más pequeña pudiera reunir más fácilmente sólo a estudiantes jóvenes o sólo mayores. En un capítulo posterior aprenderemos a calcular e informar estadísticos con un "intervalo de confianza" con una cantidad exacta de error para cualquier tamaño de muestra dada. Con una muestra de 1 000 encontraríamos que la edad promedio en el campus es de 22.4 años, más o menos 0.3 años, lo que sugiere que el promedio de edad se ubica entre 22.7 años (esto es 22.4 + 0.3) y 22.1 años (es decir, 22.4 – 0.3). El cálculo de "más menos algún error de muestreo" está basado en probabilidades matemáticas de la teoría de la probabilidad.

La teoría de la probabilidad también nos permite señalar exactamente qué tan a menudo un estadístico predecirá el parámetro incorrectamente, es decir, qué tan a menudo los errores pueden causar una respuesta incorrecta. Por ejemplo, podemos advertir que 5 por ciento de las veces nuestros procedimientos generan una conclusión falsa. Al especificar este nivel de error, sin embargo, estamos percibiendo también nuestro nivel de confianza. Si nuestra estimación es incorrecta sólo el 5 por ciento de las veces, entonces es correcta el 95 por ciento del total; así, tenemos 95 por ciento de certeza.

Un segundo factor que afecta la exactitud del muestreo es hasta qué punto todos los segmentos de una población realmente están incluidos en la muestra: la representatividad de la muestra. Una **muestra representativa** es *aquella en la que todos los segmentos de la población están incluidos en la muestra en sus proporciones correctas respecto a la población.* Por ejemplo, si una población del campus realmente es 54 por ciento hombres y 46 por ciento mujeres, una muestra representativa tendrá que acercarse a esos porcentajes.

> **Muestra representativa** Muestra en la que todos los segmentos de la población están incluidos en la muestra en sus proporciones correctas respecto a la población.

Una **muestra no representativa** es aquella *en la que algunos segmentos de la población están representados en exceso o con defecto en la muestra.* Éste es un tipo riesgoso de error de muestreo porque puede generar resultados totalmente engañosos. Supongamos, por ejemplo, que la administración del campus desea encuestar a estudiantes sobre su apoyo para ampliar el estadio de fútbol. Los voluntarios de la asociación estudiantil de enfermería llevan a cabo la encuesta y se les pide registrar el voto de cada décimo estudiante; en cambio, ellos registran los votos de cada décimo estudiante que sale del edificio de enfermería. Sin sorpresa, los resultados muestran que sólo 23 por ciento de estudiantes están a favor de la ampliación. ¿Por qué? Porque los miembros de la asociación en realidad encuestaron a la población de estudiantes de enfermería, que en su gran mayoría son mujeres y, por tanto, no es representativa del campus en conjunto. Diríamos que esta muestra está *sesgada* por una porción muy desproporcionada de mujeres. Tal muestra no representativa permitió que un segmento de la población tuviera más "votos" de lo que les correspondía sobre una cuestión.

Hay una variedad de diseños de muestreo, pero uno de los más empleados es la muestra aleatoria simple. Una **muestra aleatoria simple** es *aquella en la cual cada persona (u objeto) de la población tiene la misma oportunidad de ser seleccionado(a) para formar parte de la muestra.* (En términos técnicos, decimos que todos en la población tienen *una misma probabilidad de inclusión* en la muestra.) Este diseño es como una rifa o lotería, en la que cada persona de la población sólo entraría una vez. Una muestra aleatoria de tamaño suficiente producirá normalmente una muestra representativa.

> **Muestra aleatoria simple** Muestra en la cual cada persona (u objeto) de la población tiene la misma oportunidad de ser seleccionado(a) para formar parte de la muestra.

Control del error de medición

Además de evitar los errores de muestreo, debemos definir con precisión cómo se harán las mediciones y cómo se codificarán las respuestas una vez que se recopilen los datos. *El conjunto de procedimientos u operaciones para medir una variable* se llama **definición operacional**. Por ejemplo, supongamos que utilizamos datos del censo de Estados Unidos para dirigir un estudio sobre la pobreza urbana, con una muestra de 300 ciudades. Existen varias formas de *operacionalizar* una medida de la pobreza. El desafío consiste en seleccionar la manera que represente con mayor precisión cuántos hogares en una ciudad están habitados por familias pobres. Una medida es el porcentaje de hogares que reciben vales de alimentos. Una segunda es la tasa de desempleo en la ciudad. Una tercera sería el porcentaje de hogares que viven abajo del nivel de pobreza que se define en el ámbito federal (ingreso específico justo para el tamaño de la familia). De hecho, la tercera opción generalmente se reconoce como la mejor aproximación hacia la pobreza para una comunidad, y por ello la escogeríamos como nuestra definición operacional. Una guía eficaz para la elección de una definición operacional consiste en identificar los tipos comunes de error de medición y hacer todo lo posible para minimizarlos.

Niveles de medición: selección cuidadosa de los procedimientos estadísticos

Medición

La **medición** es *la asignación de símbolos, tanto nombres como números, a las diferencias que observamos en las cualidades o cantidades de una variable. La medición de un sujeto particular de la muestra en una variable* es la **puntuación** del sujeto para esa variable o, para usar terminología computacional, un código. Supongamos por un momento que la clase de estadística constituye una muestra. Podríamos registrar las variables de edad, semestre, género, promedio y raza. Para una estudiante, Juana, estas puntuaciones son 20 en edad, primer ingreso en semestre, femenino en género, 3.25 en promedio y blanca en raza; para otro,

Rubén, las codificaciones respectivas son 19 años, estudiante de segundo semestre, masculino, 3.48 en promedio y afroamericano. Utilizaremos los términos *puntuación* y *código* indistintamente. El *valor* de una puntuación es su cantidad.

Como esta simple ejemplificación revela, no todas las variables se miden de la misma forma. Algunas se registran con nombres o categorías que identifican diferencias en tipo o calidad, como afroamericano y blanco para la variable raza. Otras variables permiten distinciones de grado o distancia entre cantidades, como las variables de edad y promedio. Estas variables tienen una **unidad de medición**, un *intervalo determinado o distancia entre las cantidades de las variables.* Las anotamos numéricamente, como las marcas numeradas en una regla como en una cinta métrica. La unidad de medición para una escala de temperatura es un grado; para el peso, un kilogramo; para la altura, un centímetro, y así sucesivamente.

Para comprender las finas distinciones entre las propiedades de medición de las variables, usamos un esquema llamado niveles de medición. El **nivel de medición de una variable** *identifica las propiedades de medición, las cuales determinan el tipo de operaciones matemáticas (suma, multiplicación, etc.) que pueden usarse apropiadamente con dicho nivel, así como las fórmulas estadísticas que se utilizan para probar las hipótesis teóricas.* Estos niveles se llaman nominal, ordinal, de intervalo y de razón. El nivel de medición de una variable es una guía importante para seleccionar fórmulas estadísticas y procedimientos.

> **Nivel de medición de una variable** Identifica las propiedades de medición de la variable y determina el tipo de operaciones matemáticas (suma, multiplicación, etc.) que puede usarse apropiadamente con dicho nivel, así como las fórmulas estadísticas que utiliza para probar las hipótesis teóricas.

Variables nominales

Las **variables nominales** son aquellas en las que los códigos *sólo indican una diferencia en categoría, clase, calidad o tipo.* La palabra *nominal* viene del vocablo latín para *nombre* y estas variables tienen categorías de nombre. Algunos ejemplos incluyen lugar de nacimiento (Chicago, Atlanta, Monterrey, etc.), sabor favorito de helado (vanilla, chocolate, galletas y crema, etc.), marca de automóvil (Ford, Lexus, Pontiac, etc.) y carrera académica (psicología, química, ingeniería eléctrica).

Las variables nominales no admiten puntuaciones numéricas ordenadas significativamente. No obstante, gracias a que las computadoras procesan números con mayor eficacia, a veces numeramos las categorías de estas variables en códigos computacionales. Por ejemplo, para la variable *género* asignaríamos los códigos como 0 = hombre y 1 = mujer. La elección de números para tales códigos es arbitraria; también hubiéramos podido codificar 0 = mujer y 1 = hombre. Además, las categorías de una variable nominal no pueden clasificarse significativamente en orden de magnitud (de elevado a bajo) aun cuando se asignen códigos a los números ordenados. Por ejemplo, codificar mujer como 1 y hombre como 0 no implica que las mujeres tengan una puntuación de 1 o más que los hombres. No existe ningún sentido de grado con las variables nominales. Una persona es hombre o es mujer, y en cualquier caso no tiene un grado. Incluso algunas puntuaciones numéricas en apariencia son realmente variables nominales. Por ejemplo, el número del seguro social es, de hecho, una categoría y no tiene sentido calcular su promedio.

Puesto que muchas variables nominales tienen sólo dos categorías, existe un nombre especial para ellas. Una **variable dicotómica** tiene *sólo dos categorías.* Una variable dicotómica común en las encuestas es cualquiera con las respuestas "sí" y "no" y, en diseños de investigación de laboratorio, aquella que distingue la "presencia" (el grupo experimental) o la "ausencia" (el grupo de control). Por ejemplo, al probar la efectividad de un nuevo medicamento contra la fiebre de heno, al grupo experimental se le administra la nueva droga y se registra como 1. Al grupo de control se le da una droga de imitación (o placebo) y se registra como cero. En la computadora llamamos GRUPO a esta variable. Cuando deseamos aislar al grupo experimental para el análisis, damos instrucciones a la computadora para que busque los códigos de GRUPO y seleccione dichos casos con el código 1.

Variables ordinales

Al igual que las variables nominales, las **variables ordinales** *designan categorías, pero tienen la propiedad adicional de permitir clasificar las categorías* desde la mayor hasta la menor, de la mejor a la peor o de la primera a la última. Las variables ordinales comunes incluyen clasificación de clase social (alta, media, baja, indigente), nivel de clase educativa (último año, primer ingreso, etc.) y calidad de vivienda (estándar, insuficiente, en ruinas). Las preguntas de estudio que miden actitudes y opiniones a menudo emplean puntuaciones ordenadas. Por ejemplo, la variable "actitud hacia el aborto legal" podría ordenar el grado de acuerdo mediante el uso de categorías de respuesta: *totalmente de acuerdo, de acuerdo, no sabe, en desacuerdo, totalmente en desacuerdo.* Este conjunto de códigos ampliamente utilizado se denomina escala de Likert, en honor a su creador, Rensis Likert (1932).

Variables de intervalo

Las **variables de intervalo** tienen las características de las variables nominales y ordinales y además *una unidad numérica de medición definida.* Las variables de intervalo identifican las diferencias en monto, cantidad, grado o distancia y se les asignan puntuaciones numéricas muy útiles. Los ejemplos incluyen la temperatura (registrada al grado térmico más cercano) y el coeficiente de inteligencia (CI), que va desde cero hasta 200 puntos. Con las variables de intervalo, los intervalos o distancias entre las puntuaciones son las mismas entre cualquier par de puntos en la escala de medición. Por ejemplo, con la variable temperatura, la diferencia entre 10 y 11 grados Fahrenheit es la misma que entre 40 y 41. Un conjunto de unidades de medición ordenadas hace posible sumar, restar, multiplicar y dividir puntuaciones y calcular promedios.

Las variables de intervalo dan un sentido de "cuánto" o "de qué tamaño", qué tan caliente, qué tan obstinado, qué tan conservador, qué tan deprimido, qué tan largo y qué tan pesado. Con las variables de intervalo pensamos en términos de distancia entre las puntuaciones sobre una línea recta. Por ejemplo, si el promedio de las calificaciones de un grupo en una prueba es 80 y Carlos obtuvo 85 y Berta 90, entonces la puntuación de Berta estuvo dos veces más arriba del promedio que la puntuación de Carlos. Además, los márgenes de error con las variables de intervalo están más definidos y son más fáciles de manejar porque las puntuaciones numéricas pueden redondearse.

Comparar las propiedades de variables de intervalo y variables ordinales es informativo. A diferencia de las variables de intervalo, las variables ordinales carecen de una unidad de medición determinada, aun cuando las categorías ordenadas sean numeradas. Por ejemplo, la posición final en una carrera de caballos (1, 2, 3, etc.) es sólo ordinal; simplemente indica qué caballos cruzaron la línea final en primero, segundo y tercer lugar, y así sucesivamente, pero no aclara qué tan separados terminaron unos de otros. Además, la resta entre números de posición de una variable ordinal proporciona sólo diferencias entre los lugares que ocupan, no distancias entre sus posiciones. Por ejemplo, si los caballos llamados "Piernas Largas" y "Problemas en el Puente" terminan en tercero y sexto lugar, respectivamente, entonces "Piernas Largas" llegó tres posiciones adelante. Estos caballos podrían haber llegado a la meta separados por unas cuantas pulgadas o cientos de yardas. Mientras las variables ordinales permiten algunos cálculos, como diferencias en posiciones y posición promedio, tienen utilidad matemática limitada. Las variables de intervalo poseen utilidad matemática mucho mayor que las variables ordinales.

Variables de razón

Las **variables de razón** poseen las características de las variables de intervalo y *un punto cero verdadero, donde una puntuación cero significa "ninguno" o ausencia de atributo.* Peso, estatura, edad, distancia, tamaño de la población, duración en tiempo y promedio son ejemplos de variables de razón.

Comparar variables de razón con variables de intervalo resulta informativo porque ambas tienen intervalos establecidos en su unidad de medición; pero sólo las variables de razón incluyen un punto cero con significado. Algunas variables de *intervalo* pueden tener una puntuación de cero, pero el punto cero es arbitrario; es decir, podría colocarse en cualquier punto dentro del rango posible de una variable porque el cero no significa "ninguno". Por ejemplo, la temperatura cero no significa ausencia de temperatura. Así, en la escala Fahrenheit está ubicado en 32 grados abajo del punto de congelación, mientras que en la escala Celsius se encuentra en el punto de congelación.

Los puntos cero verdaderos de las variables de razón permiten incluso mayor flexibilidad en los cálculos y el análisis estadístico. Al igual que las variables de intervalo, las variables de razón pueden multiplicarse y dividirse, pero también podemos calcular razones, de ahí su nombre. Una **razón** es *la cantidad de una observación con respecto a otra.* Por ejemplo, si Juan come tres rebanadas de pizza y Esther come una, la razón es tres a uno, que se escribe como 3:1. Con una variable de nivel de razón, la respuesta para una razón calculada tiene sentido, mientras que con una variable de intervalo no la tiene. Por ejemplo, un joven de 40 kilogramos es dos veces más pesado que uno de 20 kilogramos, una razón de 2:1. Pero no tiene sentido afirmar que una variable de intervalo *temperatura* en Miami, donde hay 80 grados, es cuatro veces más calurosa que en Nueva York, donde hay 20 grados. Nueva York no es calurosa en absoluto. Entonces, una manera de determinar si una variable tiene un cero verdadero es intentar interpretarlo como una razón. Si la razón no tiene sentido, la variable está, si acaso, en un nivel de intervalo y su punto cero es arbitrario.

Debido a las similitudes de los procedimientos estadísticos aplicados a las variables de intervalo y a las de razón, a menudo agrupamos estas distinciones refiriéndonos a estas variables como intervalo/razón. De igual modo, nos referimos a las variables nominales/ordinales. La tabla 2-2 resume las propiedades de los cuatro niveles convencionales de medición.

En resumen, para determinar el nivel de medición de una variable, formula estas preguntas y sigue el diagrama de árbol que presentamos a continuación:

1. ¿Se marca la variable si se usan nombres de categoría como "masculino" o "femenino"? Si es así, entonces el nivel de medición es nominal. Estos nombres de categoría ¿pueden clasificarse de bajo a alto, por ejemplo clase baja, clase obrera, clase media y clase alta? Si es así, entonces el nivel de medición es ordinal.
2. ¿Se marca la variable si se usan valores numéricos, por ejemplo 1, 2, 3, etc., pero las puntuaciones simplemente designan posiciones, por ejemplo primero, segundo y tercero? Si es así, entonces el nivel de medición es ordinal.

TABLA 2-2 | Características de los cuatro niveles de medición

Nivel de medición	Ejemplos	Cualidades	Operaciones matemáticas permitidas
Nominal	Género, raza, preferencia religiosa, estado civil	Clasificación en dos categorías; denominación de categorías	Conteo del número de casos (es decir, frecuencia) de cada categoría de la variable; comparación de tamaños de categorías
Ordinal	Rango de clase social, preguntas de actitud y opinión	Clasificación de categorías; ordenamiento de rangos de categorías de bajo a alto	Todo lo anterior más juicios de mayor que y menor que, y cálculos de diferencias y promedios de rangos
Intervalo	Temperatura, índices resumidos, escalas de actitud y opinión	Todo lo anterior más distancias entre puntuaciones tiene una unidad fija de medida	Todo lo anterior más operaciones matemáticas como suma, resta, multiplicación, división y raíces cuadradas
Razón	Peso, ingresos, edad, escolaridad, tamaño de población	Todo lo anterior y un punto cero real	Todo lo anterior más el cálculo de razones significativas

3. ¿Se puntúa la variable mediante valores numéricos que tienen un intervalo dado o unidad de medida como pulgadas, millas o grados? Si es así, entonces el nivel de medida es de intervalo. ¿La variable también tiene un punto cero real? Si es así, entonces el nivel de medida es una razón.

Determinación del nivel de medida de una variable

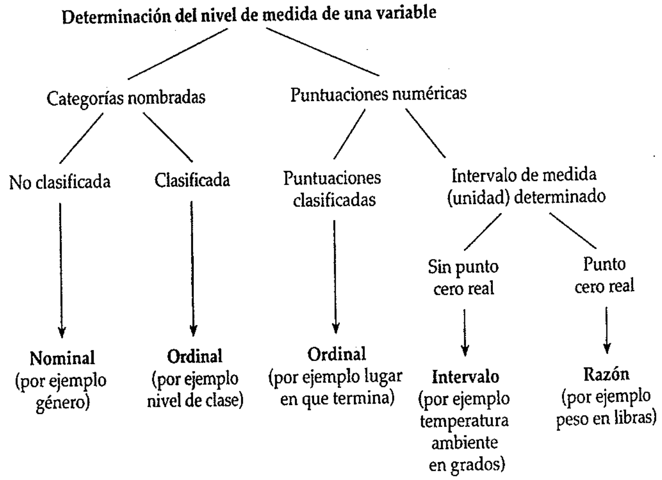


TABLA 2-3 | Creación de un índice para transformar diversas variables nominales en una variable de razón

Número y nombre de variable	Definición operacional (cómo se mide la variable)	Nivel de medición	Código (cómo se registra)	
1. FUMA	¿Es fumador habitual?	Nominal	0 = no	1 = sí
2. ALCOHÓLICO	Ha consumido cinco o más bebidas alcohólicas en el último mes	Nominal	0 = no	1 = sí
3. EJERCICIO	Hace ejercicio regularmente	Nominal	0 = no	1 = sí
4. DROGAS	Ha usado una droga ilícita en el último mes	Nominal	0 = no	1 = sí
5. CONDEBR	Ha conducido en estado de ebriedad	Nominal	0 = no	1 = sí
6. RIESGO	Número de conductas de riesgo que reportó	Razón	Suma de respuestas "sí" para las variables 1 a 5	

Cómo mejorar el nivel de medición

Para aprovechar las unidades de medida establecidas, es frecuente que los científicos inventen formas de cambiar variables nominales/ordinales a variables de intervalo/razón. La tabla 2-3 presenta diversas variables nominales que se transformaron en una variable de nivel de razón llamada índice de comportamiento de riesgo de salud. Las variables nominales se registran como cero para *no* y 1 para *sí*. Esto se llama codificación prototipo porque las puntuaciones numéricas son artificiales; 0 y 1 no distinguen montos o cantidades. En cambio, cero significa que el factor de riesgo no está presente y 1 que sí está presente. La variable de razón RIESGO es el número total de factores de riesgo de un individuo y esta variable tiene un punto cero real. Si Jeremías fuma, bebe, conduce en estado de ebriedad y consume drogas, mientras que Adán sólo fuma, entonces Jeremías tiene una conducta cuatro veces más riesgosa que Adán, una razón de 4:1. Esperamos que tu puntuación de RIESGO sea más baja.

Distinción del nivel de medida y unidad de medida

El lector debe tener cuidado de distinguir los términos *nivel de medición* de *unidad de medida.* El nivel de medición se aplica a toda la variable y da información sobre los puntos fuertes y débiles de la medición de una variable. Por ejemplo, ¿podemos calcular promedios de una variable? Si el nivel de medición de la variable es intervalo o razón, por ejemplo la variable *edad*, entonces sí. Por otro lado, si el nivel de medición es nominal (por ejemplo, género), entonces la respuesta es no. La unidad de medida, no obstante, es un término *que se emplea sólo* con variables de intervalo/razón. Fija el intervalo determinado para los valores numéricos empleados como puntuaciones para una variable de intervalo/razón. Por ejemplo, podemos elegir medir el ancho de un escritorio con una pulgada como unidad de medida. En cambio, si escogemos medir el escritorio con una regla métrica, entonces un centímetro es nuestra unidad de medida. La falta de atención a las unidades de medida puede resultar en equivocaciones costosas. Por ejemplo, en 1999, el sistema de guía del Orbitador del Clima de Marte, de la National Aeronautics and Space Administration (NASA), envió inadvertidamente a la nave espacial a la atmósfera marciana causando la destrucción del orbitador. Un equipo de ingeniería de proyectos utilizó unidades métricas de medida (es decir, partes de metros) para comunicarse con otro equipo que supuso que los números estaban en unidades inglesas (partes de pulgadas). Esta confusión acerca de las unidades de medida costó a la NASA $125 millones de dólares (http://www.cnn.com/TECH/space/9909/30/mars.metric/).

Codificación y conteo de observaciones

Una vez completa la recolección de datos, el siguiente paso en el manejo de datos consiste en codificar y registrar todas las mediciones en una hoja de cálculo o en un archivo de datos de computadora. La tabla 2-4 presenta un ejemplo de un **registro o guía de codificación**, que es *una descripción concisa de símbolos que describen el significado de cada puntuación para cada variable.* Tales datos vienen de un estudio ficticio sobre estudiantes del Instituto Apple Pond. En este registro de codificación sustituimos símbolos numéricos por las categorías de masculino y femenino y por niveles de escolaridad. Esto se hace porque a las computadoras se les facilita contar y seleccionar números (en lenguaje computacional, símbolos *numéricos*) que palabras (*caracteres* o símbolos de *cadena*).

En un registro de codificación, se debe tener cuidado en ser muy preciso porque la codificación de respuestas podría introducir un error de medición. Cada variable se codifica siguiendo dos principios básicos: inclusividad y exclusividad. El **principio de inclusividad** establece que para una variable determinada *debe haber una puntuación o un código para cada observación realizada.* Dicho de otra manera, ¿incluimos una categoría de respuesta o puntuación para toda respuesta posible? Por ejemplo, con la variable nominal *raza* podríamos indicar las categorías de blanco, afroamericano, nativo americano, asiático-americano, hispano y otra(s). La categoría de respuesta *otra(s)* evita la necesidad de ocupar espacio en el cuestionario para las categorías que se espera tengan pocas respuestas en el lugar del estudio (esquimales en Kansas). El código otra(s) es una *categoría residual* que abarca los remanentes (piensa en la palabra *residuo*).

Aun cuando ignoramos la cuestión de cómo codificar a las personas de ascendencia mixta, si sólo usamos blanco y afroamericano, la categoría de raza no será inclusiva de, por decir, asiático-americano. Sin su propia categoría u otra(s) no es factible suponer que todos los asiático-americanos se registrarán de la misma forma. Algunos anotarán blanco, pero otros quizá dejen sin contestar el espacio de la pregunta. Después de calcular los totales quizá no

TABLA 2-4 | Registro de codificación para respuestas de cuestionarios de especialistas en justicia en el Instituto Apple Pond

Nombre de la variable	Descripción de códigos de la variable
NOMBRE DEL ESTUDIANTE	Registre el nombre, apellidos paterno y materno; espacio en blanco = faltante
EDAD	Registre la edad informada hasta 97 años; 98 = 98 años o más; 99 = faltante
GÉNERO	0 = hombre, 1 = mujer, 9 = faltante
PROMEDIO	Promedio en una escala de cuatro puntos = número de puntos de calidad ganados por hora crédito (redondeado a dos lugares decimales); 9.99 = faltante
ESCOLARIDAD	1 = nuevo ingreso; 2 = segundo año; 3 = intermedio inicial; 4 = intermedio avanzado; 8 = otro; 9 = faltante

seamos capaces de señalar exactamente cuántos tenemos de *cualquier* raza. Perdimos a los asiático-americanos que no contestaron la pregunta, y algunos de nuestros "blancos" son asiático-americanos pero no tenemos noción de *cuántos.* Semejante descuido de pérdida de control sobre el error de medición puede hacer que los datos de raza sean inservibles.

El **principio de exclusividad** sostiene que para una variable determinada *a cada observación se asigna una y sólo una puntuación.* Así, cada categoría debe excluir puntuaciones que no le pertenezcan y cualquiera de las dos categorías no debe compartir una respuesta. Por ejemplo, con la variable *afiliación religiosa en la niñez,* las categorías de respuesta protestante, bautista, católico, judío y otra(s) no serían mutuamente excluyentes porque un bautista quizá se anote como bautista, mientras otro lo haga como protestante. Cuando se sumen los totales de cada categoría, no podremos decir cuántos bautistas había en la muestra. Algunos quizá se registraron como "protestante", pero no tenemos forma de especificar quiénes y cuántos lo hicieron. La tabla 2-5 muestra los resultados del Estudio Social General de 1994 para esta variable. La exclusividad se asegura formulando a los protestantes las preguntas adicionales necesarias para conocer sus denominaciones específicas. (Para ayudar a la comprensión sobre la información de las tablas de este texto, modificaremos las tablas para diferenciar claramente las "Especificaciones" de los datos disponibles y los "Cálculos". En los reportes publicados de estas tablas, dichos términos no aparecerían.)

TABLA 2-5 | Distribución de la afiliación religiosa en la niñez, respuesta a las preguntas: ¿en qué religión fue educado? Si fue protestante, ¿en qué denominación específica, si la hay?

Especificaciones		Cálculos
Categoría de respuesta	*a)* Número	*b)* Porcentaje de la muestra total
Protestante		
Bautista	706	23.73
Metodista	319	10.72
Luterano	220	7.39
Presbiteriano	139	4.67
Episcopal	68	2.29
Otra	309	10.39
Ninguna denominación o iglesia sin denominación	69	2.32
Total protestante	1 830	61.51
Católica	882	29.65
Judía	55	1.85
Ninguna	127	4.27
Otra	74	2.49
Sin respuesta	7	0.23
Total	2 975	100.00

Fuente: National Opinion Research Center, General Social Survey 1994.
www.icpsr.umich.edu/gss/codebook/relig16.htm
www.icpsr.umich.edu/gss/codebook/denom16.htm

TABLA 2-6 | Hoja de cálculo de respuestas al cuestionario de 10 especialistas en justicia delictiva en el Instituto Apple Pond (datos ficticios)

	Especificaciones			
Nombre del estudiante	**Edad**	**Género**	**Promedio**	**Escolaridad**
Jessica A Cortland	19	1	3.21	2
Mark E Pippin	22	0	2.75	4
Stayman V Winesap	19	0	2.43	1
Barry D McIntosh	21	0	3.39	3
Harriet G Smith	20	1	3.87	3
Antonio B Rome	22	0	2.32	3
Robert J Cox	18	0	3.25	1
Rodney I Greening	20	0	9.99	2
Thomas R York	22	0	2.47	4
Goldie D Licious	19	1	3.68	2

Regresa al registro de codificación de la tabla 2-4 y observa que el principio de inclusividad se cumple proporcionando un *código para los datos perdidos*, llamados **valores perdidos**. Decimos, por ejemplo, que el género y escolaridad tienen un valor perdido de 9. En algunos estudios, los valores perdidos se presentan cuando por accidente el entrevistador se salta una pregunta o un encuestado no contesta. Al calcular los estadísticos para una variable, pasamos por alto los casos que resulten en un valor perdido.

La tabla 2-6 es una hoja de cálculo de los resultados de la aplicación del cuestionario empleado en una encuesta aplicada a 10 especialistas en justicia delictiva del Instituto Apple Pond. Una hoja de cálculo es una matriz que muestra las puntuaciones de todas las variables organizadas en columnas; y todos los casos, en filas.

Una hoja de cálculo es útil para ordenar y resumir datos de una forma eficaz. Por ejemplo, si contamos rápidamente el número de mujeres en la muestra sumando las unidades citadas bajo "Género". Mediante esta simple hoja de cálculo podemos ver rápidamente que la muestra se compone de siete hombres y tres mujeres; existen dos estudiantes de primer año, tres de segundo año, tres de tercer año y dos de último año; el rango de edades oscila entre los 18 y los 22 años, y el rango del promedio va de 2.43 a 3.87 con un caso no reportado. Por supuesto, para una muestra grande, un procedimiento eficaz implica tanto la especificación de estos códigos de la hoja de cálculo en un archivo de datos de la computadora como hacer que el programa computacional se encargue de los cálculos. Los archivos de datos de computadora están organizados como estas hojas de trabajo.

Distribuciones de frecuencias

Una vez que todos los datos están organizados en una hoja de cálculo o en un archivo de datos de computadora, el siguiente paso en el análisis consiste en enfocarse por separado en cada variable y contestar la pregunta: ¿cuántos sujetos caen en cada categoría o puntuación? Organizamos los datos de cada variable en una **distribución de frecuencias**, que es *una lista de todas las puntuaciones observadas de una variable y la frecuencia (f) de cada puntuación (o categoría)*. Utilizamos letras mayúsculas para representar una variable. Si X se define como la variable *género*, la distribución de frecuencias de X simplemente muestra cuántos hombres y mujeres hay en la muestra. La tabla 2-5 mostrada anteriormente proporciona la distribución de frecuencias para la afiliación religiosa en la niñez.

Distribución de frecuencias Lista de todas las puntuaciones observadas de una variable y la frecuencia (f) de cada puntuación (o categoría).

Estandarización de distribuciones de puntuaciones

El conocimiento de la frecuencia de una categoría no resulta muy informativo por sí mismo. Por ejemplo, alguien nota que existen cinco millonarios que viven en una ciudad. Cinco no son muchos para la ciudad de Nueva York, pero lo serían para un pueblo de 800 personas. Así, es más informativo reportar la frecuencia de una categoría como una proporción o porcentaje con respecto al número total de sujetos de la muestra. La imaginación estadística nos impulsa a expresar la frecuencia de una categoría en un contexto mayor, como una parte en relación con un todo. Planteamos la pregunta: cinco millonarios ¿de cuántas personas? Como observamos en el capítulo 1, las fracciones, las proporciones y los porcentajes ofrecen denominadores comunes o "medidas estándar" para facilitar la comparación de categorías y muestras. Para una muestra como un todo, la **distribución de frecuencias con proporciones** consiste en *una lista de la proporción de respuestas para cada categoría o puntuación de una variable.* La **distribución de frecuencias de distribución** es *una lista del porcentaje de respuestas para cada categoría o puntuación de una variable.*

Distribución de frecuencias con proporciones Lista de la proporción de respuestas para cada categoría o puntuación de una variable.

Distribución de frecuencias de porcentajes Lista del porcentaje de respuestas para cada categoría o puntuación de una variable.

Para obtener estas distribuciones para cada categoría de respuesta o puntuación de una variable, escribimos una fracción y después la dividimos para obtener la proporción y el porcentaje. Para facilitar la interpretación, la distribución de frecuencias de porcentajes es la que comúnmente se reporta. Por ejemplo, en los datos de la hoja de cálculo del Instituto Apple Pond de la tabla 2-6 podemos observar que el porcentaje de hombres es

$$p\text{ [de la muestra de estudiantes que son hombres]} = \frac{\#\text{ hombres}}{n} = \frac{7}{10} = 0.7000$$

$$\%\text{ [de la muestra de estudiantes que son hombres]} = (p)(100) = (0.7000)(100) = 70.00\%$$

donde p es la proporción y n es el tamaño de la muestra. Después de hacer lo mismo para las mujeres, tenemos la distribución de frecuencias de los porcentajes de la variable *género* en la columna de la derecha de la tabla 2-7. En ésta también se incluyen la frecuencia y las distribuciones de frecuencias proporcionales. El total de todas las proporciones y porcentajes para la distribución será igual a 1.0000 y 100.00 por ciento, respectivamente, considerando el error de redondeo.

TABLA 2-7 | Frecuencia, frecuencia proporcional y distribuciones de frecuencias porcentuales de la variable género para una muestra de 10 estudiantes del Instituto Apple Pond

Especificaciones		Cálculos	
Género (X)	Frecuencia (f)	Frecuencia proporcional	Frecuencia porcentual (%)
Hombre	7	0.7000	70.00
Mujer	3	0.3000	30.00
Total	10	1.0000	100.00

Cálculo de las frecuencias proporcionales y porcentuales de una categoría

$$p\ [\text{de la muestra total } (n) \text{ en una categoría}] = \frac{f \text{ de una categoría}}{n} = \frac{\#\text{ en categoría}}{n}$$

$$\%\ [\text{de la muestra total } (n) \text{ en una categoría}) = (p\ [\text{de la muestra total } (n) \text{ en una categoría}])\ (100)$$

donde

p [de la muestra total (n) en una categoría] = proporción de todos los casos que caen en la categoría,

f = frecuencia de casos (o número de casos) en la categoría,

n = tamaño de la muestra

La tabla 2-5 (en la página 49) muestra la frecuencia y la distribución de frecuencias con porcentajes para la variable afiliación religiosa en la niñez.

Codificación y conteo de datos de intervalo/razón

Las variables con niveles de medición de intervalo/razón se distinguen de las variables nominales/ordinales por sus cualidades numéricas, sobre todo por sus unidades de medición, como millas, kilómetros, pulgadas, segundos y kilogramos. Tales variables "cuantitativas" nos permiten imaginar una regla y pensar linealmente en términos de la distancia entre puntos sobre una línea recta. Es más, podemos hacer mediciones muy precisas.

Una **medición precisa** es *aquella en la que el grado de error de medición es suficientemente pequeño para la tarea en cuestión.* La precisión depende de circunstancias prácticas y se controla al especificar el error de redondeo. Por ejemplo, al cortar dos troncos de dos pies de largo para una chimenea, un pequeño grado de precisión será suficiente porque podemos

darnos el lujo de tener un elevado grado de error, por ejemplo, "medio pie de más o de menos". En contraste, para una prueba de calidad de tarjetas de circuitos de microcomputadoras puede exigirse una precisión de una milésima de pulgada. El grado de precisión es cuestión de tolerancia. Preguntamos: ¿qué error de medición podemos tolerar (o soportar) sin encontrar problemas prácticos o sacar conclusiones científicas con falla?

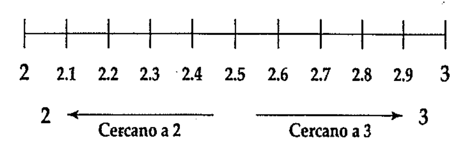
Redondeo de las observaciones de intervalo/razón

La observación de una variable de intervalo/razón tal vez no nos ofrezca la puntuación verdadera porque sus mediciones a menudo pueden hacerse de manera infinitamente más precisa. Por ejemplo, podemos medir distancias hasta la unidad más cercana en kilómetros, metros, centímetros, y así sucesivamente. Por tanto, redondeamos las puntuaciones de intervalo/razón hasta cierto grado de precisión elegido y especificado. De esta manera, reconocemos que el código registrado para la puntuación tiene algún error de medición. El **error de redondeo** es *la diferencia entre la puntuación real o perfecta (que quizá nunca conozcamos) y nuestra puntuación observada y redondeada.* El error de redondeo depende de qué *posición decimal elegimos como nuestro nivel de precisión,* nuestra unidad de redondeo. (Si es necesario, el estudiante debe revisar en el apéndice A la ubicación de las posiciones decimales.) Si decidimos medir el tiempo a la centésima de segundo más cercana, como se efectúa en los eventos olímpicos de pista, entonces nuestra unidad de redondeo es la posición de las centésimas.

El procedimiento para redondear una puntuación de una variable de intervalo/razón es como sigue:

1. Especifica la unidad de redondeo según su posición decimal.
2. Observa el número *a la derecha* de la unidad de redondeo y sigue estas reglas:
   A. Si es 0, 1, 2, 3 o 4, redondea hacia el entero inferior.
   B. Si es 6, 7, 8 o 9, redondea hacia el entero superior.
   C. Si es 5, observa la siguiente posición decimal a la derecha y, si el número es 5 o mayor, redondea hacia el entero superior; si no existe algún número en esa siguiente posición decimal, deja el redondeo en ese número.

Piensa en el redondeo como un movimiento hacia el punto más cercano sobre una línea. Por ejemplo, si redondeamos al *entero más cercano* (la posición de las unidades), simplemente estamos moviendo al entero más cercano.



Por tanto, aquí 2.1, 2.2, 2.3 y 2.4 se redondean hacia abajo a 2 sólo porque están más cerca de 2 que de 3. Se ofrecen ejemplos adicionales en el apéndice A.

Los límites reales de puntuaciones redondeadas

Una vez que conocemos las puntuaciones redondeadas, los números *en la posición decimal de la unidad de redondeo* se consideran estimaciones. El valor real de una puntuación podría ser cualquiera de las puntuaciones que se redondean para obtener la puntuación registrada.

Por ejemplo, supongamos que redondeamos la estatura de Jonathan a la pulgada más cercana y registramos 69 pulgadas. Varias horas después, con Jonathan ausente, Alan nos pregunta cuál es la estatura de Jonathan. Observamos nuestra hoja de cálculo de datos y vemos el código de 69 pulgadas. En este momento, podemos afirmar que la estatura real de Jonathan está entre 68½ y 69½ pulgadas, o 69 más menos media pulgada. Este *rango de posibles valores reales de una puntuación (ya) redondeada* se llama **límite real** o **límite verdadero** de la puntuación.

Los límites reales de una puntuación redondeada especifican el rango de números que podrían redondearse para obtener la puntuación registrada. En este sentido, calcular los límites reales es la inversa del redondeo. Por ejemplo, supongamos que registramos cuánto tiempo toma a cada uno de los 150 estudiantes completar un proyecto de laboratorio de química y lo redondeamos a la hora más cercana. También considera que 56 de estos estudiantes reciben una puntuación de dos horas. Algunos de ellos tomaron un poco menos de dos horas y otros un poco más. Precisamente, un estudiante que registra dos horas pudo haber tomado sólo 90 minutos (1½ horas), el *límite real inferior*, o hasta 150 minutos (2½ horas), el *límite real superior.* Una puntuación de dos horas *en realidad* significa entre 1½ y 2½ horas; por eso llamamos límites reales a este rango de tiempos.

Límite real inferior de dos horas 1½ (90 minutos)	← 2 horas (120 minutos) →	Límite real superior de dos horas 2½ (150 minutos)

Calculamos los límites reales moviéndonos media unidad de redondeo en cada dirección, utilizando el siguiente procedimiento:

Cálculo de límites reales de una puntuación de intervalo/razón

1. Observa la puntuación e identifica la "unidad de redondeo", el lugar decimal al que la puntuación se redondeó (como en la columna B que sigue). (Para ubicaciones del lugar decimal, revisa la figura A-1 del apéndice A.)
2. Divide entre 2 esta unidad de redondeo (como en la columna C que sigue). *Atención:* no dividas el número del lugar decimal de la unidad de redondeo entre 2.
3. Resta el número del paso 2 de la puntuación redondeada observada, para obtener el límite real inferior (LRI, como en la columna D que sigue).
4. Suma el resultado del paso 2 a la puntuación redondeada observada, para obtener el límite real superior (LRI, como en la columna E que sigue).

Ejemplos:

	(A) Calcula los límites reales de:	(B) Identifica la unidad de redondeo	(C) Divide entre 2 la unidad de redondeo	(D) LRI Resta (C) de (A)	(E) LRS Suma (C) a (A)
a)	.48	.01	$\frac{.01}{2} = .005$	.475	.485
b)	17	1	$\frac{1}{2} = .5$	16.5	17.5
c)	4000	1000	$\frac{1000}{2} = 500$	3500	4500
d)	0.7	.1	$\frac{.1}{2} = .05$	.65	.75

Por ejemplo, para los 56 estudiantes que calificaron dos horas en el proyecto del laboratorio de química, redondeamos a la hora más cercana (la posición de las unidades). Dividimos *esta unidad de redondeo de una hora* entre 2 para obtener hora y media. Entonces restamos este resultado de la puntuación redondeada observada de dos horas para obtener el límite real inferior (1½ horas) y lo sumamos a la puntuación observada de dos horas, para obtener el límite real superior (2½ horas). Incluso es improbable que uno de estos 56 estudiantes tomara exactamente dos horas para completar el proyecto; dos horas es una estimación redondeada. Podemos tener la certeza, sin embargo, de que cada uno de los 56 terminara entre 1½ y 2½ horas. Nuestro grado de precisión es la unidad de redondeo de una hora.

Los principios de inclusividad y exclusividad también se aplican a las variables de intervalo/razón. Para una variable como la edad, apegarse al principio de inclusividad parecería razonable; sólo registramos la "edad en el último cumpleaños". No obstante, para garantizar la inclusividad, un cuestionario de investigación debe incluir las respuestas "se negó" y "no sabe". La exclusividad es razonable en cuanto a que todas las mediciones se realicen de la misma manera, en este caso la edad en el último cumpleaños. Si un encuestado dice que tiene 26 años, entonces registra 26, no 27 ni 25.

Distribuciones de frecuencias de proporciones y de porcentajes para variables de intervalo/razón

Las distribuciones de frecuencias de proporciones y porcentajes para variables de intervalo/razón se calculan de la misma forma que para variables nominales/ordinales, excepto que en lugar de categorías tenemos puntuaciones. Por ejemplo, si la Universidad Smithville tiene 10 000 estudiantes y 3 000 tienen 19 años, las frecuencias proporcionales y porcentuales para la puntuación de 19 años son

$$p\,[\text{de 19 años en la Universidad Smithville}] = \frac{f \text{ de 19 años}}{n} = \frac{3\,000}{10\,000} = 0.3000$$

$$\%\,[\text{de 19 años en la Universidad Smithville}] = (p)\,(100) = 30.00\%$$

TABLA 2-8 | Ilustración de una distribución de frecuencias porcentuales acumuladas: años de escolaridad entre cuidadores de pacientes ancianos con Alzheimer

Especificaciones		Cálculos	
Años de educación formal (X)	Frecuencia (f)	Frecuencia porcentual	Porcentaje acumulado (%) (f)
5	1	5%	5
6	1	5	10
7	1	5	15
9	2	10	25
10	1	5	30
11	1	5	35
12	10	50	85
14	2	10	95
16	1	5	100
Total	20	100%	100

Si estos cálculos se realizan para todas las edades, los resultados se presentan como la distribución de frecuencias de porcentaje de la variable *edad* para la población de estudiantes de la Universidad de Smithville.

Distribuciones de frecuencias de porcentajes acumulados

La tabla 2-8 presenta la frecuencia, la frecuencia de porcentaje y las distribuciones de frecuencias de porcentajes acumulados de los niveles de escolaridad de 20 cuidadores, parientes que acompañan a pacientes con Alzheimer en una clínica (Clair, Ritchey y Allman, 1993). Estas tres piezas de información son partes típicas de los resultados obtenidos por computadora porque juntos generan respuestas rápidas a una serie de preguntas. Obviamente, la frecuencia de puntuación bruta (*f*) proporciona una respuesta sobre cuántos sujetos recibieron una puntuación específica, y la frecuencia porcentual estandariza la frecuencia de acuerdo con el tamaño de la muestra. La información adicional de la tabla 2-8, la frecuencia de porcentajes acumulados, es una valiosa forma para observar las frecuencias de las puntuaciones en una distribución hasta, e inclusive, una puntuación de interés. Ésta es la **frecuencia de porcentajes acumulados**, que es *la frecuencia porcentual de una puntuación y además la de todas las puntuaciones que la preceden en la distribución.* Por ejemplo, en el caso de los cuidadores de la tabla 2-8, ¿qué porcentaje tiene un nivel de escolaridad hasta e inclusive el nivel de preparatoria? Para obtener las frecuencias de porcentajes acumulados hacemos una lista con las puntuaciones, de la más baja a la más alta, y calculamos la frecuencia de porcentaje de cada puntuación. Entonces sumamos las frecuencias de porcentaje de la puntuación que nos interesa y todas las puntuaciones menores. En la tabla 2-8, 85 por ciento tenían 12 años de escolaridad o menos. Restando esta frecuencia de porcentajes acumulados del 100 por ciento, rápidamente podemos ver que sólo 15 por ciento de la muestra fueron más allá de la escuela preparatoria.

El siguiente cuadro es una guía sobre cómo elaborar distribuciones de frecuencias.

Para elaborar distribuciones de frecuencias

Supongamos que deseamos elaborar una tabla de distribución de frecuencias para la variable de edad de la tabla 2-6, para una muestra de ficción del Apple Pond Institute. La tabla 2-6 presenta los datos en formato de hoja de cálculo. Debemos completar la siguiente tabla A para presentar la distribución de frecuencias, distribución de frecuencias porcentuales y distribución de frecuencias de porcentajes acumulados para la variable *edad*. Sigue estos pasos:

1. Construye una plantilla con el título y encabezados apropiados (la información arriba de las columnas de números en la tabla A).
2. Observa las puntuaciones para edades entre los 10 estudiantes de la tabla 2-6 y haz una lista, sólo una vez, de cada uno de los valores de X del más bajo al más alto. Esto va bajo "Edad (X)" en la tabla A.
3. Cuenta el número de estudiantes para cada edad de la tabla 2-6 e inserta esta cantidad bajo "Frecuencia (f)" en la tabla A. Comprueba ver que la frecuencia total es el tamaño muestral, $n = 10$, y registra este total.
4. Calcula la frecuencia proporcional para cada valor de X al dividir cada frecuencia entre el total n de 10. Estos resultados van bajo "Frecuencia porcentual" en la tabla A. Comprueba que el total de la frecuencia porcentual suma 1.0000. Si los cálculos son correctos y el total no suma 1.0000, entonces inserta una nota al pie que indique "El total no sumó 1.0000 por error de redondeo".
5. Calcula la frecuencia de porcentaje de cada uno de los valores de X al multiplicar por 100 la frecuencia proporcional (es decir, al mover el punto decimal dos lugares a la derecha). Estos resultados van bajo "Frecuencia porcentual" en la tabla A. Si los cálculos son correctos y el total no suma 100.00%, entonces inserta una nota al pie que indique "El total no sumó 100 por ciento por error de redondeo".
6. Para la tabla A, calcula la frecuencia de porcentajes acumulados, la frecuencia de porcentaje de una puntuación más la de todas las puntuaciones que la preceden. Empieza por registrar 10.00 por ciento para $X = 18$. Ahora suma esta frecuencia de porcentajes acumulados de $X = 18$ al *porcentaje* de frecuencia de $X = 19$ para obtener el porcentaje de frecuencia acumulada de $X = 19$, que es 10.00 por ciento + 30.00 por ciento = 40.00 por ciento. Ahora suma esta frecuencia de porcentajes acumulados de $X = 19$ a la frecuencia *porcentual* de $X = 20$ para obtener la frecuencia de porcentajes acumulados de $X = 20$, que es 40.00 por ciento + 20.00 por ciento = 60.00 por ciento, y así sucesivamente. Asegúrate de que la frecuencia de porcentajes acumulados del valor más alto de X sume 100.00 por ciento.

TABLA A. | Distribuciones de frecuencia, de frecuencia proporcional y de frecuencia porcentual de la variable *edad* de 10 estudiantes del Apple Pond Institute

Edad (X)	Frecuencia (f)	Frecuencia proporcional	Frecuencia porcentual (%)	Frecuencia porcentual acumulativa (%)
18	1	.1000	10.00	10.00
19	3	.3000	30.00	40.00
20	2	.2000	20.00	60.00
21	1	.1000	10.00	70.00
22	3	.3000	30.00	100.00
Totales	10	1.0000	100.00	

Percentiles y cuartiles

Con frecuencia visualizamos una distribución de puntuaciones como fraccionada o "fracturada" en grupos que están arriba y debajo de una puntuación, o en grupos con iguales porcentajes de casos. Las distribuciones de frecuencias acumuladas proporcionan una herramienta para identificar **cuantiles,** *puntuaciones que separan una fracción de los casos de una distribución.* Los rangos percentilares (o simplemente percentiles) son un cuantil común. *Entre los casos en una distribución de puntuaciones,* el **rango percentilar** es *el porcentaje de casos que caen en o están debajo de un valor específico de X.* Por ejemplo, en las frecuencias de porcentajes acumulados de la tabla 2-8, vemos que un cuidador con 14 años de escolaridad tiene un nivel educativo igual o superior al 95 por ciento de la muestra, un rango percentilar de 95. Con frecuencia los percentiles se emplean en círculos de educación como una manera de ordenar notas o calificaciones de exámenes. Por ejemplo, en un examen de admisión a la universidad, un estudiante con una calificación que corresponda al percentil 90 o mayor calificaría para la admisión en una universidad de prestigio, ya que significa que está por encima del 90% del resto de los alumnos.

Pasos para calcular percentiles

El cálculo de un percentil analiza la siguiente pregunta: ¿una calificación particular es igual o más alta que cuál porcentaje de calificaciones? A continuación aparecen datos fraccionales para el examen de un curso. Nótese que las calificaciones están ordenadas en forma ascendente (es decir, de menor a mayor). *Antes de calcular un percentil, las calificaciones deben ordenarse de menor a mayor o de mayor a menor.*

Calculemos el rango percentilar de Taylor en este examen. Su calificación de 78 es igual o más alta que la de 14 de los 27 estudiantes. Primero, calculamos la proporción de casos iguales o menores a 78 y luego el porcentaje.

$$p\,[\text{de calificaciones} \le 78] = \frac{14}{27} = 0.5185 \quad \%\,[\text{de calificaciones} \le 78] = (p)\,(100) = 51.85 = 52\%$$

Entonces, el rango percentilar es 52. Ella alcanzó una calificación igual o mayor a la del 52 por ciento de los estudiantes. Nótese que redondeamos el porcentaje a dos lugares porque los rangos percentilares se reportan en porcentajes enteros (es decir, sin lugares decimales).

Calculemos ahora el rango percentilar de John:

$$p\,[\text{de calificaciones} \le 91] = \frac{23}{27} = .8518$$

$$\%\,[\text{de calificaciones} \le 91] = (p)\,(100) = 85.18 = 85\%$$

Entonces, el rango percentilar es 85. Nótese que la calificación de Barry se incluyó en el cálculo porque es igual a la de John.

Lugar de estudiante	Nombre de estudiante	Calificación de examen (ordenado)	Lugar de estudiante	Nombre de estudiante	Calificación de examen (ordenado)
1	Kevin	54	15	Shannel	79
2	Carl	58	16	William	80
3	Robert	61	17	Angie	82
4	Brian	61	18	Akilah	83
5	Maria	65	19	Daniel	85
6	Sean	69	20	Kaitlin	88
7	Jim	70	21	Marcy	90
8	Jessica	72	22	John	91
9	Carol	73	23	Barry	91
10	Brooke	75	24	Wnda	93
11	Kia	75	25	Sarah	95
12	Terry	77	26	Charles	96
13	Jackie	77	27	Elisa	97
14	Taylor	78			

RESUMEN DE PASOS PARA CALCULAR PERCENTILES:

Paso 1. Ordenar las calificaciones.
Paso 2. Calcular la proporción y porcentaje de casos con calificaciones iguales o menores que el caso de interés.
Paso 3. Indicar el percentil en porcentajes enteros.
Nota: Recordar que los percentiles se obtienen fácilmente de una distribución de porcentajes acumulada.

Los **cuartiles** son *cuantiles que identifican las puntuaciones que dividen una distribución en cuatro grupos de igual tamaño (es decir, 25 por ciento de los casos en cada grupo).* Cuando una distribución tiene un rango grande de puntuaciones, los cuartiles se obtienen fácilmente a partir de distribuciones de frecuencias de porcentajes acumulados. El primer cuartil, Q_1, es el 25o. percentil; el segundo, Q_2, es el 50o. percentil; y el tercero, Q_3, es el 75o. percentil. Un software computarizado de estadística por lo general está programado para identificar cuartiles y otros cuantiles, por ejemplo los deciles, que dividen una distribución en 10 grupos de igual tamaño.

La tabla 2-9 presenta la distribución de notas en un examen de mitad de curso (X) e ilustra la utilidad de los cuartiles. En este grupo de 20 estudiantes, el 25 por ciento más bajo (o las cinco notas más bajas) son $X = 69$ y menos, el siguiente cuarto de estudiantes es de $X = 72$ a 84, el tercer cuarto son de $X = 85$ a 91, y el cuarto más alto es de $X = 93$ y más. También podemos ver que un cuarto de los estudiantes obtuvo calificaciones de 69 o menos y no obtuvo una C; la mitad obtuvo calificaciones arriba de 84, tres cuartos calificaron 91 o menos, la mitad obtuvo calificaciones entre 72 y 91, y así sucesivamente.

TABLA 2-9 | Cuartiles de una distribución de calificaciones de un examen de mitad de curso

	Especificaciones		Cálculo		
	Calificación de examen (X)	f	Porcentaje f	Porcentaje acumulado (%) f	Ubicación de los cuartiles (Q)
	31	1	5.0%	5.0	
	58	1	5.0	10.0	
	63	1	5.0	15.0	
	68	1	5.0	20.0	
Q_1 →	69	1	5.0	25.0	← Q_1 = 25o. percentil
	72	1	5.0	30.0	
	76	1	5.0	35.0	
	77	1	5.0	40.0	
	82	1	5.0	45.0	
Q_2 →	84	1	5.0	50.0	← Q_2 = 50o. percentil
	85	1	5.0	55.0	
	86	2	10.0	65.0	
	88	1	5.0	70.0	
Q_3 →	91	1	5.0	75.0	← Q_3 = 75o. percentil
	93	2	10.0	85.0	
	94	1	5.0	90.0	
	95	1	5.0	95.0	
	97	1	5.0	100.0	
	Total	20	100.0%		

Por último, es importante recordar ordenar las calificaciones de una distribución antes de calcular los cuartiles.

Cuantiles Puntuaciones que separan una porción de los casos de una distribución.

Rango percentilar Entre los casos en una distribución de puntuaciones, es el porcentaje de casos que caen en o debajo de un valor especificado de *X*.

Cuartiles Cuantiles que identifican los valores de puntuaciones que dividen una distribución en cuatro grupos de igual tamaño.

Agrupación de datos de intervalo/razón

A veces, para lograr mayor claridad en la presentación de una tabla o gráfica, las distribuciones de intervalo/razón se agrupan o "colapsan" en un número más pequeño de categorías ordinales. Por ejemplo, en un estudio de adultos en Estados Unidos, los valores de la variable edad variarán de 20 a alrededor de 100, lo cual da 80 puntuaciones. Es confuso presentar una tabla que presente la frecuencia y frecuencia de porcentaje de las 80 edades. Para mayor claridad, combinamos edades en categorías de 10 años, como se ve en la tabla 2-10.

TABLA 2-10 | La variable de razón de edad, agrupada en categorías ordinales de 10 años

Especificaciones			Cálculos
Código ordinal	Grupo de edad	f	Porcentaje
1	20-29	47	10.49
2	30-39	68	15.18
3	40-49	106	23.66
4	50-59	96	21.43
5	60-69	53	11.83
6	70-79	45	10.04
7	80-89	24	5.36
8	90 y mayor	9	2.01
	Total	448	100.00

Hay algo importante qué ver respecto del agrupamiento de datos de intervalo/razón. Cuando agrupamos datos, eliminamos detalles, lo cual produce un *error de agrupamiento.* Por ejemplo, la tabla 2-10 muestra que hubo 106 encuestados entre 40 y 49 años que contestaron en la encuesta. Pero ¿la mayoría estaba más cerca de 40 o de 49? No tenemos manera de saberlo con sólo observar las puntuaciones agrupadas. Si se dispone de los datos no agrupados, podemos usarlos para obtener cálculos de promedios más precisos. En general, en cualquier momento nos movemos de un nivel de medición "superior" a otro "inferior" (esto es, de razón a intervalo, de intervalo a ordinal, de ordinal a nominal), perdemos información y ello limita lo que matemáticamente puede hacerse.

No obstante, al leer el trabajo de otros, tal vez se nos presenten datos agrupados sin los estadísticos correspondientes. En tales situaciones, el estudiante debe intentar comunicarse con el autor para obtener los estadísticos descriptivos. Si esto no es posible, entonces pueden calcularse promedios y otros estadísticos a partir de los datos agrupados, pero dichos estadísticos incluirán error de agrupamiento. Véase un estudio de Freund y Simon (1991:70) acerca del cálculo de estadísticos con datos agrupados.

Insensatez y falacias estadísticas: la importancia de tener una muestra representativa

El tamaño muestral y la representatividad muestral son cosas separadas. Una muestra grande no garantiza una muestra representativa. Una equivocación repetitiva y sistemática en el muestreo puede producir una muestra grande pero sesgada. Un caso clásico de error sistemático de muestreo ocurrió en la campaña presidencial de 1936, en la que la revista *Literary Digest* seleccionó una muestra grande de números telefónicos y propietarios de automóvil. Los resultados mostraron un abrumador apoyo para el candidato republicano, Alf Landon, sobre Franklin D. Roosevelt, el candidato demócrata. Cuando llegó el día de elecciones, no fue Landon sino Roosevelt el que ganó y nada menos que con un triunfo aplastante. La encuesta *Literary Digest* sistemáticamente no hizo caso de electores sin teléfono ni automóvil y así falló en encuestar en forma adecuada a los pobres, quienes fueron el grueso del apoyo de Roosevelt (Babbie 1992: 192-93). Hay métodos para verificar la representatividad de una muestra, y los trataremos en el capítulo 10. Por ahora basta decir que una pequeña muestra representativa es mejor que una grande no representativa. Una cucharadita de condimento picante con todos los ingredientes es una mejor prueba de gusto que una cucharada sacada de sólo una parte de la olla.

RESUMEN

1. El error estadístico es el grado conocido de imprecisión en los procedimientos empleados para recolectar y procesar información. El análisis estadístico comprende controlar el error y, así, saber si una conclusión es acertada.
2. Las dos principales fuentes de error son el error de muestreo y el error de medición.
3. Una población es un grupo grande de personas de interés particular que deseamos estudiar y entender. Muestreamos la población, reunimos datos de ella y calculamos estadísticos sobre estos datos para obtener estimaciones de parámetros poblacionales y sacar conclusiones acerca de la población.
4. Las estimaciones estadísticas están basadas en métodos científicos lógicos en los que los errores de muestreo y de medición se toman en cuenta cuando los resultados se presentan. Las estimaciones apresuradas son conclusiones mal tomadas de una población.
5. Para controlar el error de muestreo, debemos concentrarnos en sus dos fuentes específicas. La primera es el tamaño muestral. Cuanto mayor es la muestra, menor es el rango de error. Dado un tamaño muestral, la teoría de probabilidad nos permite decir exactamente con qué frecuencia es que una estadística muestral predecirá correctamente un parámetro. La segunda fuente de error de muestreo es la representatividad muestral. Una muestra no representativa puede llevar a conclusiones erróneas.
6. Para controlar el error de medición, primero debemos especificar una definición operacional y determinar el nivel de medición de una variable.
7. El nivel nominal es el nivel más sencillo de medición. *Nominal* proviene de la palabra latina para referirse al *nombre*. Una variable nominal es aquella que se mide simplemente al dar nombre a categorías.
8. Una variable ordinal es aquella con categorías con nombre y la propiedad adicional de permitir que las categorías se ordenen de más alta a más baja, de la mejor a la peor, o de la primera a la última. Las variables con lugares numéricos, por ejemplo: 1o., 2o. o 3o., también son variables ordinales.
9. Las variables de intervalo tienen las características de las variables nominales y ordinales, más una unidad numérica definida o "intervalo" de medida. Con su unidad establecida de medida, por ejemplo pulgadas, libras o millas, una variable de intervalo se califica numéricamente y está sujeta a numerosos cálculos matemáticos informativos.
10. Las variables de razón tienen las características de las variables de intervalo, más un punto cero real, donde una calificación de cero significa *ninguno*. El punto cero real permite una mayor flexibilidad de cálculos que la que se logra con las variables de intervalo. Con una variable de razón, podemos calcular razones, es decir, la cantidad de una observación en relación con otra.

11. Mientras que los códigos de las variables nominales y ordinales simplemente indican una diferencia en categoría, grupo, calidad o clase, los códigos de intervalo y las variables de razón identifican diferencias en cantidad, cantidad, grado o distancia.
12. Es importante distinguir lo que es el nivel de medición (nominal, ordinal, intervalo o razón) de lo que es la unidad de medida (pulgadas, libras, etcétera).
13. Un registro o guía de codificación es una descripción concisa de los símbolos que significa cada puntuación de cada variable en un conjunto de datos. Toda variable se codifica siguiendo los principios de inclusividad y exclusividad.
14. Una forma eficiente de dominar el carácter de los datos para una variable, es decir, la forma en que las puntuaciones difieren entre sujetos de una población, es organizar los datos en una distribución de frecuencia, una distribución de frecuencia de proporción y una distribución de frecuencia de porcentaje. Los porcentajes son más fáciles de interpretar que las proporciones. Para obtener la distribución de frecuencia de porcentaje, multiplica por 100 cada una de las frecuencias proporcionales.
15. Una frecuencia de porcentaje acumulada se utiliza para identificar cuantiles, es decir, puntuaciones que separan una porción de los casos de una distribución. Los rangos percentilares y los cuartiles son cuantiles que por lo general se presentan. Es importante ordenar las puntuaciones de una variable antes de calcular cuantiles.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las Extensiones del capítulo 2 del material de texto disponible en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2 incluye: *a*) un estudio de cuándo una variable ordinal puede ser tratada como si fuera un nivel de medición de intervalo. Estas variables ordinales "semejantes a intervalos" se pueden usar con procedimientos estadísticos más avanzados para variables de intervalo/razón. *b*) Un estudio de cómo establecer la validez y confiabilidad de una medición son consideraciones importantes para reducir al mínimo el error de medición. *c*) Una ilustración de por qué es obligatorio que el error de medición y el de muestreo se minimicen. Las mediciones perfectas son inútiles con un muestreo deficiente y viceversa.

FÓRMULAS PARA EL CAPÍTULO 2

Cálculo de la frecuencia proporcional de una categoría o puntuación:

$$p\text{ [de la muestra total }(n)\text{ en una categoría]} = \frac{f\text{ de categoría}}{n} = \frac{\#\text{ en categoría}}{n}$$

Cálculo de la frecuencia porcentual de una categoría o puntuación:

$$\%\text{ [de una muestra total }(n)\text{ en una categoría]}$$
$$= (p\text{ [de una muestra total }(n)\text{ en una categoría]})\,(100)$$

Cálculo de percentiles y cuartiles:

Elaborar una hoja de cálculo con los siguientes encabezados:

Puntuación (X)	f	Porcentaje f	Porcentaje acumulado f

El percentil de una puntuación es su frecuencia de porcentaje acumulada. Los cuartiles están ubicados en los percentiles 25o., 50o. y 75o.

PREGUNTAS PARA EL CAPÍTULO 2

1. Explica la diferencia entre una observación y un estadístico.
2. Señala la diferencia entre un estadístico y un parámetro.
3. ¿Cómo se puede demostrar satisfactoriamente que los estadísticos de una sola muestra son sólo estimaciones de los parámetros de una población?
4. Aun cuando Karen creció lejos de cualquier barrio minoritario, se consideraba una persona abierta y sin prejuicios. Se convirtió en trabajadora social y consiguió empleo en el área de recursos humanos que lleva el seguimiento de niños de grupos minoritarios que han sufrido abusos por padres drogadictos. Karen desarrolló abrumadores sentimientos racistas y se convenció de que las minorías son moralmente inferiores. En términos de sesgo de muestreo, ¿qué le sucedió a Karen?
5. Una autoproclamada experta en sexualidad femenina informó que tres cuartas partes de las mujeres declaran que odian a los hombres. Los conductores del programa de entrevistas desafían dicho estadístico. La experta responde diciendo que su muestra de más de 6 000 mujeres es la más grande jamás investigada sobre el tema y que sus resultados tienen un rango de error de menos de un décimo de 1 por ciento. Su método de muestreo fue un cuestionario enviado por correo dentro de una popular revista femenina. Señala las maneras en las que su muestra podría estar sesgada.
6. Por medio de una encuesta telefónica, tú deseas evaluar la opinión del público sobre la propuesta del gobierno para aumentar los impuestos para construir un nuevo zoológico. Si en forma aleatoria seleccionas cada centésimo número del directorio telefónico del área, ¿obtendrías una muestra representativa? Explícalo en términos de sesgo de la muestra.
7. El doctor Burnett prueba una nueva droga antidepresiva en un hospital para enfermos mentales. Se encuentra que la droga es eficaz. ¿Tendrá la droga necesariamente la misma efectividad para todas las personas que padecen depresión? Explica en términos de sesgo de la muestra.
8. Supongamos que estamos midiendo el peso entre una muestra de fisicoculturistas y registramos que Sam pesa unos 114 kilogramos. Él en realidad pesa entre 113.5 y 114.5 kilogramos. Estos pesos son los límites_____ y _____de una puntuación de 114 kilogramos.
9. Jacqueline y Evelin realizaron estudios de especialización en ciencias sociales en una universidad, y a los estudiantes les preguntaron cuántos de ellos esperaban ser empleados en la industria después de graduarse. Jacqueline, con una muestra de 650, obtuvo un estimado de 31 por ciento. Evelin, con una muestra de 45, obtuvo un estimado de 23 por ciento. ¿En cuál estimación podemos tener más confianza?, ¿por qué? Por lo que se refiere a controlar el error de muestreo, ¿qué más necesitamos saber sobre las muestras de Jacqueline y Evelin? ¿Por qué?

10. Bob compró leña para su chimenea de 90 centímetros de ancho. Para asegurarse de que los leños cupieran, tomó una vara de 60 centímetros de largo y se aseguró de que el largo de cada leño variara a lo más 6 pulgadas respecto a la vara. Shara diseña circuitos de microcomputadora. Su instrumento de medición mide al décimo de pulgada más cercano, pero este instrumento es menos preciso que la vara de Bob. Explica.
11. Distingue entre el nivel de medición de conceptos, unidad de medida y unidad de redondeo. Ilustra con ejemplos.

EJERCICIOS PARA EL CAPÍTULO 2

Conjunto de problemas 2A

2A-1. Indica el nivel de medición de las siguientes variables.

Nombre de variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) PROMEDIO	Promedio de calificaciones: número de puntos de calidad académica ganados, dividido entre el número de horas-crédito cubiertas	
b) ESTATURA	Estatura física en pulgadas	
c) PENAMUER	Escala de actitud de 10 reactivos acerca del apoyo a la pena de muerte, con una suma total que va de 0 a 40	
d) AUTOPLAC	Número de placa del automóvil: el número de 7 dígitos	
e) ESTLAB	Situación laboral: 1 = inexperto; 2 = semiexperto; 3 = experto	
f) POBCIUD	Población dentro de límites de la ciudad	
g) MARCAUTO	Marca de automóvil: Ford, Buick, Toyota, etcétera.	

2A-2. En un estudio de mujeres que sufrieron abuso en Sudáfrica (Jewkes y otros, 2002), 1 279 mujeres son encuestadas en un perfil de muestra. Identifica los niveles de medición de las siguientes variables sociodemográficas a partir del estudio.

Nombre de variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) PROVINCIA	Provincia de residencia: 1 = Este del Cabo; 2 = Mpumalanga; 3 = Provincia del Norte	
b) EDUCACIÓN	Educación: 1 = ninguna; 2 = primaria; 3 = principia secundaria; 4 = termina secundaria; 5 = posgraduado (cualquiera)	
c) EDAD	Edad en años: confidencial	
d) OCUPACIÓN	Situación laboral: 1 = desempleado; 2 = comerciante; 3 = doméstica; 4 = profesional; 5 = etcétera	
e) CASADA	Estado civil: 1 = casada por la iglesia; 2 = casada en ceremonia tradicional; 3 = viuda/divorciada/separada; 4 = soltera	
f) PROMISCUA	Índice de promiscuidad familiar: número de personas por cuarto	

2A-3. Imagina que a una muestra de médicos se les pregunta lo siguiente en un cuestionario. Con las categorías de respuesta que se dan, ¿cada variable sigue los principios de inclusividad y exclusividad? Si no es así, ¿cómo pueden mejorarse para estar de acuerdo con estos principios?

a) ¿Cuál es su especialidad médica? (Marque una.)
___General ___Endocrinología ___Oncología
___Reumatología ___Otolaringología ___Psiquiatría

b) ¿Cuántos pacientes ve en un día hábil promedio?
___menos de 10 ___10-20 ___20-30 ___30-40
___40-50

c) Por favor marque la categoría que más cerca se ajuste a su ingreso total anual por su práctica médica.
___\$0-\$25 000 ___\$26 000-\$40 000 ___\$51 000-\$75 000
___\$76 000-\$100 000 ___\$100 000-\$150 000
___\$151 000 o más

d) ¿En qué región completó su educación médica? (Por favor seleccione una.)
___Sur ___Noreste ___Oeste
___Sudoeste ___Medio Oeste

2A-4. Redondea los números siguientes a la unidad de redondeo estipulada.

a) 39.7246 a la décima más cercana
b) 26.3194 a la décima más cercana
c) 29.1352 a la centésima más cercana
d) 29.1352 a la milésima más cercana
e) 692 a la centésima más cercana
f) 41.755 a la centésima más cercana
g) 24.721 a la décima más cercana

2A-5. Especifica los límites reales de los siguientes números redondeados.

a) 0.059
b) 0.0060
c) 2 890 redondeado a la decena más cercana
d) 5 400 redondeado a la centena más cercana
e) Edad 3 redondeada al cumpleaños más cercano
f) Edad 3 redondeada al último cumpleaños
g) 12 pulgadas
h) 12.0 pulgadas

2A.6. La siguiente es la distribución del parentesco de cuidadores de familia a parientes con Alzheimer. Calcula la razón entre el número de mujeres y cuidadores hombres.

Parentesco del familiar cuidador con el paciente	*f*
Esposa	114
Esposo	17
Hija	37
Hijo	4
Hermana	8
Hermano	1
Madre	2
Nuera	13
Otro pariente femenino	24
Otro pariente masculino	2

2A-7. A continuación se presentan datos sobre el número de vehículos registrados para una muestra aleatoria de 20 hogares en el condado de Madison: 2, 1, 2, 4, 2, 3, 4, 2, 1, 4, 2, 1, 0, 3, 2, 4, 3, 4, 2, 2.

a) Compila los datos en una tabla de distribución de frecuencias con columnas para la frecuencia, la frecuencia proporcional, la frecuencia porcentual y la frecuencia de porcentajes acumulada. (No se requiere mostrar las fórmulas.)
b) Si una familia tiene tres vehículos registrados, ¿cuál es el rango percentilar de la familia? Interpreta tu respuesta.

2A-8. A continuación veamos una lista de calificaciones de examen de medio curso para un grupo de 14 estudiantes graduados.

Nombre del estudiante	Calificación de examen de medio curso
Jonathan	76
Susan	82
Jason	95
Andrea	52
Kelli	64
Jennifer	94
James	79
Brian	88
William	69
Caroline	95
Patricia	90
Kevin	98
Mark	88
Jeffrey	92

a) Calcula el rango percentilar de Jeffrey con base en el conjunto de calificaciones del examen presentado.
b) Calcula el rango percentilar de Brian.

Conjunto de problemas 2B

2B-1. Indica el nivel de medición de las variables siguientes.

Nombre de variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) PESO	Peso físico en libras	
b) DENSIDAD	Población (número de personas que residen en una zona definida) por milla cuadrada de área	
c) TASAMORT	Tasa de mortalidad infantil: número de muertes en el primer año de vida por 1 000 nacimientos	
d) ESTUDIANTE	Situación estudiantil: 1 = pasante; 2 = graduado; 3 = especial	
e) ESTIMA	Evaluación de autoestima: escala de resumen de 15 renglones con calificaciones de 0 a 60	
f) SATISTRAB	Satisfacción en el trabajo: 0 = muy insatisfecho; 1 = insatisfecho; 2 = satisfecho; 3 = muy satisfecho	
g) EXPECVIDA	Expectativa de vida: número promedio de años que recién nacidos pueden esperar vivir (usualmente ajustado a sexo y edad)	

2B-2. En un estudio de cuidadores de miembros mayores de familia, 293 mujeres son entrevistadas respecto al estrés de cuidar y mantener empleo al mismo tiempo. Identifica los niveles de medición de cada una de las variables siguientes de este estudio.

Nombre de variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
Variables de cuidador		
a) DISTANCIA	Distancia desde casa al trabajo en millas	
b) EDADCUID	Edad del cuidador en años	
c) SALCUID	Puntuación de salud indicada: 1 = mala; 2 = regular; 3 = buena; 4 = excelente	
Variables de quien recibe cuidados		
d) GÉNERO	Género: 0 = masculino; 1 = femenino	
e) ACTDIA	Actividades diarias: número de actividades que pueden hacer sin ayuda, p.e., bañarse y vestirse	
f) SATVIDA	Satisfacción con la vida: escala de 12 ítems con un rango de puntuaciones de 0 a 36	

2B-3. Imagina que a un grupo de cuidadores se les hacen las siguientes preguntas en un cuestionario. Con las categorías de respuesta que se dan, ¿cada variable sigue los principios de inclusividad y exclusividad? Si no es así, ¿cómo pueden mejorarse para estar de acuerdo con estos principios?

a) ¿Tiene usted preferencia religiosa? Si es así, ¿cuál es? (Por favor marque una.)
___ Protestante ___ Católico ___ Judío
___ Bautista ___ Ninguna

b) ¿Cuál es su estado civil? (Por favor marque uno.)
____Soltero ____Casado

c) Por favor marque la categoría que más cerca se ajusta al ingreso total anual de su familia, de todas las fuentes.
___$0-$10 000 ___ $11 000-$20 000 ___ $21 000-$30 000
___$31 000-$40 000 ___ $41 000-$50 000
___$50 000 o más

d) ¿Tiene empleo actualmente fuera de casa? (Por favor marque uno.)
___ A tiempo completo ___ Tiempo parcial
___ Desempleado ___ Empleado

2B.4. Redondea los siguientes números a la unidad de redondeo señalada:
a) 28.349 a la décima más cercana
b) 31.666 a la centésima más cercana
c) 587 al centenar más cercano
d) 25.6385 a la milésima más cercana
e) 25.6558 a la centésima más cercana
f) 25.6388 a la décima más cercana
g) 25.6388 a la decena más cercana

2B-5. Especifica los límites reales de los siguientes números redondeados:
a) 4.0 pulgadas
b) 4 pulgadas
c) Edad de 5 redondeado al último cumpleaños
d) Edad de 5 redondeado al cumpleaños más cercano
e) 3 300 redondeado al centenar más cercano
f) 3 360 redondeado a la decena más cercana
g) 0.0030
h) 0.068

2B-6. La siguiente es la distribución del número de veces que cada curso está programado durante el año en una universidad importante por disciplina académica. Calcula la razón entre el número de ofertas de cursos de psicología y de sociología.

	Cursos	*f*
Psicología		
	PY 101	5
	PY 109	5
	PY 212	2
	PY 214	3
	PY 217	4
Sociología		
	SOC 100	7
	SOC 120	2
	SOC 220	2
	SOC 235	1
	SOC 240	1

2B-7. Funk y cols. (2003) examinaron la relación entre jugar en juegos de video violentos y los sentimientos de empatía en una muestra de niños. Supongamos que los datos siguientes representan puntuaciones de escala de empatía para una muestra de 20 niños, después de jugar un violento juego de video: 4, 1, 3, 1, 2, 5, 3, 1, 2, 4, 2, 3, 5, 0, 2, 3, 4, 2, 1, 2.

a) Compila los datos en una tabla de distribución de frecuencia con columnas para la frecuencia, la frecuencia proporcional, la frecuencia de porcentaje y la frecuencia de porcentaje acumulada. (No se requiere mostrar fórmulas.)

b) Si un niño obtiene calificación de 2 en la escala experimental de empatía, ¿cuál es el rango de percentil de ese niño? Interpreta tu respuesta.

2B-8. A continuación aparece una lista de calificaciones de examen de medio curso para un grupo de 14 estudiantes graduados.

Nombre del estudiante	Calificación de examen de medio curso
Jeremy	76
Susanna	82
Jack	95
Andrew	52
Anika	64
Jennifer	94
James	79
Brad	88
William	69
Carol	95
Patricia	90
Kevin	98
Mark	88
Jeffrey	92

a) Calcula el rango percentil de Jennifer con base en el conjunto de calificaciones de examen de la lista.

b) Calcula el rango percentilar de Carol.

Conjunto de problemas 2C

2C-1. Indica el nivel de medición de las siguientes variables.

Nombre de la variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) EDAD	Edad numérica en años	
b) EDUCACIÓN	Escolaridad: 0 = sin diploma de secundaria; 1 = secundaria; 2 = universidad inconclusa; 3 = título universitario	
c) REGNACI	Región de nacimiento: 1 = Sur; 2 = Noreste; 3 = Oeste; 4 = Sudoeste; 5 = Medio Oeste; 6 = fuera de Estados Unidos	
d) ACTFIN	Actitudes hacia reforma de financiamiento de campañas: escala de resumen de 15 renglones con calificaciones de 0 a 65	
e) SITELECT	Situación del elector: 1 = no ciudadano; 2 = ciudadano, no registrado; 3 = ciudadano, elector registrado; 4 = ciudadano, sin derecho a voto (es decir, delincuente condenado)	
f) POBCON	Población votante sin distrito electoral	
g) INGPROM	Promedio de ingreso (de electores registrados que residen dentro de una zona geográfica definida)	

2C-2. En un estudio dirigido por Macassa y otros (2003), los datos de encuesta se usan para examinar desigualdades en mortalidad infantil en Mozambique, con base en la posición socioeconómica de padres. Identifica los niveles de medición de cada una de las siguientes variables del estudio.

Nombre de la variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) OCUPMAD	Ocupación de la madre: 1 = no trabaja; 2 = manual no capacitada; 3 = manual capacitada; 4 = profesional; 5 = etcétera	
b) GENNI	Género del niño: 0 = masculino; 1 = femenino	
c) EDADMA	Edad de la madre en años	
d) GRUEDAD	Grupo de edad de la madre al nacer: 1 = 15-18; 2 = 19-23; 3 = 24-28; 4 = 29-33; 5 = 34 o más	
e) ORDNACI	Orden al nacer: 1 = primer nacimiento; 2 = segundo nacimiento; 3 = tercer nacimiento; 4 = cuarto nacimiento o más	
f) RESIDE	Lugar de residencia: 1 = urbano; 2 = rural	

2C-3. Franks y cols. (2003) estudiaron la relación entre algunas características sociodemográficas, la estimación propia de salud, y la mortalidad en Estados Unidos, usando las variables siguientes. Con las categorías hipotéticas de respuesta dadas aquí,

¿sigue cada una de las variables los principios de inclusividad y exclusividad? Si no es así, ¿cómo pueden mejorarse para estar de acuerdo con estos principios?

a) ¿A qué grupo de edad perteneces?
___25-34 ___ 35-44 ___45-54 ___ 55-64
___65-74 ___ 74 o más

b) ¿Cuál es tu raza/etnia?
___ blanca ___ negra ___ latina ___ asiática

c) ¿Cuál es tu nivel educativo (en años de escolaridad)?
___ 9 o menos ___ 9-11 ___ 12 ___ 13-15 ___ 15 o más

d) ¿Cuál es tu nivel actual de ingreso?
___ \$0-\$10 000 ___ \$11 000-\$15 000 ___ \$21 000-\$30 000
___ \$31 000-\$40 000 ___ \$41 000-\$50 000
___ \$50 000 y más

2C-4. Redondea los siguientes números a la unidad de redondeo estipulada:
a) 62.5982 a la decena más cercana
b) 62.5982 a la décima más cercana
c) 62.5982 a la centésima más cercana
d) 62.5982 a la milésima más cercana
e) 723 a la centena más cercana
f) 54.7652 a la centésima más cercana
g) 17.823 a la décima más cercana

25-C. Especifica los límites reales de los siguientes números redondeados:
a) 0.327
b) 0.02381
c) 6 720 redondeado a la decena más cercana
d) 63 000 redondeado a la centena más cercana
e) Edad 7 años redondeado al cumpleaños más cercano
f) Edad 7 años redondeado al último cumpleaños
g) 8 kilogramos
h) 8.00 kilogramos

2C-6. La siguiente es la distribución de estudiantes por nivel de grupo en una importante universidad urbana. Calcula la razón entre el número de estudiantes pasantes y el número de estudiantes graduados.

Nivel de grupo	*f*
Pasante	
De primer año	65
De segundo año	73
Secundaria	61
Preparatoria	48
Graduado	
Título de licenciatura	18
Título de doctorado	22

27C. Reynolds (1997) examinó los efectos del empleo industrial y las condiciones de trabajo en los niveles de estrés psicológico de los trabajadores, usando una versión modificada de la Escala de Depresión (CES-D) del Centro de Estudios Epidemiológicos. Supongamos que los datos siguientes representan puntuaciones del CES-D para una muestra de 20 personas que respondieron: 3, 2, 2, 5, 4, 6, 4, 3, 7, 6, 5, 2, 4, 3, 4, 5, 7, 6, 4, 5.
a) Compila los datos en una distribución de frecuencia con columnas para la frecuencia, la frecuencia proporcional, la frecuencia de porcentaje y la frecuencia de porcentaje acumulada. (No se requiere mostrar las fórmulas.)
b) Si un trabajador obtiene una calificación de 6 en la escala CES-D, ¿cuál es el rango percentil del trabajador? Interpreta tu respuesta.

2C-8. A continuación se presenta una lista de calificaciones del Graduate Record Examination (GRE) para un grupo de 20 estudiantes graduados potenciales que solicitan inscribirse en una universidad importante.

Nombre del estudiante	Calificaciones del GRE
Jack Jones	1380
Valerie Jackson	1400
Robin Schmidt	1220
Jerome González	1410
Richard Roper	1100
James Filer	1190
Rashan Miller	1200
Jeff Wong	1470
Kevin McMillan	1420
Joseph Polanski	1510
Stephanie Nicholson	1450
Alexandra Zimmerman	1520
Jennifer Fitzsimmons	1210
William van der Bergh	1180
Nicholas Andropov	1550
Jacqueline Sheets	1110
Chris Chang	1400
Michael McKee	1450
Sharon Johnson	1380
Ronald Lucie	1220

a) Calcula el rango percentil de Kevin McMillan con base en el conjunto de calificaciones del GRE anotadas.
b) Calcula el rango percentilar de Michael McKee.

Conjunto de problemas 2D

2D-1. Indica el nivel de medición de las variables siguientes.

Nombre de la variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) AFILPAR	Afiliación a partido político: 1 = republicano; 2 = demócrata; 3 = independiente; 4 = otro; 5 = ninguno	
b) EDCOM	Años de educación formal terminada	
c) RELAT	Actitudes hacia la exhibición de iconos religiosos en edificios legislativos/judiciales federales y estatales: escala de resumen de 12 renglones con puntuaciones de 0 a 48	
d) RELDOM	Denominación religiosa que debes identificar con: 1 = bautista; 2 = luteranismo; 3 = catolicismo; 4 = judaísmo; 5 = islamismo; 6 = episcopal; 7 = otra	
e) POLSAT	Satisfacción con el actual sistema político: 1 = muy insatisfecho; 2 = insatisfecho; 3 = satisfecho; 4 = muy satisfecho	
f) CESD	Escala de depresión del Centro de Estudios Epidemiológicos: instrumento de 20 renglones que detecta posible depresión clínica en adolescentes y adultos, con puntuaciones de 0 a 60	

2D-2. En un estudio de cuidadores de niños incapacitados, 141 mujeres fueron entrevistadas respecto al estrés producido por cuidar personas (datos ficticios). Identifica los niveles de medición de cada una de las siguientes variables del estudio.

Nombre de la variable	Definición operacional y codificación (cómo se mide y registra la variable)	Nivel de medición
a) CAL	Calidad percibida de cuidado a paciente: 1 = excelente; 2 = buena; 3 = regular; 4 = mala	
b) INGRESO	Ingreso anual familiar de todas las fuentes, en dólares de Estados Unidos.	
c) DIAGN	Diagnóstico médico de paciente: códigos numéricos de industria de seguros médicos estándar	
d) BAÑO	¿La persona cuidada puede bañarse?	
e) TEMP	Temperatura corporal del paciente en grados Fahrenheit.	

2D-3. Sibicky y cols. (1995) estudiaron la preocupación empática como motivación para ayudar en conductas. Con las categorías de respuesta que se dan, ¿sigue cada variable los principios de inclusividad y exclusividad? Si no es así, ¿cómo pueden mejorarse para estar de acuerdo con estos principios?

a) ¿Fuiste motivado para ayudar por (marca una):
______ cordialidad ______ compasión ______ simpatía?

b) Tienes ______ más de 25 años ______ menos de 25 años?
¿Cuántas veces estarías dispuesto a ayudar a la misma persona?
______0-4 ______ 6-10 ______ 11-15 ______ Ninguna

2D-4. Redondea los siguientes números a la unidad de redondeo especificada:

a) 5.455 a la décima más cercana
b) 5.455 a la centésima más cercana
c) 20.821 a la centésima más cercana
d) 381 a la centena más cercana
e) 467 988 al millar más cercano
f) 467 988 al centenar de millares más cercano
g) 0.00051 a la milésima más cercana

2D-5. Especifica los límites reales de los siguientes números redondeados:

a) 5.00 kilogramos
b) 5 kilogramos
c) Edad 9 redondeado al último cumpleaños
d) Edad 9 redondeado al cumpleaños más cercano
e) 71 000 redondeado a la centena más cercana
f) 9 680 redondeada a la decena más cercana
g) 0.01605
h) 0.248

2D-6. La siguiente es la distribución de trabajadores por título de posición en una empresa de comunicaciones. Calcula la razón entre el número de otros empleados y el número de gerentes.

Título	Número de trabajadores
Posiciones gerenciales	
Presidente	1
Director	1
Vicepresidente	3
Asistente de vicepresidente	8
Asistente administrativo	8
Jefe de personal de piso	12
Otras posiciones	
Secretaria	16
Adjunto de ventas	42
Oficinista	18
Técnico/profesional	14
Aseo	3

2D-7. Pearson y otros (1990) demostraron que cuando una abuela vive con una familia, es probable que participe en actividades de los padres. Supongamos que los siguientes datos representan el número de órdenes paternales dadas a 25 niños por sus abuelas: 5, 4, 3, 3, 6, 5, 3, 2, 4, 7, 5, 6, 2, 3, 4, 8, 7, 5, 6, 4, 2, 1, 5, 7, 3.

a) Compila los datos en una tabla de distribución de frecuencia con columnas para la frecuencia, frecuencia proporcional, la frecuencia de porcentaje y la frecuencia de porcentaje acumulada. (No se requiere mostrar fórmulas.)

b) Si una abuela dio dos órdenes, ¿cuál es su rango percentil? Interpreta tu respuesta.

2D-8. A continuación aparece una lista de calificaciones del Graduate Record Examination (GRE) para un grupo de 20 estudiantes graduados potenciales que solicitan inscribirse en una universidad importante.

Nombre del estudiante	Calificaciones del GRE
Jack Jones	1380
Valerie Jackson	1400
Robin Schmidt	1220
Jerome Gonzalez	1410
Richard Roper	1100
James Filer	1190
Rashan Miller	1200
Jeff Wong	1470
Kevin McMillan	1420
Joseph Polanski	1510
Stephanie Nicholson	1450
Alexandra Zimmerman	1520
Jennifer Fitzsimmons	1210
William van der Bergh	1180
Nicholas Andropov	1550
Jacqueline Sheets	1110
Chris Chang	1400
Michael McKee	1450
Sharon Johnson	1380
Ronald Lucie	1220

a) Calcula el rango percentilar de Jeff Wong con base en el conjunto de calificaciones del GRE de la lista.

b) Calcula el rango percentilar de Chris Chang.

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 2

Si en tu clase se utilizan las aplicaciones de computadora optativas que acompañan a este texto, descarga los ejercicios del capítulo 2 en el sitio web de *The Statistical Imagination*, www.mhhe.com/ritchey2. Los ejercicios comprenden (1) crear y guardar archivos de datos (*.SAV) usando el *SPSS para Windows*; (2) producir distribuciones de frecuencias, cuartiles y percentiles, y (3) administrar y guardar archivos de resultados. Otras instrucciones aparecen en el apéndice D de este texto.

El control de calidad de la codificación y entrada de los datos es muy importante para reducir al mínimo los errores. Para alcanzar estándares científicos y éticos, un investigador debe tener aptitud para mirar a sus colegas a los ojos y asegurarles honestamente que el conjunto de datos no tiene ningún error aleatorio. Además del aspecto ético, el cometer errores en el acceso de los datos representa un desperdicio de tiempo y energía en las posteriores fases del análisis. Un investigador ético pero descuidado quizá pase meses analizando un conjunto de datos y después encuentre datos que son incorrectos. El descubrimiento de un solo dato incorrecto requiere de un nuevo y completo análisis de los mismos. Por tanto, debemos verificar diligentemente la exactitud de un conjunto de datos antes de empezar el análisis.

A continuación se ofrecen algunos lineamientos para la detección de errores de codificación y de entrada de datos autocodificados o archivos de datos existentes.

Lineamientos de control de calidad para la entrada de datos

1. Asegúrate de que los valores de los códigos ingresados sean consistentes con el registro de codificación y los instrumentos de medición (por ejemplo, cuestionarios).
2. Si tienes habilidades limitadas en mecanografía o tu vista no es buena, solicita que un asistente lea y verifique dos veces los datos.
3. Haz una nueva verificación de los códigos de una salida impresa que presente una lista de todas las variables y sus códigos.
4. Utiliza salidas impresas de distribuciones de frecuencias para verificar códigos raros (es decir, códigos que no deben estar presentes en los datos).

CAPÍTULO

3

Tablas y gráficas: una imagen dice más que mil palabras

RESUMEN DEL CAPÍTULO

Introducción: representación gráfica de datos

En tiempos recientes, el análisis estadístico se refiere a simplificar y resumir grandes cantidades de información. Si el lector es un científico o sólo es un observador informal de lo que pasa a su alrededor, es probable que sus intereses sean entender el comportamiento de una gran población de sujetos u objetos. Por ejemplo, recuerda tu primer día en el campus. Es probable que hayan surgido muchas preguntas en tu mente cuando tratabas de ajustarte a todas las nuevas actividades relacionadas con este interesante entorno. Quizá te preguntaste: ¿qué tan grande es el campus? ¿Cuántos estudiantes están inscritos aquí? ¿Hay muchos de mi pueblo o que estudian una especialización en mi materia? ¿En lo que se refiere a mi capacidad para tener éxito al menos soy un estudiante promedio? ¿Es éste un lugar en donde me sentiré cómodo?

Es natural que los seres humanos deseen simplificar y organizar sus percepciones del mundo que les rodea con sólo resumirlas. Para orientarnos a una nueva situación, rápidamente buscamos nuevas generalizaciones que describen esta gran imagen. Deseamos simplificar las complejidades que nos asaltan para que nos comportemos de una manera apropiada y eficiente y no ser engañados. Tal como lo vimos en los capítulos 1 y 2, hay tanta información que debemos tener cuidado de mantener un sentido de proporción y equilibrio en la forma en que la interpretamos y reaccionamos a ella. Un sentido de proporción acerca de la realidad puede medirse con números o imágenes.

Para transmitir un sentido de proporción, describimos numéricamente la distribución de las puntuaciones de una variable con frecuencias porcentuales, como lo hicimos en el capítulo 2. Las distribuciones numéricas, sin embargo, tienen sentido sólo si una persona tiende a pensar de manera proporcional. Los gráficos constituyen un excelente soporte a la célebre máxima:

"Una imagen dice más que mil palabras." Cuando algunas de esas palabras son numéricas, una imagen hace más claros cientos de palabras y cálculos.

Debido a lo fáciles que son actualmente los programas de cómputo, los medios de comunicación masiva nos bombardean con gráficos, gráficos de pastel y pictográficos (fotografías de objetos, iconos o mapas sombreados). A veces, los programas de cómputo toman vida propia y crean un gráfico de forma incorrecta. En términos del control del error, es importante conocer los cálculos matemáticos que hay detrás de la construcción gráfica y no sólo confiar en los programadores.

Los diseños gráficos y pictóricos se eligen con base en (1) el nivel de medición de una variable, (2) los objetivos y los aspectos relevantes del estudio y (3) el público a quien se dirigen. Para las audiencias públicas, los gráficos sencillos y a todo color funcionan mejor y brindan una perspectiva global de los estadísticos descriptivos, tales como porcentajes y promedios. En contraste, los públicos compuestos por especialistas están acostumbrados a los estadísticos inferenciales, aquéllos diseñados para explicar y probar hipótesis. Junto con las tablas estadísticas, los gráficos nos ayudan a discernir las formas de las distribuciones de frecuencias. Incluso los gráficos descriptivos alertan a un analista sobre fuentes de error potenciales que puedan influir en el análisis realizado.

Lineamientos para graficar

Las presentaciones gráficas deben cumplir con algunas reglas y lineamientos simples, los cuales también se aplican a las tablas y a la elaboración de reportes.

Lineamientos para graficar

1. Elige el diseño con base en *a*) el nivel de medición de una variable, *b*) los objetivos del estudio y *c*) el público a quien se dirige.
2. Ante todo, una buena presentación gráfica tiene que ser clara y entendible. Debe simplificar, no complicar.
3. Un gráfico o diagrama requiere explicarse por sí mismo y transmitir información, sin hacer referencia a un texto o a alguien que lo explique. La selección cuidadosa de títulos, descripción de la escala, subtítulos y otras leyendas contribuyen a lograr este objetivo. Somete cada gráfico o tabla a la prueba de "perdido en el estacionamiento". Pregúntate: si este gráfico fuera abandonado en un estacionamiento, ¿podría tomarlo un perfecto extraño e interpretarlo?
4. Antes de decidirte sobre el tipo de presentación pictórica (por ejemplo, gráfico de pastel contra gráfico de barras), elabora bosquejos con varias opciones. Los programas de cómputo hacen esto en forma relativamente fácil. Para ampliar las alternativas, solicita opiniones y consulta otros materiales, tales como informes organizacionales.
5. Adhiérete a los principios de inclusividad (capítulo 2). Anota al pie de página cualquier excepción.
6. Si los datos no son tuyos, indica la fuente de los mismos al final de la tabla.

Graficación de datos nominales/ordinales

Gráficos de pastel

Un estilo sencillo de presentación para datos nominales/ordinales es el gráfico de pastel. Un **gráfico de pastel** es *un círculo que se divide (o rebana) desde su punto central, donde cada rebanada representa la frecuencia proporcional de determinada categoría.* Todos hemos rebanado pasteles y, a veces, no obtenemos una porción justa. Cuando un investigador quiere ofrecer un sentido de proporción respecto de una variable nominal/ordinal, los gráficos de pastel son especialmente útiles, pues representan de manera equitativa el tamaño relativo o desigualdad entre las categorías. El tamaño relativo de una rebanada de pastel es una forma de pensamiento proporcional, con la cual todos estamos familiarizados.

> **Gráfico de pastel** Círculo que se divide (o rebana) desde su punto central, donde cada rebanada representa la frecuencia proporcional de determinada categoría de una variable nominal/ordinal. Es especialmente útil para transmitir un sentido de equidad, tamaño relativo o desigualdad entre las categorías.

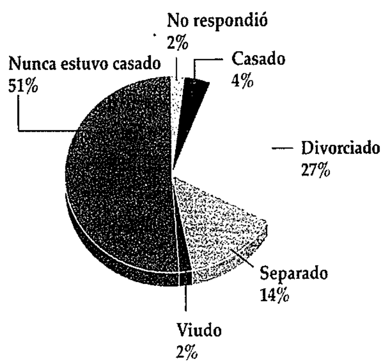
La figura 3-1 muestra la distribución del estatus marital en una muestra de 161 personas sin hogar. El área dentro del círculo entero representa el 100 por ciento de los sujetos en la muestra. El área de una rebanada indica el porcentaje en una categoría específica. Es fácil percibir que más de la mitad de los encuestados nunca estuvieron casados y que una porción sustancial se había divorciado. La revelación más sorprendente es la parte pequeña correspondiente a "casado".

Mientras un programa de cómputo sin dificultad produjo el gráfico de pastel de la figura 3-1, para establecer la relación entre el gráfico de pastel y la distribución de frecuencias del estado civil elaboremos el gráfico a mano. El primer paso para elaborar cualquier gráfico consiste en determinar la distribución de frecuencias de la variable. Del mismo modo, con los gráficos de pastel calculamos la frecuencia proporcional y la frecuencia porcentual de cada categoría. Las frecuencias proporcionales, junto con el conocimiento sobre las dimensiones de un círculo, sirven para calcular el tamaño de las rebanadas.

La división correcta del pastel depende de saber que los ángulos que cortan un círculo desde su centro se miden en grados con un transportador (regla circular con forma de media luna). Sin importar el tamaño de un círculo, su circunferencia se define por tener una distan-

FIGURA 3-1

Diagrama del estatus mental de personas sin hogar, $n = 161$



cia circular total de 360 grados (°). Medio círculo tiene 180° o una proporción de 0.5 por 360 grados; y un cuarto de círculo, 90° o ¼ de 360 grados. Estos puntos de referencia del círculo se ilustran en la figura 3-2. Cualquier parte del pastel se rebana multiplicando la proporción en una categoría por 360 grados.

La tabla 3-1 muestra los cálculos para el gráfico de pastel que aparece en la figura 3-1. ¿Cómo se le asigna a cada categoría su porción del pastel? La porción de una categoría es la frecuencia proporcional p por 360°, la circunferencia total del pastel. Si 0.51 (51 por ciento) de los encuestados nunca habían sido casados, deben "obtener" 51 por ciento del pastel, es decir un valor de 185°, o simplemente más de la mitad. Después que los 360 grados hayan sido asignados a todas las categorías, se utiliza un transportador para trazar en el papel las porciones correctamente proporcionadas. Para lograr mayor claridad, se anotan los porcentajes en el gráfico de pastel. Si el gráfico se presenta ante una audiencia pública, redondeamos a un porcentaje entero (es decir, la posición de las unidades).

Interpretación de los gráficos de pastel Regresemos a la figura 3-1, que presenta la distribución de estado civil para una muestra de 161 personas sin hogar, y sistemáticamente interpretemos su significado. Primero nos concentramos en las rebanadas más grandes, que representan categorías que ocurren con las mayores frecuencias. En la figura 3-1 es fácil ver que más de la mitad de quienes respondieron y no tienen casa nunca habían sido casados y muchos de ellos eran divorciados o separados. Segundo, comparemos entre sí los tamaños de las rebanadas. Un porcentaje de personas sin casa que nunca habían sido casados era más alto que todas las otras categorías combinadas. Las personas que no tenían pareja, es decir que nunca se habían casado, divorciado o separado, constituían un segmento grande de personas sin hogar. Tercero, comparemos los resultados con otras poblaciones. Busquemos rebanadas inesperadamente pequeñas o grandes de esta población en comparación con otras poblaciones. Lo más sorprendente de la figura 3-1 es la pequeña pieza correspondiente a "casados". Por ejemplo, en la población general de adultos, 56.7 por ciento de los adultos de más de 18 años son casados en comparación con sólo el 4 por ciento de esta muestra de personas sin hogar (U.S. Bureau of the Census 2003). En resumen, no tener casa no es benéfico para el estado civil.

En la tabla 3-1 nótese que el porcentaje de la categoría "Nunca estuvo casado" se redondea hacia abajo a 51 por ciento en lugar de hacia arriba a 52 por ciento. Es más, asignamos a esta categoría 185° en lugar de 186°. Estos ajustes por error de redondeo son necesarios para impedir que el total de grados del círculo exceda los 360°, pues un círculo tiene un espacio definido de esa magnitud. Ajustar la categoría "Nunca estuvo casado" tiene el efecto más pequeño de error en comparación con el ajuste en otras categorías.

FIGURA 3-2

Grados de un círculo: un cuarto, medio y círculo completo.

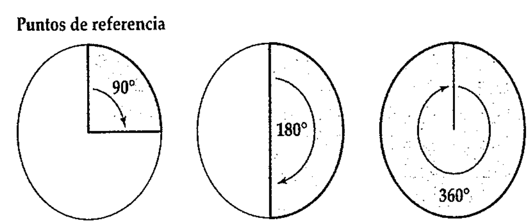


TABLA 3-1 | Hoja de cálculo para construir un gráfico de pastel: distribución de estado civil para una muestra de 161 personas sin hogar

Especificaciones		Cálculos		
Estado civil	**f**	**p**	**(p)(360°)**	**Porcentaje (%)**
Nunca estuvo casado	83	.5155	185°	51
Divorciado	43	.2671	96	27
Separado	22	.1366	49	14
Casado	7	.0435	16	4
Viudo	3	.0186	7	2
No respondió	3	.0186	7	2
Totales	161	.9999*	360°	100

*El total no sumó 1.0000 por el error de redondeo.

Los paquetes de software (por ejemplo el *SPSS for Windows*, que es una opción con este texto) ofrecen una amplia gama de estilos para gráficos de pastel. Una o más rebanadas pueden presentarse en relieve o "piezas desarmadas", y se pueden presentar pares de gráficas de pastel para comparar grupos o periodos.

Forma de construir e interpretar una gráfica de pastel

Para construir una gráfica de pastel:

1. Elabora una tabla de distribución de frecuencia con los siguientes encabezados:

Categoría	f	p	(p)(360°)	(%)

donde

"Categoría" = nombre de la categoría de una variable nominal/ordinal,

f = frecuencia de casos (o número de casos) de una categoría,

p = p [del n total en una categoría] = (f de categoría)/n, con n = tamaño muestral,

$(p)(360°)$ = grados para cada rebanada,

% = porcentaje [del n total en una categoría] = $(p)(100)$.

2. Traza un círculo y pon un punto en su centro. Traza una recta del punto al círculo. Pon un transportador sobre esta recta, marca el número de grados para la primera categoría y traza una recta para crear la rebanada del pastel. Pon un transportador sobre esta segunda recta, marca el número de grados para la segunda categoría, y así sucesivamente. Asegúrate de que haya exactamente 360°.

3. Marca con toda claridad cada rebanada e indica el porcentaje de casos que representa. Asegúrate de que las leyendas sean horizontales (es decir, no las ajustes a la forma circular de la rebanada). Utiliza tu juicio de dibujante para poner las leyendas dentro o fuera de las rebanadas.
4. Titula adecuadamente el gráfico de pastel. Identifica la fuente de datos en la parte inferior del gráfico.

Para interpretar un gráfico de pastel:

1. Concéntrate en las rebanadas más grandes del pastel. Estipula las categorías que se presentaron con mayor frecuencia.
2. Compara entre sí los tamaños de rebanadas.
3. Si es apropiado, compara los resultados con otras poblaciones. Busca rebanadas inesperadamente pequeñas o grandes (por ejemplo, el pequeño porcentaje de personas casadas entre adultos sin hogar de la figura 3-1).

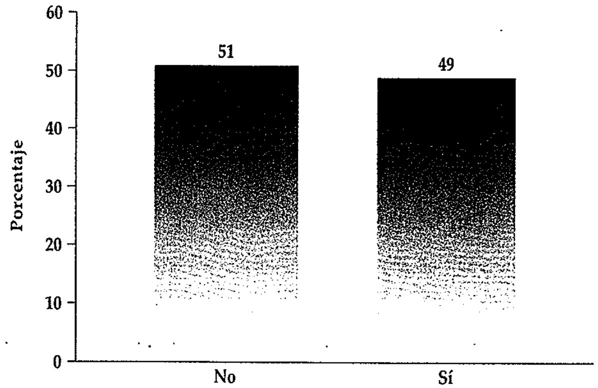
Gráficos de barras

Otra manera de graficar datos nominales/ordinales consiste en utilizar un gráfico de barras. Un **gráfico de barras** *se compone de una serie de barras verticales u horizontales, donde la longitud de la barra representa la frecuencia porcentual de una categoría de una variable nominal/ordinal.* Al igual que una rebanada de un gráfico de pastel, el área de una barra determinada por su longitud transmite un sentido de frecuencia proporcional de una categoría. Los gráficos de barras se construyen sobre dos ejes: uno trazado horizontalmente (el de las *abscisas*) y el otro colocado en forma vertical (el de las *ordenadas*). En otras palabras, las dos líneas se unen en un ángulo de 90° o ángulo recto. Las categorías de una variable se sitúan en un eje y las marcas para los porcentajes en el otro. Sólo necesitamos calcular la frecuencia porcentual de cada categoría para elaborar un gráfico de barra.

Gráfico de barras Serie de barras verticales u horizontales, donde la longitud de la barra representa la frecuencia porcentual de una categoría de una variable nominal/ordinal. Los gráficos de barras son especialmente eficaces para ilustrar una competencia entre categorías.

Interpretación de gráficos de barras La figura 3-3 presenta un gráfico de barras de la distribución de frecuencia porcentual de empleo de adultos sin hogar en la semana anterior a una entrevista. Primero, observa la altura de las barras. La barra más alta es la categoría con la frecuencia más alta. Haz algún comentario sobre el orden de las categorías. En la figura 3-3 vemos que más personas sin hogar no trabajaron en la semana previa que las que sí trabajaron. Segundo, compara las barras y comenta sobre algunas de estas que sean especialmente altas o cortas. En la figura 3-3 notamos que las barras tienen más o menos la misma altura. Tercero, si es apropiado, compara los resultados con otras poblaciones. Busca barras inesperadamente altas o cortas. Un investigador experto señalaría que los resultados contradicen

FIGURA 3-3
Gráfico de barras de adultos sin hogar en la semana anterior a una entrevista, n = 161



la idea comúnmente admitida de que las personas sin hogar son vagos o indolentes. Si se consideran las circunstancias de quienes no tienen hogar, uno esperaría un porcentaje mucho más bajo de trabajo reciente que el que revela o se deja ver en esta muestra. A este respecto, entonces, el porcentaje de quienes trabajan es más alto de lo esperado. En resumen, la similitud de las alturas de las barras rápidamente da a entender que hay casi tantos sin hogar que trabajan como los que no trabajan.

Forma de construir e interpretar un gráfico de barras

Para construir un gráfico de barras:

1. Elabora una tabla de distribución de frecuencias con los encabezados siguientes:

Categoría	f	p	(%)

donde

"Categoría" = nombre de la categoría de una variable nominal/ordinal,

f = frecuencia de casos (o número de casos) en una categoría,

$p = p$ [del n total de una categoría] = (f de categoría)/n, con n = tamaño muestral,

% = porcentaje [del n total de una categoría] = $(p)(100)$.

2. Traza el eje horizontal del gráfico de barras, con un ancho apropiado para el número de barras.
3. Traza el eje vertical. Observa la frecuencia más alta (f) o frecuencia porcentual de la tabla de distribución de frecuencia y escribe marcas en el eje que vayan desde cero hasta un poco más de la frecuencia más alta.

4. Traza las barras con un ancho que sea visualmente atractivo. Puedes escoger el orden de las barras de la más alta a la más baja. Separa las barras de cada categoría. Utiliza frecuencias (f) o frecuencias porcentuales sobre el eje vertical como marcadores de altura de barras. Bajo las barras incluye una leyenda clara con los nombres de las categorías.
5. Escribe un título apropiado para gráfico de barras. Asegúrate de que los títulos que apliques a los ejes sean precisos y claros. En la parte inferior del gráfico identifica la fuente de los datos.

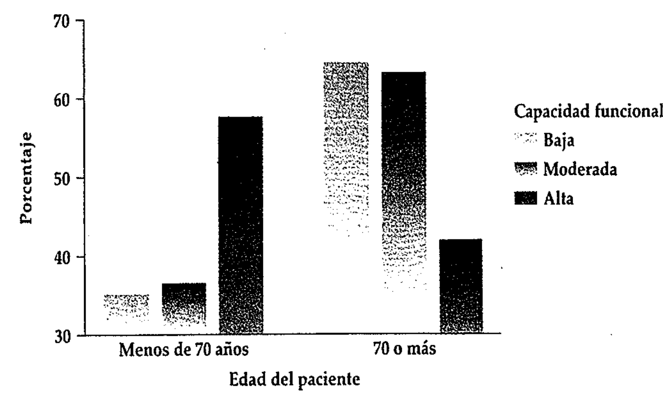
Para interpretar un gráfico de barras:

1. Observa la altura de las barras. La barra más alta indica la categoría que tiene la frecuencia más alta. Haz comentarios sobre el orden de las categorías.
2. Compara las barras y comenta sobre cualquiera que sea especialmente alta o corta.
3. Si es apropiado, compara los resultados contra otras poblaciones. Busca barras especialmente altas o cortas.

La figura 3-4 presenta un gráfico de barras "agrupadas", que es muy útil para comparar dos o más grupos en una variable nominal/ordinal. Esta figura compara la habilidad funcional de 104 pacientes de un hospital de veteranos y deja ver cómo la baja capacidad funcional es tan característica en veteranos enfermos mayores de 70 años.

Por último, el eje vertical de un gráfico de barras no siempre mide simplemente números o porcentajes. Entre las categorías de una variable nominal/ordinal los gráficos de barras se pueden usar para expresar cantidades relativas de cualquier variable. Por ejemplo, el ejercicio 3B-2, que aparece en los ejercicios al final de este capítulo, se refiere a un gráfico de barras sobre el consumo de alcohol para cinco países europeos. La variable nominal es el país, y los nombres de países se aplican por todo el eje horizontal del gráfico de barras. En el eje vertical, en lugar de indicar un número o porcentaje, la escala será "litros de alcohol consumido". Los valores a graficar se ven directamente en la tabla de ese ejercicio. Del mismo modo, podríamos construir un gráfico de barras que haga una comparación de los ingresos medios de estos cinco países con las cantidades en dólares indicadas en el eje vertical.

FIGURA 3-4
Gráfico de barras agrupadas referente a la capacidad funcional de pacientes de un hospital de veteranos por edad, n = 104



Graficación de variables de intervalo/razón

Histogramas

Un histograma es un tipo de gráfico que se utiliza con variables de intervalo/razón. Una de estas variables de razón, que definimos como X, es la de evaluaciones de rendimiento de combustible proporcionadas por la Environmental Protection Agency (U.S. Department of Energy, 2004). Estas evaluaciones se estiman en millas por galón (MPG), que se fijan en modelos nuevos de vehículos. De nueva cuenta, el primer paso para cualquier gráfico es elaborar una distribución de frecuencias. La hoja de trabajo de cálculo de la tabla 3-2 presenta la distribución de frecuencia de evaluaciones de rendimiento de combustible, para conducción en la ciudad de modelos de autos compactos de cuatro cilindros del año 2004 (excluyendo modelos híbridos de gasolina/eléctricos). Nuestro interés está en cómo se agrupan las puntuaciones y en cómo se dispersan. Con toda facilidad podemos ver, por ejemplo, que la evaluación mínima fue 18 MPG, y la máxima, de 38 MPG. Si observamos las evaluaciones de rendimiento de combustible con alta frecuencia (es decir, aquellas para las que f es grande),

TABLA 3-2 | Hoja de cálculo para construir histogramas y polígonos: distribución de frecuencia de evaluaciones de rendimiento de combustible para conducción en ciudades, en millas por galón (MPG); 106 modelos de autos compactos; modelo 2004

Especificaciones		Cálculos
X Evaluación de rendimiento de combustible (MPG)	f	Límites reales
18	1	17.5-18.5
20	4	19.5-20.5
21	6	20.5-21.5
22	19	21.5-22.5
23	10	22.5-23.5
24	17	23.5-24.5
25	9	24.5-25.5
26	13	25.5-26.5
27	5	26.5-27.5
28	5	27.5-28.5
29	4	28.5-29.5
30	1	29.5-30.5
31	1	30.5-31.5
32	3	31.5-32.5
33	3	32.5-33.5
35	2	34.5-35.5
36	1	35.5-36.5
38	2	37.5-38.5
Total	106	

Fuente: U.S. Department of Energy, 2004.

podemos ver que muchos modelos de autos compactos están proyectados para rendir entre 22 y 26 MPG en condiciones de conducción en ciudades. (Para personas acostumbradas a medir el rendimiento de combustible en unidades métricas, multiplica MPG por 0.42 para obtener el equivalente en kilómetros por litro de gasolina.)

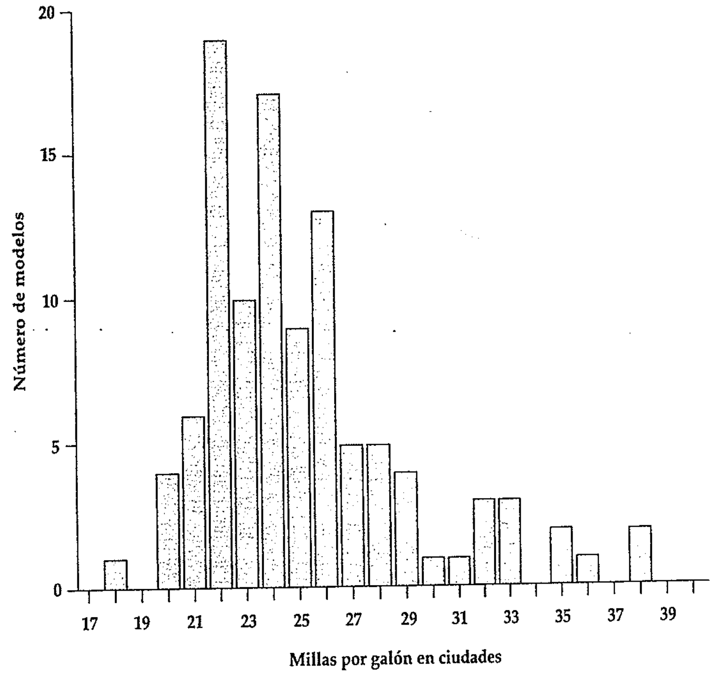
Al hacer un gráfico de datos, obtenemos un sentido de proporción incluso mejor acerca de cómo las evaluaciones están distribuidas para autos compactos. La figura 3-5 presenta las evaluaciones de rendimiento de combustible de la tabla 3-2 en forma de un histograma de frecuencia. Un **histograma de frecuencias** es *un gráfico de 90° que presenta las puntuaciones de una variable a lo largo del eje horizontal, y la frecuencia de cada puntuación en una columna paralela al eje vertical.* En otras palabras, se grafica X sobre el eje horizontal y f en el vertical. Un histograma es semejante a un gráfico de barras, excepto que las columnas de un histograma se tocan entre sí (a menos que una puntuación tenga una frecuencia de cero casos, en cuya circunstancia habrá una columna faltante). Los puntos graficados sobre el eje horizontal reciben el nombre de marcas y representan las puntuaciones X (en este ejemplo, evaluaciones en MPG). Nótese que las marcas más bajas y más altas, 17 y 39 MPG, respectivamente, se prolongan un poco por debajo de las puntuaciones mínima y máxima observadas de 18 y 38 MPG. Para trazar y aplicar una leyenda al eje vertical, en la tabla de distribución de frecuencia identifica la frecuencia más alta (f) registrada. En la tabla 3-2 ésta es una frecuencia de 19 para una puntuación de 22 MPG. Aplica una leyenda al eje vertical de 0 a un poco más de 19. El ancho de cada columna del histograma es igual. Las columnas que se tocan entre sí son los límites reales de cada puntuación. Por ejemplo, las puntuaciones de la tabla 3-2 están redondeadas al entero más cercano; en otras palabras, cada modelo evaluado a 22 MPG no rindió exactamente 22 MPG en los cálculos de la Environmental Protection Agency. Tomar en cuenta los límites reales ensancha las columnas en forma tal que las columnas adyacentes se tocan entre sí y satisfacen el principio de inclusividad.

> **Histograma de frecuencia** Es un diagrama de 90 grados que presenta las puntuaciones de una variable de intervalo/razón a lo largo del eje horizontal, y la frecuencia de cada puntuación en una columna paralela al eje vertical.

Interpretación de histogramas ¿Qué mensaje expresa el histograma de la figura 3-5? Hay varias características de los histogramas que transmiten información. Primero, observa la altura de las columnas. Podemos ver que la evaluación de rendimiento de combustible que se presenta con más frecuencia es 22 MPG, y las evaluaciones de 24 y 26 millas también son bastante comunes. Segundo, busca grupos de puntuaciones y ve si hay una "tendencia central", es decir, un valor de puntuación alrededor del que se centra la distribución. Podemos ver que el grueso de los modelos rinden entre 20 y 29 MPG. Además, con excepción de unas pocas puntuaciones especialmente altas, las evaluaciones tienden a centrarse alrededor de 24 MPG. Tercero, busquemos la simetría o equilibrio en la distribución de puntuaciones. Podemos ver que este histograma no es simétrico. El grueso de las puntuaciones X se apoya sobre el extremo inferior de la distribución, y unos cuantos modelos, los que rinden arriba de 30 MPG, tienen evaluaciones de rendimiento de combustible especialmente altas. (En el capítulo 4 definiremos la forma de esta distribución de puntuación como un sesgo a la derecha.)

FIGURA 3-5

Histograma de frecuencias de las evaluaciones de rendimiento de combustible de 106 modelos de autos compactos; modelo 2004



Fuente: U.S. Department of Energy, 2004.

Forma de construir e interpretar histogramas

Para construir un histograma:

1. Elabora una tabla de distribución de frecuencias con los encabezados siguientes:

Puntuación (X)	f	Límites reales de puntuación

donde

Puntuación = puntuación de una variable de intervalo/razón

f = frecuencia de casos (o número de casos) en una puntuación

2. Calcula los límites reales de cada puntuación. (En el capítulo 2 puedes revisar los límites reales.)
3. Traza un eje horizontal del histograma. Observa las puntuaciones más baja y más alta en la tabla de distribución de frecuencias. Escribe marcas sobre el eje y aplica de conformidad los valores de X; deja espacio adicional en cada extremo del eje fuera de los valores de las puntuaciones más baja y más alta.

4. Traza el eje vertical. Observa la frecuencia más alta (f) en la tabla de distribución de frecuencias y escribe marcas en el eje que vayan desde cero hasta un poco más de la frecuencia más alta.
5. Traza las columnas usando límites reales como marcadores para anchos de columna y frecuencias (f) como marcadores para alturas de columna.
6. Aplica un título preciso al histograma. Asegúrate de que las leyendas de ejes sean correctas y claras. Identifica la fuente de datos en la parte inferior del gráfico.

Para interpretar un histograma:

1. Observa la altura de las barras. La columna más alta indica el valor de la puntuación de X que tenga la frecuencia más alta (f).
2. Busca grupos de puntuaciones y ve si hay una "tendencia central", un valor de puntuación de X alrededor del que se centra la distribución.
3. Busca la simetría o equilibrio en la distribución de las puntuaciones. ¿Las puntuaciones tienden a ubicarse en forma homogénea alrededor de una puntuación central, o son puntuaciones especialmente bajas o altas (como en la figura 3-5)?

Polígonos y gráficos de líneas

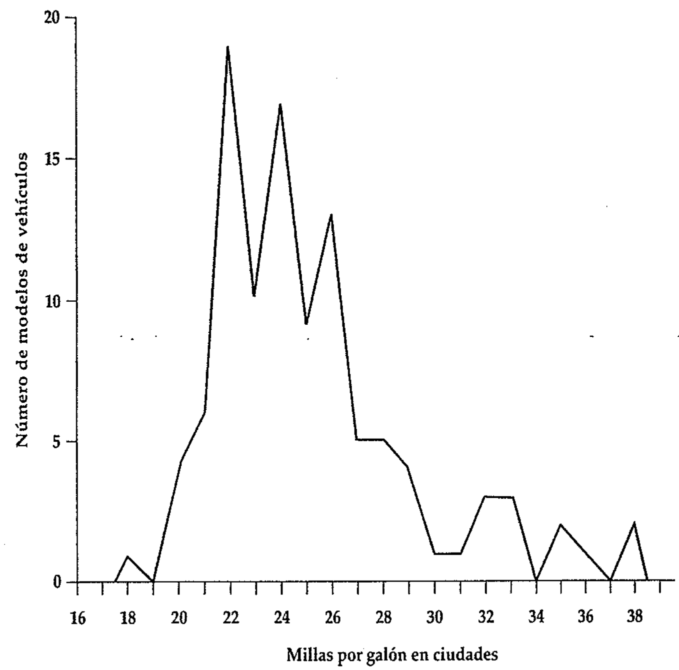
Otra técnica gráfica para representar variables de intervalo/razón es el polígono de frecuencia o *gráfico de líneas*. Un **polígono de frecuencias** es *un diagrama de 90 grados con puntuaciones de intervalo/razón señaladas en el eje horizontal o línea base, y las frecuencias de las puntuaciones están representadas por las alturas de puntos localizados sobre las puntuaciones y enlazados mediante líneas rectas.* Los ejes de un polígono se diseñan como los de un histograma. Los valores de X están marcados en el eje horizontal o línea base. Las frecuencias (f) se grafican en el eje vertical, pero, para indicar la frecuencia de la variable en una puntuación particular, usamos puntos en lugar de columnas y enlazamos los puntos para obtener una línea gráfica. La figura 3-6 es el polígono para las evaluaciones de rendimiento de combustible de la tabla 3-2, que usamos para el histograma de la figura 3-5. Mientras que los histogramas atraen la atención a las columnas más altas, donde se encuentra el grueso de las puntuaciones, los polígonos comunican un sentido de tendencia o movimiento. Esto es, observamos el flujo de picos y valles en la línea gráfica cuando comparamos las evaluaciones de rendimiento de combustible de la más baja a la más alta por toda la línea de base.

Polígono de frecuencias (gráfico de líneas) Gráfico de 90 grados con la puntuación de intervalo/razón trazada sobre el eje horizontal, y las frecuencias de puntuación descritas por las alturas de puntos localizados sobre puntuaciones y enlazados por líneas rectas.

Interpretación de los polígonos (gráficos de líneas) ¿Qué mensaje expresa el polígono de la figura 3-6? Las notables características de los polígonos son semejantes a las de los histogramas. Primero, identifica el pico más alto del polígono. Este pico se manifiesta para una evaluación de rendimiento de combustible de 22 MPG y ésta es la puntuación que

FIGURA 3-6

Polígono de frecuencia de la distribución de evaluación de combustible en la ciudad en millas por galón (MPG) para autos compactos, modelo 2004



Fuente: U.S. Department of Energy, 2004.

ocurre con más frecuencia. Los picos ligeramente más bajos se muestran para evaluaciones de rendimiento de combustible de 24 y 26 MPG, revelando estos valores que también se presentan con frecuencia. Segundo, busca una extensión de espacio bajo la línea gráfica y ve si hay una tendencia central. Podemos ver que casi toda el área bajo la línea gráfica se apoya entre las puntuaciones de 20 y 29 MPG, y las evaluaciones tienden a centrarse alrededor de 24 MPG. Hay unas pocas puntuaciones especialmente altas, con más de una puntuación de 30 MPG. Tercero, busca la simetría o equilibrio en la distribución de las puntuaciones y busca una tendencia en la forma de la línea gráfica. Ve si las colas de la línea gráfica se extienden de manera uniforme. En la figura 3-6 la "cola de dragón" que se prolonga a la derecha hace resaltar el desequilibrio de la distribución. En tanto que casi todos los modelos de autos compactos tienen evaluaciones de rendimiento debajo de 30 MPG, unos cuantos modelos son especialmente eficientes. En general, el polígono expresa el mensaje de que conforme nos movemos hacia arriba en la escala de evaluaciones de rendimiento de combustible, hacia un mayor rango de MPG, existen menos modelos.

Los polígonos de frecuencia son especialmente útiles para comparar dos o más muestras. Por ejemplo, comparemos las distribuciones de evaluaciones de rendimiento de combustible para autos compactos contra los vehículos utilitarios de tracción en las cuatro ruedas (SUV). La tabla 3-3 da las distribuciones de frecuencia y frecuencia porcentual para ambos tipos de vehículos. Nótese que difieren los tamaños muestrales de los tipos de vehículos. Hay 106 modelos de autos compactos pero sólo 68 modelos de los SUV. Si usamos frecuencias sin elaborar para construir los polígonos, el polígono para los autos compactos más numerosos hará empequeñecer el polígono para los SUV. Entonces, usamos un común denominador de

Forma de construir e interpretar polígonos (gráficos de línea)

Para construir un polígono:

1. Elabora una tabla de distribución de frecuencias con los siguientes encabezados (exactamente como la tabla empleada para construir histogramas):

Puntuación (X)	f	Límites reales de puntuación

donde

Puntuación = puntuación de una variable de intervalo/razón

f = frecuencia de casos (o número de casos) para una puntuación

2. Calcula los límites reales de cada puntuación X. (En el capítulo 2 puedes consultar sobre los límites reales.)
3. Traza un eje horizontal o "línea de base" del polígono. Observa las puntuaciones más baja y más alta en la tabla de distribución de frecuencia. Escribe marcas sobre el eje y aplica de conformidad los valores de X; deja espacio adicional en cada extremo del eje fuera de los valores de las puntuaciones más baja y más alta.
4. Traza el eje vertical. Observa la frecuencia más alta (f) en la tabla de distribución de frecuencia y escribe marcas en el eje que vayan desde cero hasta un poco más de la frecuencia más alta.
5. A partir del valor más bajo de X y avanzando hasta el más alto, traza puntos arriba de cada valor de X hasta la altura de su frecuencia (f).
6. Une los puntos con líneas rectas. *Atención:* si un valor de X tiene una frecuencia de cero (por ejemplo, la calificación de 34 MPG en la tabla 3-2), la línea se prolonga por debajo de la línea de base.
7. Cierra los extremos de la línea del gráfico. Con una línea entre su punto y su límite real más bajo, une el valor más bajo de X a la línea de base. Con una línea entre su punto y su límite real superior, une el valor más alto de X a la línea de base.
8. Asigna un título apropiado al polígono. Asegúrate de que las leyendas de los ejes sean correctas y claras. Identifica la fuente de los datos en la parte inferior del gráfico.

Para interpretar un polígono:

1. Busca picos. El pico más alto indica el valor de X con la frecuencia más alta.
2. Busca una extensión de espacio bajo la línea gráfica para ver si hay agrupaciones de puntuaciones y para ver si hay una tendencia central.
3. Busca simetría o equilibrio en la distribución de puntuaciones. Busca una tendencia en la forma de la línea del gráfico. Ve si las colas del gráfico están situadas de manera homogénea alrededor de una puntuación central. Si no es así, observa puntuaciones especialmente bajas o altas al identificar cuál cola se prolonga.

100 al calcular las frecuencias porcentuales. La figura 3-7 presenta los dos polígonos con "Porcentaje de modelos de vehículos" estipulado en el eje vertical. Una vez que estas dos distribuciones sean descritas gráficamente juntas, sus diferencias quedan bien claras. Los picos de las dos líneas expresan una diferencia en tendencia central o promedio, que es el tema de nuestro siguiente capítulo. Los vehículos SUV tienden a evaluaciones de rendimiento relativamente bajas, en una franja más bien angosta que va de 15 a 29 MPG. En contraste, los autos compactos tienen rendimientos más bien altos y una franja más ancha de 20 a 29 MPG. Una de las características más sorprendentes de este gráfico es el pequeño traslape entre las dos líneas. Los SUV *más* eficientes tienen evaluaciones de rendimiento iguales a los autos compactos *menos* eficientes.

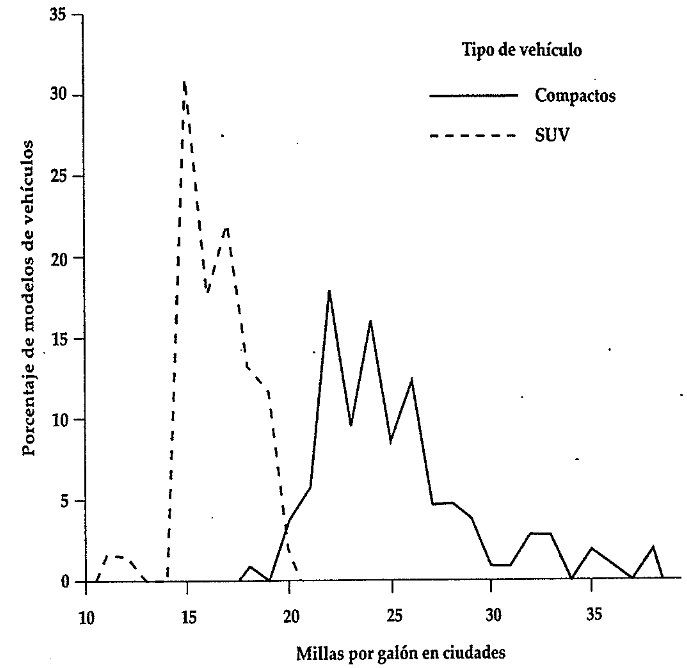
TABLA 3-3 | Comparación de las evaluaciones de rendimiento de combustible para operación en ciudades, en millas por galón (MPG), para modelos de autos compactos y vehículos utilitarios con tracción en las 4 ruedas (SUV); modelo 2004.

	Modelos de autos compactos		Modelos SUV	
Rendimiento de combustible (MPG)	*f*	Porcentaje (%)	*f*	Porcentaje (%)
11	0	0.0	1	1.5
12	0	0.0	1	1.5
15	0	0.0	21	30.9
16	0	0.0	12	17.6
17	0	0.0	15	22.1
18	1	0.9	9	13.2
19	0	0.0	8	11.8
20	4	3.8	1	1.5
21	6	5.7	0	0.0
22	19	17.9	0	0.0
23	10	9.4	0	0.0
24	17	16.0	0	0.0
25	9	8.5	0	0.0
26	13	12.3	0	0.0
27	5	4.7	0	0.0
28	5	4.7	0	0.0
29	4	3.8	0	0.0
30	1	0.9	0	0.0
31	1	0.9	0	0.0
32	3	2.8	0	0.0
33	3	2.8	0	0.0
35	2	1.9	0	0.0
36	1	0.9	0	0.0
38	2	1.9	0	0.0
Totales	106	99.8	68	100.1*

*Los porcentajes totales no suman 100 por error de redondeo.

Fuente: U.S. Department of Energy, 2004.

FIGURA 3-7
Comparación de las evaluaciones de rendimiento de combustible para conducción en ciudades, en millas por galón, para 106 modelos de autos compactos y 68 vehículos SUV; modelo 2004



Fuente: U.S. Department of Energy, 2004.

Por último, histogramas y polígonos son útiles para identificar puntuaciones inusualmente bajas o altas en una distribución. Por ejemplo, en la figura 3-7 observa que hay dos modelos SUV con rendimientos de combustible especialmente bajos (11 y 12 MPG). Del mismo modo, hay unos pocos modelos de autos compactos con rendimientos especialmente altos, arriba de 34 MPG. Estas puntuaciones poco comunes reciben el nombre de *valores extremos*. Además de atraer nuestra atención a puntuaciones "extremosas", es importante identificar resultados extremos y considerar sus efectos sobre el análisis estadístico inferencial.

Uso de gráficos en la estadística inferencial y su aplicación en la investigación

Los gráficos sirven principalmente para propósitos descriptivos en audiencias públicas. En la investigación científica y en la estadística inferencial, no obstante, a veces graficamos una variable para familiarizarnos con su distribución y prepararla para un análisis posterior. Para agilizar esta fase de preparación, por medio de una computadora generamos histogramas y polígonos (gráficos de líneas). En estadística inferencial, tales gráficos son muy útiles para detectar puntuaciones atípicas en una distribución. Por ejemplo, un gráfico que ilustra las tasas de divorcio en 50 estados de la Unión Americana revela que Nevada, con sus leyes que favorecen el divorcio, es notablemente diferente. Un caso atípico se llama **puntuación desviada o valor extremo**, que es *una puntuación notablemente diferente de las otras en la distribución de puntuaciones.* Como lo veremos más adelante, los valores extremos distorsionan los cálculos estadísticos, como —por ejemplo— los promedios. Si las distorsiones son grandes, es necesario desechar o ajustar estas puntuaciones desviadas. Por ejemplo, en el caso de los porcentajes de divorcio el procedimiento adecuado podría ser desechar los datos de Nevada. Además, informamos al lector que se trata de un estado excepcional, digno de un estudio de caso (análisis individualizado), y manifestamos que nuestras conclusiones sólo se aplican a los restantes 49 estados. Otra manera de modificar los efectos distorsionantes de los valores extremos consiste en ajustar matemáticamente las puntuaciones extremas. Un método requiere tomar el logaritmo de las puntuaciones, una transformación matemática que comprime las puntuaciones en un rango menor. Un segundo método radica simplemente en reducir el valor extremo a la siguiente puntuación más baja o más alta, que es un procedimiento llamado "truncamiento". Una recomendación final: es impropio omitir o ajustar valores extremos simplemente porque no se ajustan al patrón esperado. El investigador debe explicar con toda claridad las razones teóricas y prácticas para realizar tales ajustes. En próximos capítulos se abundará sobre los valores extremos.

Puntuación desviada o valor extremo Puntuación que es notablemente diferente de las otras en la distribución de puntuaciones.

Insensatez y falacias estadísticas: distorsión gráfica

Los gráficos y los diagramas ofrecen mapas mentales de conjuntos grandes de datos. Los procedimientos para diseñar gráficos son normativos, es decir, diferentes personas tienen ideas distintas respecto de aquello que agrada al ojo. En otras palabras, la presentación gráfica es casi un arte.

Puesto que las normas de presentación de datos son tanto estéticas como técnicas, a menudo son poco claras. Por ejemplo, ¿qué tan amplias deben ser las barras en un gráfico? ¿Debemos utilizar varios colores en las barras? Las convenciones que se aplican a éstas y otras cuestiones similares son flexibles y con frecuencia siguen una moda. El arte inspira la creatividad y la individualidad.

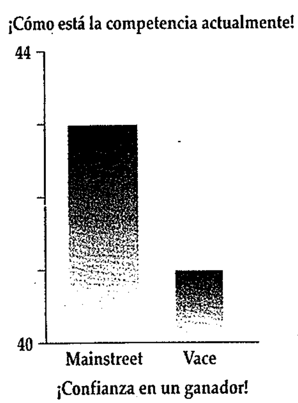
No obstante, cuando las reglas son poco claras es fácil transgredirlas, intencional o involuntariamente. Por ejemplo, con la amplia variedad de programas gráficos de computadora que existen en la actualidad, los usuarios a menudo están dispuestos a dejar los detalles del gráfico en manos del individuo anónimo que diseñó el programa. En consecuencia, los medios de comunicación masiva (al contrario de los estrictamente científicos) nos abruman con gráficos generados de manera instantánea por la computadora. Cuando usted llegue a ser experto en el pensamiento estadístico, empezará a notar que muchos, si no la mayoría de estos "instantáneos", son poco confiables en el mejor de los casos (es decir, abiertos a múltiples interpretaciones), y en el peor, incorrectos.

La siguiente parábola ilustra una distorsión gráfica común. El candidato (ficticio) a gobernador Mortimer Mainstreet tenía un cómodo margen de dos a uno en las primeras encuestas electorales sobre su único serio contendiente, Harry Vace. Poco después, sin embargo, su delantera empezó a revertirse. Se rumora que Mainstreet está a punto de despedir a su personal de campaña, cuyos miembros ya temen que sus sueños de gobernar el estado se desvanezcan.

La última encuesta señala que la ventaja de Mortimer disminuyó 2 puntos porcentuales, de 43 a 41 por ciento, con el 16 por ciento aún indeciso y un margen de error de más menos 3 por ciento. La competencia se ha vuelto muy cerrada. En un intento por evitar perder sus empleos, los miembros del personal de campaña informan a Mortimer las puntuaciones que muestra la figura 3-8 (¡Si Mortimer Mainstreet se deja llevar por esto, no merece ser gobernador! ¿Puedes identificar todas las cuestiones erróneas en este gráfico?

FIGURA 3-8

Cartel presentado a Mortimer Mainstreet por su personal de campaña, para informar sobre los resultados de la más reciente encuesta en la campaña para gobernador.



RESUMEN

1. Las tablas y gráficos dan un sentido de proporción acerca de una distribución de puntuaciones sin que haya necesidad de que el lector tenga profundos conocimientos de matemáticas.
2. Selecciona un diseño gráfico con base en *a*) el nivel de medición de la variable, *b*) los objetivos del estudio, y *c*) la audiencia a quien va dirigido.
3. Hay lineamientos razonados para construir gráficos y tablas. Un gráfico debe simplificar, no complicar. Una tabla debe explicarse por sí misma y tener sentido, sin tener que referirse a un texto.
4. Generalmente se emplean gráficos de pastel y gráficos de barras para ilustrar la distribución de categorías de una variable nominal/ordinal.
5. Los gráficos de pastel son especialmente útiles para expresar un sentido de claridad, tamaño relativo o desigualdad entre categorías.
6. Los gráficos de barras son especialmente útiles para expresar un sentido de competencia entre categorías.
7. Los gráficos de barras agrupadas son buenos para comparar dos o más grupos para una variable nominal/ordinal.
8. Para variables de intervalo/razón, utiliza un histograma de frecuencia o polígono de frecuencia (gráfico de línea).
9. Los histogramas de frecuencias atraen la atención hacia donde cae el grueso de puntuaciones en una distribución.
10. Los polígonos de frecuencias (o gráficos de línea) describen un sentido de tendencia o movimiento en una distribución de puntuaciones.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 3 del material de texto disponibles en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2, incluyen *a*) graficación de histogramas y polígonos con datos agrupados, *b*) graficación de *ojivas*, es decir, distribuciones de frecuencias acumuladas, y *c*) graficación de gráficos de caja y manijas, que son útiles para identificar valores extremos.

FÓRMULAS PARA EL CAPÍTULO 3

Cálculos para gráficos de pastel:
Elabora una hoja de cálculo con los siguientes encabezados:

Categoría	*f*	*p*	(*p*)(360°)	%

Cálculos para gráficos de barras:
Elabora una hoja de cálculo con los siguientes encabezados:

Categoría	*f*	*p*	%

Cálculos para histogramas y polígonos:
Elabora una hoja de cálculo con los siguientes encabezados:

Puntuación	*f*	Límites reales de la puntuación

PREGUNTAS PARA EL CAPÍTULO 3

1. ¿Cuáles son las tres cosas a considerar al escoger el tipo o diseño de un gráfico?
2. ¿Cuál es el principal objetivo de graficar datos?
3. ¿En qué consiste la prueba del "estacionamiento"?
4. ¿Con variables de qué niveles de medición se usan gráficos de pastel y gráficos de barras?
5. ¿Con variables de qué niveles de medición se usan histogramas y polígonos de frecuencias?
6. ¿Bajo qué circunstancias es especialmente útil un gráfico de pastel?
7. ¿Bajo qué circunstancias es especialmente útil un gráfico de barras?
8. Explica la relación entre las puntuaciones redondeadas y los límites reales de las puntuaciones.
9. La señora Barker está en un autobús con los 24 alumnos de su grupo de quinto grado; va charlando con el conductor del autobús, Kevin Braughn. Si fueras a construir un histograma de frecuencias con las edades de los pasajeros del autobús, ¿qué aspecto tendría? ¿Qué sería peculiar respecto de las edades de la señora Barker y de Kevin? ¿Qué término estadístico sirve para describir estas dos puntuaciones?

EJERCICIOS PARA EL CAPÍTULO 3

Conjunto de problemas 3A

3A-1. Según la U.S. Federal Bureau of Prisons (2003), la distribución porcentual de internos por nivel de seguridad de prisión es como sigue. Construye un gráfico de pastel.

Nivel de seguridad	Porcentaje (%)
Mínimo	19.4
Bajo	38.9
Medio	24.8
Alto	10.7
Sin nivel de seguridad	6.1

3A-2. Para países europeos, Lueschen y otros (1995) examinaron los gastos de atención médica de 1990 como porcentaje del producto interno bruto (PIB). Construye un gráfico de barras y comenta lo que hallaron.

País	Porcentaje del PIB gastado en atención médica en 1990 (%)
Bélgica	7.6
Francia	8.8
Alemania	8.3
Países Bajos	8.2
España	6.6

3A-3. A continuación aparecen datos del U.S. Bureau of the Census (2000) sobre género y nivel educativo de personas de 25 años de edad o más.

a) Construye un gráfico de barras agrupadas de género y nivel educativo.

b) Comenta sobre la naturaleza de género y nivel educativo en Estados Unidos.

	Género	
Nivel educativo	Masculino (%)	Femenino (%)
Menos de diploma de secundaria	19.9	19.3
Diploma de secundaria	27.6	29.6
Parte de universidad	20.6	21.5
Título de licenciatura/adjunto	21.9	21.8
Título de *master*	6.0	5.8
Título profesional	2.6	1.4
Doctorado	1.4	0.6
Totales	100.0	100.0

3A-4. Alba, Logan y Crowder (1997) examinaron la conformación de vecindarios étnicos de raza blanca en la ciudad de Nueva York. Un área de interés es la migración de las ciudades del centro del país entre 1980 y 1990. La tabla siguiente presenta poblaciones de grupos étnicos para un vecindario para esos dos años.

	1980			1990		
	Alemanes	Irlandeses	Italianos	Alemanes	Irlandeses	Italianos
Población	46 920	9 570	50 773	18 300	9 436	41 429

a) Construye gráficos de pastel por separado para cada uno de estos dos años para describir las poblaciones de cada grupo étnico.

b) Construye un gráfico de barras agrupadas para estos dos años para describir la población de cada grupo étnico.

c) Compara los dos tipos de gráficos. En general, ¿qué expresan los gráficos? ¿Qué estilo gráfico es mejor para describir el fenómeno? Explica.

3A-5. Supongamos que las siguientes son las edades de estudiantes de un equipo universitario de debate: 20, 19, 20, 21, 20, 21, 22, 24, 23, 22, 19, 20, 21, 21, 22, 23, 22, 20, 21, 21, 23, 29.

a) Elabora una distribución de frecuencias y construye un histograma de frecuencias de estos datos.

b) Construye un polígono de frecuencias de estos datos.

c) ¿Cuál de las dos gráficas escogerías para presentar a una audiencia pública? ¿Por qué?

d) Una de las puntuaciones es peculiar. ¿Cómo se llama esta peculiaridad?

3A-6. Las siguientes son distribuciones de frecuencia de distancias (en kilómetros) recorridas diariamente por estudiantes de secundaria en suburbios y distritos escolares rurales (datos ficticios).

a) Construye polígonos de frecuencias para las dos distribuciones en el mismo gráfico. (*Atención*: los tamaños muestrales difieren.)

b) ¿Cuál es la conclusión obvia que se saca al comparar distancias recorridas por los dos distritos escolares?

Kilómetros	Suburbano f	Rural f
1	2	0
2	4	1
3	9	0
4	13	3
5	14	5
6	8	6
7	6	9
8	5	13
9	4	17
10	2	24
11	0	15
12	0	8
13	0	7
14	0	2
15	0	1

Conjunto de problemas 3B

3B-1. La tabla siguiente presenta la distribución porcentual de asesinatos en Estados Unidos en 2002, por la relación de la víctima y el delincuente (Federal Bureau of Investigation 2002a). Construye un gráfico de pastel.

Relación de la víctima y el delincuente	Porcentaje (%)
Familia	12.7
Extraño	14.0
Otro conocido	30.5
Desconocido	42.8
Total	100.0

3B-2. Lueschen y cols. (1995) analizaron el consumo de alcohol de cinco países europeos. Construye un gráfico de barras de sus datos y comenta sobre lo que encuentres.

País	Litros consumidos por persona de más de 24 años edad en 1990
Bélgica	12.4
Francia	16.7
Alemania	12.3
Países Bajos	9.9
España	15.5

3B-3. El Federal Bureau of Investigation (2002a) proporciona datos sobre porcentajes de arresto por venta y elaboración de drogas, así como posesión simple de drogas, por región, según se muestra a continuación.

a) Elabora un gráfico de barras agrupadas por arresto por posesión de drogas, por tipo y región.

b) Comenta sobre los porcentajes de arrestos por posesión de drogas, por tipo y región.

Abuso de drogas	Noreste (%)	Medio oeste (%)	Sur (%)	Oeste (%)
Venta/elaboración	27.9	23.1	17.2	16.4
Posesión	72.1	76.9	82.8	83.6
Totales	100.0	100.0	100.0	100.0

3B-4. Los datos referentes a actividades principales y logros del programa de investigaciones del servicio de inmigración y naturalización son compilados por la Oficina de Immigration Statistics of the U.S. Department of Homeland Security (2003). Se presentan los datos de los años seleccionados.

a) Construye gráficos de pastel para estos dos años (en la misma página) para describir las actividades y logros del programa de investigaciones del INS.

b) Construye un gráfico de barras agrupadas para estos dos años, para describir las actividades y logros del programa de investigaciones del INS.

c) Compara los dos tipos de gráficos. En general, ¿qué expresan los gráficos? ¿Qué estilo gráfico es mejor para describir el fenómeno? Explica.

	Investigaciones 1992			Investigaciones 2002		
Casos	Criminal	Empleador	Contrabando	Criminal	Empleador	Contrabando
Completados	38 716	7 053	7 073	78 841	2 061	2 395

3B-5. Supongamos que una universidad pequeña está interesada en aumentar la participación de actividades del *campus*. A una muestra aleatoria de estudiantes se le pidió marcar, en una lista para el curso previo, eventos a los que asistieron. Los resultados para el número de eventos a los que asistieron fueron como sigue: 2, 2, 4, 8, 5, 2, 3, 1, 6, 5, 4, 12, 1, 4, 2, 7, 6, 3, 2, 4, 7, 4, 2, 3.

a) Construye un histograma de frecuencias de estos datos.

b) Construye un polígono de frecuencias de estos datos.

c) ¿Cuál de los dos gráficos escogerías para presentarlos a una audiencia pública? ¿Por qué?

d) Una de las puntuaciones es peculiar. ¿Cómo se llama esta peculiaridad?

3B-6. Supongamos que las siguientes son distribuciones de frecuencias de las edades de pacientes en un hospital para el tratamiento por abuso de sustancias, por los diagnósticos principales de adicción a cocaína y alcohol.

a) Construye polígonos sobrepuestos de frecuencia de estos datos. (*Atención:* los tamaños muestrales difieren.)

b) ¿Qué revela el gráfico?

	Adictos a cocaína	Adictos al alcohol
Edad	f	f
26	2	1
27	5	2
28	6	2
29	11	3
30	13	5
31	8	6
32	4	9
33	4	15
34	0	17
35	1	15
36	0	7
37	0	5
38	0	2

Conjunto de problemas 3C

3C-1. La United Status Bureau of the Census (2000) proporciona datos sobre logros educativos (entre personas de 25 años de edad o más) en Estados Unidos. Utiliza la siguiente distribución porcentual para construir un gráfico de pastel.

Logro educativo	Porcentaje (%)
Doceavo grado o menos, sin diploma	19.6
Graduado de secundaria	28.5
Universidad, sin título	21.1
Grado de licenciatura/adjunto	21.8
Grado de *master*	5.9
Título profesional	2.0
Doctorado	1.0

3C-2. El Central Bank of the Russian Federation (2000, 2002, 2003) publica diversos indicadores macroeconómicos acerca del estado de la economía rusa. A continuación se muestra un resumen de años seleccionados respecto al producto interno bruto (PIB) en miles de millones de rublos. Construye un gráfico de barras y comenta sobre lo que encuentres.

Año	PIB (en miles de millones de rublos)
1998	2696
1999	4545
2000	7302
2001	9039
2002	10863

3C-3. La tabla siguiente contiene la distribución porcentual de la población de Estados Unidos por género y edad para el año 2000 (U.S. Bureau of the Census 2000).

a) Construye un gráfico de barras agrupadas para comparar las distribuciones de edad de hombres y mujeres.

b) Comenta sobre la naturaleza de género y edad en Estados Unidos.

	Género	
Grupo de edades	Masculino (%)	Femenino (%)
Menos de 18	26.8	24.6
18-24	10.0	9.3
25-44	30.8	29.6
45-64	21.8	22.2
65 y más	10.4	14.4

3C-4. Nishi y otros (2004) examinaron los efectos de indicadores socioeconómicos sobre diversos indicadores de salud entre empleados civiles japoneses. Algunas características seleccionadas de su población de estudio se presentan aquí, específicamente relacionadas con el grado de empleo y género.

a) Construye gráficos de pastel para estos dos años (en la misma página), para describir la distribución de sujetos de estudio por grado de empleo y género.

b) Construye un gráfico de barras agrupadas para hombres y mujeres, para describir la distribución de grados de empleo por género. (Utiliza números de casos más que porcentajes.)

c) Compara los dos tipos de gráficos. En general, ¿qué expresan los gráficos? ¿Qué estilo gráfico es mejor para describir la naturaleza de género y grado de empleo entre los empleados civiles japoneses seleccionados? Explica tu respuesta.

Grados de empleo	Hombres *n*	Mujeres *n*
No manual de nivel más alto	239	95
No manual de nivel más bajo	585	174
Manual	135	118

3C-5. Supongamos que las siguientes son las edades de estudiantes en un grupo de jugadores del equipo de fútbol intramuros: 18, 19, 22, 20, 22, 21, 19, 24, 28, 23, 22, 19, 18, 19, 22, 21, 20, 24, 23, 18, 19, 21.

a) Construye un histograma de frecuencias de estos datos.

b) Construye un polígono de frecuencias de estos datos.

c) ¿Cuál de los dos gráficos escogerías para presentarlos a una audiencia pública? ¿Por qué?

d) Una de las puntuaciones es peculiar. ¿Cómo se llama esta peculiaridad?

3C-6. Supongamos que las siguientes son distribuciones de frecuencia de las edades de hombres y mujeres de un grupo numeroso de introducción a la sociología.

a) Construye polígonos de frecuencias de estos datos, superpuestos. (*Atención:* los tamaños muestrales difieren.)

b) ¿Qué revela el gráfico?

Edad	Hombres f	Mujeres f
18	4	3
19	5	5
20	4	4
21	3	6
22	2	7
23	5	4
24	4	3
25	1	5
26	2	2
27	3	1
28	4	3
29	3	1
30	1	2

Conjunto de problemas 3D

3D-1. Cockerham, Snead y DeWaal (2002) examinaron estilos de vida sanos en la antigua Unión Soviética. A continuación se ilustra la distribución educativa de la muestra de estos autores. Construye un gráfico de pastel para estos datos. Este gráfico ha de presentarse a una audiencia de profesionales, de modo que no hay necesidad de redondear los porcentajes.

Nivel de educación	Porcentaje (%)
Sin cursos profesionales	24.4
Con cursos profesionales	12.0
Capacitación profesional sin educación secundaria	8.2
Capacitación profesional con educación secundaria	13.3
Escuela técnica	21.4
Universidad	20.6

3D-2. Pikhart y otros (2004) examinaron factores psicosociales en el trabajo y el papel que éstos desempeñaron en los resultados de salud mental. Se examinaron tres países y se presentan datos muestrales del estado civil de ciudadanos de la República Checa. Construye un gráfico de barras y comenta sobre la naturaleza de esta distribución.

Estado civil	n
Casado	255
Soltero	11
Divorciado	37
Viuda(o)	3

3D-3. Abbotts y otros (2004) examinaron la relación existente entre religiosidad y salud mental en niños en dos denominaciones cristianas principales en el oeste de Escocia. A continuación se muestra la distribución de frecuencia de asistencia religiosa para grupos religiosos seleccionados.

Asistencia	Iglesia de Escocia	Católica	Otros	Ninguna
Todos los días	7	17	0	0
Casi todos los días	27	63	9	3
Semanalmente	202	450	50	20
Con menos frecuencia	229	143	13	64
Nunca	481	71	20	249

a) Construye un gráfico de barras agrupadas que compare la frecuencia de la asistencia religiosa entre denominaciones religiosas, o falta de denominaciones especificadas.

b) Comenta sobre la distribución de la asistencia religiosa para los grupos seleccionados.

3D-4. Cardano, Costa y Demaria (2004) examinaron la movilidad social y salud en un estudio longitudinal de hombres en Turín, Italia. La distribución porcentual de la clase social de residentes hombres en Turín, entre 25 y 49 años en 1981 y 1991, se muestra en la tabla siguiente (en forma modificada).

Clase social	1981 (%)	1991(%)
Media alta	8.8	16.6
Asalariado	25.3	25.5
Autoempleado	15.3	16.2
Clase trabajadora	50.6	41.7

a) Construye gráficos de pastel para clase social por cada uno de estos dos años.

b) Construye un gráfico de barras agrupadas para estos dos años, para describir la distribución de la clase social mostrada.

c) Compara los dos tipos de gráficos. En general, ¿qué expresan los gráficos? ¿Qué estilo gráfico es mejor para describir el fenómeno? Explica.

3D-5. Supongamos que las siguientes son las edades de estudiantes de un programa local de graduados de sociología: 24, 25, 26, 25, 24, 22, 27, 33, 25, 22, 24, 25, 26, 23, 27, 25, 26, 25, 24, 22.

a) Construye un histograma de frecuencias para estos datos.

b) Construye un polígono de frecuencias para estos datos.

c) ¿Cuál de los dos gráficos escogerías para presentar a una audiencia pública?

d) Una de las puntuaciones es peculiar. ¿Cómo se llama esta peculiaridad?

3D-6. Supongamos que las siguientes son distribuciones de frecuencia de las edades de adultos mayores en dos centros de retiro diferentes.

a) Construye polígonos de frecuencia sobrepuestos de estos datos. (*Atención:* los tamaños muestrales difieren.)

b) ¿Qué revela el gráfico?

Edad	Centro I *f*	Centro II *f*
67	6	2
68	5	1
69	5	0
70	3	1
71	4	2
72	4	3
73	3	2
74	2	2
75	1	4
76	1	3
77	2	5
78	1	3
79	0	4
80	2	3

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 3

En el sitio web www.mhhe.com/ritchey2, en *The Statistical Imagination*, están disponibles ejercicios computarizados del capítulo. Estos ejercicios indican cómo elaborar gráficos y tablas con el uso de *SPSS for Windows* y cómo seleccionar estilos gráficos apropiados. Además, el apéndice D de este texto, *An Introduction to SPSS*, proporciona instrucciones básicas para elaborar gráficos y tablas.

Como es el caso para software gráfico, los gráficos en *SPSS* tienen características preestablecidas (ajustes por *default*) que pueden construir un gráfico diferente del que se pretenda. Por tanto, las salidas impresas gráficas suelen requerir editarse. Es oportuno un aviso de precaución. Trata a un programa computarizado de gráficos simplemente como herramienta de dibujo. Procura examinar con todo cuidado un gráfico para cerciorarte de que es preciso. Sigue los Lineamientos de Construcción de Gráficos y Tablas que se ven al inicio de este capítulo. El lector, y no el paquete de software, es responsable en última instancia del producto final.

CAPÍTULO 4

Estimación de promedios

RESUMEN DEL CAPÍTULO

Introducción

Todos estamos familiarizados con el concepto general de promedio, en situaciones tales como una calificación promedio, un ingreso promedio, una puntuación promedio en el boliche o un promedio de bateo. Si alguien tiene un "promedio" de alguna manera, por ejemplo altura, peso, inteligencia, etc., esta persona no es atípica. Poseer un promedio significa ser como la mayoría de las personas.

En una distribución de puntuaciones, un promedio caerá entre las puntuaciones extremas, en alguna parte del área media de la distribución de puntuaciones. Por ejemplo, la mayoría de los hombres no son demasiado altos o bajos, están "sobre el promedio". A esta puntuación típica o promedio la llamamos la tendencia central de la variable. Un **estadístico de tendencia central** *proporciona una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones en bruto.* Por ejemplo, las estaturas de los hombres estadounidenses tienden a agruparse alrededor de cinco pies con ocho pulgadas, y los pequeños saludables pesan alrededor de siete libras al nacer. Si Bob tiene un promedio de 165 en el boliche, no esperamos que obtenga esta puntuación exacta en cada juego, pero conseguirá cercanamente esa puntuación la mayoría de las veces.

> **Estadístico de tendencia central** Estadístico que proporciona una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones en bruto.

Existen tres estadísticos de tendencia central comunes: la media, la mediana y la moda. ¿Por qué tres? Porque cada uno tiene aspectos fuertes, pero también debilidades potenciales, dependiendo de la forma particular de la distribución de puntuaciones de una variable. Según sea la forma de una distribución, una medición del promedio puede resultar más exacta que otra, y, en ocasiones, informar cualquier estadístico de tendencia central sólo conduciría a errores o no proporcionaría información suficiente.

La media

La media aritmética de una distribución de puntuaciones (o, simplemente la media) consiste en un estadístico de tendencia central que es familiar a cualquier estudiante que haya calculado el promedio de sus calificaciones para algún curso. La **media** es *la suma de todas las puntuaciones dividida entre el número de puntuaciones observadas* (*es decir, el tamaño de la muestra*). Para calcular la media de una variable, simplemente sumamos todas las puntuaciones y dividimos el resultado entre el tamaño de la muestra.

> **La media** Suma de todas las puntuaciones dividida entre el número de puntuaciones observadas (es decir, el tamaño de la muestra).

La media es el estadístico de tendencia central más útil. Con un cálculo matemático rápido, ofrece un resumen de las puntuaciones típicas o promedio en una distribución. Puesto que emplea la operación matemática de división, la media se aplica a las variables de intervalo/razón.

En las fórmulas matemáticas el símbolo convencional utilizado para representar el nombre de una variable es una letra mayúscula. Las letras X y Y se emplean con frecuencia. Por ejemplo, podríamos emplear X para simbolizar la edad y Y para la altura. A menudo, Y se usa para la variable dependiente y X para la variable independiente. Por ejemplo, pondríamos Y = calificación promedio (CP) de la universidad con el siguiente conjunto de variables predictoras: X_1 = calificación promedio (CP) de la preparatoria, X_2 = puntuaciones en el examen de admisión a la universidad, X_3 = habilidad en la comprensión de lectura y X_4 = año de escolaridad.

Para una variable X, cualquier cosa que definamos, el símbolo para la media *calculada con datos de la muestra* es $\overline{X}$, que se llama "X barra". Por ejemplo, si X = edad, y la edad media del grupo de estadística es 20.5 años, decimos "X barra es igual a 20.5 años". Recuerda especificar las unidades de medida de la variable, en este caso, años. La media se calcula como sigue (Σ se lee como "la suma").

> **Cálculo de la media**
>
> $$\overline{X} = \frac{\Sigma X}{n}$$
>
> donde
>
> $\overline{X}$ = la media de la variable X de intervalo/razón con datos de la muestra
>
> ΣX = la suma de todas las puntuaciones individuales para la variable X
>
> n = el número de observaciones (es decir, el tamaño muestral)

Si hay 12 niños en una muestra, cuyas edades son 6, 12, 5, 10, 9, 10, 8, 7, 9, 11, 8 y 10 años, su edad media es

$$\overline{X} = \frac{\Sigma X}{n} = \frac{6 + 12 + 5 + 10 + 9 + 10 + 8 + 7 + 9 + 11 + 8 + 10}{12}$$

$$= \frac{105 \text{ años}}{12} = 8.75 \text{ años}$$

Técnicamente, la media es 8.75 años *por niño*, pero omitimos la unidad del denominador. Conceptualmente, el valor de la media nos dice cuáles serían las puntuaciones X de una muestra *si es que* cada sujeto de la muestra tuviera la misma puntuación. En el ejemplo anterior, 8.75 años (es decir, 8 años nueve meses) sería la edad de cada niño si todos los niños tuvieran exactamente la misma edad. Es útil, entonces, pensar en la media como una medición de "partes iguales". Por ejemplo, si quisiéramos saber la cantidad media de dinero en efectivo que llevan consigo los estudiantes de un salón de clases, pondríamos todo el dinero en efectivo en un recipiente y lo dividiríamos equitativamente. (¿Algún voluntario?) La cantidad que recibiría cada persona sería el valor medio del dinero en efectivo. La media también puede ser considerada como un punto de equilibrio, es decir, el punto en el que se equilibran las *diferencias entre* la media de X y las puntuaciones individuales X de la distribución. En el capítulo 5 ampliaremos esta noción.

Por último, al calcular los estadísticos de tendencia central, particularmente la media, debe tenerse cuidado para no incluir las puntuaciones codificadas como casos perdidos. Al determinar la media sólo se incluyen los casos "válidos". Por ejemplo, si en una muestra de 49 personas 2 de ellas no informaron sus edades, la suma de las edades se dividiría entre 47, que es el número de puntuaciones válidas, en lugar de dividirla entre el tamaño de la muestra 49. Además, con archivos de computadora, debe tenerse cuidado de no sumar los códigos de "valor perdido" (como 999) a la suma de las puntuaciones.

Pensamiento proporcional sobre la media

Combinación de las medias de dos muestras de tamaño diferente La media es el estadístico de tendencia central más ampliamente usado de variables de intervalo/razón. Así, es importante que tengamos un buen sentido de proporción respecto de su cálculo. Primero, examinemos una situación donde se comete un error común: combinar las medias de dos grupos sumando las dos medias y dividiendo el resultado entre 2. [El único momento en que esto no representa un error es cuando los dos grupos tienen los mismos tamaños de muestra (es decir, cuando las n son iguales).] Por ejemplo, observa el número medio de días de vacaciones por año (X) para el grupo 1, las ocho secretarias de un banco local, y, para el grupo 2, los tres vicepresidentes. Para las ocho secretarias:

$$\bar{X}_{(grupo1)} = \frac{\Sigma X_{(grupo1)}}{n_{(grupo1)}} = \frac{7 + 10 + 7 + 12 + 16 + 7 + 14 + 10}{8}$$

$$= \frac{83 \text{ días}}{8} = 10.38 \text{ días de vacaciones}$$

Para los tres vicepresidentes:

$$\bar{X}_{(grupo2)} = \frac{\Sigma X_{(grupo2)}}{n_{(grupo2)}} = \frac{60 + 30 + 30}{3}$$

$$= \frac{120 \text{ días}}{3} = 40.00 \text{ días de vacaciones}$$

Si calculamos *incorrectamente* la media de la oficina completa sumando estas dos medias y dividiendo el resultado entre 2, obtendríamos la respuesta errónea de 25.19 días de vacaciones. El cálculo correcto para esta media combinada es

$$\bar{X}_{(grupos\ 1\ y\ 2\ combinados)} = \frac{\Sigma X_{(grupo1)} + \Sigma X_{(grupo2)}}{n_{(grupo1)} + n_{(grupo2)}}$$

$$= \frac{83 + 120}{8 + 3} = \frac{203}{11} = 18.45 \text{ días de vacaciones}$$

Analizando un poco veremos que esta formulación es equivalente a tratar a los 11 empleados como una muestra. Para ejemplificar casos al "promediar" erróneamente las medias de un grupo, véase el apartado de "Insensatez y falacias estadísticas" al final de este capítulo.

Cálculo de la media combinada de dos muestras de tamaño diferente

Dadas las medias y tamaños muestrales de dos grupos:

$$\bar{X}_{(grupos\ 1\ y\ 2\ combinados)} = \frac{\Sigma X_{(grupo1)} + \Sigma X_{(grupo2)}}{n_{(grupo1)} + n_{(grupo2)}}$$

donde

$\bar{X}$ = la media de la variable de intervalo/razón, X, calculada en datos muestrales

$\Sigma X_{(grupo)} = (n)_{(grupo)}(\bar{X})_{(grupo)}$

y

n = el número de observaciones (es decir, el tamaño muestral)

Ejemplo: supongamos X = calificación en examen final; la calificación media para los 13 estudiantes del último año del grupo es 87, y la calificación media para los 16 estudiantes del penúltimo año es 79. ¿Cuál es la calificación media para los dos grupos combinados?

1. Calcula la ΣX para cada grupo:

$$\Sigma X_{(\text{último año})} = (13)\,(87) = 1\,131 \text{ puntos de examen}$$
$$\Sigma X_{(\text{penúltimo})} = (16)\,(79) = 1\,264 \text{ puntos de examen}$$

2. Sustituye las sumas en la ecuación precedente:

$$\bar{X}_{(de\ último\ y\ penúltimo)} = \frac{\Sigma X_{(último)} + \Sigma X_{(penúltimo)}}{n_{(último)} + n_{(penúltimo)}}$$

$$= \frac{1\,131 + 1\,264}{13 + 16} = \frac{2\,395}{29} = 82.59 \text{ puntos de examen}$$

Debilidades potenciales de la media: situaciones en las que reportarla sola puede conducir a errores

Cuando se reporta un estadístico de tendencia central, tendemos a suponer que su valor es representativo de puntuaciones típicas en la parte central de una distribución. En ocasiones, sin embargo, cuando se informa la media puede conducir a errores al respecto. Éste es el caso porque el cálculo de la media puede inflarse (aumentarse) o desinflarse (disminuir) debido a puntuaciones o valores extremos. Puntuaciones muy altas, o valores extremos positivos, inflan el valor de la media "agrandando" la suma de X (es decir ΣX) en el numerador de la fórmula. Puntuaciones sumamente bajas en una distribución, o valores extremos negativos, desinflan el valor de la media "encogiendo" ΣX. Por ejemplo, suponga que calculamos la cantidad media del dinero en efectivo que llevan 10 estudiantes. Idealmente, esta media debe indicarnos cuál es la cantidad típica. Pero supongamos que un estudiante cobró un cheque por \$400 y nuestro cálculo es el siguiente, donde X = la cantidad de dinero en efectivo de cada estudiante (para simplificar, se redondea al dólar más cercano):

$$\bar{X} = \frac{\Sigma X}{n} = \frac{5 + 2 + 6 + 10 + 8 + 3 + 9 + 11 + 5 + 400}{10}$$

$$= \frac{\$459}{10} = \$45.90 \approx \$46$$

Por obvias razones, esta media de \$46 no representa la cantidad de dinero promedio típica, o la tendencia central que los alumnos pueden llevar en efectivo. La mayoría de los estudiantes tiene menos de \$10, y reportar una media de \$46 es engañoso. El cálculo de la media se distorsiona por la presencia de un valor aislado. Para obtener un sentido de proporción sobre cómo se calcula la fórmula de la media, examina la relación entre el numerador (ΣX) y el denominador (n). Cuando ΣX es grande y n es pequeño, la media será grande. Cuando ΣX es grande debido a la presencia de uno o dos valores extremos de alto valor, la media se "inflará" hasta un valor grande.

Recuerda que nuestro objetivo es usar estadísticos de muestras para estimar los parámetros de una población. Si se reporta una media *muestral* inflada o disminuida, se presentará un resumen distorsionado de las puntuaciones que obtienen los sujetos *de una población.* Esta limitación de la media es un problema especial con muestras pequeñas; cuanto menor sea la muestra, mayor será la distorsión que genere un valor extremo. Por ejemplo, calcula la edad media de la siguiente muestra de cinco estudiantes de la universidad local, donde un estudiante de la muestra tiene una edad extremadamente alta: 19, 19, 20, 21, y 54 años. La respuesta dejará la impresión que esta muestra está bastante arriba de la edad típica en la universidad, cuando, en realidad, cuatro de los cinco estudiantes *tienen* la edad típica. También observa lo que sucede cuando existe una puntuación sumamente baja, como con esta muestra de edades: 8, 19, 20 y 21 años. En tales casos, los valores extremos deben eliminarse y la media debe calcularse de nuevo sin ellos. Al reportar esta "media ajustada", notamos por qué se realizó el ajuste.

En cualquier momento que calculemos una media, en especial con una muestra pequeña, primero examinamos la distribución de frecuencias de la puntuación en bruto para los valores extremos. Un recurso práctico para esto es un gráfico de caja (capítulo 3). Ya que la media es más útil que la mediana y la moda, con frecuencia ajustamos las puntuaciones de una distribución para reducir los efectos de los valores extremos en su cálculo. Los efectos deformantes de los valores extremos se mencionan en todo el texto.

La mediana

La **mediana** (Mdn) es *la puntuación central en una distribución ordenada*, es decir, el valor de una variable que divide en mitades a la distribución de las puntuaciones, *la puntuación por arriba de la cual queda la mitad de los casos y por debajo queda la otra mitad.* Por ejemplo, si la media del ingreso familiar en la ciudad Cornbelt es \$26 000, la mitad de las familias de esta ciudad tienen ingresos mayores a \$26 000 y la otra mitad ingresos menores a \$26 000. Conceptualmente, la mediana es un punto de localización, la puntuación de la mitad. La mediana trae a colación una posición geográfica entre áreas iguales, como la mediana de una carretera. La puntuación mediana también es igual al percentil 50, el punto bajo en el cual caen el 50% de las observaciones. Entre los tres estadísticos de tendencia central, la mediana es más útil cuando una distribución está sesgada (es decir, tiene pocas puntuaciones hacia un lado). Por ejemplo, la mediana del precio de las ventas recientes de viviendas es preferible a la media del precio, porque unas cuantas ventas de alto precio incrementarían el valor de la media.

> **La mediana** Para una variable ordinal o de intervalo/razón, es la puntuación central en una distribución ordenada, la puntuación que deja debajo de sí la mitad de los casos y, por arriba, la otra mitad.

Para calcular la mediana de una distribución, primero debes ordenar las puntuaciones para una variable X, es decir, las puntuaciones deben colocarse en orden de tamaño, de menor a mayor o de mayor a menor. Divide entre 2 el tamaño de la muestra n para acercarte a la puntuación de la mitad de la distribución. Si n es un número impar, la mediana será un caso real en la muestra. Supongamos, por ejemplo, que tenemos una muestra de cinco familias con los siguientes ingresos mensuales (X):

Ubicación de la Mdn
↓

Orden de los casos	1	2	3	4	5
Valores de X	\$3 540,	\$4 675,	\$7 350,	\$9 860,	\$19 000

↑
Mdn = un valor de X

La mediana de ingreso es \$7 350, el valor de X para la tercera puntuación ordenada.

Si n es un número par, la mediana se localiza entre las dos puntuaciones de la mitad y se calcula tomando la media de esas dos puntuaciones. Por ejemplo, si una sexta familia con ingreso de \$20 000 se inserta en la muestra anterior,

Ubicación de la Mdn
↓

Orden de los casos	1	2	3	4	5	6
Valores de X	\$3 540,	\$4 675,	\$7 350,	\$9 860,	\$19 000	\$20 000

↑
Mdn = \$8 605

La mediana se sitúa entre el tercero y cuarto casos. Se calcula sumando las puntuaciones de \$7 350 y \$9 860 y dividiendo entre 2.

Con una muestra pequeña, localizar la mediana es un trabajo sencillo; con una muestra grande (y con ayuda de un programa de cómputo), la mediana se sitúa matemáticamente al dividir entre 2 el tamaño muestral y sumar .5. Nótese que este resultado dará la *ubicación ordenada* de la mediana, no la mediana en sí. Ordena las puntuaciones y luego cuenta hasta esta posición. La puntuación X de esta posición es la mediana. Después de hallar la mediana, revisa de nuevo y comprueba si de verdad tu respuesta divide los casos por la mitad. La mediana puede usarse con variables de intervalo/razón, así como con variables ordinales. Finalmente, no confundas la mediana con otro estadístico llamado *rango medio*, que es el punto situado a la mitad entre los valores mínimo y máximo de X.

Cálculo de la mediana (Mdn)

1. Ordena de menor a mayor la distribución de puntuaciones.
2. Ubica la posición de la mediana. Divide entre 2 el tamaño de la muestra, *n*, para ubicarte cerca de la puntuación que está a la mitad de la distribución. Si *n* es un número impar, la mediana será un caso real de la muestra; si *n* es par, la mediana se localizará entre las dos puntuaciones que están a la mitad, y se calculará tomando la media de esas dos puntuaciones. (Matemáticamente, la posición de la mediana se encuentra dividiendo entre 2 el tamaño de la muestra y sumando .5.)

Debilidades potenciales de la mediana: situaciones en las que reportarla sola puede conducir a errores

La mediana se basa en la ubicación ordenada de puntuaciones de una distribución. Es *insensible a valores de puntuaciones* en una desviación, es decir, cualesquiera que sean los valores de X que la rodean, la mediana es la puntuación central determinada por el número de puntuaciones (n) de la muestra. Por ejemplo, las siguientes dos distribuciones de puntuaciones de un examen tienen la misma mediana aun cuando estén compuestas de puntuaciones muy diferentes.

Aula 1: 39, 51, 77, 78, 81

↓

Mdn

↑

Aula 2: 74, 75, 77, 94, 98

Afirmar que la calificación promedio del examen de ambos grupos es 77 sería impreciso porque sugiere que los dos tuvieron igual desempeño. (En realidad, el aula 2 lo hizo mucho mejor, con una *media* de 83.6, comparado con la *media* de 65.2 para el aula 1.) La mediana no se afecta por los valores de X.

Aun cuando es insensible a valores de puntuaciones, la mediana es sensible a (o se ve afectada por) cualquier cambio en el tamaño de la muestra. Por ejemplo, supongamos que en el aula 1 dos estudiantes se presentan tarde al examen y lo hacen mal, lo cual es común en estudiantes que llegan tarde a un examen. Cuando sus puntuaciones se incluyen en la distribución, la mediana cambia de manera drástica de 77 a 51:

Aula 1 (incluye puntuaciones tardías): 34, 36, 39, 51, 77, 78, 81

↑

Mdn

La mediana, entonces, tiene dos debilidades potenciales: (1) es insensible a los valores de las puntuaciones de una distribución y (2) es sensible (o afectada por) cualquier cambio en el tamaño de la muestra. Antes de presentar la mediana, asegúrate que ninguna de estas debilidades potenciales te lleve a conclusiones erróneas.

La moda

La **moda** (Mo) es *la puntuación que se presenta con mayor frecuencia en una distribución*. Conceptualmente, la moda es la puntuación "más popular". La tabla 4-1 muestra la distribución de edades para una muestra de estudiantes universitarios. La moda es 19 años porque la mayoría de las personas (49 de ellas) tiene esta edad. Observa que la moda es una puntuación X (19 años), *no* una frecuencia, f (49 casos).

TABLA 4-1 | Distribución de edades para una muestra de 125 estudiantes universitarios

	Especificaciones		Cálculos
	Edad	*f*	*Porcentaje (%)*
	18	31	24.8
Mo ⟶	19	49	39.2
	20	20	16.0
	21	18	14.4
	22	7	5.6
	Total	125	100.0

La moda Puntuación que ocurre con mayor frecuencia en una distribución.

Cálculo de la moda (Mo)

1. Agrupa las puntuaciones en una distribución de frecuencias.
2. Identifica la moda, que es el valor de X con la mayoría de los casos (es decir, la mayor frecuencia, f).

Es oportuna una nota de precaución. No confundas la moda (la "puntuación que ocurre con mayor frecuencia") con la "mayoría de puntuaciones". Una mayoría simple sería "más de la mitad" o 50 por ciento de los casos de una muestra más uno, por lo menos. Observa que en esta distribución, aunque la puntuación que ocurre más frecuentemente es 19 años, la mayoría de la muestra no tiene 19 años; sólo 39.2 por ciento de la muestra tiene esa edad. Ninguna edad de esta distribución tiene mayoría.

La moda es útil con variables de todos los niveles de medición. La moda es fácil de reconocer en gráficos. En un gráfico de pastel, es la categoría con la rebanada más grande; en un gráfico de barras, es la barra más alta; en un histograma, es la columna más alta; y en un polígono, la puntuación del punto más alto, o pico.

Debilidades potenciales de la moda: situaciones en las que reportarla sola puede conducir a errores

En general, llamada *por sí misma*, la moda es el estadístico de tendencia central menos útil, porque tiene un alcance informativo limitado. Si bien identifica la puntuación que ocurre con más frecuencia, no sugiere nada sobre las puntuaciones que ocurren alrededor de este valor de la puntuación. Así, la moda es muy útil cuando se presenta en conjunción con la mediana y la media. Como veremos más adelante, reportar los tres estadísticos de tendencia central es bastante informativo.

FIGURA 4-1

Distribuciones de puntuaciones de varias formas con la misma moda

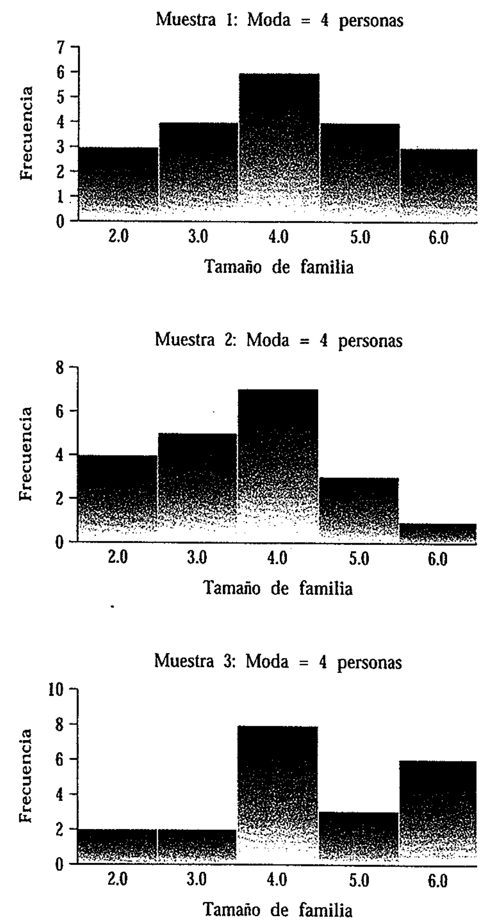


TABLA 4-2 | Distribución de sueldos en un restaurante de comida rápida

Sueldo $	*f*	Clasificación de empleados
5.75	13	Empleados regulares
10.50	2	Gerentes nocturnos
18.90	1	Gerente en Jefe
Total	16	

La moda puede ser engañosa cuando se usa sola porque es insensible tanto a los valores de las puntuaciones de una distribución como al tamaño de la muestra. Esto significa que puedes tener cualquier número de distribuciones con formas totalmente diferentes, y aun así todas podrían tener la misma moda, como se ilustra en la figura 4-1. Además, una variable puede tener más de una moda o ninguna moda significativa en absoluto.

Existe al menos una situación en la cual la moda es un estadístico de tendencia central apropiado por sí mismo e informar la media y la mediana es confuso. Esto ocurre cuando las puntuaciones de X son en esencia del mismo valor para todos los casos, excepto para unos cuantos. Un ejemplo es la estructura de sueldos en un restaurante de comida rápida, donde todos, excepto los gerentes, tienen un mismo sueldo bajo. Esta distribución se muestra en la tabla 4-2, donde X es el sueldo por hora y f es la frecuencia de las puntuaciones. La media aquí es $7.17, y está "inflada" por las puntuaciones extremas de los sueldos de los gerentes. Para alguien que busca empleo, esta media deja la falsa impresión de que el restaurante, en promedio, ofrece un sueldo un tanto arriba del salario mínimo. La mediana es $5.75, igual que la moda, pero reportar esta mediana lleva a la interpretación incorrecta de que la mitad de los empleados ganan más que esa cantidad, lo cual no es el caso. Informar la moda, $5.75, significa que a muchos empleados se les paga este sueldo bajo. Ésta es la ilustración más exacta de esta distribución de sueldos.

Estadísticos de tendencia central y el nivel apropiado de medición

Recuerda, como vimos en el capítulo 2, que el nivel de medición de una variable nos dice qué fórmulas matemáticas y estadísticas son apropiadas para dicha variable. La media y la mediana son claramente apropiadas con variables de intervalo/razón. Tiene sentido hablar sobre el peso, la estatura o el ingreso medios. Los estadistas novicios deben evitar el uso de la media y la mediana con variables ordinales. Con variables nominales las medias y las medianas no tienen sentido. La variable nominal *género* es un caso oportuno. Una persona no puede ser un promedio de tanto hombre y tanto mujer; se es o el uno o el otro. Recuerda la tabla 2-5, donde se presenta la distribución de afiliaciones religiosas en la niñez para una muestra de adultos estadounidenses. No tiene sentido preguntar cuál es la media de religión.

Mientras que la media y la mediana se aplican mejor a las variables de intervalo/razón, la moda puede usarse con variables de todos los niveles de medición. De la tabla 2-5 podríamos reportar que la moda de religión es "Protestante total" para las principales religiones, "Católico" para cualquier denominación particular, o "Bautista" para cualquier denominación protestante particular.

Curvas de distribución de frecuencias: relaciones entre la media, la mediana y la moda

Puesto que cada uno de los tres estadísticos de tendencia central tiene debilidades potenciales, vale la pena observarlos como un conjunto de estadísticos que se van a interpretar juntos. Estos tres estadísticos son especialmente útiles cuando se examinan de manera gráfica. Una forma imaginativa de entender la relación entre estos tres estadísticos consiste en localizar los valores de cada uno en una curva de distribución de frecuencias.

Una **curva de distribución de frecuencias** es *un sustituto de un histograma de frecuencias o polígono donde reemplazamos estos gráficos con una curva suavizada.* Esta sustitución es apropiada porque la curva suavizada no se ve tanto como una ilustración de la distribución de la muestra, sino más bien como una estimación de la manera en que se distribuyen las puntuaciones en la *población.* Al igual que con un histograma, las puntuaciones de una variable se ilustran de izquierda (el más bajo) a derecha (el más alto), es decir, las puntuaciones se ordenan sobre el eje horizontal. El área bajo una curva de distribución de frecuencias representa *el número total de sujetos en la población y es igual a una proporción de 1.00 o a un porcentaje de 100 por ciento.* Nuestro interés está en evaluar la forma de una distribución y *examinar las posiciones relativas de la media, la mediana y la moda, para estimar la forma de una distribución de frecuencias.* Las curvas de distribución de frecuencias aplican sólo a niveles de medición de variables de intervalo/razón.

> **Curva de distribución de frecuencias** Es sustituto de un histograma o polígono de frecuencias donde reemplazamos estos gráficos con una curva suavizada. El área bajo la curva representa el número total de sujetos en la población y es igual a una proporción de 1.00 a un porcentaje de 100 por ciento.

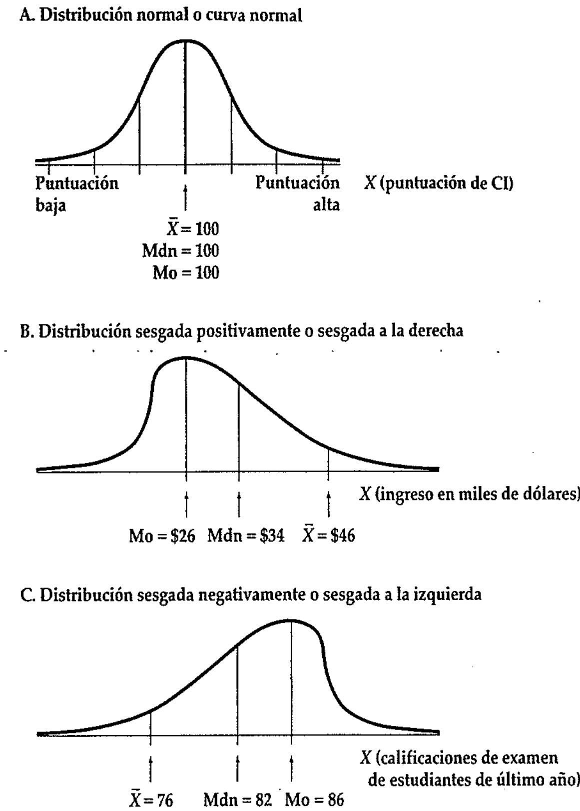
La figura 4-2 presenta tres formas muy comunes de curvas de distribución de frecuencias de puntuaciones. Al igual que con nuestros histogramas, el eje horizontal de las curvas representa las puntuaciones de una variable X. El eje vertical (el cual a veces no nos molestamos en dibujar) representa la frecuencia proporcional o frecuencia porcentual; así, la altura de la curva en cualquier valor de X representa la proporción de una muestra o población con esa puntuación.

La distribución normal

Una **distribución normal** es aquella donde *la media, la mediana y la moda de una variable son iguales entre sí y la distribución de la puntuaciones tiene forma de campana.* También nos referimos a esto como una "curva normal". La figura 4-2A ilustra puntuaciones de CI, que están normalmente distribuidos con una media de 100. Una distribución normal es simétrica (es decir, equilibrada en cada lado). Su media, mediana y moda se localizan en el centro de la distribución. La presencia de la mediana aquí asegura la simetría porque, por definición, la mediana divide por la mitad una distribución ordenada de puntuaciones. Puesto que la moda está en el punto central de una distribución normal, el pico de la curva se localiza allí.

FIGURA 4-2

Curvas de distribución de frecuencias comunes y posiciones relativas de la media, la mediana y la moda, donde X es una variable de intervalo/razón (datos ficticios)



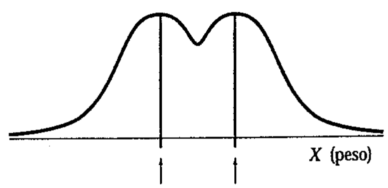
> **Distribución normal** Curva de distribución de frecuencias donde la media, la mediana y la moda de una variable son iguales entre sí y la distribución de las puntuaciones tiene forma de campana.

Distribuciones sesgadas

Una **distribución sesgada** es *aquella en la cual la media, la mediana y la moda de una variable son desiguales y algunos de los sujetos tienen puntuaciones sumamente altas o bajas.* Cuando éste es el caso, la distribución se alarga hacia un lado, como la hoja de una espada o de una brocheta (*skewer*), de ahí el nombre de *sesgada* (*skewed*) (figura 4-2B y C).

> **Distribución sesgada** Curva de distribución de frecuencias en la cual la media, la mediana y la moda de una variable son desiguales y algunos de los sujetos tienen puntuaciones sumamente altas o bajas.

FIGURA 4-3
Distribución bimodal de pesos de hombres y mujeres



Las posiciones de la media, la mediana y la moda son predecibles para las curvas de distribución sesgadas. Un **sesgo a la derecha (o positivo)** *tiene puntuaciones extremas en el extremo positivo de la distribución de puntuaciones* (figura 4-2B). Por ejemplo, el ingreso familiar en Estados Unidos está sesgado positivamente; la mayoría de las familias ganan bastante dinero, pero pocas son sumamente ricas. Las puntuaciones extremas altas inflan la media, "jalándola" en dirección positiva. La moda es la medida de tendencia central con la menor puntuación calculada. La mediana será igual a la media o a la moda o, más probablemente, caerá entre éstas.

El **sesgo a la izquierda (o negativo)** tiene *puntuaciones extremas en el final bajo o negativo de la distribución de puntuaciones* (figura 4-2C). Por ejemplo, las puntuaciones del examen en un curso del último año en la universidad tienden a estar sesgadas a la izquierda. La mayoría de los estudiantes de último año obtiene altas puntuaciones, pero pocos se quedan en la dirección negativa. Estas pocas puntuaciones extremas bajas desinflan la media, jalándola en la dirección negativa. La moda es la mayor puntuación calculada, y la mediana cae entre la media y la moda.

Ya sea con un sesgo a la izquierda o la derecha, si la mediana no cae entre la media y la moda esto sugiere que la distribución está singularmente formada. Una distribución así es una distribución bimodal, la cual tiene dos modas o picos. Por ejemplo, la variable peso para una muestra que incluye a hombres y mujeres produciría una distribución bimodal, con la moda más alta que resulta del hecho de que en promedio los hombres son más pesados que las mujeres (figura 4-3).

Uso de los datos de una muestra para estimar la forma de una distribución de puntuaciones en una población

Ya sea con variables intervalo/razón cuando al calcular los estadísticos de tendencia central e histogramas para datos de una *muestra*, los datos para una variable con frecuencia aparecen ligeramente sesgados. Esto no garantiza, sin embargo, que las puntuaciones de la variable estén sesgadas *en la población* de la que se tomó la muestra. El sesgo en los datos de la muestra puede deberse al error muestral. En otras palabras, una segunda muestra de la población parecería normal o ligeramente sesgada en la otra dirección.

Los estadísticos de sesgo se emplean para determinar si los datos de la muestra están tan sesgados que sugieren que las puntuaciones de la población están sesgadas. No vamos a calcular un estadístico de sesgo a mano. Los programas de cómputo, sin embargo, proporcionan estadísticos de sesgo, y uno común está disponible con las aplicaciones de cómputo opcionales que acompañan este texto. Cuando el valor absoluto de este estadístico de sesgo (su valor ignorando el signo de más y de menos) es mayor a 1.2, la distribución podría estar significativamente sesgada, dependiendo de la forma de la distribución, así como del tamaño de la muestra. Unos pocos valores extremos de una muestra grande tendrán poco efecto en los

estadísticos. Si el valor absoluto de este estadístico de sesgo es mayor que 1.6, sin embargo, sin importar el tamaño de la muestra, la distribución probablemente esté sesgada; entonces reportar la media de X de la muestra como un estimado de la media de la población sería engañoso, a causa de la distorsión potencial de la media por las puntuaciones extremas. Aparte de la cuestión de describir con precisión la forma de una distribución, el sesgo es una preocupación con la estadística inferencial. Como veremos en capítulos posteriores, al probar una hipótesis sobre la relación entre dos variables, una variable sesgada exige trabajo adicional para evitar conclusiones incorrectas. Se identificarán tales casos conforme se encuentren.

Como veremos en el capítulo 5, cuando una distribución no esté sesgada o de otra manera tenga una forma particularmente extraña, la media es el estadístico de tendencia central a elegir. Esto es especialmente válido para reportes dirigidos al público en general, cuyos miembros pueden sentirse abrumados con más de un estadístico. Sin embargo, si una distribución está sesgada, la mediana es el estadístico que deba reportarse. La mediana minimiza el error al describir una distribución sesgada, porque cae entre la media y la moda, como se ilustra en la figura 4-2B y C. Como la más central de los tres estadísticos, la mediana es el mejor de las tres pobres opciones para una distribución sesgada, cuando sólo un estadístico debe reportarse.

Para audiencias científicas, las distribuciones sesgadas se registran informando los tres estadísticos de tendencia central y quizás incluyendo un gráfico para transmitir con precisión la forma de la distribución. A veces una distribución sesgada es muy informativa. Por ejemplo, las *estancias en el hospital* están positivamente sesgadas. En un año dado, la mayoría

TABLA 4-3 | Características, aspectos fuertes, y debilidades potenciales de la media, mediana y moda

		Aspectos fuertes y aplicaciones		
Estadístico de tendencia central	**Definición**	**Niveles apropiados de medida**	**Aplicación a formas de distribución de puntuaciones**	**Debilidades potenciales**
Media	Valor de X si todas las puntuaciones son iguales	Intervalo/razón	Abierta a operaciones matemáticas cuando una distribución tiene forma normal	Su cálculo es distorsionado por valores extremos o un sesgo en la curva de la distribución
Mediana	Puntuación en la mitad de una distribución ordenada; puntuación por arriba de la cual queda la mitad de las puntuaciones y, por debajo, la otra mitad	Intervalo/razón	Preferible cuando la distribución es sesgada	Insensible a los valores de X de la distribución, pero sensible a cambios en el tamaño muestral
Moda	Puntuación que ocurre con más frecuencia en una distribución; la puntuación "más popular"	Nominal, ordinal, intervalo/razón	Preferida cuando prácticamente todas las puntuaciones (o categorías) de la distribución son iguales	Insensible a los valores de X e insensible a cómo se distribuyen las puntuaciones alrededor de X

de las personas no pasan algún día o pasan muy pocos en el hospital. Pero un porcentaje sustancial pasa mucho tiempo, y unos pocos "se sesgan" al permanecer semanas o meses en el hospital. Tal sesgo estimula la reflexión sobre los predictores de estancias largas. ¿Puedes pensar en hipótesis que expliquen el sesgo de estancias en el hospital?

Como veremos en el capítulo 5, en general la media es el estadístico de tendencia central más valioso porque permite mayor flexibilidad en los cálculos matemáticos. Casi siempre, la mediana y la moda representan callejones sin salida porque no ofrecen operaciones matemáticas adicionales que valgan la pena. Se gana poco con informar de ellas. Siempre que sea posible, la media es la medición sumaria que debe usarse, en especial con estadísticos inferenciales. Debido a esto, con frecuencia ajustamos distribuciones sesgadas para "hacerlas normales", de manera que podamos usar la media. Más adelante en este texto se discuten las especificaciones de este tipo de control del error. La tabla 4-3 resume las propiedades de los tres estadísticos de tendencia central.

Organización de los datos para calcular los estadísticos de tendencia central

Existen dos formatos comunes para organizar los datos y calcular los estadísticos de tendencia central en tales datos. Un formato es una hoja de cálculo que por lo general se usa para la entrada de datos para computadora, pero las hojas de cálculo también se usan en negocios, gobierno y grupos comunitarios para conservar registros de la organización. Programas computarizados de hojas de cálculo, como *Lotus 1-2-3*, *Excel*, y *Corel Quattro Pro*, están especialmente diseñados para este fin. Estos formatos de hoja de cálculo evolucionaron a partir de la manera lógica de resolver problemas manualmente, con sólo poner en lista las puntuaciones de una variable en una columna vertical.

El segundo formato común para realizar cálculos es un formato de distribución de frecuencias. En éste, las puntuaciones de una variable se escriben en una columna y la frecuencia de cada puntuación en otra (como las distribuciones de frecuencias del capítulo 2). Este formato es típico de una *salida de resultados* de una computadora. Ahora resolvamos un problema sencillo para ilustrar el uso de ambos formatos.

Formato de hoja de cálculo para calcular estadísticos de tendencia central

Supongamos que nos interesa saber con qué frecuencia es que los estudiantes de cinematografía, de un departamento de comunicaciones de la universidad, estudian su disciplina asistiendo a cines con películas de estreno. Recolectamos una muestra aleatoria de 19 estudiantes y pedimos a cada uno mencionar las nuevas películas que vieron en el cine el mes pasado, y registramos los siguientes resultados: 2, 6, 4, 5, 2, 3, 4, 3, 6, 4, 3, 3, 5, 4, 5, 2, 3, 4, 3. La tabla 4-4 presenta estos datos en un formato de hoja de cálculo con los cálculos necesarios para calcular la media. Las puntuaciones se ordenan para facilitar el cálculo de la mediana y la moda.

Primero, calculemos la media:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{72}{19} = 3.79 \text{ películas}$$

En segundo término, calculemos la mediana. Ya ordenamos las puntuaciones, lo cual es necesario para calcular la mediana. El tamaño muestral ($n = 19$) dividido entre 2 es alrededor de nueve casos, y como n es impar, determinamos que el décimo caso es la mediana. En la

TABLA 4-4 | Datos organizados en hoja de cálculo: número de películas de estreno vistas en el último mes (X)

Especificaciones		
Número del sujeto	**Iniciales del sujeto**	***X***
1	BH	2
2	KP	2
3	JN	2
4	TW	3
5	JD	3
6	WA	3
7	KM	3
8	BC	3
9	CR	4
10	ML	4
11	MW	4
12	MF	4
13	JS	4
14	BY	4
15	LL	5
16	WF	5
17	CM	5
18	BL	6
19	SH	6
$n = 19$		$\Sigma X = 72$ películas

hoja de cálculo contamos hacia abajo hasta el décimo caso y descubrimos que la mediana son cuatro películas:

$$\text{Mdn} = 4 \text{ películas}$$

Por último, calculamos la moda. La observación de los datos ordenados en la tabla 4-4 revela que la puntuación que ocurre con más frecuencia es 4:

$$\text{Mo} = 4 \text{ películas}$$

Obviamente, el empleo de una hoja de trabajo para hacer cálculos manualmente con un gran número de casos sería difícil. El propósito de un ejercicio hecho a papel y lápiz es entender las características fundamentales de la estadística. En un trabajo real de investigación, se emplean paquetes de software de estadística y hojas de cálculo computarizadas para ahorrar tiempo y reducir errores en cálculos.

Formato de distribución de frecuencias para calcular la moda

La tabla 4-5 presenta los mismos datos sobre los 19 estudiantes de cinematografía, pero emplea un formato de distribución de frecuencias. Trabajando a partir de la hoja de cálculo de

TABLA 4-5 | Datos organizados en un formato de distribución de frecuencias: número de películas de estreno vistas el último mes (*X*)

Especificaciones	
X	**f**
2	3
3	5
4	6
5	3
6	2
$n = 19$	

la tabla 4-4 (como lo haría una computadora), en la tabla 4-5 vemos que hay una frecuencia de tres estudiantes que reportan dos películas, cinco que reportan tres películas, y así sucesivamente.

El cálculo de la moda es muy fácil con el formato de distribución de frecuencias. En la tabla 4-5 simplemente observamos la columna que indica las frecuencias (es decir, la columna *f*) y vemos cuál puntuación se presentó con más frecuencia. Más estudiantes (seis de ellos) vieron cuatro películas que ningún otro número de películas para el mes:

$$\text{Mo} = 4 \text{ películas}$$

La salida de distribución de frecuencias y los estadísticos descriptivos básicos son características estándar de los paquetes de software de estadística.

Insensatez y falacias estadísticas: mezcla de subgrupos en el cálculo de la media

Debido a que la media es susceptible de distorsión por valores y puntuaciones extremos, debemos describir con claridad qué casos o qué sujetos se incluyen en su cálculo. Organizaciones tales como son por ejemplo empresas o instituciones escolares, intencionalmente o no, por lo general reportan medias que son irreales. Por ejemplo, el vocero de un distrito escolar público puede reportar que el sueldo medio de sus maestros es $45 000. Cuando esto ocurra, es probable que los maestros se reúnan en el aula de descanso de la facultad y se pregunten entre sí: ¿Quién entre nosotros gana tanto dinero? Por supuesto, los maestros no son tontos. Ellos saben de inmediato que quien realizó los cálculos "mezcló los rangos de estatus" e incluyó al personal de mayor salario, por ejemplo, consejeros académicos, auxiliares de los directores, y directores, todos ellos certificados como docentes pero rara vez dan clases. Estos administradores quizá hayan sido incluidos porque el "estadista" simplemente pidió a la computadora calcular el sueldo medio para todos los maestros certificados sin tener en cuenta el rango. Cuando se incluyó este personal bien pagado, sus altos salarios sesgaron la media. Para evitar tal insensatez estadística, deben informarse por separado las medias para subgrupos distintos.

La mezcla de rangos de estatus resulta a veces en una media que no se ajusta a ningún grupo. Por ejemplo, una compañía puede tener sólo dos rangos de empleados: obreros que

promedian cerca de 30 000 dólares al año, y gerentes que promedian cerca de 70 000 dólares al año. Si estos dos grupos son aproximadamente del mismo tamaño, el sueldo medio para la compañía entera estaría cercano a $50 000. Curiosamente, ningún empleado de la compañía gana un sueldo cercano a esa cantidad.

Otro ejemplo es la edad media de asistentes de una clase nocturna de tercer grado de una escuela primaria. La edad media se calculará en 20 años más menos, aunque todos ahí tendrán 8 o 9 años (los niños) o alrededor de 30 (los padres). La media es ciertamente impropia para resumir esta distribución de edades.

RESUMEN

1. Un estadístico de tendencia central es el que proporciona una estimación típica, usual, normal o promedio que se encuentre en una distribución de puntuaciones brutas.
2. Hay tres medidas de tendencia central: la media, la mediana y la moda. Cada una de ellas tiene puntos fuertes y débiles.
3. Los valores relativos de las tres estadísticas de tendencia central nos informan acerca de la forma de una distribución de puntuaciones.
4. La media y la mediana son apropiadas con variables de intervalo/razón. Con variables nominales, las medias y las medianas no tienen sentido. La moda se puede usar con variables de todos los niveles de medida.
5. La media es el estadístico de tendencia central más útil.
6. El cálculo de la media resulta afectado por valores extremos y por un sesgo en la distribución de las puntuaciones.
7. La mediana es una puntuación posicional, la puntuación central en una distribución ordenada. Es igual al 50o. percentil.
8. Cuando una distribución está sesgada, la mediana es el estadístico a elegir porque su valor caerá entre la media y la moda y, así, minimiza el error.
9. Recuerda ordenar las puntuaciones de menor a mayor antes de calcular la mediana.
10. La mediana es insensible a los valores de las puntuaciones de una distribución. La mediana es sensible a un cambio en el tamaño muestral.
11. La moda es la puntuación o categoría que ocurre con más frecuencia en una distribución. La moda puede verse como la puntuación o categoría más popular, pero no debe confundirse la moda con "la mayoría de puntuaciones".
12. La moda es fácil de ubicar en tablas y gráficos. Al identificar la moda, debes tener cuidado en recordar que es una puntuación (*X*), no una frecuencia (*f*).
13. La moda es la menos útil de las medidas de tendencia central por sí misma, porque tiene un alcance informativo limitado, esto es, nos dice poco. La moda es insensible a los valores de puntuaciones de una distribución e insensible al tamaño muestral. Dos distribuciones de puntuaciones pueden tener formas radicalmente diferentes, pero tener la misma moda.

14. Una curva de distribución de frecuencias es un sustituto para un histograma o polígono de frecuencias en donde sustituimos estos gráficos con una curva suavizada.
15. Las ubicaciones relativas de la media, mediana y moda sobre el eje X son predecibles para ciertas formas de curvas de distribución. En una distribución normal o "curva normal", la media, mediana y moda de la variable son iguales. En una distribución sesgada negativamente, la media tendrá el valor de X más bajo, la moda más alta, y la mediana caerá entre ellas. En una distribución sesgada positivamente, la media tendrá el valor de X más alto, la moda el más bajo y la mediana caerá entre ellos.
16. Por lo general, la media de dos grupos combinados *no es* simplemente la suma de las medias dividida entre 2. Esto sólo funciona cuando los dos grupos son del mismo tamaño muestral.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 4 del material de texto disponibles en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2, incluyen ilustraciones donde la media y la mediana se pueden usar con variables ordinales ordenadas con ciertas características.

FÓRMULAS PARA EL CAPÍTULO 4

Cálculo de la media:

Trabajando con una hoja de cálculo:

$$\overline{X} = \frac{\Sigma X}{n}$$

Cálculo de la media combinada de dos grupos (a partir de puntuaciones individuales):

$$\overline{X}_{(grupos\ 1\ y\ 2\ combinados)} = \frac{\Sigma X_{(grupo1)} + \Sigma X_{(grupo2)}}{n_{(grupo1)} + n_{(grupo2)}}$$

Cálculo de la media combinada de dos grupos (a partir de medias de grupo):

$$\text{Como } \overline{X} = \frac{\Sigma X}{n}, \ \Sigma X = (n)(\overline{X})$$

Sustituye para obtener:

$$\overline{X}_{(grupos\ 1\ y\ 2\ combinados)} = \frac{n_{(grupo1)}\overline{X}_{(grupo1)} + n_{(grupo2)}\overline{X}_{(grupo2)}}{n_{(grupo1)} + n_{(grupo2)}}$$

Cálculo de la mediana:

1. Ordena la distribución de puntuaciones de la más baja a la más alta.
2. Ubica la posición de la mediana. Divide entre 2 el tamaño muestral, n, para obtener la puntuación central de la distribución. Si n es un número impar, la mediana será un caso real en la muestra. Si n es un número par, la mediana estará entre las dos puntuaciones centrales y se calcula tomando la media de estas dos puntuaciones. (Matemáticamente, la posición de la mediana se encuentra dividiendo entre 2 el tamaño de la muestra y sumando .5.)

Cálculo de la moda:

1. Agrupa las puntuaciones en una hoja de cálculo de puntuaciones ordenadas sin elaborar o formato de distribución de frecuencias.
2. Identifica Mo = valor de X con la frecuencia mayor.

PREGUNTAS PARA EL CAPÍTULO 4

1. Para cada estadístico de tendencia central, ¿variables de qué niveles de medición son apropiadas?
2. Define la media, la mediana y la moda. Especifica las limitaciones potenciales de cada una.
3. ¿Por qué es mejor calcular las tres medidas: la media, mediana y la moda, que confiar en una de ellas?
4. Como regla general, es incorrecto calcular la media para dos grupos combinados dividiendo simplemente entre 2 la suma de sus medias separadas. ¿Cuál es la excepción a esta regla?
5. Si una distribución de puntuaciones está sesgada, ¿qué único estadístico de tendencia central es más apropiado para una audiencia pública? ¿Por qué?
6. En general, la moda de una distribución es el estadístico de tendencia central menos útil. ¿Bajo qué circunstancias, no obstante, es el estadístico de tendencia central más apropiado a reportar?
7. Si la edad modal de una distribución es 22 años, ¿significa esto que una mayoría de las personas de esta población tiene 22 años de edad? Explica.
8. ¿Cómo se ubica la moda en un histograma, un polígono y una curva de distribución de frecuencias?
9. En una curva de distribución de frecuencias, ¿qué representan los ejes horizontal y vertical?
10. Describe las características de una curva de distribución de frecuencias normal.
11. Expresa en términos generales cómo un sesgo a la izquierda, en una distribución de frecuencias, afecta a los tres promedios comunes: media, mediana y moda.
12. Expresa en términos generales cómo un sesgo a la derecha, en una distribución de frecuencias, afecta a los tres promedios comunes: media, mediana y moda.
13. Supongamos que una distribución de edades tiene una media de 55 años, una moda de 28 años y una mediana de 34 años. ¿Cuál es la forma probable de la curva de distribución de frecuencias de esta variable?
14. Como se ilustra en "Insensatez y falacias estadísticas" de este capítulo, la media de una variable puede ser una mala medida de tendencia central cuando se mezclan los rangos de estatus dentro de una población. Da un ejemplo de cómo mezclar rangos puede resultar en una media que no se ajusta en absoluto a ningún rango.

EJERCICIOS PARA EL CAPÍTULO 4

Conjunto de problemas 4A

Recuerda incluir las fórmulas, estipular las unidades de medida y contestar la pregunta.

4A-1. Dados los datos siguientes con X = edad, calcula la edad modal, edad mediana y edad media. Empieza por organizar los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenados)".

X	X (cont.)
14	14
15	17
19	19
19	22
22	28

4A-2. Siete trabajadores de oficinas entraron a un concurso de reducción de peso. Tras unas semanas de someterse a dieta, los pesos que bajaron (en libras) fueron como sigue: 5, 7, 3, 0, 2, 4 y 3. Calcula la pérdida de peso mediana y media. Los datos de X = libras perdidas.

4A-3. Los expertos en demografía estudian las poblaciones de varios estados. Un sujeto de interés es el crecimiento o disminución del tamaño de una población, que es afectado por el índice de natalidad, longevidad (tiempo que viven las personas con base en edades en que normalmente mueren), y cuántos se establecen en un lugar o se van de éste (emigración). Una variable es la edad de mortalidad (es decir, la edad a su fallecimiento). Supongamos que en la nación A, la edad de mortalidad modal es 55, la mediana es 60 y la media es 65. En la nación B, la media también es 65, pero la moda es 75 y la mediana es 70.

a) Con esta información, construye las curvas de frecuencias de edad de mortalidad para cada nación.

b) ¿Qué nación parece encontrarse mejor en términos de longevidad?

4A-4. Los cinco miembros de una familia trabajan. Sus salarios por hora son: $30, $10.50, $5.15, $12, y $6. Los datos de X = salario por hora.

a) Calcula la media y la mediana.

b) En comparación con las otras puntuaciones, ¿cómo llamaríamos al salario de $30 por hora?

c) ¿Cuál es su efecto en el cálculo de la media?

d) Ajusta esta peculiaridad, recalculando la media sin ella.

4A-5. Los siguientes son precios de colegiaturas anuales para cinco universidades norteamericanas importantes: $10 000, $29 000, $8 000, $12 500, $11 300. Los datos Y = precio de colegiatura.

a) Calcula la media y la mediana.

b) En comparación con las otras puntuaciones, ¿cómo llamaríamos al precio de colegiatura de $29 000?

c) ¿Cuál es su efecto en el cálculo de la media?

d) Ajusta esta peculiaridad, recalculando la media sin ella.

4A-6. La edad media de los 47 hombres del club de bridge de Sparkesville es 54.8 años. La edad media de las 62 mujeres del club es 56.4 años. ¿Cuál es la edad media de los 109 miembros? Los datos X = edad.

4A-7. En un experimento para ver si los pollos pueden distinguir colores, le dan premios de granos de maíz a un pollo cuando pica correctamente un cojincillo de igual color. Los tiempos de reacción se miden al centésimo de segundo más cercano. Los tiempos de reacción de Flossy son como sigue: 1.32, 1.45, 1.21, 1.05, .97, .91, .93, .93, .96, .93, .88, .94, .98.

Los datos X = tiempo de reacción.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenada)".

b) Calcula los tiempos de reacción de media, mediana y modal de Flossy.

c) Describe la forma de la distribución de los tiempos de reacción de Flossy.

4A-8. Dados los siguientes estadísticos y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de curva	Trazo de curva
Edad (años)	30	35	39		
Tamaño de la familia	4.1	3.0	2.0		
Años empleado	11	8	7		
Peso (libras)	160	132	134		

Conjunto de problemas 4B

Recuerda incluir fórmulas, estipular unidades de medida y contestar la pregunta.

4B-1. Los datos siguientes son para la variable Y = distancia desde el lugar de trabajo (en millas) para los empleados de un vendedor de copiadoras. Calcula las puntuaciones de la media y la mediana. Empieza por organizar los datos en una tabla de hoja de cálculo con puntuaciones ordenadas bajo una columna marcada "Y (ordenados)".

Y	Y (cont.)
13	10
9	11
6	14
3	5
12	7

4B-2. Las puntuaciones de la parte analítica del examen de registros de graduados (GRE) de cinco candidatos a un programa de graduados fueron como sigue: 700, 625, 640, 590 y 600. Calcula las puntuaciones de la media y la mediana. Los datos X = puntuación GRE.

4B-3. Al evaluar los porcentajes de delincuencia entre dos ciudades, un criminalista calcula X = número promedio de vehículos robados por día (en un periodo de seis meses). Para la ciudad A, la moda de X es 15 vehículos, la mediana es 20 y la media es 25. Para la ciudad B, la media es también 25, pero la moda es 35 y la mediana es 30.

a) A partir de esta información, construye las curvas de frecuencias para cada una de las ciudades.

b) ¿En cuál ciudad te sentirías más seguro de estacionar tu auto en la calle? ¿Por qué?

4B-4. Los siguientes son promedios puntuales de calificaciones (GPA en una escala de 4 puntos) de estudiantes en un programa de clases prácticas: 1.7, 2.6, 2.3, 3.9, 2.2, 1.9, 2.1. Los datos Y = GPA.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenada)".

b) Calcula la media y la mediana de Y.

c) En comparación con las otras puntuaciones, ¿cómo llamaríamos al GPA de 3.9?

d) ¿Cuál es su efecto en el cálculo de la media?

e) Ajusta esta peculiaridad, recalculando la media sin ella.

4B-5. Las siguientes son calificaciones de examen final para cinco estudiantes, pasantes de ciencias sociales, en una importante universidad urbana: 90, 88, 64, 92, 87. Los datos X = calificación de examen.

a) Calcula la media y mediana de calificación del examen.

b) En comparación con las otras puntuaciones, ¿cómo llamaríamos a la puntuación de 64?

c) ¿Cuál es su efecto en el cálculo de la media?

d) Ajusta esta peculiaridad, recalculando la media sin ella.

4B-6. Supongamos que las siguientes son las edades medias de pacientes adictos a sustancias, en un hospital local para su tratamiento, separadas por tipo de adicción. Calcula la edad media de todos los pacientes adictos a sustancias del hospital. Los datos X = edad.

	Adicción primaria			
	Cocaína (n = 44)	Cocaína *crack* (n = 29)	Heroína (n = 24)	Alcohol (n = 69)
Edad media (años)	29.8	23.4	34.6	42.9

4B-7. Los promedios de bateo de la alineación inicial del equipo de ligas pequeñas, los *Dodgers Bola Rápida*, son como sigue: .360, .200, .350, .355, .230, .345, .360, .380, y .400. Los datos X = promedio individual de bateo.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas.

b) Calcula los porcentajes de bateo de media, mediana y moda del equipo.

c) Describe la forma de la distribución.

4B-8. Dados los siguientes estadísticos y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de la curva	Trazo de la curva
Estatura (pulgadas)	70	68	66		
Exámenes este semestre	10	13	15		
Puntuación de espiritualidad	30	30	30		
Presupuesto de abarrotes	$130	$109	$104		

Conjunto de problemas 4C

Recuerda incluir las fórmulas, estipular las unidades de medida y contestar la pregunta.

4C-1. A partir de la siguiente serie de mediciones de estaturas (en pulgadas), calcula las estaturas modal, mediana y media. Los datos X = estatura (en pulgadas). Empieza por organizar los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "X (ordenada)".

X	X (cont.)
60	78
70	59
68	67
72	74
69	70

4C-2. Supongamos que los siguientes son números de victorias de conferencia entre siete equipos colegiales de baloncesto: 12, 8, 7, 9, 11, 5 y 4. Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas. Calcula el número de mediana y media de victorias entre estos equipos. Los datos X = número de victorias o "ganes".

4C-3. Al comparar las puntuaciones de un examen entre los estudiantes, el director de un departamento calcula las calificaciones del examen de dos grupos. Para el grupo A, la moda es 75 puntos, la mediana es 80 y la media es 85. Para el grupo B, la media es también 85, pero la moda es 95 y la mediana es 90. Los datos X = calificación de examen.

a) A partir de esta información, construye las curvas de frecuencias de las calificaciones del examen para cada grupo.

b) ¿Cuál grupo parece estar mejor en este examen en particular?

4C-4. Los siguientes son salarios anuales (Y) entre siete médicos empleados en una zona urbana importante: $88 000, $94 000, $86 000, $110 000, $212 000, $115 000 y $97 000.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*Y* (ordenada)".

b) Calcula el salario medio y mediano.

c) En comparación con otras puntuaciones, ¿cómo llamaríamos al salario de $212 000?

d) ¿Cuál es su efecto en el cálculo de la media?

e) Ajusta esta peculiaridad, recalculando la media sin ella.

4C-5. Las siguientes son evaluaciones del rendimiento de empleados (completadas por sus supervisores) de una importante empresa fabricante de software. Cada empleado evaluado en una escala de 0 a 10, con base en su rendimiento de varios indicadores establecidos. Los datos Y = evaluación de empleado.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*Y* (ordenada)".

b) Calcula la media y la mediana.

c) En comparación con otras puntuaciones, ¿cómo llamaríamos a la evaluación de 3?

d) ¿Cuál es su efecto en el cálculo de la media?

e) Ajusta esta peculiaridad, recalculando la media sin ella.

Y
8
3
9
7
8
6
9
8
7

4C-6. Supongamos que el peso medio de 34 hombres que participan en un programa local de reducción de peso es 228 libras, en tanto que el peso medio de 46 mujeres que participan en el mismo programa es de 194 libras. ¿Cuál es el peso medio de los 80 participantes? Los datos X = peso (en libras).

4C-7. Supongamos que nueve amigos están compitiendo todos contra todos en una carrera cronometrada de 40 yardas. El tiempo de cada uno de los participantes (en segundos) es como sigue: 4.8, 5.2, 4.7, 4.9, 5.4, 4.8, 4.9, 4.8 y 5.3. Los datos X = tiempo (en segundos).

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas.

b) Calcula los tiempos de media, mediana y moda de la carrera de 40 yardas del grupo.

c) Describe la forma de la distribución.

4C-8. Dadas las siguientes estadísticas y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de la curva	Trazo de la curva
Peso (libras)	195	205	215		
Dinero para gastar	$150	$140	$125		
Puntuación escala de depresión	25	25	25		
Edad (años)	30	40	50		

Conjunto de problemas 4D

Recuerda incluir fórmulas, estipular unidades de medida y contestar la pregunta.

4D-1. Dadas las siguientes mediciones de peso de varios amigos, calcula los pesos modal, mediano y medio. Los datos X = peso (en libras). Empieza por organizar los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*X* (ordenada)".

X	X (cont.)
158	180
180	195
169	200
190	195
180	160

4D-2. Bajo la supervisión de sus maestros, un pequeño grupo de estudiantes adolescentes decidió evaluar su crecimiento general en estatura (en pulgadas), para un periodo de 18 meses. Las diferencias en sus estaturas entre el inicio (tiempo 1) y terminación (tiempo 2) son como sigue: 4.4, 6.0, 3.6, 2.9, 4.3, 3.6, 2.9, 4.2 y 2.8. Calcula el crecimiento mediano y medio. Los datos X = crecimiento en 18 meses (en pulgadas).

4D-3. Un investigador está interesado en comparar patrones de ingreso familiar en dólares de Estados Unidos (X) entre dos comunidades de clase media alta. Para la Comunidad 1, el ingreso modal familiar es $80 000, la mediana es $90 000 y la media es $100 000. Para la Comunidad 2, la media es también $100 000, pero la moda es $120 000 y el ingreso mediano familiar es $110 000.

a) De esta información, construye las curvas de frecuencias de ingreso familiar para cada comunidad.

b) ¿Qué comunidad parece ser más acaudalada con respecto al ingreso familiar?

4D-4. Los siguientes son tamaños de grupo para siete grupos de introducción a la sociología en una importante universidad urbana: 65, 79, 72, 115, 84, 87 y 78. Los datos Y = número de estudiantes.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*Y* (ordenada)".

b) Calcula la media y la mediana.

c) En comparación con otras puntuaciones, ¿cómo llamaríamos al tamaño del grupo de 115?

d) ¿Cuál es su efecto en el cálculo de la media?

e) Ajusta esta peculiaridad, recalculando la media sin ella.

4D-5. Los siguientes son números de empleados pagados entre nueve subsidiarias de una importante empresa financiera internacional. Los datos X = número de empleados.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*Y* (ordenada)".

b) Calcula la media y la mediana.

c) En comparación con otras puntuaciones, ¿cómo llamaríamos a la subsidiaria con sólo 67 empleados?

d) ¿Cuál es su efecto en el cálculo de la media?

e) Ajusta esta peculiaridad, recalculando la media sin ella.

X
212
198
283
176
191
254
67
187
193

4D-6. La puntuación media de examen de registro de graduados (GRE) de los 39 solicitantes hombres al departamento de sociología de la universidad central es 1 140 puntos GRE. La puntuación media para las 54 solicitantes mujeres es 1 210. ¿Cuál es la puntuación media GRE para los 93 solicitantes? Los datos X = puntuación GRE.

4D-7. Nueve amigos compiten todos contra todos en una liga de fútbol americano. Las yardas de pase para los mariscales de campo estrellas de cada competidor de la semana anterior son como sigue: 283, 205, 183, 197, 296, 315, 304, 227 y 296. Haz X = yardas de pase.

a) Organiza los datos en una tabla de hoja de cálculo con las puntuaciones ordenadas bajo una columna marcada "*Y* (ordenada)".

b) Calcula las yardas de pase media, mediana y moda.

c) Describe la forma de la distribución.

4D-8. Dados los siguientes estadísticos y lo que sabemos acerca de cómo están relacionadas dentro de una distribución de puntuaciones, describe la probable forma de la distribución para cada una de las variables citadas. Traza la curva indicando las ubicaciones relativas de la media, mediana y moda.

Variable	$\bar{X}$	Mdn	Mo	Forma de la curva	Trazo de la curva
Gastos de entretenimiento	$163	$154	$139		
Puntos de la escala de religiosidad	30	30	30		
Niveles de colesterol	182	207	219		
Índice de masa corporal	25	30	33		

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 4

En el sitio web www.mhhe.com/ritchey2, en *The Statistical Imagination*, están disponibles ejercicios computarizados opcionales del capítulo. Estos ejercicios incluyen la generación de estadísticos de tendencia central con *SPSS Windows* y el uso de la salida para incrementar el sentido de proporción acerca de las formas de distribuciones de puntuaciones para variables de intervalo/razón. Las estadísticas de tendencia central se pueden calcular usando el comando *Descriptives* o el comando *Frequencies*, que se describen en el apéndice D.

CAPÍTULO

5

Medición de la dispersión o variación en una distribución de puntuaciones

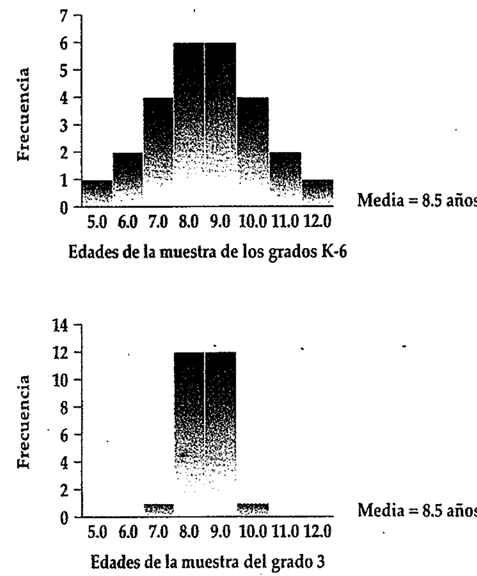
RESUMEN DEL CAPÍTULO

Introducción

Para una variable de intervalo/razón, reportar un estadístico central por sí mismo no es suficiente para comunicar la forma de una distribución de puntuaciones. Dos muestras con las mismas medias pueden tener formas sumamente diferentes. La figura 5-1 presenta dos distribuciones de edades: para una muestra de alumnos de escuela primaria (desde jardín de niños hasta sexto grado, o K-6) y un grupo de tercer grado de otra escuela. La edad media de los alumnos de ambas escuelas es de 8.5 años. En la escuela K-6, sin embargo, los niños tienen entre 5 y 12 años; en el tercer grado de la otra escuela ninguno de los alumnos es menor de 7 años ni mayor de 10 años. Aun cuando estas dos distribuciones de edades tienen la misma tendencia central, sus puntuaciones se dispersan de manera muy diferente, con una mayor dispersión de edades en la escuela K-6.

El tema de este capítulo es la **dispersión**, es decir, *cómo se dispersan las puntuaciones de una variable de intervalo/razón de menor a mayor y la forma de la distribución de éstas.* Existe un número infinito de posibles formas de distribución para una variable con una media dada. Todas las puntuaciones podrían agruparse alrededor de la media con la clara forma de una curva de campana, aunque la curva podría ser de diferentes tamaños según el tamaño de la muestra, o bien, las puntuaciones podrían estar ligeramente o muy sesgadas hacia un lado. Además, una sola variable puede tener dispersiones muy diferentes de

FIGURA 5-1
Comparación de la dispersión de las edades de los alumnos de dos muestras con las mismas medias



una población a otra. Por ejemplo, el ingreso familiar anual de residentes en Estados Unidos varía desde cero hasta decenas de millones de dólares, mientras que el ingreso familiar de los pobres que viven en proyectos habitacionales va de cero a unos pocos miles de dólares.

> **Dispersión** Forma en que se dispersan las puntuaciones de una variable de intervalo/razón de menor a mayor y la forma de la distribución entre éstas.

Los **estadísticos de dispersión** *describen cómo se dispersan las puntuaciones de una variable de intervalo/razón a lo largo de su distribución.* Los estadísticos de dispersión permiten descripciones precisas de la frecuencia de casos en cualquier punto de una distribución. Por ejemplo, si el gobierno federal decide aumentar los impuestos para los "ricos", empleando estadísticos de dispersión podemos identificar el nivel de ingresos del 5 por ciento más ricos de todas las familias del país. Del mismo modo, si un programa de asistencia social se planea para cubrir sólo 10 000 familias de la ciudad, podemos establecer qué nivel de ingreso familiar satisface los requisitos para recibir la asistencia. Estudiar la dispersión es como ir y venir en un paseo por el eje *X* de un histograma y observar dónde se concentran los casos. ¿La mayor parte de los casos caen alrededor de la media o están cargados hacia algún lado? ¿Cuántos casos caen entre dos puntos? ¿Qué valor de la variable se lleva el 10 por ciento de los casos? Los dos estadísticos de dispersión que más se emplean son el rango y la desviación estándar.

> **Estadísticos de dispersión** Son estadísticos que describen cómo se dispersan las puntuaciones de una variable de intervalo/razón a lo largo de su distribución.

El rango

El **rango** es *una expresión de cómo las puntuaciones de una variable de intervalo/razón se distribuyen de menor a mayor*, es decir, es la distancia entre las puntuaciones mínima y máxima de una muestra. Se calcula como la diferencia entre las puntuaciones máxima y mínima, más el valor de la unidad de redondeo. El valor de la unidad de redondeo (1, por ejemplo, si las puntuaciones se redondean al número entero más cercano, 0.1 si las puntuaciones se redondean al décimo más cercano, y así sucesivamente) se suma para considerar el límite real inferior de la puntuación más baja y el límite real superior de la puntuación más alta.

Cálculo del rango de una variable X de intervalo/razón

1. Ordena las puntuaciones de la distribución de menor a mayor.
2. Identifica las puntuaciones mínima y máxima.
3. Identifica el valor de la unidad de redondeo (véase el apéndice A como repaso).
4. Calcula el rango:

Rango = (puntuación máxima – puntuación mínima) + valor de la unidad de redondeo

El rango Es una expresión de la forma en que las puntuaciones de una variable de intervalo/razón se distribuyen de menor a mayor.

Calculemos el rango en un ejemplo. Supongamos que X = edad (redondeada al año más cercano) y tenemos la siguiente distribución de puntuaciones:

21, 23, 43, 26, 20, 21, 25

Empieza por ordenar las puntuaciones:

20, 21, 21, 23, 25, 26, 43

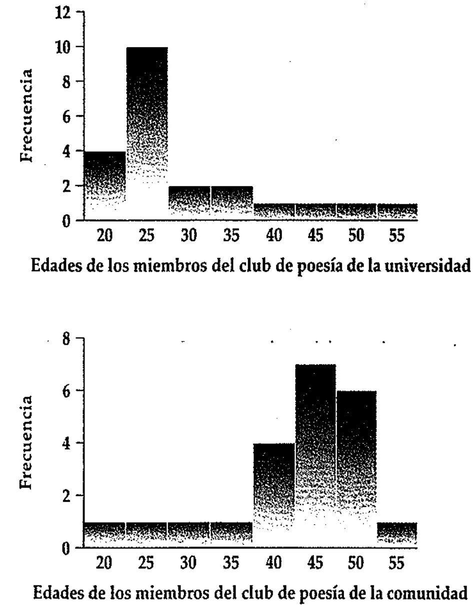
Identifica las puntuaciones mínima y máxima de 20 y 43, respectivamente, y distingue que la unidad de redondeo es 1.

Calcula el rango:

Rango = (puntuación máxima – puntuación mínima) + valor de la unidad de redondeo
= (43 – 20) + 1 = 24 años

Como resultado del redondeo, el individuo que reportó 20 podría tener 19.5 años; y el de 43 años podría tener 43.5 años. El rango de 24 años es la distancia entre estos límites reales menor y mayor de las puntuaciones, es decir, 43.5 años – 19.5 años = 24 años.

FIGURA 5-2
Comparación de dos distribuciones con formas diferentes que tienen el mismo rango



A veces resulta más informativo reportar las puntuaciones mínima y máxima por sí mismas, señalando que estas edades varían desde 20 hasta 43. De esta manera, indirectamente indicamos que en la muestra no hay menores de 20 años ni mayores de 43 años de edad.

Limitaciones del rango: situaciones en las que reportarlo solo puede conducir a errores

Puesto que el rango utiliza las puntuaciones más extremas de una distribución, un valor aislado inflará enormemente su cálculo. Esto sucedió para las siete edades indicadas anteriormente. Los 43 años hicieron que el rango pareciera estar extendido por encima de los 24 años. Reportar esto daría la impresión de que la muestra tiene un número considerable de sujetos de 30 y 40 años. Un reporte más exacto estipularía que, con excepción del estudiante de 43 años, las edades tenían un rango de 7 años (26 – 20 + 1 = 7 años). Omitir el valor aislado e indicarlo como excepción es una forma razonable de ajustar esta limitación del rango.

El rango también está limitado por su estrecho alcance informativo. No nos dice nada sobre la forma de la distribución entre las puntuaciones extremas. Por ejemplo, las dos distribuciones descritas en la figura 5-2 tienen el mismo rango, lo que sugiere formas similares, pero de hecho sus formas son radicalmente diferentes. Por último, hay poco que pueda hacerse matemáticamente con el rango. En suma, el rango tiene utilidad limitada, en especial cuando se reporta solo.

La desviación estándar

La desviación estándar es otra medida sumaria de la dispersión o variación de las puntuaciones de una distribución. Este estadístico de dispersión es muy diferente del rango. Al concentrarse en los extremos de la distribución, el rango se aproxima a la dispersión desde "fuera" o desde los extremos de la distribución. Observar el rango es como ver un juego de baloncesto desde lo alto de las tribunas; la cancha parece encajonada por los tableros de cada extremo. En contraste, **la desviación estándar** *describe la forma en que las puntuaciones de una variable de intervalo/razón se dispersan a lo largo de la distribución en relación con la puntuación media.* La media es un estadístico de tendencia central y como tal proporciona un punto de enfoque que se centra "dentro" de la distribución. Observar la dispersión a partir de la media con su desviación estándar es como mirar desde el centro de la cancha; el centro de atención está en la distancia del centro de la cancha a otros puntos en cualquier dirección. Al igual que la media, la desviación estándar es muy apropiada con variables de intervalo/razón.

> **La desviación estándar** Describe la forma en que las puntuaciones de una variable de intervalo/razón se dispersan por la distribución en relación con la puntuación media.

Pensamiento proporcional y lineal sobre la desviación estándar

Para una variable de intervalo/razón, la desviación estándar se calcula determinando qué tan alejada está cada puntuación de la media, es decir, cuánto *se desvía* de la media. En este sentido, la desviación estándar es una derivada (o producto) de la media, y las dos medidas siempre se reportan juntas. De hecho, la frase "la media y la desviación estándar" es una de las más empleadas por los estadísticos. La desviación estándar, como una medida sumaria de todas las puntuaciones de una distribución, nos dice con qué amplitud se agrupan las puntuaciones alrededor de la media. Como brevemente lo analizaremos, la desviación estándar también es útil en conjunción con la curva normal.

La siguiente es la fórmula para calcular la desviación estándar:

Cálculo de la desviación estándar

$$s_X = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n - 1}}$$

donde

s_X = desviación estándar para la variable X de intervalo/razón

$\bar{X}$ = media de X

n = tamaño muestral

Merece la pena seguir un método de paso a paso al cálculo de la desviación estándar. Esto elimina el misterio de la fórmula (con sus símbolos de Σ, cuadrado y raíz cuadrada) y nos ayuda a apreciar que la desviación estándar es parte esencial de la curva normal.

Identifica especificaciones Comenzamos por identificar la información dada.

Especificación: X = una variable de intervalo/razón, n = tamaño muestral, y una distribución de puntuaciones en bruto para X.

Calcula la media Calculamos la media porque la desviación estándar está diseñada para medir la dispersión alrededor de la media.

$$\bar{X} = \frac{\Sigma X}{n}$$

Calcula las puntuaciones de desviación: pensamiento lineal A continuación determinamos qué tan alejada está la puntuación de cada individuo respecto a la media. La diferencia entre una puntuación y su media se llama **puntuación de desviación**, es decir, *cuánto difiere o se "desvía" de la media una puntuación individual:*

$$X - \bar{X} = \text{puntuación de desviación para un valor de } X$$

Considere una puntuación de desviación como una medida de distancia en el eje X. ¿Qué nos dice la puntuación de desviación? Supongamos que X es la variable *peso*, y el peso medio es de 138 libras para una muestra de jugadoras de voleibol de la Universidad de Elmstown. La jugadora estrella, Sandra "Mil Amores" Carson, pesa 173 libras; ésta es su puntuación en bruto o "puntuación X". Su puntuación de desviación es 35 libras:

$$\text{Puntuación de desviación} = X - \bar{X} = 173 - 138 = 35 \text{ libras}$$

La puntuación de desviación nos dice dos cosas sobre una puntuación de la distribución: (1) la cantidad o distancia a la que la puntuación X se aleja de la media y (2) la dirección de la puntuación X: si está abajo o arriba de la media. Cuando una puntuación X es mayor que la media, la puntuación de desviación resultará un valor positivo, como el de Sandra, lo cual significa que la puntuación X se encuentra a la derecha de una curva de distribución. Cuando una puntuación X es menor que la media, la puntuación de desviación resultará negativa, lo que significa que la puntuación X queda a la izquierda de la media. La puntuación de desviación de Sandra de +35 libras nos indica que ella está 35 libras *por encima* del peso medio del equipo.

> **Puntuación de desviación** Indica cuánto es que una puntuación individual difiere o "se desvía" de la media.

La puntuación de desviación es el cálculo matemático central para determinar la desviación estándar. Como una medida breve para toda la muestra, la desviación estándar es una suma y promedio del cuadrado de estas puntuaciones de desviación, como en los pasos siguientes.

Suma las puntuaciones de desviación El siguiente paso para calcular la desviación estándar es sumar las puntuaciones de desviación. Esta suma siempre será igual a cero (dentro del error de redondeo):

$$\Sigma (X - \bar{X}) = 0 = \text{suma de las puntuaciones de desviación}$$

La suma de las puntuaciones de desviación es una verificación respecto a la exactitud de los cálculos, porque la suma de las puntuaciones de desviación *siempre* será igual a cero (dentro del error de redondeo). En el capítulo 4 vimos la forma en que la media es un punto de equilibrio en la distribución. Lo que hace la media es balancear las desviaciones, para que se cancelen entre sí y resulten en una suma de puntuaciones de desviación igual a cero. De hecho, otra definición matemática de la **media** es *aquel punto en una distribución donde las puntuaciones de desviación suman cero.*

Eleva al cuadrado las puntuaciones de desviación y suma los cuadrados La dispersión de una variable a menudo se compara para dos o más muestras. El hecho de sumar las puntuaciones de desviación no detectará una diferencia en la dispersión entre dos muestras, porque la suma para ambas será cero. Esto potencialmente nos deja en un callejón sin salida. Si las puntuaciones de una muestra se dispersan ampliamente y si las de la otra lo hacen de manera estrecha, ¿qué beneficio implica informar que ambas tienen una suma de puntuaciones de desviación de cero? ¡Ninguno! Por consiguiente, al comparar dos muestras, debemos encontrar una manera de sumar las puntuaciones de desviación para que la suma sea más grande para una muestra con una dispersión mayor. La solución más útil consiste en elevar al cuadrado cada puntuación de desviación y después sumar los cuadrados. Al elevar al cuadrado se eliminan los signos negativos en las puntuaciones de desviación. La *suma de las puntuaciones de desviación al cuadrado* es la **variación** (a menudo se denomina **suma de cuadrados**), *un estadístico que resume las desviaciones para toda la muestra:*

$$\Sigma (X - \overline{X})^2 = \text{la variación (o “suma de cuadrados”)}$$

> **La variación o suma de cuadrados** Es la suma de las puntuaciones de desviación al cuadrado; un estadístico que resume las desviaciones para la muestra entera.

Divide la suma de cuadrados entre *n* – 1 para ajustar el tamaño y el error de la muestra: pensamiento proporcional La suma de cuadrados, o variación, constituye una buena medida de la dispersión de una distribución, pero este estadístico presenta dos problemas. Primero, supongamos que deseamos comparar las distribuciones de dos muestras de tamaños diferentes. Por ejemplo, podemos comparar las distribuciones de los promedios para muestras de estudiantes de la Crosstown University (n = 88) y de la universidad local (n = 104). Cuando sumamos los cuadrados para cada muestra, podría suceder que obtuviéramos una suma más alta para la universidad estatal, simplemente porque sumamos más números, 104 casos en lugar de sólo 88. Cada puntuación X añade cierta cantidad al cálculo. En otras palabras, todo lo demás es igual, y cuanto más observaciones existan, mayor será la suma de cuadrados. Para realizar una comparación equilibrada de dos muestras de tamaño diferente, entonces, necesitamos ajustar el número de observaciones en cada muestra dividiendo cada uno entre su tamaño muestra (n). Esto nos da la variación promedio (la media de la suma de cuadrados) en cada muestra. De esta forma ajustamos la suma de cuadrados en proporción al número de casos en la muestra.

Una segunda consideración respecto al tamaño de la muestra es que incluirá el error de muestreo; cuanto mayor sea la muestra, menor será el error de muestreo. Los estadistas han determinado que si restamos 1 de n, este pequeño ajuste produce un estadístico de la muestra que estima con mayor precisión el parámetro de la población. Dicho en otras palabras, si restamos 1 del tamaño de la muestra, realizamos un ajuste para el error de muestreo. (Considera

que con muestras grandes este ajuste tendría poco efecto en el cálculo, mientras que con muestras pequeñas tendría un gran efecto.)

En resumen, dividimos la variación (suma de cuadrados) entre n –1 para compensar tanto los efectos del tamaño muestral de la suma como el error de muestreo. El resultado se llama **varianza**, y su símbolo es s_X^2:

$$s_X^2 = \frac{\Sigma (X - \overline{X})^2}{n - 1} = \text{varianza de una muestra}$$

La **varianza** es *la variación promedio de las puntuaciones en una distribución*. Para evitar confundir la varianza y la variación nota el sonido acentuado en “varianza” y advierte que n está en su denominador. (Finalmente, debemos hacer notar que si la desviación estándar se calcula para las puntuaciones de una población entera el error de muestreo no constituirá un problema. Por consiguiente, no necesitamos restar 1 de n para obtener la variación de una población, que se simbolizaría como σ_X^2.)

> **La varianza** Es la variación promedio de las puntuaciones en una distribución (es decir, la media de la suma de cuadrados).

Saca la raíz cuadrada de la varianza para obtener la desviación estándar Para producir una buena medida de dispersión se requiere un último paso. La varianza es perfectamente aceptable para cálculos, pero no se interpreta de manera directa porque las unidades de medida están elevadas al cuadrado. Así, podríamos calcular la varianza de peso para el equipo de fútbol de la universidad local, y encontraríamos que es de 1 391.45 libras cuadradas. Bien, ¿qué es una “libra cuadrada”? Es una libra multiplicada por una libra, pero excepto quizás un matemático ¿quién sabe lo que realmente significa? Necesitamos una unidad de medida directamente interpretable, libras en lugar de libras al cuadrado. Para “regresar” a libras, sacamos la raíz cuadrada de la varianza. (La raíz cuadrada de una unidad de medida al cuadrado es la unidad de medida en sí.) El resultado es la desviación estándar:

$$s_X = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n - 1}} = \sqrt{s_X^2}$$

En el caso del peso del equipo local la desviación estándar sería 37.30 libras:

$$s_X = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n - 1}} = \sqrt{s_X^2}$$

$$= \sqrt{1\,391.45} = 37.30 \text{ libras}$$

Los elementos de la ecuación de la desviación estándar —es decir, las puntuaciones de desviación, la suma de los cuadrados o variación y la varianza— son importantes por sí mismos. Estos elementos aparecen por sí mismos en numerosas fórmulas estadísticas (véase, por ejemplo, el capítulo 12). Los pasos para calcular la desviación estándar se resumen en la tabla 5-1, que el lector encontrará sumamente útil en capítulos posteriores.

Es buena práctica elaborar una hoja de trabajo para estos cálculos. La tabla 5-2 presenta una hoja de trabajo para calcular la desviación estándar de los pesos de 12 de los 98 jugadores del equipo de fútbol de Crosstown.

Para calcular las puntuaciones de desviación, $X - \bar{X}$, calculamos la media y restamos cada puntuación de ella para obtener la tercera columna de la hoja de cálculo:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{2\,856}{12} = 238 \text{ libras}$$

TABLA 5-1 | Para comprender la desviación estándar por medio de su cálculo

Pasos en el cálculo de la desviación estándar	Lo que obtiene el paso
1. Identifica las especificaciones.	**1.** X debe ser una variable de intervalo/razón.
2. Calcula la media: $\bar{X} = \frac{\Sigma X}{n}$	**2.** Porque la desviación estándar está basada en las desviaciones desde la media.
3. Calcula las puntuaciones de desviación: $X - \bar{X}$	**3.** Para determinar la distancia de cada puntuación desde la media.
4. Suma las puntuaciones de desviación: $\Sigma(X - \bar{X})$	**4.** Asegúrate de que $\Sigma(X - \bar{X}) = 0$
5. Eleva al cuadrado las puntuaciones de desviación y súmalas para obtener la variación de la suma de cuadrados: Variación $= \Sigma(X - \bar{X})^2$	**5.** Las puntuaciones de desviación se elevan al cuadrado para eliminar signos negativos y obtener una suma diferente de cero.
6. Calcula la varianza: $s_X^2 = \frac{\Sigma(X - \bar{X})^2}{n-1}$	**6.** Divide entre $n-1$ la suma de cuadrados para ajustar el tamaño muestral y el error de redondeo.
7. Calcula la desviación estándar, s_x: $s_X = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{s_X^2}$	**7.** Toma la raíz cuadrada de la varianza para obtener directamente unidades de medida que se puedan interpretar (unidades en lugar de unidades al cuadrado).

TABLA 5-2 | Hoja de trabajo para calcular la desviación estándar: peso de jugadores de fútbol de Crosstown ($n = 12$)

Especificaciones		Cálculos	
(1) Jugador	(2) X	(3) $X - \bar{X}$	(4) $(X - \bar{X})^2$
1	165	−73	5 329
2	200	−38	1 444
3	216	−22	484
4	217	−21	441
5	226	−12	144
6	236	−2	4
7	239	1	1
8	244	6	36
9	261	23	529
10	268	30	900
11	283	45	2 025
12	301	63	3 969
$n = 12$	$\Sigma X = 2\,856$ libras	$\Sigma(X - \bar{X}) = 0$	$\Sigma(X - \bar{X})^2 = 15\,306$ libras al cuadrado

Por último, elevamos al cuadrado las puntuaciones de desviación de la columna 3 para obtener la columna 4. La suma de la columna 4 de la tabla 5-2 y el tamaño de la muestra n son todo lo que necesitamos para calcular la desviación estándar:

$$s_X = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{15\,306}{11}} = \sqrt{1\,391.45} = 37.30 \text{ libras}$$

Limitaciones de la desviación estándar

Como la desviación estándar se calcula a partir de la media, al igual que ésta se infla por los valores extremos. Éstos generan puntuaciones con grandes desviaciones. Cuando se elevan al cuadrado, estas grandes desviaciones, ya sean positivas o negativas, producen un alto resultado positivo e inflado. Así, la desviación estándar puede ser muy confusa cuando se reporta para una distribución sesgada, en la que pocas puntuaciones se extienden en una dirección. Para convencerse del efecto de las puntuaciones extremas tanto en la media como en la desviación estándar, completa la hoja de cálculo de la tabla 5-2; pero agrega los dos casos siguientes para obtener una nueva muestra con $n = 14$: el jugador 13 que pesa 115 libras y el jugador 14 que pesa 125 libras. A continuación compara las respuestas de las muestras original y nueva.

Para calcular el rango y la desviación estándar

Problema: calcula el rango, media y desviación estándar de los impuestos de gasolina cobrados por 10 estados seleccionados del oeste de Estados Unidos. Estos estados y sus impuestos se presentan en la tabla 5-3, donde X = impuesto de gasolina por galón.

TABLA 5-3 | Impuestos estatales a la gasolina en estados seleccionados del oeste en mayo de 1996.

Especificaciones		Cálculos	
Estado	Impuesto (¢) por galón X	$X - \bar{X}$	$(X - \bar{X})^2$
Nuevo México	17	-4.7	22.09
California	18	-3.7	13.69
Arizona	18	-3.7	13.69
Utah	19	-2.7	7.29
Colorado	22	0.3	0.09
Washington	23	1.3	1.69
Nevada	23	1.3	1.69
Oregón	24	2.3	5.29
Idaho	25	3.3	10.89
Montana	28	6.3	39.69
	$\Sigma X = 217¢$		$\Sigma(X - \bar{X})^2 = 116.10¢$ al cuadrado
$n = 10$		$\Sigma(X - \bar{X}) = 0$	

Fuente: Tarifas de impuestos de http://www.api.org/news/596sttax.htm. Copyright© 1996 por American Petroleum Institute. Reimpreso con permiso del Instituto.

1. Asegúrate de que la variable sea de nivel de medición de intervalo/razón (como en el caso de los impuestos a la gasolina).
2. Organiza los datos en una hoja de cálculo con puntuaciones ordenadas de menor a mayor y con los siguientes encabezados de columna:

Caso	X	$X - \bar{X}$	$(X - \bar{X})^2$

donde "Caso" = número o nombre del caso,

X = puntuación observada, sin elaborar, de la variable de intervalo/razón,

$X - \bar{X}$ = puntuación de desviación,

$(X - \bar{X})^2$ = cuadrado de la puntuación de desviación.

3. Calcula el rango. Con estas puntuaciones ordenadas, vemos que la puntuación mínima es 17 centavos y la máxima es 28 centavos. Nuestra unidad de redondeo es un número entero.

Rango = (puntuación máxima – puntuación mínima) + valor de unidad de redondeo

$$= (28 - 17) + 1 = 12¢$$

4. Calcula la media.

$$\bar{X} = \frac{\Sigma X}{n} = \frac{217}{10} = 21.7¢$$

5. A partir de la media, resta cada una de las puntuaciones en bruto (X) para obtener su puntuación de desviación. Suma las puntuaciones de desviación para asegurarte de que totalicen cero (dentro del error de redondeo).
6. Resta cada una de las puntuaciones de desviación y suma los cuadrados.
7. Calcula la desviación estándar:

$$s_X = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{116.10}{9}} = 3.59¢$$

La desviación estándar como parte integral de la estadística inferencial

Las características de la media y de la desviación estándar las hacen muy útiles para alcanzar un sentido de proporción respecto de las variables individuales que se estudian. La desviación estándar y las puntuaciones de desviación, a partir de las cuales se calculan, también son esenciales para examinar las relaciones entre dos variables. El foco de la estadística inferencial consiste en desarrollar una comprensión de por qué las puntuaciones individuales de una variable dependiente se desvían de su media.

Supongamos, por ejemplo, que estamos estudiando el abuso en el consumo de alcohol. Para una muestra de bebedores adultos, encontramos que la media del consumo de bebidas alcohólicas es de 4.3 galones por año. Gary consumió 7.3 galones el último año, 3 galones arriba de la media. Sam consumió sólo 1 galón, 3.3 galones abajo de la media. ¿Qué sucede con estas desviaciones alta y baja? Quizá podríamos generar hipótesis acerca de algunas variables predictoras (independientes) que creamos que estén relacionadas con esta variable dependiente. Por ejemplo, la hipótesis del consumo a la hora de la comida podría explicar, en parte, la puntuación de desviación positiva de Gary, los bebedores de familias que consumen vino con sus alimentos tienen un consumo de alcohol medio más alto. Existe también la hipótesis del bebedor social, la cual podría explicar, en parte, la puntuación de desviación negativa de Sam, los bebedores que sólo consumen alcohol en reuniones sociales tienen un consumo de alcohol medio más bajo.

Para una muestra completa, nuestro interés radica en explicar la variación, la suma de puntuaciones de desviación al cuadrado. Las puntuaciones de desviación, la variación y la desviación estándar simplemente son medidas de diferencias en las puntuaciones para una variable entre los sujetos de una población. ¿Es más alta la cantidad media de consumo de alcohol anual para las personas de ciertas regiones, entre diferentes edades o grupos religiosos o entre sexos? Las respuestas a tales preguntas dependen de las propiedades matemáticas de la media, la desviación estándar y la curva normal.

¿Por qué se llama desviación "estándar"?

La desviación estándar recibe su nombre por el hecho de que proporciona una *unidad de medida común* (un estándar) para comparar variables con *unidades observadas de medida* muy diferentes. Por ejemplo, imagine que Mary Smith y Jason Jones solicitan una beca con base en su desempeño en los exámenes de admisión a la universidad. Mary contestó la prueba académica de la universidad (ACT) y obtuvo 26 puntos ACT. Jason hizo lo propio con la prueba de admisión Stanford (SAT) y obtuvo 900 puntos SAT. Estos dos resultados de las pruebas tienen unidades de medida muy diferentes: los puntos de la prueba ACT van de cero a 36; y los de la prueba SAT, de 200 a 1 600. Las puntuaciones en bruto para las dos pruebas no pueden compararse directamente, pero con el uso de las medidas y las desviaciones estándar para ambas pruebas podemos crear una manera para compararlas. Con los siguientes estadísticos, encontramos que, en comparación con otros aspirantes que contestan las pruebas, Mary obtuvo la puntuación más alta:

X = puntuación de la prueba ACT $\bar{X} = 22$ puntos ACT $s_X = 2$ puntos ACT

Y = puntuación de la prueba SAT $\bar{Y} = 1\,000$ puntos SAT $s_Y = 100$ puntos SAT

La puntuación de ACT de 26 que obtuvo Mary tiene una desviación estándar de 2 arriba de la media de aquellos que toman la prueba ACT, es decir, su puntuación está 4 puntos ACT, esto es, 2 por 2 desviaciones estándar sobre el promedio de 22. La puntuación de Jason es de *1 desviación estándar abajo de la media* de aquellos que contrastan la prueba SAT, es decir, su puntuación está 100 puntos SAT, 1 desviación estándar abajo del promedio de 1 000. Sin lugar a dudas podemos otorgarle la beca a Mary. Utilizando las desviaciones estándar como unidades de medida en lugar de "puntos de prueba ACT" y de "puntos de prueba SAT", tenemos una norma común o estándar para ambas variables, de ahí el nombre de desviación *estándar*. ¿Quién te dijo que no podías comparar peras con manzanas?

Puntuaciones estandarizadas (puntuaciones Z)

El ejemplo anterior ilustra el hecho de que la puntuación de un sujeto de la investigación en cualquier variable de intervalo/razón puede expresarse de diversas maneras. Primero, lo *expresamos en sus unidades de medida observadas, originales,* como una **puntuación en bruto**. Por ejemplo, la puntuación en bruto X de Mary es 26 puntos ACT. Segundo, lo *expresamos como una desviación de la media*, es decir, la puntuación de desviación ($X - \bar{X}$); la puntuación de desviación de Mary es +4 y significa que ella obtuvo 4 puntos ACT arriba de la media de aquellos que tomaron el ACT. Tercero, expresamos su puntuación como *un número de desviaciones estándar de la media* de la puntuación ACT. Llamamos a esto su **puntuación estandarizada** (o **puntuación Z**), que para la variable X se calcula como sigue:

Cálculo de puntuaciones estandarizadas (puntuaciones Z)

$$Z_X = \frac{X - \bar{X}}{s_X}$$

donde

Z_X = puntuación estandarizada para un valor de X
= número de desviaciones estándar que una puntuación en bruto (puntuación X) se desvía de la media

X = una variable de intervalo/razón

$\bar{X}$ = la media de X

s_X = la desviación estándar de X

Si hacemos que la puntuación X = ACT con $\bar{X} = 22$ puntos ACT y $s_X = 2$ puntos ACT, la puntuación Z de Mary es

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{26 - 22}{2} = \frac{4}{2} = 2.00 \text{ SD}$$

donde SD significa "desviaciones estándar". Una puntuación Z es la distancia de una puntuación X hacia la media (es decir, su puntuación de desviación) dividida entre la desviación estándar de las distancias.

Una clave para tener claras estas tres maneras de expresar la puntuación consiste en enfocarse en las unidades de medida. Las puntuaciones en bruto y las puntuaciones de desviación para una variable se presentan en la unidad de medida original observada, que, por supuesto, es definida por una variable. Por ejemplo, la unidad de medida observada para edad es años; para peso, libras o kilogramos; para altura, pulgadas o centímetros; y así sucesivamente. Pero cualquiera que sea la unidad de medida de una variable, sus puntuaciones Z se miden en SD. La tabla 5-4 resume estas distinciones.

Aquí aparecen algunos ejemplos de una muestra aleatoria de mujeres estudiantes en la universidad local:

1. Donde X = peso, $\bar{X}$ = 120 libras, s_X = 10 libras:

Caso	X(peso)	$X - \bar{X}$(puntuación de desviación)	Z_X (puntuación estandarizada)
Cheryl Jones	110 libras	−10 libras	−1 SD
Jennifer Smith	125 libras	5 libras	.5 SD
Terri Barnett	107 libras	−13 libras	−1.3 SD

2. Donde Y = estatura, $\bar{Y}$ = 65 pulgadas, s_Y = 3 pulgadas:

Caso	Y(estatura)	$Y - \bar{Y}$(puntuación de desviación)	Z_Y (puntuación estandarizada)
Cheryl Jones	64 pulgadas	−1 pulgada	−.33 SD
Jennifer Smith	65 pulgadas	0 pulgada	0 SD
Terri Barnett	68 pulgadas	3 pulgada	1 SD

TABLA 5-4 | Diferentes formas en las que pueden presentarse puntuaciones de una variable

Forma de puntuación para una variable y su símbolo	Unidades de medida de la variable	Ejemplo: X = estatura
Puntuación en bruto (puntuación X): X	Unidad de medida de la variable	Pulgadas
Puntuación de desviación = $X - \overline{X}$	Unidad de medida de la variable	Pulgadas
Puntuación estandarizada (Z_X) o "puntuación Z": $Z_X = \dfrac{X - \overline{X}}{s_X}$	Desviaciones estándar de la variable (SD)	SD

Recuerda que las puntuaciones de desviación y las puntuaciones Z son medidas de la distancia desde la puntuación en bruto de la variable hasta su media. La puntuación de desviación se obtiene restando la media de la puntuación en bruto (es decir, $X - \overline{X}$). Al dividir esta puntuación de desviación entre la desviación estándar, cortamos esta puntuación de desviación en las partes y múltiplos de las desviaciones estándar desde la media. Recuerda que después de calcular la media, calcular las puntuaciones de desviación es lo siguiente que hacemos cuando calculamos la desviación estándar. La esencia de la desviación estándar está en ver una puntuación en bruto individual como una desviación desde la media.

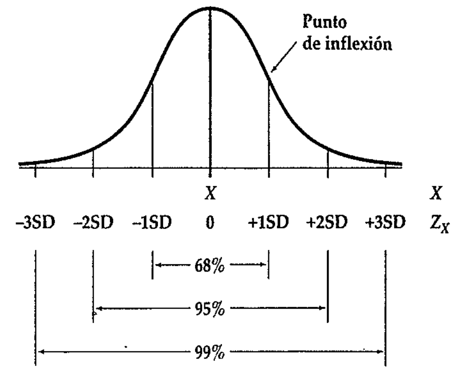
Para obtener un buen sentido de proporción sobre las fórmulas para las puntuaciones de desviación y las puntuaciones Z, examinemos las relaciones entre los tamaños de las puntuaciones en bruto, las puntuaciones de desviación y las puntuaciones Z. Primero, cuanto más lejana de la media esté una puntuación X mayores serán su puntuación de desviación y su puntuación Z. Primero, cuanto más lejana de la media esté una puntuación X mayores serán su puntuación de desviación y su puntuación Z. Es más, el signo de cualquier puntuación de desviación y puntuación Z indica la **dirección de una puntuación**: *ya sea que la observación caiga arriba de la media (la dirección positiva) o debajo de la media (la dirección negativa).* El signo "–" (signo menos) indica que una puntuación en bruto está debajo de la media; el signo "+" (signo más), que está implícito, no escrito, indica que está encima de la media. En los ejemplos anteriores Cheryl y Terri están abajo del promedio en peso, y Terri está arriba del promedio en estatura. De hecho, a partir de estas puntuaciones Z podemos decir que Terri es una persona alta, delgada más de 1 SD abajo en peso, pero 1 SD arriba en estatura. Jennifer tiene estatura media; así, su puntuación de desviación y su puntuación Z para Y son cero: ella no se desvía de la estatura media.

Puesto que usaremos puntuaciones Z o medidas similares de desviación en cada capítulo en el resto del texto, es prudente practicar cómo calcular puntuaciones de desviación y puntuaciones Z, así como estudiar las direcciones (signos) de esas puntuaciones. Se recomienda una doble verificación. Si una puntuación en bruto queda debajo de la media, su desviación y sus puntuaciones Z son negativas. También debes tener presente que las puntuaciones Z son simplemente otra manera de expresar puntuaciones en bruto. Cada puntuación en bruto tiene una puntuación Z correspondiente, y viceversa.

La desviación estándar y la distribución normal

Además de proporcionar un estándar de comparación entre variables y muestras diferentes, bajo condiciones apropiadas la media y la desviación estándar ofrecen gran riqueza de información. Éste es el caso cuando una variable tiene una distribución de puntuaciones que es normal, formada como la curva de distribución normal. Como lo definimos en el capítulo 4, una distribución normal es simétrica, con su media, mediana y moda iguales entre sí y

FIGURA 5-3
Relación entre la desviación estándar y la curva normal



localizadas en el centro de la curva. Sin embargo, la simetría o equilibrio en la curva no es toda la imagen. La curva normal también tiene una forma de campana inconfundible, que no es muy plana ni demasiado puntiaguda. Muchas variables se distribuyen normalmente (por ejemplo, la estatura, peso e inteligencia). Cualquiera que sea la variable que se examine, si está normalmente distribuida tendrá las propiedades de una curva normal.

Lo que hace que una desviación estándar sea una herramienta estadística tan valiosa es que es una parte matemática de la curva normal. Cuando se sigue la curva desde su centro (es decir, su pico) en cualquier dirección, la curva cambia de forma para aproximarse al eje X. Desde el pico, el punto en el que la curva empieza a desplazarse hacia fuera es 1 desviación estándar desde la media. Este punto recibe el nombre de **punto de inflexión** y se destaca en la figura 5-3. Esto indica que la media y la desviación estándar son aspectos matemáticos de un fenómeno natural: la tendencia a que una distribución normal en forma de campana se presente para numerosos eventos naturales.

Comprender el fenómeno de normalidad es un aspecto importante de la imaginación estadística. Muchos fenómenos que ocurren naturalmente tienen distribuciones de frecuencias en forma de campana como la curva normal. La curva normal ilustra el hecho de que cuando nos desviamos más allá de la media esperamos encontrar cada vez menos casos. Para muchas variables, existe un promedio alrededor del cual cae la mayoría de las puntuaciones, y cuando nos alejamos de este promedio, las frecuencias del caso disminuyen. Por ejemplo, la estatura física se distribuye normalmente; la mayoría de las personas están cerca del promedio, con unas cuantas personas muy altas o muy bajas.

Uno de los rasgos más sobresalientes del fenómeno de normalidad, que ocurre naturalmente, es que ofrece predicciones precisas sobre cuántas puntuaciones de una población caen dentro de cualquier rango de puntuaciones. Como se ilustró en la figura 5-3, *para cualquier variable normalmente distribuida:*

1. Cincuenta por ciento de las puntuaciones caen encima de la media; 50 por ciento, debajo. Esto se debe al hecho de que la mediana es igual a la media.
2. Prácticamente todas las puntuaciones caen dentro de 3 desviaciones estándar a partir de la media en ambas direcciones. Ésta es una distancia de 3 puntuaciones Z abajo de 3 puntuaciones Z arriba de la media, una amplitud total de 6 desviaciones estándar. La cantidad precisa es 99.7 por ciento. El restante 0.3 por ciento de casos (es decir, 3 casos de cada 1 000) caen fuera de 3 desviaciones estándar y, teóricamente, la curva se extien-

de hacia el infinito en ambas direcciones. (Prácticamente hablando, las puntuaciones para algunas variables, como el peso corporal, tienen límites finitos.)

3. Cerca del 95 por ciento de las puntuaciones de una variable normalmente distribuida caen dentro de una distancia de 2 desviaciones estándar en ambas direcciones de la media. Esto es más menos 2 puntuaciones Z de la media.
4. Alrededor de 68 por ciento de las puntuaciones de una variable normalmente distribuida caen dentro de una distancia de 1 desviación estándar (más menos 1 puntuación Z) en ambas direcciones de la media.

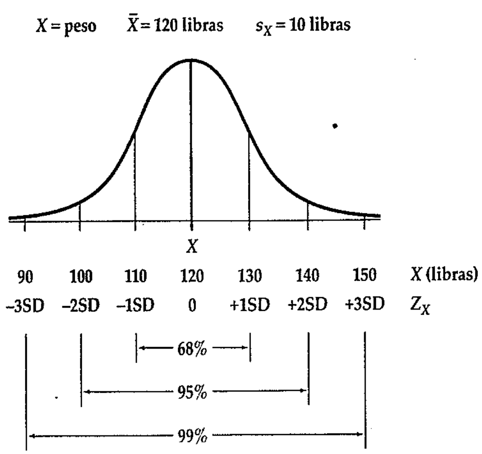
Recuerda que la distribución normal tiene características muy predecibles. *Si una variable se distribuye en esta peculiar forma de campana*, podemos utilizar los estadísticos de la muestra y lo que sabemos respecto de la curva normal para estimar cuántas puntuaciones *en una población* caen dentro de cierto rango.

Para ilustrar la utilidad de la curva normal, sigamos nuestro ejemplo: una muestra de mujeres estudiantes de la Universidad Crosstown, donde X = peso, el peso medio es de 120 libras y s_X = 10 libras. Primero, necesitamos asegurarnos de que la distribución de las puntuaciones es, de hecho, normal, es decir que tenga forma de campana. Esto podría hacerse elaborando un histograma de las puntuaciones de una muestra (no se ilustra). Si la forma de este gráfico se aproxima a la de una campana, suponemos que esta variable no sólo está normalmente distribuida en la muestra sino también en la población. Nos referimos a este hecho como "suponiendo normalidad". (La forma de un histograma de la muestra puede ser ligeramente fuera de lo normal como resultado del error muestral.) Como se grafica en la figura 5-4, suponiendo normalidad, podemos hacer las siguientes estimaciones de los pesos de la población de mujeres estudiantes de la Universidad Crosstown.

1. La mitad de estas estudiantes pesa más de 120 libras.
2. Cerca del 68 por ciento de las mujeres estudiantes de la Universidad Crosstown pesan entre 110 y 130 libras.

FIGURA 5-4

Uso de la curva normal para estimar la distribución de peso (X) en mujeres estudiantes de la Universidad Crosstown



3. Alrededor del 95 por ciento de las mujeres estudiantes de la universidad local pesan entre 100 y 140 libras.
4. Muy pocas pesan menos de 90 libras o más de 150 libras.

Recuerda, una puntuación Z simplemente es otra forma de expresar una puntuación en bruto (es decir, la puntuación X para una observación individual). Si Susana pesa 110 libras, ella está 1 SD debajo del peso medio y tiene una puntuación Z de −1.00 SD.

Presentación tabular de resultados

En artículos de investigación, una tabla básica de estadística descriptiva es la que lista todas las variables y sus medias y desviaciones estándar. La tabla 5-5 presenta una tabla de estadística descriptiva de un estudio del bienestar psicológico de personas sin hogar en dos puntos en el tiempo.

TABLA 5-5 | Estadísticos descriptivos para síntomas psicológicos, satisfacción con la vida y autoestima.

	Tratamiento 1		Tratamiento 2	
Subescalas	**M**	**SD**	**M**	**SD**
Síntomas psicológicos				
Enojo	4.17	0.80	4.14	0.85
Ansiedad	3.97	0.79	3.97	0.80
Depresión	3.60	0.76	3.68	0.77
Manía	3.59	0.87	3.68	0.90
Psicosis	4.51	0.72	4.52	0.72
Satisfacción con la vida				
Vestido	4.33	1.59	4.49	1.60
Alimento	4.79	1.53	4.98	1.42
Salud	4.81	1.38	4.77	1.41
Vivienda	4.37	1.49	4.51	1.54
Diversión	3.74	1.53	3.84	1.56
Dinero	2.98	1.57	3.19	1.67
Social	4.42	1.44	4.51	1.79
Autoestima				
Estima 1	3.21	0.85	3.24	0.84
Estima 2	3.36	0.87	3.28	0.85

Nota: n = 298. Puntuaciones más altas reflejan mayor bienestar subjetivo.
Fuente: Modificado de Marshall y otros, 1996: 49. Reimpreso con permiso de la American Sociological Association.

Insensatez y falacias estadísticas: ¿qué indica cuando la desviación estándar es más grande que la media?

Como vimos en el capítulo 4, la media es susceptible de distorsión por la presencia de puntuaciones extremas, valores extremos y distribuciones sesgadas. Debido a que se basa en desviaciones desde la media, la desviación estándar es susceptible al mismo problema. La distorsión está determinada por el hecho de que las puntuaciones de desviación están elevadas al cuadrado.

Un tipo común de distribución sesgada es un sesgo positivo (o a la derecha), en el que la mayoría de las personas tienen bajas puntuaciones, pero algunas tienen puntuaciones altas. Por ejemplo, "estancias en el hospital", o el número de veces que una muestra aleatoria de personas de más de 65 años hayan estado en el hospital el año pasado, es un sesgo a la derecha. Casi todas las personas registrarán cero en la estancia; algunas, uno; otras reportarán dos; y pocas personas muy enfermas anotarán estancias frecuentes. Este tipo de distribución se presenta en la tabla 5-6.

TABLA 5-6 | Distribución sesgada de estancias en el hospital, durante el último año, entre personas mayores de 65 años (datos ficticios)

Especificaciones		Cálculos	
(1) Caso	(2) X	(3) $X - \bar{X}$	(4) $(X - \bar{X})^2$
1	0	–2.41	5.81
2	0	–2.41	5.81
3	0	–2.41	5.81
4	0	–2.41	5.81
5	0	–2.41	5.81
6	0	–2.41	5.81
7	0	–2.41	5.81
8	0	–2.41	5.81
9	1	–1.41	1.99
10	1	–1.41	1.99
11	1	–1.41	1.99
12	2	–0.41	0.17
13	2	–0.41	0.17
14	5	2.59	6.71
15	9	6.59	43.43
16	10	7.59	57.61
17	10	7.59	57.61
$n = 17$	$\Sigma X = 41$ veces	$\Sigma(X - \bar{X}) = 0.03$*	$\Sigma(X - \bar{X})^2 = 218.15$ veces

*No totalizó cero debido al error de redondeo.

Incluso sin un histograma, los valores relativos de la media y de la desviación estándar para esta distribución proporcionan una señal de que la distribución está sesgada. Estos estadísticos se calculan como sigue:

X = estancias en el hospital = durante el último año, el número de veces que una persona ingresa en un hospital y pasa ahí por lo menos una noche

$$\bar{X} = 2.41 \text{ veces} \qquad s_X = 3.69 \text{ veces} \qquad n = 17 \text{ casos}$$

Observa que la desviación estándar es mayor que la media, lo cual sugiere que una o más puntuaciones extremas inflaron la media y la desviación estándar. Además, desde el momento en que se elevan al cuadrado los números de la desviación estándar, unas cuantas puntuaciones extremas pueden hacer "explotar" su valor. Toma nota, por ejemplo, de la enorme contribución a la suma de cuadrados que los tres casos más grandes hicieron con sus estancias de 9, 10 y 10 veces.

¿Por qué una desviación estándar más grande que la media indica un sesgo? Recuerda que si una distribución no está sesgada (es decir, tiene una forma normal de campana), su rango tendrá una amplitud de entre 4 y 6 desviaciones estándar. Cuando la curva se traza, la amplitud de 2 o 3 desviaciones estándar se ajustará en cada lado de la media. Si el límite inferior de las puntuaciones X de una variable es cero, por lo menos la distancia de 2 desviaciones estándar debería ajustarse entre una puntuación X de cero y la media. Cuando la desviación estándar es mayor que la media, como en el caso de las estancias en el hospital, ni una sola amplitud de la desviación estándar puede lograr este ajuste. Otra forma de explicarlo es que la desviación estándar debería ser alrededor de la mitad del tamaño de la media, o menos.

Dos reglas generales se aplican a los tamaños relativos de la media y de la desviación estándar:

1. Si la desviación estándar es más grande que la media, esto probablemente indica un sesgo, es decir, la presencia de valores extremos u otra peculiaridad en la forma de la distribución, por ejemplo una distribución bimodal.
2. Si la desviación estándar no es de la mitad del tamaño de la media o menos, debe tenerse cuidado al examinar la distribución para analizar la posible existencia de sesgos o valores extremos.

Como veremos en capítulos posteriores, cuando una variable sesgada está correlacionada con otras variables, los resultados pueden ser erróneos (capítulo 14). En tales casos, deben realizarse ajustes a los estadísticos para evitar tales errores.

RESUMEN

1. La dispersión se refiere a la forma en que las puntuaciones de una variable de intervalo/razón se dispersan, desde la menor hasta la mayor, y a la forma de la distribución entre ellas. Los estadísticos de dispersión miden esta diseminación.
2. Los estadísticos de dispersión que más se utilizan son el *rango* y la *desviación estándar*.
3. El rango es una expresión de la forma en que las puntuaciones de una variable de intervalo/razón se distribuyen de la menor a la mayor. Es la distancia entre las puntuaciones mínima y máxima de una muestra.

4. El rango tiene limitaciones. Es afectado en gran medida por valores extremos. Además tiene un estrecho alcance de información. Indica el ancho de una distribución de puntuaciones, pero no nos dice nada acerca de cómo es que las puntuaciones se dispersan entre las puntuaciones máxima y mínima.
5. La desviación estándar describe el modo en que las puntuaciones de una variable de intervalo/razón se dispersan a lo largo de la distribución, en relación con la puntuación media. La desviación estándar se calcula al determinar qué tan alejada está cada puntuación respecto a la media, es decir, cuánto "se desvía" de la media. Entonces, la desviación estándar está basada en puntuaciones de desviación.
6. La desviación estándar tiene limitaciones. Los valores extremos la inflan en gran medida. Puede ser errónea si la distribución de puntuaciones es sesgada.
7. La desviación estándar indica una unidad estándar de comparación, es decir, una unidad común de medida para comparar variables con unidades de medida muy diferentes. Las puntuaciones estandarizadas (puntuaciones Z) expresan una puntuación en bruto como un número de desviaciones estándar (SD) a partir de la puntuación media. Dos variables con diferentes unidades de medida pueden compararse si ambas están estandarizadas al calcular las puntuaciones Z.
8. La dirección de una puntuación Z está determinada por su signo. Una puntuación Z positiva se presenta cuando una puntuación en bruto es mayor que la media; una puntuación Z negativa ocurre cuando una puntuación en bruto es menor que la media.
9. Hay tres formas de expresar el valor de cualquier puntuación de una variable de intervalo/razón:
   a) Como una *puntuación en bruto*, el valor observado de X en su unidad de medida original (por ejemplo, pulgadas o libras).
   b) Como una *puntuación de desviación*, la diferencia entre la media y una puntuación en bruto. Las puntuaciones de desviación también se expresan en la unidad de medida original de la variable.
   c) Como *puntuación* Z (es decir, la *puntuación estandarizada*), la diferencia entre la media y la puntuación en bruto, pero expresada como un número de desviaciones estándar (SD).
10. La desviación estándar es una parte matemática de la curva normal. Es la distancia en el eje X de la media a la puntuación directamente bajo el punto de inflexión de la curva.
11. Si una variable está normalmente distribuida, podemos usar estadísticos muestrales y lo que sabemos acerca de la curva normal para estimar cuántas puntuaciones de una población caen dentro de cierto rango. *a*) 50% de las puntuaciones caen arriba de la media y 50% caen abajo. *b*) Prácticamente todas las puntuaciones (99.7%) caen a no más de tres desviaciones estándar de la media en ambas direcciones. *c*) Alrededor del 95% de las puntuaciones caen a no más de dos desviaciones estándar de la media en ambas direcciones. *d*) Alrededor del 68% de las puntuaciones de una variable normalmente distribuida caen a no más de 1 desviación estándar de la media en ambas direcciones.
12. Si la desviación estándar es mayor que la media, la distribución de puntuaciones *no puede* tener forma normal. Es probable que un histograma de la variable deje ver un sesgo o una distribución de puntuaciones de forma extraña.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB THE STATITICAL IMAGINATION

Las extensiones del capítulo 3 del material de texto disponibles en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2, incluyen la forma en que una estimación de la desviación estándar basada en el rango puede usarse para detectar si una distribución de puntuaciones está sesgada.

FÓRMULAS PARA EL CAPÍTULO 5

Organiza una hoja de cálculo con casos en orden:

Especificaciones		Cálculos	
(1) Caso	(2) X	(3) $X-\bar{X}$	(4) $(X-\bar{X})^2$
•	•	...	...
•	•	...	...
•	•	...	...
$n = \dots$	$\Sigma X = \dots$	$\Sigma(X-\bar{X}) = 0$	$\Sigma(X-\bar{X})^2 = \dots$

Cálculo del rango:

1. Ordena las puntuaciones de la distribución de menor a mayor.
2. Identifica las puntuaciones mínima y máxima.
3. Identifica el valor de la unidad de redondeo (véase apéndice A).
4. Calcula el rango:

Rango = (puntuación máxima – puntuación mínima) + valor de unidad de redondeo

Cálculo de la desviación estándar:

1. Empieza por calcular la media de X y completar una hoja de cálculo semejante a la de la tabla 5-2.
2. Calcula la desviación estándar:

   Trabajando con una hoja de cálculo

$$s_X = \sqrt{\frac{\Sigma (X-\bar{X})^2}{n-1}}$$

Calculando puntuaciones estandarizadas (puntuaciones Z):

$$Z_X = \frac{X-\bar{X}}{s_X}$$

PREGUNTAS PARA EL CAPÍTULO 5

1. Los estadísticos de dispersión se calculan ¿sólo en variables de qué niveles de medición?
2. Tanto el rango como la desviación estándar son medidas de la dispersión de las puntuaciones en una distribución. Explica las diferencias en perspectiva entre estos dos estadísticos.
3. ¿Qué efecto tiene una puntuación extrema o valor aislado sobre el cálculo del rango?
4. La desviación estándar se "deriva" de la media. ¿Qué significa esto?
5. Al calcular el rango, el valor de la unidad de redondeo de la variable se suma a la diferencia entre las puntuaciones máxima y mínima. ¿Por qué se suma el valor de la unidad de redondeo?
6. Al calcular la desviación estándar, ¿por qué es necesario elevar al cuadrado las puntuaciones de desviación?
7. Al calcular la desviación estándar para datos de una muestra, ¿por qué debemos dividir entre $n - 1$?
8. Al calcular la desviación estándar, ¿por qué se requiere sacar la raíz cuadrada?
9. ¿Cuál es la relación matemática entre la varianza y la desviación estándar?
10. Menciona otro nombre para la variación.
11. ¿Cuál es el significado de la palabra *estándar* en el término *desviación estándar*?
12. Una expresión de qué tan lejos está una puntuación en bruto de la media de una distribución, en las unidades de medida originales de la variable X, se llama una puntuación __________.
13. Una expresión de qué tan lejos está una puntuación en bruto de la media de una distribución, en unidades de medida de desviaciones estándar (SD), se llama una puntuación _______.
14. ¿Cuáles son las propiedades de una distribución normal?
15. En una distribución normal, ¿qué porcentaje de puntuaciones caen aproximadamente dentro de 1 desviación estándar de la media en ambas direcciones?, ¿y dentro de 2 desviaciones estándar de la media en ambas direcciones?, ¿y dentro de 3 desviaciones estándar de la media en ambas direcciones?
16. En una distribución normal ¿qué porcentaje exacto de puntuaciones caen sobre la media? ¿Qué estadístico de tendencia central, además de la media, explica este fenómeno?
17. En una distribución normal la curva alcanza su máximo en el valor de la media. ¿Qué estadístico de tendencia central, además de la media, justifica este fenómeno?
18. Si una puntuación en bruto cae debajo de la media en una distribución, el signo de la puntuación Z, ¿será positivo o negativo? Ilustra tu respuesta utilizando la fórmula para calcular una puntuación Z.
19. En cualquier distribución de puntuaciones de intervalo/razón hay una puntuación en la cual las desviaciones de ésta suman cero. ¿Qué estadístico de tendencia central se localiza en ese punto?
20. Para su grupo de edad Charles está 1 desviación estándar debajo de la estatura media, pero 1.5 desviaciones estándar arriba del peso medio. Describe su complexión corporal general.

21. Daniel está 3 desviaciones estándar arriba de la media en términos de su coeficiente intelectual (CI). Describe su intelecto general.
22. Explica por qué es probable que una distribución no sea normal cuando la desviación estándar es más grande que la media.

EJERCICIOS PARA EL CAPÍTULO 5

Conjunto de ejercicios 5A

5A-1. Utiliza la fórmula de la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta cálculos en variables de intervalo/razón para diferentes muestras de tamaño n.

Suma de cuadrados	n	Varianza	Desviación estándar
11 828.52	88	135.96	____
3 120.00	21	____	____
893.49	____	30.81	____
____	347	124.65	11.16

5A-2. Hughes y Waite (2002) estudiaron las condiciones de vida y salud en los años tardíos de la edad media del ciclo vital. Supongamos que la siguiente es una serie de edades del estudio que realizaron estos investigadores: 74, 81, 83, 77, 76, 79, 79.

a) Organiza una hoja de cálculo con casos ordenados con X = edad.
b) Calcula la edad media, mediana y modal.
c) Calcula el rango de las edades.
d) Calcula la desviación estándar de las edades.

5A-3. Hoff (2003) examinó las vidas laborales de médicos empleados por organizaciones de conservación de la salud (OCS). Supongamos que los datos siguientes describen la carga diaria de pacientes (es decir, el número de pacientes revisados por día) de siete médicos de las OCS: 8, 7, 11, 4, 5, 13, 7.

a) Organiza una hoja de cálculo con casos ordenados con Y = número de pacientes revisados por día.
b) Calcula la media, mediana y modal de las cargas de pacientes para este grupo de siete médicos.
c) Calcula el rango de pacientes revisados por día.
d) Calcula la desviación estándar.

5A-4. Takao y otros (2003) examinaron la relación entre grupo ocupacional y actividad física entre empleados japoneses. Supongamos que los datos siguientes representan una muestra de puntuaciones en una escala que mide las posiciones de individuos dentro de la jerarquía ocupacional japonesa: 27, 26, 28, 30, 31, 29, 27, 31, 29. Los datos X = puntuación de escala de grupo ocupacional.

a) Organiza los datos usando un formato de hoja de cálculo con las puntuaciones de X ordenadas.

b) Calcula la media y desviación estándar.

5A-5. Es semana de regreso a casa y están ocurriendo locuras en el *campus*. Una de estas locuras es una carrera a pie entre el club femenino de estudiantes. Una muestra aleatoria de hermanas que corren, prendas de garantía y alumnas produce las edades siguientes: 19, 18, 20, 19, 29, 18, 20, 18, 22, 21. Los datos X = edad.

a) Organiza los datos usando un formato de hoja de cálculo con las puntuaciones de X ordenadas.

b) Calcula la media y desviación estándar.

c) ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.

d) Comenta sobre las diferencias entre los estadísticos original y ajustado.

5A-6. Ellickson y cols. (2003) examinaron la conducta de fumar en adolescentes y la subsiguiente conducta después de fumar. Supongamos que los siguientes datos son de una muestra de fumadores de 16 a 20 años de edad.

Y = número de cigarrillos fumados por día.

$\bar{Y}$ = 15 cigarrillos $\qquad s_Y$ = 5 cigarrillos

a) Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.

b) ¿Quién destaca como fumador?

Caso	Y (cigarrillos por día)	$Y - \bar{Y}$ (desviación estándar)	Z_Y (puntuación estandarizada)
Bob Smith	17		
Spencer Byrd	30		
Sonya Turnham	4		
Chuck Martin	20		

5A-7. Ferraro y Yu (1995) estudiaron la relación entre el peso corporal y la autoestima en cuanto a salud. Supongamos que te dan los siguientes breves estadísticos acerca del peso obtenido en este estudio.

X = peso $\qquad \bar{X}$ = 169 libras $\qquad s_X$ = 18 libras

a) Traza la curva normal y ubica estos pesos en ésta.

b) La tabla siguiente incluye datos para unas pocas de las observaciones. Completa la columna central al estimar cada una de las Z_X visuales (es decir, con sólo observar X en la curva).

c) Para cada puntuación X, calcula la puntuación Z exacta e insértala en la columna de la derecha. (Demuestra la fórmula y cálculo para X = 128 libras.)

X (libras)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
169		
128		
192		
177		
151		
109		

Conjunto de problemas 5B

5B-1. Usa la fórmula para la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta cálculos sobre variables de intervalo/razón de diferentes muestras de tamaño *n*.

Suma de cuadrados	*n*	Varianza	Desviación estándar
38.76	7	____	____
347 295.92	1 041	____	18.27
____	91	40.89	____
5 865.04	____	17.56	____

5B-2. Goesling (2001) examinó el fenómeno de la desigualdad de ingresos en todo el mundo, tanto dentro como entre las naciones. Supongamos que lo siguiente es una muestra de ingresos mensuales de residentes de Estados Unidos: $2 347; $2 434; $1 636; $1 963; $2 358; $1 968; $2 683.

a) Organiza una hoja de cálculo con los casos en orden con X = ingreso mensual.

b) Calcula el ingreso mensual medio, mediano y modal.

c) Calcula el rango.

d) Calcula la desviación estándar.

5B-3. Wiesner (2003) examinó relaciones recíprocas entre síntomas depresivos y conducta delictiva entre adolescentes hombres y mujeres. Supongamos que las siguientes son las edades de adolescentes comprendidos en este estudio: 10, 8, 9, 11, 12, 9, 13.

a) Organiza una hoja de cálculo con los casos en orden con Y = edad.

b) Calcula las edades media, mediana y modal para los adolescentes seleccionados.

c) Calcula el rango de edades.

d) Calcula la desviación estándar.

5B-4. Groome y Soureti (2004) estudiaron la relación entre el desorden de estrés postraumático y síntomas de ansiedad en niños, después del terremoto de 1999 cerca de Atenas, Grecia. Entre los datos que presentan en su estudio analizan la magnitud de diversos terremotos en la escala Richter ocurridos en el mar Mediterráneo a fines del siglo veinte: 5.8, 2.4, 2.2, 6.0, 3.1, 2.4, 2.2, 5.8, 2.4. Los datos X = puntos de magnitud en la escala de Richter.

a) Organiza los datos usando un formato de hoja de cálculo con puntuaciones X ordenadas.

b) Calcula la media y desviación estándar.

5B-5. Betts y Morell (1999) analizaron los efectos de los antecedentes personales (es decir, de familia, de secundaria, recursos, grupo paritario, etc.) sobre el promedio de puntos de calificación de pasantes (GPA). Supongamos que las siguientes fueron puntuaciones GPA para una muestra de estudiantes universitarios pasantes: 3.6, 3.8, 3.6, 3.9, 2.6, 3.8, 3.8, 3.9. Los datos X = GPA.

a) Organiza los datos usando un formato de hoja de cálculo con puntuaciones ordenadas.

b) Calcula la media y desviación estándar.

c) ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.

d) Comenta sobre las diferencias entre los estadísticos original y ajustado.

5B-6. Green y cols. (2001) estudiaron el fenómeno de delitos con violencia y discutieron las dificultades prácticas asociadas con la recolección de datos sobre este fenómeno. No obstante, supongamos que tú has podido asegurar datos confiables acerca de delitos con violencia en Estados Unidos. Los datos Y = porcentaje de delitos con violencia = número de delitos con violencia reportados por 100 000 habitantes cubiertos por agencias periodísticas. Los siguientes son porcentajes para una muestra seleccionada de estados:

$\bar{Y}$ = delitos con violencia por 100 000 habitantes
s_Y = .32 delitos con violencia por 100 000 habitantes

a) Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.

b) ¿Qué resalta por tener un porcentaje relativamente alto de delitos con violencia?

Estado	Y (porcentaje de delitos con violencia)	$Y - \bar{Y}$ (puntuación de desviación)	Z_Y (puntuación estandarizada)
Florida	1.15		
Indiana	1.08		
Iowa	1.02		
Mississippi	.97		
Texas	1.75		

5B-7. Slater y cols. (2003) examinaron la relación entre el contenido violento de medios masivos de comunicación y la conducta agresiva entre adolescentes. Para replicar los resultados obtenidos por estos investigadores, supongamos que llevamos a cabo un análisis de conducta agresiva similar entre muchachos de 13 a 16 años de un centro de detención juvenil. La variable se operacionaliza como el número de actos agresivos, es decir, insultos verbales y amenazas, actos de violencia física y destrucción de propiedades cometidos en la semana previa. Los actos se tabulan al observar videocintas de salas y secciones de recreación, biblioteca, sanitarios y cafetería del centro de detención. Calculamos los estadísticos descriptivos sobre esta variable y obtenemos los siguientes resultados con X = número de actos agresivos:

$$\bar{X} = 16.8 \text{ actos} \qquad s_X = 4.4 \text{ actos}$$

a) Traza la curva normal y ubica estos actos agresivos en la curva.

b) La tabla siguiente incluye información para unas pocas observaciones. Completa la columna central al estimar cada Z_X visual (es decir, con sólo observar X en la curva).

c) Para cada puntuación X, calcula la puntuación Z exacta e insértala en la columna derecha. (Muestra la fórmula y cálculo para X = 9 actos.)

X (actos agresivos)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
9		
12		
19		
26		
3		
14		

Conjunto de problemas 5C

5C-1. Usa la fórmula para la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta los cálculos para variables de intervalo/razón de diferentes muestras de tamaño n.

Suma de cuadrados	n	Varianza	Desviación estándar
12654.27	97	131.82	_____
2876.54	18	_____	_____
975.46	_____	34.82	_____
_____	526	142.53	_____

5C-2. Para los sujetos de investigación en comunidades de bajos ingresos en Kenya, Molyneux y otros (2004) examinaron la comprensión de documentos de un acuerdo informado. Como parte del componente cuantitativo del estudio, supongamos que los siguientes son ingresos de residentes que viven dentro de la zona de investigación de los autores: $ 627, $435, $569, $615, $796, $715, $615.

a) Organiza los datos usando un formato de hoja de cálculo con casos ordenados con X = ingresos.

b) Calcula el ingreso medio, mediano y modal.

c) Calcula el rango de ingresos.

d) Calcula la desviación estándar de ingresos.

5C-3. Siebert (2004) estudió los determinantes de la depresión entre trabajadoras sociales en Carolina del Norte. Supongamos que los datos siguientes representan el número de contactos en la semana pasada para cada trabajadora social de la muestra: 10, 8, 13, 7, 6, 15, 6.

a) Organiza una hoja de cálculo con los casos en orden con Y = contactos en la semana pasada.

b) Calcula la media, mediana y moda del número de contactos para la muestra de trabajadoras sociales.

c) Calcula el rango.

d) Calcula la desviación estándar.

5C-4. Roose y cols. (2004) trataron de determinar la eficacia de medicamentos para el tratamiento de síntomas depresivos en pacientes de 75 años de edad o más. Supongamos que los datos siguientes son puntuaciones de la escala de depresión del Centro de Estudios Epidemiológicos (CESD), para un pequeño número de participantes de estudio: 38, 31, 42, 27, 19, 49, 31, 19, 38. Los datos X = puntos de la escala CESD.

a) Organiza los datos usando un formato de hoja de cálculo con puntuaciones ordenadas.

b) Calcula la media y desviación estándar.

5C-5. Ebrahim y cols. (2004) estimaron la asociación entre la posición socioeconómica y la incapacidad autorreportada en hombres de edad avanzada. Supongamos que las siguientes son las edades de varios hombres en este estudio: 74, 69, 76, 72, 72, 78, 87, 74, 69, 74. Los datos X = edad.

a) Organiza los datos usando un formato de hoja de cálculo con puntuaciones ordenadas.

b) Calcula la media y desviación estándar.

c) ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.

d) Comenta sobre las diferencias entre los estadísticos original y ajustado.

5C-6. Ramstedt (2004) examinó el consumo de alcohol y la mortalidad relacionada con el alcohol en Canadá. Supongamos que los datos siguientes son de una submuestra de sujetos adictos al alcohol, tomados de la investigación del autor. Los datos Y = número de copas de alcohol ingeridas por día.

$$\bar{Y} = 6 \text{ copas}$$
$$s_Y = 2 \text{ copas}$$

a) Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.

b) ¿Quién destaca como el más bebedor?

Sujeto	Y (copas por día)	$Y - \bar{Y}$ (puntuación de desviación)	Z_Y (puntuación estandarizada)
Jill Williams	4		
Thomas Wilke	8		
Jason Schmidt	12		
Jenny Pence	7		

5C-7. El índice de masa corporal (IMC) es una medida del nivel saludable de peso que toma en cuenta la estatura de personas. Se calcula en kilogramos de peso por el cuadrado de la estatura. Xiaoxing y Baker (2004) investigaron la relación entre el IMC, la actividad física y el riesgo de ambos en la disminución de la salud general y el funcionamiento físico. Supongamos que te dan los siguientes y breves estadísticos sobre el IMC de este estudio. Los datos X = puntuación IMC.

$$\bar{X} = 31 \text{ kg/m}^2$$
$$s_X = 9 \text{ kg/m}^2$$

a) Traza la curva normal y ubica estos índices de masa corporal en la curva.

b) La tabla siguiente incluye datos para algunas de las observaciones. Completa la columna central al estimar cada una de las Z_X visuales (es decir, con sólo observar X en la curva).

c) Para cada puntuación X, calcula la puntuación Z exacta e insértala en la columna de la derecha. (Muestra la fórmula y cálculo para $X = 11$.)

X (libras)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
31		
11		
42		
38		
22		
2		

Conjunto de problemas 5D

5D-1. Usa la fórmula de la desviación estándar para completar los espacios en blanco de la tabla siguiente. La tabla presenta los cálculos sobre variables de intervalo/razón de diferentes muestras de tamaño *n*.

Suma de cuadrados	n	Varianza	Desviación estándar
29.57	5	_____	_____
426113.21	1986	_____	14.65
_____	82	35.43	_____
8450.35	_____	22.12	_____

5D-2. Garroutte y cols. (2004) examinaron la relación entre la identidad étnica y la satisfacción entre pacientes indios norteamericanos crónicamente enfermos, en una clínica de la nación Cherokee. Supongamos que la siguiente es una serie de edades del estudio de estos investigadores: 56, 64, 62, 57, 64, 59, 58.

a) Organiza una hoja de cálculo con los casos en orden con X = edad.

b) Calcula la media, mediana y moda para la edad.

c) Calcula el rango de edades.

d) Calcula la desviación estándar de las edades.

5D-3. Henning y Feder (2004) compararon las diversas características de hombres y mujeres encarcelados por violencia doméstica. Supongamos que la muestra siguiente representa el número de incidentes informados de violencia doméstica entre los hombres y mujeres incluidos en este estudio: 5, 2, 3, 4, 3, 6, 1.

a) Organiza una hoja de cálculo con los casos en orden con Y = incidentes de violencia doméstica.

b) Calcula la media, mediana y moda de los incidentes.

c) Calcula el rango.

d) Calcula la desviación estándar.

5D-4. Tohill y cols. (2004) revisaron la evidencia epidemiológica que evalúa la relación entre consumo de frutas y verduras y el peso corporal. Supongamos que los siguientes son pesos corporales promedio de un pequeño número de participantes: 168, 181, 144, 159, 181, 204, 168, 144, 181. Donde X = peso corporal (libras).

a) Organiza los datos usando un formato de hoja de cálculo con las puntuaciones ordenadas.

b) Calcula la media y desviación estándar.

5D-5. Grove y Wasserman (2004) examinaron patrones de ciclos de vida del promedio de puntos de calificaciones (GPA) entre cinco grupos colegiados en una importante universidad privada. Supongamos que las siguientes son una muestra de las GPA colegiadas de varios sujetos de estudio en esta investigación: 3.8, 3.4, 2.4, 3.8, 3.7, 3.4, 3.4, 3.8. Donde X = GPA.

a) Organiza los datos usando un formato de hoja de cálculo con las puntuaciones ordenadas.

b) Calcula la media y desviación estándar.

c) ¿Hay algo peculiar en esta distribución? Ajústalo al recalcular los estadísticos.

d) Comenta sobre las diferencias entre los estadísticos original y ajustado.

5D-6. Varano y cols. (2004) exploraron la correlación entre el consumo de drogas y homicidios. Supongamos que los estadísticos siguientes representan los porcentajes de homicidios de las zonas de investigación incluidas en este análisis. Donde Y = porcentaje de homicidios = número de homicidios informados por 100 000 habitantes.

$$\bar{Y} = 6.59 \text{ homicidios por } 100\,000 \text{ habitantes}$$
$$s_Y = 1.74 \text{ homicidios por } 100\,000 \text{ habitantes}$$

a) Completa las columnas de la tabla siguiente. Asegúrate de especificar las unidades de medida.

b) ¿Qué zona destaca por tener un porcentaje relativamente alto de homicidios?

Estado	Y (porcentaje de homicidios)	$Y - \bar{Y}$ (puntuación de desviación)	Z_Y (puntuación estandarizada)
Área 1	4.97		
Área 2	8.99		
Área 3	5.99		
Área 4	6.95		
Área 5	6.29		

5D-7. Boardman (2004) evaluó la relación entre la estabilidad residencial y la salud física entre adultos de raza negra y blanca. Parte de la variación en los niveles de salud se debía a diferencias en los niveles de estrés entre vecindarios. Supongamos que te dan los siguientes estadísticos de resumen de una escala empleada para analizar niveles de estrés en vecindarios. Donde X = puntuación en la escala de estrés.

$$\bar{X} = 11.3 \text{ puntos de la escala de estrés}$$
$$s_X = 3.2 \text{ puntos de la escala de estrés}$$

a) Traza la curva normal para estas puntuaciones de escala de estrés y ubícalas en la curva.

b) La tabla siguiente incluye datos para algunas de las observaciones. Completa la columna central al estimar cada una de las Z_X visuales (es decir, con sólo observar X en la curva).

c) Para cada puntuación X, calcula la puntuación Z exacta e insértala en la columna de la derecha. (Demuestra la fórmula y cálculo para X = 7 puntos de escala de estrés.)

X (puntos de la escala de estrés)	Estimación visual de puntuación Z (SD)	Puntuación Z calculada (SD)
7		
10		
14		
19		
1		
16		

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 5

En el sitio web www.mhhe.com/ritchey2, en *The Statistical Imagination*, están disponibles ejercicios computarizados opcionales del capítulo. Además, el apéndice D de este texto, *Guide to SPSS for Windows*, contiene instrucciones básicas para calcular estadísticos de dispersión y puntuaciones estandarizadas. Tal como se destaca en este capítulo, la desviación estándar por lo general se reporta con la media. De este modo, estos "estadísticos descriptivos" se encuentran juntos en el software. La media, rango y desviación estadística se pueden calcular desde varios lugares del *SPSS* y son parte de estadísticos opcionales para numerosos procedimientos estadísticos de prueba, como los de los capítulos 9-12.

CAPÍTULO 6

Teoría de la probabilidad y la distribución normal de probabilidad

RESUMEN DEL CAPÍTULO

Introducción: el impulso humano para predecir el futuro

La capacidad mental humana se distingue de la de otras especies por una habilidad para predecir el futuro y concebir lo que ocurrirá "a la larga". Predecir eventos futuros significa comprenderlos, y el desarrollo de la cultura humana depende de la predicción. El campo de la estadística es acerca de realizar predicciones con medidas muy precisas. Los estadísticos adquieren posición y autoridad a partir de la predicción y el entendimiento exitoso. Con sus aplicaciones en las ciencias, negocios, industria, predicción meteorológica, medicina, salud, servicios públicos, gobierno, industria de juegos de azar, entretenimiento y deportes, el trabajo estadístico es un ejemplo de la predicción del futuro.

Los deseos humanos para anticipar eventos no son nuevos. Desde los primeros tiempos de la organización social humana, los "videntes" (como los chamanes sacerdotes-doctores)

adquirieron gran autoridad prediciendo ciclos climáticos y eventos importantes como la lluvia. Aplicar las matemáticas para la resolución de un problema práctico también tiene una historia larga. Las técnicas de medición y cálculo son tan antiguas como la cultura humana y ya estaban muy bien establecidas hace 4 000 años. Las pirámides del Egipto antiguo ejemplifican una organización matemática precisa del mundo físico. De hecho, muchos eruditos argumentan que la magnitud del conocimiento egipcio se subestima bastante (Gillings 1972, Neugebauer 1962, Struik 1948). Como Tompkins (1971: XIV-XV) anota:

> Quienquiera que haya construido la Gran Pirámide [...] conocía la circunferencia precisa del planeta y la duración del año con varios decimales-datos que no se redescubrieron hasta el siglo XVII.

Las matemáticas, la trigonometría y las "ciencias exactas" modernas como la física tienen sus orígenes en antiguas culturas mediterráneas, en las cuales se veneró el estudio de la naturaleza. Subyacente al desarrollo de la ciencia es el reconocimiento que la naturaleza física es altamente cíclica y, por tanto, predecible hasta el punto que obedece leyes científicas estrictas. Por ejemplo, un físico parado en el suelo puede estar seguro de que cualquier objeto más pesado que el aire caerá a la tierra. Sin embargo, el comportamiento humano no es tan predecible. Por tanto, los científicos sociales y del comportamiento a menudo deben matizar sus conclusiones, aclarando que sus predicciones se basan en grados de precisión limitados *pero calculables*. Por ejemplo, un científico social que estudia la relación, digamos, de las calificaciones en la preparatoria con las calificaciones en la universidad quizá sólo sea capaz de asegurar que existe una correlación entre ellas, declarando que hay una *posibilidad* calculable que un estudiante sobresaliente en la preparatoria obtendrá excelentes calificaciones en la universidad. Las leyes de la posibilidad son herramientas para determinar el grado de exactitud en la predicción de las ciencias sociales. Nos referimos al *análisis y a la comprensión de las ocurrencias por el azar* como **teoría de la probabilidad**.

> **Teoría de la probabilidad** Análisis y comprensión de las ocurrencias por el azar.

El descubrimiento de las leyes de la probabilidad empezó en tiempos antiguos y quizá se estimuló tanto por actividades de ocio, como las apuestas, como por ocupaciones laborales (David 1962: 4-10). Entre los artefactos de la primera dinastía de Egipto (3500 a.C.) había juegos de tablero, piezas para jugar y *astralagi* animal (huesos de las articulaciones), los precursores de los dados. En Egipto, los dados cúbicos fueron de uso común hacia el año 3000 a.C. Los juegos eran tan comunes en tiempos romanos que se prohibieron en ciertos días. En la literatura romana existen referencias a un libro de Claudio (10 a.C.-54 d.C.) titulado *Cómo ganar en los dados*. Los jugadores astutos tenían la imaginación estadística. Ellos podían pensar proporcionalmente, reconociendo que algunas "tiradas de huesos" ocurrían con una mayor proporción de las veces que otras. Aconsejando exitosamente a los miembros de las clases gobernantes sobre cómo aumentar sus ganancias en las apuestas, estos primeros estadísticos alcanzaron una posición alta. Analistas de la bolsa de valores e investigadores de encuestas exitosos, así como pronosticadores en carreras de caballos, son el equivalente moderno de aquellos consejeros estadísticos altamente respetados.

Podemos conjeturar que el interés humano en predecir los resultados de eventos futuros no se limitó a los juegos de azar. Hasta donde los seres humanos se interesan, las fuerzas de la naturaleza (sobre todo el clima) implican el azar, y la adaptación ambiental se forja con el destino y la buena o mala suerte. La evolución cultural es estimulada por la necesidad de la sociedad para anticipar lo que ocurrirá. Por ejemplo, en el periodo dinástico medio (hacia 2000 a.C.) los antiguos egipcios habían desarrollado sistemas complejos de irrigación y de canales para regular la inundación anual del río Nilo. Con datos manejados por una burocracia muy eficiente, supervisaron la profundidad del río con "nilómetros", colocados en puntos estratégicos a lo largo de la vasta longitud del Nilo de 4 145 millas. Al estudiar y anticipar los flujos, utilizaron ventajosamente los caudales para la irrigación de los cultivos (David 1962). La exactitud de sus predicciones generó una economía estable que reforzó el poder político de las dinastías gobernantes. Muchas culturas antiguas tenían una parte de empiristas: individuos que creían en los méritos de la observación y la medición. El empirismo y la popularidad de las apuestas, la religión y la adivinación atestiguan un interés humano innato en apostar en lo que ocurrirá y se preparan para ello. Todo, desde predecir la fortaleza de la tropa enemiga hasta decidirse si llevar un paraguas, invertir en acciones o proponer matrimonio, requiere medidas y estimaciones de la probabilidad de su éxito o fracaso. El análisis estadístico empleando la teoría de la probabilidad es la herramienta mediante la cual se realizan predicciones con un grado máximo de exactitud. Para los estadísticos, muchas de estas predicciones se basan en la curva normal. Las ideas detrás de la teoría moderna de las probabilidades ya estaban bien establecidas en 1800 [Laplace 1951 (1820)].

¿Qué es probabilidad?

Una **probabilidad (p)** es *una especificación de con qué frecuencia es probable que ocurra un evento de interés particular entre un gran número de ensayos (situaciones en las que el evento puede ocurrir)*. Llamamos **probabilidad de éxito** a la probabilidad de ocurrencia de este evento de interés. De igual manera, la probabilidad de que *no* ocurra el evento se llama **probabilidad de fracaso**.

Al observar los resultados o "consecuencias" de un gran número de ensayos, podemos identificar todas las consecuencias posibles. Para tomar un ejemplo muy simple, suponga que deseamos determinar la probabilidad de lanzar una moneda y obtener "cara". Para determinar los resultados posibles al lanzar una moneda, podemos lanzar una moneda muchas veces, digamos, 10 000 veces o ensayos. Veríamos de inmediato que los resultados posibles son "cara" (éxito) y "cruz" (fracaso). Para calcular la probabilidad de obtener cara (éxito), dividimos el número de veces que ocurre cara entre 10 000. Tendríamos que cara ocurre casi la mitad de las veces. Sin embargo, este ejemplo ilustra que para calcular una probabilidad con frecuencia simplemente "hacemos la cuenta". Es obvio que una moneda tiene dos resultados posibles. Estos resultados tienen una posibilidad de ocurrencia igual, por lo que la probabilidad de obtener cara es la mitad, o sea, 0.5000. De esta manera, otra forma de imaginar una probabilidad es el número de eventos posibles exitosos entre el número de resultados posibles. Cuando es posible identificar fácilmente todos los resultados posibles, como al lanzar monedas o dados, no es necesario hacer ensayos repetidos.

Para simbolizar una probabilidad, se utilizan corchetes para distinguir el evento de interés señalado; y una letra p minúscula para indicar "la probabilidad" de un cálculo específico. Observa que este símbolo es el mismo usado en capítulos anteriores para *proporción*. Esto se hace debido a que las probabilidades *son* proporciones, como analizaremos en breve.

> **Una probabilidad (p)** Especificación de la frecuencia probable de que ocurra un evento de interés particular entre un gran número de ensayos.

La fórmula general para presentar una probabilidad es como sigue:

Cálculo de una probabilidad

$$p\,[\text{de éxito}] = \frac{\#\text{ de éxitos}}{\#\text{ de ensayos}} = \frac{\#\text{ de resultados exitosos posibles}}{\#\text{ total de resultados posibles}}$$

donde p [de éxito] = probabilidad del "evento de interés".

Para la consistencia en la instrucción, en este texto las respuestas a los cálculos de probabilidad se redondearán y se presentarán con cuatro lugares decimales. Por supuesto, las probabilidades también pueden presentarse como porcentajes, multiplicando p por 100. Los siguientes son algunos ejemplos que revelan qué tan simple es la noción de probabilidad:

Ejemplo A: Al lanzar una sola moneda, ¿cuál es la probabilidad de obtener cara?

Con una moneda hay dos resultados posibles, y cara es uno de ellos. Por tanto,

$$p\,[\text{cara}] = \frac{\#\text{ de caras}}{\#\text{ de resultados posibles}} = \frac{1}{2} = .5000$$

Ejemplo B: Al sacar una sola carta al azar de un mazo estándar de juego de 52 cartas bien revueltas:

$$a)\ p\,[\text{rey}] = \frac{\#\text{ de reyes en el mazo}}{\#\text{ total de cartas en el mazo}} = \frac{4}{52} = .0769$$

$$b)\ p\,[7] = \frac{\#\text{ de 7 en el mazo}}{\#\text{ total de cartas en el mazo}} = \frac{4}{52} = .0769$$

$$c)\ p\,[\text{corazones}] = \frac{\#\text{ de corazones en el mazo}}{\#\text{ total de cartas en el mazo}} = \frac{13}{52} = .2500$$

Ejemplo C: Al sacar una sola canica de una caja bien mezclada de 300, donde 100 son color rojo y 200 son color verde:

$$a)\ p\,[\text{roja}] = \frac{\#\text{ de rojas en la caja}}{\#\text{ total de canicas en la caja}} = \frac{100}{300} = 0.3333$$

$$b)\ p\,[\text{verde}] = \frac{\#\text{ de verdes en la caja}}{\#\text{ total de canicas en la caja}} = \frac{200}{300} = 0.6667$$

Como estas ilustraciones revelan, calcular una probabilidad es simplemente una cuestión de pensamiento proporcional respecto de la ocurrencia a largo plazo de un evento de interés. Preguntar ¿cuál es la probabilidad?, es cuestionarse sobre cuántas, del total de veces, ocurre una categoría de eventos, es decir, cuántas de esas veces podemos esperar cierto resultado. Esta expectativa se expresa entonces como una proporción o un porcentaje.

Reglas básicas de la teoría de la probabilidad

Existen sólo algunas reglas básicas a seguir para calcular cualquier probabilidad. Todos los cálculos de probabilidades emplean estas reglas esenciales.

Regla de probabilidad 1: las probabilidades siempre varían entre 0 y 1

Como las probabilidades son proporciones, su límite numérico inferior es cero (el evento no puede ocurrir) y su límite numérico superior es 1.00 (el evento debe ocurrir). En otras palabras, las probabilidades siempre se calculan entre 0.00 y 1.00 (o bien 0% y 100%). Si éste no es el caso, ha ocurrido un error matemático.

Algunos eventos tienen una probabilidad cero de ocurrencia —nunca suceden (por ejemplo, permanecer vivo bajo el agua durante 24 horas sin dispositivos de supervivencia)—. Algunos eventos ocurren con una probabilidad del 100% —siempre suceden (por ejemplo, que el Sol saldrá mañana)—. Sin embargo, muchos eventos no están tan definidos; sus probabilidades de ocurrencia están en alguna parte entre nunca y siempre.

Regla de probabilidad 2: la regla de la adición para eventos alternativos

Algunas veces deseamos definir *éxito* como más que un solo evento característico. Por ejemplo, ¿cuál es la probabilidad de sacar un rey *o* un as de un mazo de cartas? Aquí tenemos dos alternativas para el éxito: un rey o bien un as. La **regla de la adición para eventos alternativos** establece que *la probabilidad de eventos alternativos es igual a la suma de las probabilidades de los eventos individuales*. Por tanto,

$$p\,[\text{rey o as}] = p\,[\text{rey}] + p\,[\text{as}]$$

$$= \frac{\#\text{ de reyes en el mazo}}{\#\text{ total de cartas en el mazo}} + \frac{\#\text{ de ases en el mazo}}{\#\text{ total de cartas en el mazo}}$$

$$= \frac{4}{52} + \frac{4}{52} = \frac{8}{52}\ 0.1538 \text{ (cerca del 15\%)}$$

Un truco simple a seguir con la regla de la adición es reemplazar la palabra *o* con un signo de adición, +.

No compliquemos esto. La regla de la adición es sólo una directriz para ayudarnos a calcular una probabilidad cuando hay varias formas de obtener el éxito. Para el caso de sacar un as o un rey, hay ocho maneras. (Si no estás convencido, cuenta los ases y los reyes en un mazo de cartas.)

En capítulos posteriores emplearemos el símbolo P (mayúscula) para representar la probabilidad de éxito y Q para representar la probabilidad de fracaso. (Estos símbolos probablemente evocaron el viejo adagio: "ocúpate de tus p y q".) La regla de adición conduce a un

punto importante: la probabilidad de éxito *o* fracaso debe ser 1.00; es decir, $P + Q = 1$. De aquí se deriva que si conocemos P entonces Q puede calcularse rápidamente. Es decir,

$$Q = 1 - P$$

De manera similar

$$P = 1 - Q$$

Por ejemplo, si $P = p$ [rey o as], entonces

$Q = p$ [cualquier carta distinta de rey o as] $= 1 - p = 1 - 0.1538 = 0.8462$ (aproximadamente 85%)

En otras palabras, si tenemos una posibilidad de aproximadamente 15% de sacar un rey o un as, entonces tenemos una posibilidad de casi 85% de no sacarlos.

Regla de probabilidad 3: ajuste para las ocurrencias conjuntas

A veces el éxito para un evento no es directo debido a que un resultado en particular es exitoso en más de una manera. Por ejemplo, al sacar una sola carta de un mazo estándar de 52 cartas, hay un problema en el siguiente cálculo en el que se usa la regla de la adición:

$$p\,[\text{rey o reina o corazón}] = p\,[\text{rey}] + p\,[\text{reina}] + p\,[\text{corazón}]$$

Incorrecto

$$= \frac{\#\text{ de reyes} + \#\text{ de reinas} + \#\text{ de corazones en el mazo}}{\#\text{ total de cartas en el mazo}} = \frac{21}{52} = .4038$$

Esta respuesta es incorrecta. Si tomamos un mazo de cartas y contamos las "cartas exitosas" (reyes, reinas y corazones), tendremos 19, no 21. Éste es el caso debido a que cuando sumamos las probabilidades separadas contamos al rey y a la reina de corazones dos veces. Siendo un rey *y* un corazón, el rey de corazones es exitoso de dos maneras (en otras palabras, las características no son mutuamente exclusivas). De igual manera, doble éxito ocurre para la reina de corazones.

Cuando tenemos *un evento que tiene doble éxito o une dos aspectos de éxito*, lo llamamos **ocurrencia conjunta**. Para calcular la probabilidad correcta, debemos restar cada ocurrencia conjunta para eliminar este doble conteo. En este caso, la reina de corazones y el rey de corazones, cada uno, son una ocurrencia conjunta. Por tanto,

p [rey o reina de corazones]

$$= p\,[\text{rey}] + p\,[\text{reina}] + p\,[\text{corazón}] - \{p\,[\text{ocurrencias conjuntas}]\}$$

$$= \frac{\#\text{ de reyes} + \#\text{ de reinas} + \#\text{ de corazones en el mazo}}{\#\text{ total de cartas en el mazo}} - \frac{2\text{ ocurrencias conjuntas}}{\#\text{ total de cartas en el mazo}}$$

$$= \frac{21}{52} - \frac{2}{52} = \frac{19}{52} = .3654$$

Regla de probabilidad 4: la regla multiplicativa para eventos compuestos

Algunos eventos tienen dos o más partes. A estos *eventos con partes múltiples* los denominamos **eventos compuestos** (de la química, donde un *compuesto* como el agua se define como una sustancia compuesta de dos o más elementos, en este caso hidrógeno y oxígeno). Por ejemplo, podemos definir éxito como sacar un par de ases del mazo, es decir, sacar un as, regresarlo, barajar una vez más (es decir, haciendo aleatorio de nuevo el evento) y luego sacar un as de nuevo. La **regla multiplicativa para eventos compuestos** estipula que *la probabilidad de un evento compuesto es igual a la multiplicación de las probabilidades de las partes separadas del evento.* Así,

$$p\,[\text{as luego as}] = p\,[\text{as}] \cdot p\,[\text{as}]$$

$$= \frac{4}{52} \cdot \frac{4}{52} = \frac{16}{2\,704} = .0059$$

Un truco simple a seguir es reemplazar la palabra "y" con el signo de multiplicación, •.

No compliquemos esto. Matemáticamente, la regla de la multiplicación simplemente extrae el número de éxitos en el numerador de la fracción y el número total de eventos posibles en el denominador. De acuerdo con esto, resulta que si pasamos meses sacando una carta, reemplazándola, barajar de nuevo, sacar una segunda carta y registrando los resultados, descubriríamos que hay 16 combinaciones posibles de pares de ases, como se muestra en la figura 6-1. ¡Gracias a Dios existen los matemáticos! Ellos con astucia notaron que, en lugar de tener que identificar estas combinaciones pieza por pieza, sólo necesitamos multiplicar las probabilidades separadas.

El simple ejercicio de lanzar una moneda reforzará aún más la simplicidad de la regla multiplicativa. Calculemos la probabilidad de lanzar una moneda dos veces y obtener cara en los dos eventos:

$$p\,[\text{cara luego cara}] = p\,[\text{cara}] \cdot p\,[\text{cara}]$$

$$= 0.5 \cdot 0.5 = 0.2500 \text{ (o 1 de 4)}$$

Como se muestra en la figura 6-2, lanzar dos monedas (o lanzar una sola dos veces) resulta en cuatro resultados posibles y sólo uno de ellos es cara luego cara.

Para comprender esto realmente, elabore su propio diagrama para la probabilidad de obtener sólo caras al lanzar tres monedas. (Matemáticamente, las probabilidades de eventos dicotómicos se calculan desarrollando la fórmula de distribución binomial; ve el capítulo 13.)

Regla de probabilidad 5: explicación del reemplazo para eventos compuestos

Cuando ilustramos la regla 4, la regla de la multiplicación para eventos compuestos, calculamos la probabilidad de sacar un par de ases y determinamos que $p = .0059$. Estipulamos que la primera carta obtenida sería devuelta al mazo antes de sacar la segunda. Esta estipulación para calcular la probabilidad de un evento común se denomina "con reemplazamiento". Si

FIGURA 6-1

Pares posibles de ases cuando se saca al azar una carta, se reemplaza y se toma al azar una segunda carta

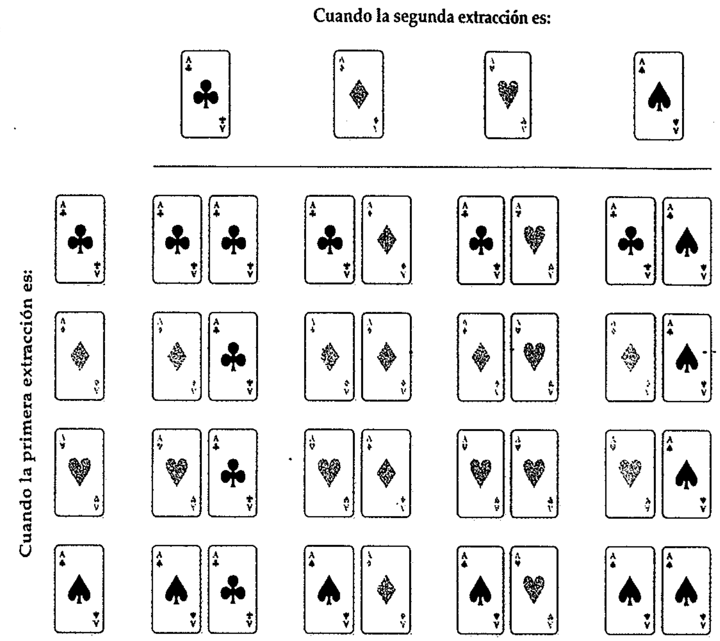
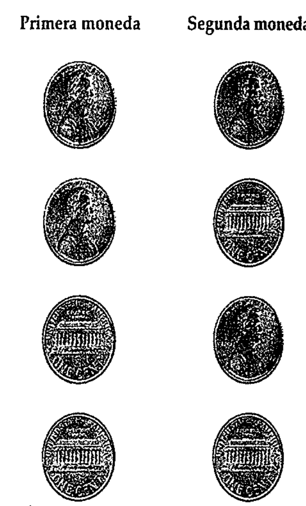


FIGURA 6-2

Resultados posibles de dos monedas lanzadas



no hubiéramos regresado la primera carta, el cálculo se habría hecho "sin reemplazamiento" y la probabilidad calculada habría sido diferente:

$$p\text{ [as y as] sin reemplazamiento} = p\text{ [as]} \cdot p\text{ [as]}$$

$$= \frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652} = .0045$$

La probabilidad del primer as es la misma con o sin reemplazamiento debido a que el evento inicia con 52 cartas y cuatro ases. Pero si la primera carta sacada es un as *y si no se reemplaza*, entonces para el segundo evento sólo hay 51 cartas en el mazo y únicamente hay 3 ases. Debemos poner mucha atención a los puntos de reemplazamiento en los eventos compuestos. Los numeradores y denominadores se ajustan como corresponde. Por ejemplo, calculemos lo siguiente:

$$p\text{ [as y rey y as] sin reemplazamiento} = p\text{ [as]} \cdot p\text{ [rey]} \cdot p\text{ [as]}$$

$$= \frac{4}{52} \cdot \frac{4}{51} \cdot \frac{3}{50} = \frac{48}{132600} = .0004$$

Finalmente, no todos los eventos compuestos implican cuestiones de reemplazamiento. Por ejemplo, el reemplazamiento no es una cuestión que tenga que ver con el lanzamiento. Las probabilidades calculadas son las mismas para "cara y cara" al lanzar dos monedas a la vez o al lanzar una moneda dos veces.

Las cinco reglas de probabilidad son fundamentales; es decir, deben considerarse al calcular la probabilidad de cualquier evento, sin importar qué tan simple o complicado éste pueda ser. Los ejemplos simples presentados en este capítulo ilustran estos principios básicos. En textos más avanzados se presentan formulaciones más complejas de probabilidades como las de Lee y Maykovich (1955). Por fortuna, para los estudiantes de investigadores de hoy, no es necesario tener habilidades matemáticas extensas para calcular probabilidades. Los escritores de software de cómputo sólo requieren que nosotros aprendamos qué teclas pulsar o qué tabla leer para obtener las respuestas a cuestiones de probabilidad. Sin embargo, la comprensión total de la teoría básica de la probabilidad es necesaria para evitar interpretaciones erróneas de la salida de la computadora. Además, la comprensión de la teoría de la probabilidad es esencial para adquirir la imaginación estadística.

Uso de la curva normal como una distribución de probabilidades

Pensamiento proporcional respecto de un grupo de casos y casos únicos

Como observamos en el capítulo 5, la desviación estándar se utiliza para examinar la forma en que las puntuaciones se dispersan en una distribución y para comparar la dispersión de dos o más muestras. Sin embargo, podemos hacer mucho más con la desviación estándar. Con una sola variable de intervalo/razón con la *que tenemos motivos para creer que está normalmente distribuida en su población*, podemos calcular puntuaciones estandarizadas (puntuaciones Z) y emplearlas para determinar la proporción (p) de puntuaciones de una distribución que cae entre dos puntuaciones en la distribución. Puesto que la curva normal tiene una forma inconfundible, podemos identificar y medir áreas debido a que éstas representan una proporción de casos.

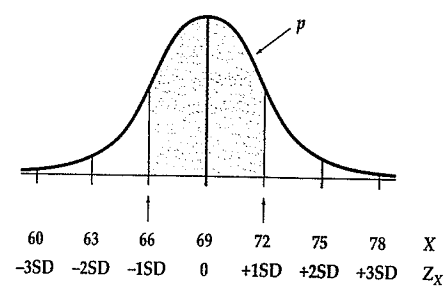
Recuerda del capítulo 5 que una puntuación Z nos indica cuántas desviaciones estándar está alejada una puntuación bruta (o puntuación X) de la media:

$$Z_X = \frac{X - \bar{X}}{s_X} = \text{número de desviaciones estándar (SD) de la media}$$

Notamos que aproximadamente 68% de los casos en una población normalmente distribuida tienen puntuaciones X dentro de una distancia de 1 desviación estándar a ambos lados de la media (es decir, entre una puntuación Z de más y menos 1). Por ejemplo, supongamos que tenemos la siguiente información donde X = estatura para una muestra de hombres en un club deportivo:

$\bar{X} = 69$ pulgadas $\quad s_X = 3$ pulgadas $\quad$ La distribución es normal

Puesto que esta distribución es normal, tracemos la curva normal para obtener un sentido de proporción respecto de cuántos hombres tienen cierta estatura. Nuestro conocimiento básico de la curva normal nos indica que aproximadamente 68% están entre 66 y 72 pulgadas, como se indica en la curva siguiente. Además, como la mediana se ubica en la media, sabemos que la mitad de los hombres miden menos de 69 pulgadas (5 pies, 9 pulgadas) y la otra mitad mide más. Y puesto que más del 99% de una población normalmente distribuida caen dentro de tres puntuaciones Z a ambos lados de la media, muy pocos son más bajos que 60 pulgadas o más altos que 78 pulgadas.



Por tanto,

$$p\text{ [de } X = 66 \text{ a } X = 72] = \text{aproximadamente } 68\%$$

De hecho, con la ayuda de una tabla estadística podemos calcular puntuaciones Z y utilizarlas para determinar cualquier área bajo la curva. Este procedimiento se denomina "partición de áreas bajo la curva normal", y en breve haremos algunas particiones.

Como resultado, las áreas bajo la curva normal representan probabilidades de ocurrencia. Observa que empleamos el símbolo p para representar proporciones *y* probabilidades. Las probabilidades son proporciones del número de veces que se tiene éxito de todas las ocurrencias posibles. Conocer la proporción de éxito para la población en conjunto nos da la probabilidad de éxito para un solo sujeto. En otras palabras, un área especificada bajo la curva normal proporciona la probabilidad de ocurrencia de cualquier puntuación individual que cae entre dos valores de puntuaciones cualesquiera.

Para ilustrar esta relación, supongamos que estamos en el club deportivo pasando el rato. Para entretenernos, jugamos un juego extraño llamado "adivine la estatura". Las reglas del juego son tales que cuando oímos a alguien acercarse desde la esquina suponemos su estatura y luego se la preguntamos cuando se acerca. Si estamos dentro de 3 pulgadas de su estatura correcta, ganamos.

¿Cómo podemos mejorar nuestras posibilidades de ganar? Sabemos que las estaturas de los miembros del club deportivo están normalmente distribuidas respecto a la media de 69 pulgadas, con una desviación estándar de 3 pulgadas. Esto nos indica que aproximadamente 68% de los hombres tienen una estatura entre 66 y 72 pulgadas. Pensemos estadísticamente; es decir, miremos a futuro. Por cada 100 hombres que se aproximan, 68 caerán en el rango de "éxito".

$$p \text{ [de que la estatura del próximo hombre se encuentre entre 66 y 72 pulgadas]}$$

$$= \frac{\text{\# de hombres con esa estatura}}{\text{100 que se aproximan}} = \frac{68}{100} = 0.6800$$

Si suponemos 69 pulgadas, nuestra oportunidad de ganar, entonces, es de aproximadamente 68%, que no es una mala probabilidad. Para obtener un sentido de proporción sobre lo anterior, imagina que los miembros del club son 100 canicas en una caja, con canicas color verde que representan a aquéllos con estaturas entre 66 y 72 pulgadas. Hay 68 canicas color verde y la probabilidad de tomar una al azar es de 0.6800 o 68%.

En una distribución normal de puntuaciones, (1) la proporción de casos entre dos puntuaciones, (2) el área bajo la curva entre estas dos puntuaciones y (3) la probabilidad de seleccionar al azar un caso entre estas puntuaciones *son todas iguales*. Por eso es que empleamos p para representar todas estas ideas. Por ejemplo, el símbolo p [de $X = 66$ a $X = 72$ pulgadas] lo podemos interpretar de tres maneras:

1. *Una interpretación distributiva que describe el resultado en relación con la distribución de puntuaciones en una población o muestra.* Por tanto, casi 0.6800 (o 68%) de los hombres del club están entre 66 y 72 pulgadas de estatura.
2. *Una interpretación gráfica que describe la proporción del área bajo una curva normal* (suponiendo que la distribución tiene forma normal). Por tanto, casi 68% del área bajo la curva normal cae entre las puntuaciones X de 66 y 72 pulgadas.
3. *Una interpretación estadística que describe la probabilidad de una sola extracción de un sujeto de esta población.* Por tanto, si un miembro del club se aproxima al azar, hay una posibilidad de cerca de 0.6800 de que esté entre 66 y 72 pulgadas de estatura.

Tres maneras de interpretar el símbolo *p*

1. Una interpretación distributiva que describe el resultado en relación con la distribución de puntuaciones en una población o muestra.
2. Una interpretación gráfica que describe la proporción del área bajo una curva normal (suponiendo que la distribución tiene forma normal).
3. Una interpretación probabilística que describe la probabilidad de una sola extracción al azar de un sujeto de esta población.

Estas tres interpretaciones indican lo mismo: casi 68% de los hombres se encuentran entre 66 y 72 pulgadas de estatura. Debido a su interpretación estadística, con frecuencia a la curva normal se le refiere como curva de probabilidad.

Estas distinciones también resaltan un punto importante sobre las probabilidades de eventos. Aunque indicada para un solo tipo de "éxito", cualquier probabilidad se basa en toda la distribución de todos los eventos posibles. Un evento singular se evalúa relativo a un conjunto mayor de ocurrencias. Este tipo de pensamiento proporcional es básico para comprender la imaginación estadística.

Partición de áreas bajo la curva normal

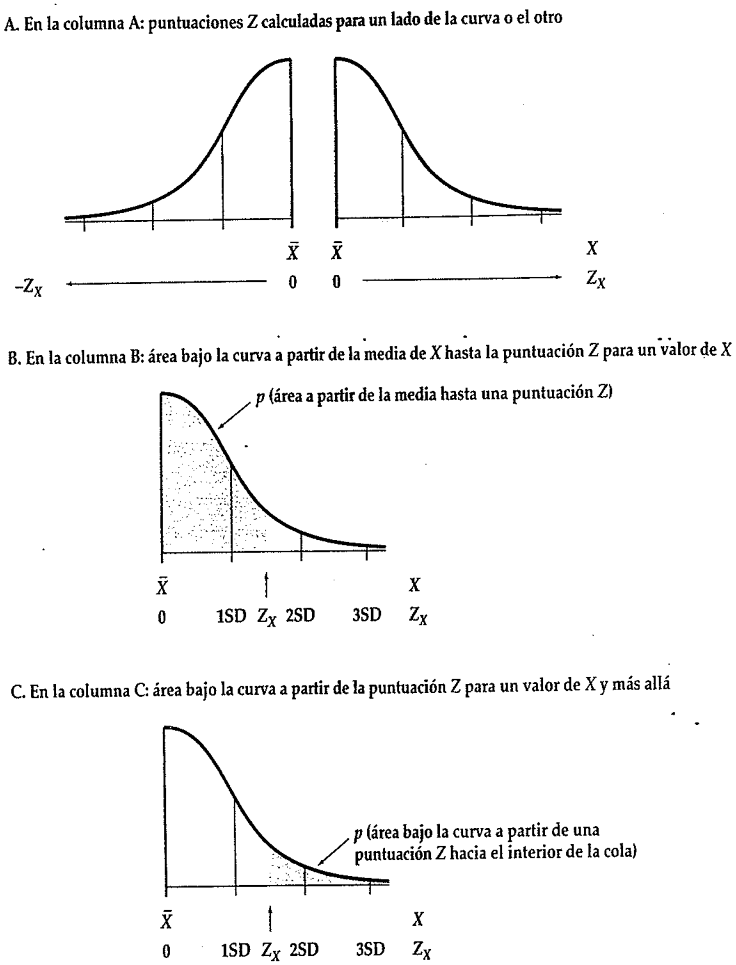
La partición de un área bajo la curva normal consiste en *identificar parte de la curva y calcular la proporción (p) de la curva total que representa dicha parte*. Empleamos la tabla de distribución normal (tabla estadística B del apéndice B) cuando hacemos la partición. ¿De dónde provienen los números de esta tabla? Hace mucho tiempo, los estadistas descubrieron cómo se ajustan las ocurrencias de muchos fenómenos naturales a la forma de campana de la curva normal. Ellos realizaron los cálculos de este fenómeno y determinaron la media, la desviación estándar y las puntuaciones Z. Luego formularon áreas o proporciones (p) bajo la curva. Estas áreas son fijas y se aplican a cualquier variable normalmente distribuida debido a que la normalidad es una ocurrencia natural, como la gravedad. La tabla de la curva normal nos proporciona áreas bajo la curva calculadas de manera precisa. En este punto debemos enfatizar una cosa: la partición de áreas empleando la media, la desviación estándar y las puntuaciones Z, y la curva de distribución normal *sólo funciona si tenemos una razón para creer que las puntuaciones en una población están normalmente distribuidas*. Si la distribución de las puntuaciones está sesgada o, de lo contrario, formada de manera singular, la tabla de la curva normal no se puede emplear en los cálculos.

Partición de un área bajo la curva normal Identificar parte de la curva y calcular la proporción (p) de la curva total que dicha parte representa.

La tabla de la curva normal proporciona lo necesario para calcular con exactitud qué extensión del área está bajo la curva entre dos puntuaciones cualesquiera o a los lados de cualquier puntuación individual. Recuerda que un área bajo la curva representa una proporción (p) de la población entre las puntuaciones brutas correspondientes a esa sección de la curva. Estas p se calculan con cuatro lugares decimales. La figura 6-3 contiene la información de la tabla de la curva normal. Como se indica en la figura 6-3A, la columna A de la tabla de la curva normal indica puntuaciones Z, donde Z_X es el número de *desviaciones estándar* que una puntuación X se desvía de la media. La columna A sólo proporciona puntuaciones *positivas* o aquellas que se aplican a la derecha de la curva normal. Pero la curva es simétrica (es decir, el lado izquierdo es una imagen en espejo del lado derecho). Por tanto, la columna A se puede emplear con puntuaciones Z negativas simplemente imaginando un signo negativo delante de las entradas.

La columna B de la tabla de la curva normal proporciona el área desde la media hasta una puntuación Z, como se representa en la figura 6-3B. Por ejemplo, observa en la tabla de la curva normal una puntuación Z de 1.00 en la columna A. La entrada en la columna B es 0.3413 (casi 34%). Para el club deportivo podemos afirmar que 34% de los miembros están entre las estaturas de 69 y 72 pulgadas. De manera similar, podemos ver esta puntuación Z

FIGURA 6-3
Información proporcionada en las columnas de la tabla de distribución normal (tabla estadística B del apéndice B)



como −1.00, y de nuevo, en la columna B se lee 0.3413, la proporción de miembros entre las estaturas de 66 y 69 pulgadas. Dos por 0.3413 es 0.6826, o casi 68%, la proporción entre 66 y 72 pulgadas, que indicamos que cae dentro de 1 desviación estándar a ambos lados de la media. De igual forma, dijimos que casi 95% de las puntuaciones en cualquier distribución normal caen dentro de casi 2 desviaciones estándar de la media. En realidad, 95% caen dentro de 1.96 SD (una puntuación Z) a ambos lados de la media. Comienza a aprender cómo emplear la tabla encontrando estas puntuaciones Z en la columna A y calculando las áreas en la columna B.

La columna C de la tabla de la curva normal proporciona el área bajo la curva a partir de una puntuación Z y más allá en la "cola" de la curva, como se muestra en la figura 6-3C. Por ejemplo, .1587 (o 15.87%) de las puntuaciones en una distribución normal caen a la derecha de la puntuación Z de 1.00 o a la izquierda de una puntuación Z de −1.00. Esto lo determinamos observando la puntuación Z de 1.00 en la columna A y después observando la entrada .1587 en la columna C.

Antes indicamos que cualquier variable normalmente distribuida tiene una mediana igual a la media. Por tanto, 50% de las puntuaciones en cualquier distribución normal caen en cualquier dirección a partir de la media. Como la tabla proporciona la mitad de la curva, observa que para cualquier puntuación Z las columnas B y C suman 0.5000 o 50%. Por último, ten en cuenta que las puntuaciones Z pueden ser positivas o negativas, dependiendo de si una puntuación bruta está arriba o debajo de la media, respectivamente. Las puntuaciones Z pueden ser infinitamente grandes, aunque en la práctica por lo común caen entre aproximadamente −3.00 y 3.00 debido a que en una distribución normal casi 100% de los casos caen dentro de 3 desviaciones estándar a ambos lados de la media. Sin embargo, las *áreas* en las columnas B y C de la tabla de la curva normal siempre son positivas; estas áreas representan un espacio. Un espacio cero es la cantidad menor que podemos tener y un espacio de 100% es el mayor.

Problemas de ejemplo empleando la curva normal

Para mostrar la utilidad de la tabla de la curva normal, solucionemos algunos **problemas de ejemplo**. Ten en cuenta que la partición tiene como base la media y la desviación estándar; por tanto, la variable debe ser de intervalo/razón. Además, para emplear la tabla de la curva normal debemos estar seguros de que la variable está normalmente distribuida en la población. La distribución no puede ser asimétrica, puntiaguda, plana, bimodal o de alguna otra forma. La normalidad se determina mejor observando el histograma de la variable para identificar la forma distintiva de campana. Sin embargo, si el histograma de una variable se hace *para una muestra* y si la distribución de las puntuaciones no tiene una forma de campana perfecta, la variable aún podría estar normalmente distribuida *en la población*. La diferencia en la forma podría deberse a un error en el muestreo. En este texto no trataremos este punto particular. Simplemente declararemos que suponemos que la variable está normalmente distribuida en la población; en pocas palabras, "suponemos normalidad".

Cómo dividir áreas bajo la curva normal

RESUMEN:

Cuando se trabaja con la curva normal hay tres elementos esenciales: *a*) puntuaciones X (puntuaciones brutas), *b*) puntuaciones Z (puntuaciones estandarizadas) y *c*) las *p* (proporciones del área tomadas de la tabla de la curva normal, apéndice B, tabla estadística B). Si cuentas con uno de estos tres elementos, puedes determinar los otros dos:

Si se da *X*, calcule *Z* y obtenga *p* de la tabla de la curva normal.
Si se da *Z*, calcule *p* de la tabla de la curva normal y calcule *X*.
Si se da *p*, calcule *Z* de la columna A de la tabla de la curva normal y calcule *X*.

1. Asegúrate de que la variable es de un nivel de medición de intervalo/razón y que la distribución de puntuaciones se supone ser normal.

2. Calcula la media y la desviación estándar de X (si estos estadísticos no se proporcionan en el ejercicio).
3. Traza la curva normal y marca la ubicación de la media y 3 desviaciones estándar en ambos lados de ella. Marca en la curva las puntuaciones Z (es decir, de –3 SD a +3 SD).
4. Marca la curva para puntuaciones X. Inicia sumando la desviación estándar a la media para obtener la puntuación X que cae 1 desviación estándar arriba de la media.
5. Identifica y sombrea el área objetivo, p, según lo determine el ejercicio. Con referencia a la tabla de la curva normal, determina si el área objetivo es un área del tipo B o una columna del tipo C. Algunas veces la respuesta se encuentra, al hacer cálculos adicionales, empleando las figuras de las columnas B o C. (Vea, por ejemplo, los problemas tipo 3, 4 y 6 siguientes.)
6. Calcula las puntuaciones Z para las puntuaciones X estipuladas en el ejercicio.
7. Consulta la tabla de la curva normal para proporciones (p) bajo la curva.

Supongamos que hemos realizado entrevistas a 500 mujeres que recibieron pagos de asistencia de ayuda familiar. Nos referiremos a estas mujeres como destinatarias de la asistencia. (En lenguaje común, se les refiere como *madres en beneficencia*, un término un tanto políticamente tendencioso.) El interés que tenemos en estas mujeres es saber cómo la pobreza afecta su autoestima, un sentimiento de un individuo de "valía, adecuación, competencia y capacidad de agradar" (Ensminger 1995: 351). Imagina que medimos la autoestima con una escala de actitud de 20 puntos, que tiene un nivel de medición de intervalo. Se determina que la puntuación media de autoestima es 8 con una desviación estándar de 2. Un histograma nos asegura que la distribución tiene una forma normal. Iniciamos estos ejercicios haciendo un inventario de la información conocida, la cual es "dada" en el problema.

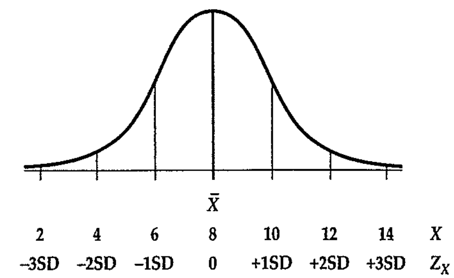
Datos: una variable de intervalo X = autoestima y sus puntuaciones brutas (no mostradas).

A partir de estas puntuaciones obtenemos los estadísticos siguientes:

$\bar{X}$ = 8 puntos de autoestima $\quad s_X = 2$ puntos de autoestima

$n = 500$ $\quad$ Suponga una distribución normal

[*Sugerencia de estudio: traza la curva de forma de campana en todos los problemas*. Es una buena práctica siempre trazar la curva normal. Marca la media y 3 desviaciones estándar a partir de ella en ambas direcciones. Titula la curva de acuerdo con X (en este caso, puntuaciones de autoestima) y para las puntuaciones Z. *Recuerda:* X es una puntuación bruta con una unidad de medida de puntos de autoestima, Z_X es una puntuación estandarizada o Z con unidades de medida de desviaciones estándar (SD) y p es una proporción del área bajo la curva.]



Nuestro conocimiento básico de la distribución normal de inmediato nos indica lo siguiente: (1) 50% de las destinatarias de asistencia tienen puntuaciones arriba de 8 y 50% tienen puntuaciones debajo de 8, (2) casi el 68% tienen puntuaciones entre 6 y 10 en la medida de autoestima, (3) casi el 95% tienen puntuaciones entre 4 y 12 y (4) casi todas, más de 99%, tienen puntuaciones entre 2 y 14.

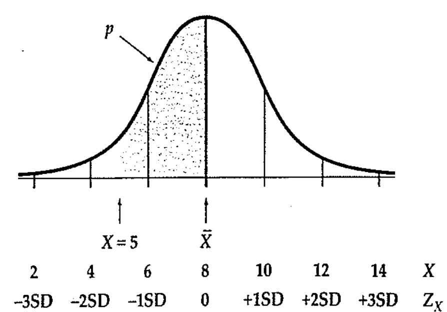
Podemos emplear la tabla de la curva normal para responder varios tipos de preguntas respecto a la distribución de la autoestima entre destinatarias de asistencia familiar. [*Sugerencia de estudio importante:* la tabla de la curva normal requiere puntuaciones Z. Cuando tengas duda respecto a cómo iniciar un problema, calcula las puntuaciones Z.]

Problema tipo 1: *p* [de casos desde la media hasta una puntuación *X*] Encuentra la proporción (p) de casos entre la media y alguna puntuación X.

Plan de solución: Traza y marca la curva normal para la variable X; sombrea el área objetivo (p) desde la media hacia la puntuación X especificada; calcula la puntuación Z para esa puntuación X; ubica la puntuación Z en la columna A de la tabla de la curva normal; obtén p de la columna B; reporta la respuesta en términos comunes.

Ilustración: Donde X = puntuación de autoestima, ¿qué porcentaje de destinatarias de asistencia tienen puntuaciones de autoestima entre 5 y 8?

Identifica esta área objetivo, p.



La columna B en la tabla de la curva normal proporciona áreas bajo la curva desde la media hacia cualquier puntuación Z. Al trazar la curva, podemos ver que el área objetivo (p) está limitada por las medias; por tanto, p es un área "tipo columna B".

El paso siguiente al resolver problemas es para transformar una puntuación bruta en una puntuación Z:

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{5 - 8}{2} = \frac{-3}{2} = -1.50 \text{ SD}$$

Recuerda que una puntuación Z sólo es otra manera para expresar una puntuación bruta. Una destinataria de asistencia con puntuación de 5 en autoestima cae 1.50 SD *debajo* de la media, la puntuación Z *negativa* de −1.50; ella está entre aquéllas con autoestima un tanto baja. En la columna A de la tabla de la curva normal, encuentra 1.5 y trátela como −1.5. Consulta la columna B y reporta la respuesta como sigue:

$$p\,[\text{de } X = 5 \text{ a } X = 8] = 0.4332 \qquad \% = p\,(100) = 43.32\%$$

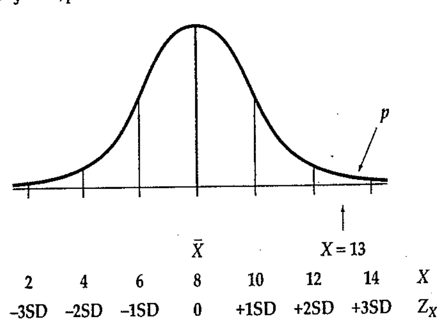
Por último, responde la cuestión en términos comunes: poco más de 43% de las destinatarias de asistencia tuvieron una puntuación entre 5 y 8 en la medida de autoestima. (Ésta es una interpretación distribucional que describe el resultado en relación con la distribución de puntuaciones de la población de destinatarias de asistencia familiar.) Si seleccionamos al azar un segundo nombre de los expedientes de casos, hay una posibilidad de casi 43% de que esta persona tendrá una puntuación entre 5 y 8 en la medida de autoestima. (Ésta es una interpretación probabilística, la probabilidad de que una sola destinataria de asistencia elegida al azar se encuentre en el área objetivo.) Calculamos porcentajes y sustituimos el término *posibilidad* por *probabilidad* para clarificar la expresión al público en general.

Problema tipo 2: p [de casos mayores que una puntuación X] Determine la proporción (p) de casos mayores que una puntuación X específica.

Plan de solución: Traza y marca la curva normal para la variable X; sombrea el área objetivo (p) desde la puntuación X hacia la cola en *la dirección positiva o "mayor que"*; calcula la puntuación Z y ubica en la columna A; obtén p de la columna C.

Ilustración: Donde X = puntuación de autoestima, ¿qué proporción de recipientes de asistencia tuvieron una puntuación de 13 o mayor en la escala de autoestima?

Sombrea el área objetivo, p:



Calcula la puntuación Z para $X = 13$:

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{13 - 8}{2} = \frac{5}{2} = 2.50 \text{ SD}$$

Encuentra 2.50 en la columna A de la tabla de la curva normal. Consulta la columna C y reporta la proporción del área mayor que o igual a 13, como sigue:

$$p\,[\text{de } X \geq 13] = 0.0062$$

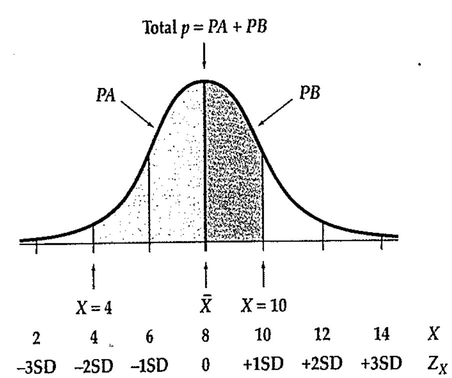
Responde la cuestión en términos comunes: sólo 62 de cada 10 000 recipientes de asistencia tuvieron una puntuación de 13 o mayor en la escala de autoestima. (En una muestra de 500 esto sería casi 3 personas.) Muy pocas destinatarias de asistencia tienen una autoestima extremadamente alta. Si se eligiera al azar un nombre entre los expedientes de casos, habría una probabilidad menor que 1% que esta persona tuviera una puntuación de 13.

Problema tipo 3: p [casos entre dos puntuaciones X en lados distintos de la media] Determina la proporción de casos entre dos puntuaciones X, una debajo de la media y la otra arriba de la media.

Plan de solución: Traza y marca la curva normal; sombrea el área objetivo (p) desde una puntuación X hasta la otra; calcula las puntuaciones Z para las dos puntuaciones X; ubícalas en la columna A de la tabla de la curva normal; obtén las áreas *PA* y *PB* (trazadas abajo) de la columna B; calcula el área (p), que será la suma de *PA* y *PB*.

Ilustración: Donde X = puntuación de autoestima, ¿qué proporción de las destinatarias de asistencia tuvieron una puntuación entre 4 y 10 en la escala de autoestima? *(Sugerencia de estudio:* sólo trazando la curva podemos ver fácilmente que este problema implica dos áreas adjuntas a la media: dos áreas de tipo columna B.)

Sombrea el área objetivo, p:



Calcula las puntuaciones Z para $X = 4$ y $X = 10$:

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{4 - 8}{2} = \frac{-4}{2} = -2.00 \text{ SD}$$

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{10 - 8}{2} = \frac{2}{2} = 1.00 \text{ SD}$$

Ahora utiliza la tabla de la curva normal. En la columna A encuentra cada una de las dos puntuaciones Z. Consulta la columna B para obtener las áreas *PA* y *PB* y reporta la respuesta como sigue:

$$PA = p \text{ [de } X = 4 \text{ a } X = 8] = 0.4772$$
$$PB = p \text{ [de } X = 8 \text{ a } X = 10] = 0.3413$$
$$p \text{ [de } X = 4 \text{ a } X = 10] = PA + PB = 0.4772 + 0.3413 = 0.8185$$
$$\% = p(100) = 81.85\%$$

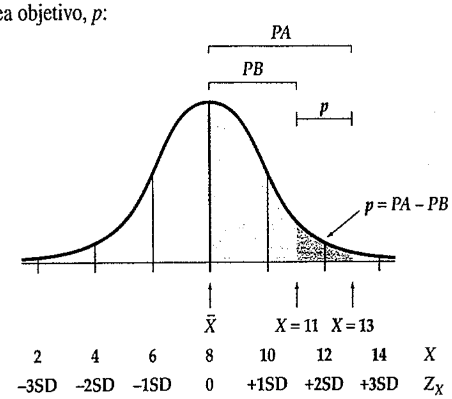
Responde la cuestión en términos comunes: casi 82% de las destinatarias de asistencia tienen puntuaciones de autoestima entre 4 y 10. Si se elige al azar un nombre entre los expedientes de casos, hay una posibilidad de 82% de que esta persona tendrá una puntuación de autoestima entre 4 y 10.

Problema tipo 4: *p* [de casos entre dos puntuaciones *X* en un lado de la media] Determina la proporción (p) de casos entre dos puntuaciones X *en un lado* de la media.

Plan de solución: Traza y marca la curva; sombrea el área objetivo (p) de una puntuación X a la otra; calcula las puntuaciones Z y ubícalas en la columna A de la tabla de la curva normal; obtén las áreas *PA* y *PB* de la columna B; calcula el área p, que es *PA* menos *PB*.

Ilustración: ¿Qué proporción de las destinatarias de asistencia tuvieron una puntuación entre 11 y 13 en la escala de autoestima? En la muestra de 500, *¿cuántas* destinatarias de asistencia son? *(Sugerencia de estudio:* al trazar la curva, observamos que el área objetivo p no toca la media. Por tanto, *no* es un área tipo columna B en la tabla de la curva normal; ni es un área con forma de cola, tipo columna C. Por tanto, para resolver esta ilustración, debemos calcular p de manera indirecta.)

Sombrea el área objetivo, p:



Calcula las puntuaciones Z para $X = 13$ y $X = 11$:

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{13 - 8}{2} = \frac{5}{2} = 2.50 \text{ SD}$$

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{11 - 8}{2} = \frac{3}{2} = 1.50 \text{ SD}$$

En la columna A encuentra cada una de las puntuaciones Z. Consulta la columna B para obtener las áreas *PA* y *PB*, y reporta la respuesta como sigue:

$$PA = p \text{ [de } X = 8 \text{ a } X = 13] = 0.4938$$
$$PB = p \text{ [de } X = 8 \text{ a } X = 11] = 0.4332$$
$$p \text{ [de } X = 11 \text{ a } X = 13] = PA - PB = 0.4938 - 0.4332 = 0.0606$$
$$\% = p(100) = 6.06\%$$

[*Sugerencia de estudio:* resta las p (es decir, las áreas bajo la curva), no las puntuaciones Z.] Para determinar cuántas de las 500 destinatarias de asistencia tuvieron puntuaciones en este rango, toma la proporción del tamaño de la muestra n como sigue:

Cálculo del número de casos de la muestra que corresponden a un área
$\# = p\,(n)$
donde
= número de casos en la muestra para el área designada, p
p = proporción del área bajo la curva
n = tamaño de la muestra

El número de destinatarias de asistencia con puntuación entre 11 y 13 en la escala de autoestima es

$$\# = p\,(n) = 0.0606\,(500) = 30.3 = 30 \text{ recipientes}$$

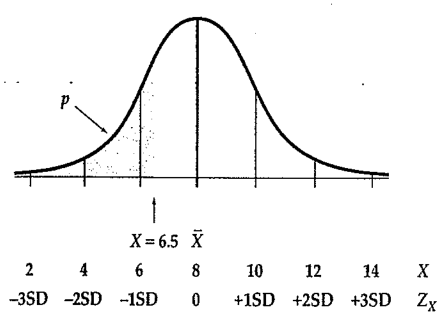
Por último, responde estas cuestiones en términos comunes: sólo 6% de los recipientes de asistencia tienen puntuaciones de autoestima entre 11 y 13. Esto es sólo 30 de las 500 destinatarias de asistencia. Si se eligiera al azar un nombre de los expedientes de casos, sólo habría una posibilidad de 6% de que esta persona tuviera una calificación entre 11 y 13.

Problema tipo 5: *p* [de casos menores que una puntuación *X* que es menor que la media] Determina la proporción (p) de casos menores que o iguales a una puntuación X especificada que es menor que la media.

Plan de solución: Traza y marca la curva normal; sombrea el área objetivo (p) de la puntuación X hacia la cola *en la dirección negativa*; calcula la puntuación Z y ubícala en la columna A de la tabla de la curva normal; obtén p de la columna C.

Ilustración: Si se eligiera al azar un nombre entre los expedientes de casos, ¿cuál sería la probabilidad de que esta destinataria de asistencia tuviera una puntuación de 6.5 o menor en la escala de autoestima?

Sombrea el área objetivo, p:



Calcula la puntuación Z para $X = 6.5$:

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{6.5 - 8}{2} = \frac{-1.5}{2} = -.75 \text{ SD}$$

En la columna A de la tabla de la curva normal encuentra .75 y trátalo como si fuera –.75. Consulta la columna C y reporta la respuesta como sigue:

$$p \text{ [de } X \leq 6.5] = .2266$$
$$\% = p(100) = 22.66\%$$

Responde la cuestión en términos comunes: la probabilidad de que un recipiente de asistencia seleccionado al azar obtuviera una puntuación de 6.5 o menor en la escala de autoestima es casi 23%.

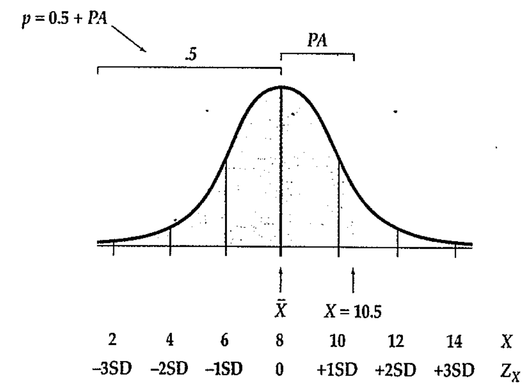
Problema tipo 6: p [de casos menores que una puntuación X que es mayor que la media] Determina la proporción (p) de casos *menores que* una puntuación X especificada que es *mayor que* la media.

Plan de solución: Traza la curva; sombrea el área objetivo (p); calcula la puntuación Z y ubícala en la columna A; obtén p de la columna B y suma 0.5000.

Ilustración: Donde X = puntuación de autoestima, ¿cuál es la probabilidad (p) de que una destinataria de asistencia seleccionada al azar tenga una puntuación de 10.5 o menor en la escala de autoestima?

[*Sugerencia de estudio:* recuerda que la tabla de la curva normal proporciona áreas sólo para un lado de la curva. Asimismo, recuerda que una curva normal tiene una mediana igual a la media; por tanto, la mitad (o una proporción de 0.5000) de las puntuaciones caen debajo de la media. Esta ilustración se resuelve trabajando con el área arriba de la media y luego sumando el área debajo de la media. (Por cierto, para encontrar la proporción, p, de casos *mayores que* una puntuación X especificada que es *menor que* la media, trabaje desde el lado izquierdo. Calcula el área debajo de la media y luego súmala a 0.5000, que es el área arriba de la media.).]

Sombrea el área objetivo, p:



Calcula la puntuación Z para $X = 10.5$:

$$Z_X = \frac{X - \bar{X}}{s_X} = \frac{10.5 - 8}{2} = \frac{2.5}{2} = 1.25 \text{ SD}$$

En la columna A de la tabla de la curva normal encuentra 1.25. Consulta en la columna B y reporta la respuesta como sigue:

$$PA = p \text{ [de } X = 8 \text{ a } X = 10.5] = 0.3944$$
$$p \text{ [de } X \leq 10.5] = PA + 0.5000 = 0.3944 + 0.5000 = 0.8944$$

Responde la cuestión en términos comunes: la probabilidad de que una destinataria de asistencia seleccionada al azar tenga una puntuación de 10.5 o menor en la escala de autoestima es mayor que 89%.

Problema tipo 7: encuentra la puntuación X que tiene una p especificada [de casos] arriba o debajo de ella Determina el valor de una puntuación bruta X para la cual un porcentaje especificado de la muestra o población cae arriba o debajo de ese valor.

Plan de solución: Mientras que los tipos de problema anteriores proporcionaban una puntuación X y requerían un área (p), este problema proporciona información sobre p y requiere una puntuación X. Traza y marca la curva normal; identifica *aproximadamente* y sombrea el área objetivo, p; encuentra esta área en la columna B o en la C de la tabla de la curva normal, cualquiera que sea la columna aparentemente apropiada del trazo; lee la columna A para obtener la puntuación Z; despeja para X como sigue:

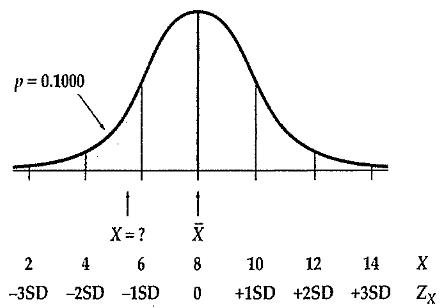
$$Z_X = \frac{X - \bar{X}}{s_X}, \quad \text{por tanto, } X = \bar{X} + (s_X)(Z_X)$$

Ilustración: El Departamento de Salud Mental tiene un programa designado para prevenir eventos de depresión psicológica aguda fomentando la autoestima (X) de las destinatarias de asistencia. El programa sólo tiene recursos para 50 personas entre las 500 a quienes se les midió la autoestima. Elijamos las 50 con la autoestima más baja debido a que presuntamente son las que están en mayor riesgo de depresión. ¿Cuál es la puntuación mayor de autoestima que una destinataria puede tener para calificar en el programa?

Para identificar el área objetivo, p, calculamos la proporción de destinatarias de asistencia que van a calificar:

$$p\,[\text{calificar para el programa}] = \frac{\#\,\text{calificando}}{n} = \frac{50}{500} = .1000$$

Al trazar el área objetivo, ten en cuenta que será una cola en la dirección *negativa* de las puntuaciones debido a que estamos buscando las 50 destinatarias de asistencia *más bajas*. Observa que el área objetivo es un área tipo columna C.



[*Sugerencia de estudio:* en este punto, estima la respuesta a partir de la gráfica. Nuestra marca de la posición de X debe estar cerca. Ahora sabemos que sólo 15.87% de los casos caen debajo de −1 SD, y por tanto, la marca de 10% debe estar justo debajo de ella. Por consecuencia, nuestra puntuación X deberá estar ligeramente debajo de 6. Estimando la respuesta de esta manera no sólo fomenta el pensamiento proporcional, sino que también proporciona una advertencia si nuestra respuesta calculada es incorrecta.]

Ahora empleamos la tabla de la curva normal. En la columna C encuentra 0.1000 o la cantidad más cercana a él, en este caso, 0.1003. Consulta la columna A para determinar la puntuación Z correspondiente de −1.28 y despeja para X:

$$X = \overline{X} + (s_X)\,(Z_X) = 8 + (2)\,(-1.28) = 8 - 2.56 = 5.44 \text{ puntos de autoestima}$$

Responde la cuestión en términos comunes: las destinatarias de asistencia que tienen una puntuación menor que o igual a 5.44 en la escala de autoestima caen en el 10% menor y por tanto califican para el programa contra la depresión.

Sugerencia de estudio: el problema tipo 7 muestra que mientras conozcamos la media y la desviación estándar de una distribución y podamos suponer que la distribución de las puntuaciones en la población tiene forma normal, sólo es necesario una información adicional para resolver cualquier problema. Esta información puede ser una puntuación X bruta, una puntuación Z estandarizada o bien un área bajo la curva normal (p).

Por tanto:

- Si se da una puntuación X, calcula Z_X y utiliza la tabla de la curva normal para obtener p.
- Si se da una puntuación Z, utiliza la tabla de la curva normal para obtener p o despeja para X, donde $X = \overline{X} + (s_X)\,(Z_X)$.
- Si se da un porcentaje o un área, p, utiliza la tabla de la curva normal para obtener la puntuación Z correspondiente y despeja para X, donde $X = \overline{X} + (s_X)\,(Z_X)$.

Cálculo de percentiles para poblaciones con distribución normal

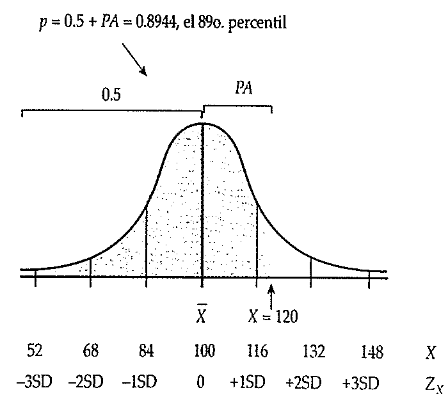
Los problemas tipo 5, 6 y 7 de la curva normal tratan de áreas bajo la curva que están debajo de una puntuación X bruta particular. Estas áreas definen **rangos percentilares,** *el porcentaje de una muestra o población que cae en o debajo de un valor especificado de una variable* (ve el capítulo 2). Por ejemplo, con respecto al problema tipo 6, quien obtenga una puntuación de 10.5 en la escala de autoestima tiene una puntuación mayor que el 89% de las destinatarias de asistencia en la muestra, un rango percentilar de 89. Cuando una variable está normalmente distribuida, podemos emplear la tabla de la curva normal para calcular de manera rápida rangos percentilares.

Muchas distribuciones, en especial el éxito, la inteligencia y las pruebas de admisión en escuelas, se diseñan de manera específica para producir una distribución de puntuaciones que esté normalmente distribuida. Todos recordamos recibir rangos percentilares además de las puntuaciones brutas para tales pruebas. Las compañías que distribuyen las pruebas las "normalizan" de manera intencional tal que las distribuciones de las puntuaciones se ajusten a la curva normal. Una vez que se logra esta normalización, se utiliza la tabla de la curva normal para generar rangos percentilares.

Una prueba estandarizada de uso común es la Escala de Inteligencia de Stanford-Binet. (Para un repaso, visita el sitio http://www.chclibrary.org/micromed/00066170.html.) Esta prueba está diseñada para evaluar el desarrollo cognitivo en niños. El nivel normal de desarrollo es la puntuación media de la prueba de 100 puntos en la escala. Supongamos que la distribución de las puntuaciones es normal y que la desviación estándar es 16 puntos en la escala.

Ilustración: Stanley Jones obtuvo una puntuación de 120 en la Escala de Inteligencia de Stanford-Binet. ¿Cuál es su rango percentilar en la escala? Es decir, ¿cuál es el porcentaje de aplicantes de la prueba que él igualó o mejoró?

Sombrea el área objetivo, p:



Calcula la puntuación Z para $X = 120$:

$$Z_X = \frac{X - \overline{X}}{s_X} = \frac{120 - 100}{16} = \frac{20}{16} = 1.25 \text{ SD}$$

En la columna A de la tabla de la curva normal encuentra 1.25. Consulta la columna B y reporta la respuesta como sigue:

$$PA = p\ [\text{de } X = 100 \text{ a } X = 120] = 0.3944$$
$$p\ [\text{de } X \leq 120] = PA + 0.5000 = 0.3944 + 0.5000 = 0.8944$$

Responde la cuestión en términos comunes: el rango percentilar de Stanley Jones en la escala de Stanford-Binet es 89.

Al emplear la curva normal para calcular rangos percentilares, cuando una puntuación X es mayor que la media, encontraremos un área tipo columna B como PA arriba y la sumamos a .5000. Éste es un cálculo de problema tipo 6. Sin embargo, si una puntuación X es menor que la media, calcularemos la puntuación Z negativa y consultamos la columna C de la tabla de la curva normal. Éste es un cálculo de problema tipo 5. Utilizando estos dos tipos de problemas, podemos fácilmente hacer los cálculos para una serie de puntuaciones de Stanford-Binet. Esto se ilustra en la tabla 6-1. Para practicar, calcula estos rangos percentilares. Observa que los rangos percentilares reportados están redondeados por simplicidad, pero están redondeados hacia abajo en cada caso. Un rango percentilar indica el porcentaje de una muestra o población que cae "en o debajo de un valor específico". Una puntuación redondeada hacia arriba no sería correcta. Por ejemplo, la persona en la tabla 6-2 que tuvo una puntuación de 126 en la escala Stanford-Binet *no* podría decir que tuvo una puntuación igual a 95% de los que tomaron la prueba. Por tanto, una puntuación Stanford-Binet de 126 está en el 94o. percentil.

Por último, debemos mencionar que los rangos percentilares se pueden determinar para distribuciones que no están normalmente distribuidas. Todo lo que se necesita para calcular cualquier percentil es determinar qué porcentaje de una distribución cae debajo de una puntuación X especificada.

TABLA 6-1 | Uso de la curva normal para obtener rangos percentilares

Datos		Cálculos		
Puntuación Stanford-Binet	Puntuación Z	De la columna C problema tipo 5	0.5000 + columna B problema tipo 6	Rango percentilar
68	−2.00	0.0228	—	2o.
80	−1.25	0.1056	—	10o.
90	−0.62	0.2672	—	26o.
100	0.00	0.5000	—	50o.
108	0.50	—	0.6915	69o.
126	1.62	—	0.9474	94o.
133	2.06	—	0.9803	98o.

TABLA 6-2 | Comparación de puntuaciones brutas (X), puntuaciones Z y rangos percentilares para obtener un sentido respecto a variables normalmente distribuidas

Datos		Cálculos	
Estudiante	X	Z_x	Rango percentilar
Ronald	24	0	50
Barry	28	1	84
Sophia	32	2	98

La mayoría de los programas de cómputo proporcionan esta información como el "porcentaje acumulado" de una distribución (vea el capítulo 2).

La curva normal como una herramienta para el pensamiento proporcional

Una vez que hemos aprendido los detalles de dividir la curva normal en áreas, debemos iniciar a apreciar realmente su utilidad. Como una herramienta descriptiva, normalizar las distribuciones de las puntuaciones en las pruebas se hace debido a que la experiencia ha demostrado que la inteligencia, el aprendizaje y el éxito están normalmente distribuidos; es decir, la mayoría de la gente tiene una inteligencia y éxito casi promedio, y ésta es la razón por la que la curva normal se "acampana" en medio. Muy pocas personas son genios o están extremadamente debajo de lo normal, y esto explica por qué la curva empieza a aproximarse al eje horizontal cuando observamos puntuaciones de más de 1 desviación estándar desde la media en cualquier dirección.

Trabajar con la curva normal también nos hace más cautos en la interpretación de datos. Por ejemplo, ahora estamos más conscientes que las puntuaciones de pruebas diferentes (como el ACT y el SAT) se pueden comparar observando las posiciones relativas de las puntuaciones dentro de sus propias distribuciones; una manera simple de hacer esto es comparar rangos percentilares.

Después de trabajar con la distribución normal, estamos al tanto que diferencias iguales entre puntuaciones no siempre indican que una puntuación está a la misma distancia de otra en términos de lo inusual que es. Por ejemplo, supongamos que los 2 000 estudiantes de nuevo ingreso a una universidad estatal tuvieron una puntuación media en el ACT (X) de 24 con una desviación estándar de 4 y que la distribución tenía una forma normal. Ronald obtuvo 24; Barry, 28, y Sophia, 32. La observación de las puntuaciones brutas sugiere que Barry cae precisamente entre Ronald y Sophia en sus rangos en estas puntuaciones. Sin embargo, nuestro sentido de una distribución normal nos deberá convencer que Barry está considerablemente arriba del promedio, aunque su puntuación de 28 sólo es 4 puntos mayor que un 24. Esto es aparente cuando se comparan las puntuaciones brutas (X), las puntuaciones estandarizadas (Z_X) y los rangos percentilares, como se hace en la tabla 6-2.

Esto ilustra la importancia de saber cómo se dispersa una distribución de puntuaciones. Barry sólo tiene 4 puntos más que Ronald en la puntuación bruta, pero tiene 34 puntos porcentuales más en términos del rango percentilar. Barry, al igual que Sophia, tuvó puntuaciones mejores que la gran mayoría de los estudiantes de nuevo ingreso. Las puntuaciones brutas por sí solas sugieren lo contrario y pueden ser muy confusas. La desviación estándar como una unidad de medida con distribuciones normales es una herramienta poderosa para obtener una visión precisa de la importancia de una puntuación bruta.

El fenómeno de normalidad es la esencia del análisis estadístico. Es muy importante que aprendamos cómo movernos por la curva normal y desarrollar las habilidades para dividir las áreas bajo ella. Un vistazo rápido al resto de este texto lo convencerá de la importancia de dominar los problemas en este capítulo. Casi cada capítulo posterior a éste tiene ilustraciones de la curva normal o de curvas similares. Además, como proporciona probabilidad convenientemente, la distribución de la curva normal y las curvas predecibles similares a ella se denominan **distribuciones de probabilidad**. Como analizaremos en el capítulo 7, los eventos de muestreo tienen patrones de ocurrencia predecibles, y sus curvas de probabilidad se utilizan para determinar qué tan usual o inusual es un evento de muestreo. Las probabilidades en general y la distribución normal de probabilidad en particular son elementos clave en el análisis estadístico.

Por último, una prueba de pensamiento interesante sobre probabilidades implica el tema de eventos de baja ocurrencia. Cualquier evento natural o humano tiene *alguna* probabilidad de ocurrir. Por ejemplo, en ocasiones alguien es golpeado por un meteorito. Sin embargo, la mayoría de nosotros no miramos al cielo de manera constante para ver estos objetos. Comprendemos las probabilidades y sabemos que "tienes que tener muy mala suerte" para que esto te suceda. De igual forma, 40 millones de personas juegan a la lotería. Una persona gana. ¡Es muy afortunada! Cae nieve en Florida una vez cada cinco años, pero sucede el día de su boda. ¡Tiene muy mala suerte! En términos de las ideas de la teoría de la probabilidad, ¿qué significa decir que alguien tiene muy buena o mala suerte? ¿Qué es la suerte?

Insensatez y falacias estadísticas: la falacia del jugador: independencia de eventos de probabilidades

Imagine que Bob y Terri están jugando a lanzar una moneda al aire. Bob gana con cara, y Terri gana con cruz. Ellos se turnan decidiendo cuánto valdrá el siguiente lanzamiento, eligiendo una cantidad entre 5 y 25 centavos. Bob ganó tres veces seguidas a 10 centavos cada lanzamiento. ¿Deberá Terri aumentar la apuesta a 25 centavos en el siguiente lanzamiento? ¿Incrementa las probabilidades de que se obtenga cruz en el siguiente lanzamiento el hecho de que se obtuvo cara tres veces seguidas?

La respuesta es no. Un error estadístico común al calcular probabilidades implica la independencia de las partes de eventos compuestos. Cada moneda se lanza de manera independiente de lo que sucedió en los lanzamientos anteriores. Si lanzamos una moneda dos veces y obtenemos caras en ambos lanzamientos, esto no aumenta la probabilidad de obtener cruz en el tercer lanzamiento. Esa probabilidad permanece igual a 0.5000.

Esta tendencia a imaginar que eventos independientes están enlazados es un tipo de falacia del jugador. Cuando un jugador tiene una racha de mala suerte, quizá comience a creer que debe seguir una racha de buena suerte. En efecto, a largo plazo la buena y mala suerte se equilibran. Pero, ¿qué es a largo plazo? ¿Son 3, 10 o 1 millón de lanzamientos? Para un apostador dado, ¿a la larga es mucho más tiempo de lo que durará su dinero? Además, el equilibrio entre la buena y la mala suerte ocurre entre todos los apostadores, no en un solo apostador. Por tanto, si 100 parejas jugaran a lanzar una moneda, en el transcurso de la tarde habría una gran oportunidad de que se obtuvieran ambos resultados, cara y cruz, de igual manera. Pero Bob y Terri podrían terminar lanzando más caras, en tanto que Joe y Maggie quizá obtengan más cruces.

Suponer que los lanzamientos de monedas están enlazados es pensar erróneamente que sabemos la longitud de una "serie", una secuencia de lanzamientos a la larga. Por desgracia, hay un número infinito de secuencias posibles, ya que cada lanzamiento es independiente del siguiente. Por ejemplo, las tres caras consecutivas de Bob podrían ser parte de cualquiera de las siguientes series en que cara y cruz caen un número igual de veces:

X,X,C,C,C,X,C,C,X,X
X,X,X,X,C,X,C,C,C,C,C,X
C,X,C,X,X,C,X,X,X,C,C,X,C,C,C,X
C,C,C,C,X,X,C,C,X,X,X,C,X,X,C,X,C,X,X,C

Para un jugador suponer que conoce de algún modo la secuencia futura de los resultados es suponer que el futuro se puede predecir a una magnitud mayor que lo que nos dicen las probabilidades básicas de ocurrencia. Es obvio que ésta no es una manera sensata de jugar.

RESUMEN

1. La teoría de la probabilidad es el análisis y la comprensión de ocurrencias fortuitas. En el campo de la estadística, la teoría de probabilidad se emplea para calcular el error estadístico.
2. Una probabilidad es una especificación de qué tan frecuentemente es probable que ocurra un evento de interés durante un número de eventos grande (es decir, situaciones en que el evento puede ocurrir).
3. La probabilidad de éxito (P) es la probabilidad de que ocurra un evento de interés. La probabilidad de fracaso (Q) es la probabilidad de que *no* ocurra un evento de interés. $P + Q = 1$.
4. Hay cinco reglas básicas de probabilidad.
5. La curva normal es una distribución de probabilidad para una variable de intervalo/razón que está normalmente distribuida. Las áreas bajo la curva normal se pueden dividir, donde puntuaciones Z y la tabla de la curva normal se utilizan para calcular proporciones de puntuaciones de una población que caen entre dos puntuaciones cualesquiera en la distribución o más allá de cualquier puntuación en sus colas. El área bajo la curva se simboliza como p debido a que se puede interpretar como una probabilidad.
6. Para una distribución normal, hay tres formas de interpretar p: (1) una interpretación distribucional que describe el resultado en relación con la distribución de puntuaciones en una población o muestra; (2) una interpretación gráfica que describe la proporción del área bajo una curva normal, y (3) una interpretación probabilística que describe la probabilidad de tomar al azar sólo un sujeto de esta población.
7. Si una distribución de puntuaciones de intervalo/razón tiene forma normal, entonces se pueden emplear la curva normal y las puntuaciones Z para calcular rápidamente rangos percentilares.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 6 del material del texto disponibles en el sitio web *The Statistical Imagination*, en www.mhhe.com/ritchey2, incluyen cómo calcular probabilidades para juegos comunes de azar y habilidad, como el póquer "Texas Hold' Em."

FÓRMULAS Y REGLAS DE PROBABILIDAD DEL CAPÍTULO 6

Cálculo de una probabilidad:

$$p\,[\text{de éxitos}] = \frac{\#\text{ de éxitos}}{\#\text{ de ensayos}} = \frac{\#\text{ de posibles resultados exitosos}}{\#\text{ total de resultados posibles}}$$

Cálculo de una puntuación Z:

$$Z_X = \frac{X - \overline{X}}{s_X}$$

Cálculo de una puntuación bruta (X) cuando se conoce Z_X:

$$X = \overline{X} + (s_X)\ (Z_X)$$

Cálculo del número de casos que corresponde a un área bajo la curva normal:

$$\# = p\ (n)$$

Reglas básicas de la teoría de las probabilidades

1. Regla de probabilidad 1: las probabilidades siempre oscilan entre 1 y 0.
2. Regla de probabilidad 2: la regla de la adición para eventos alternativos: la probabilidad de eventos alternativos es igual a la suma de las probabilidades de los eventos individuales.
3. Regla de probabilidad 3: ajuste por ocurrencias conjuntas, eventos que cuentan dobles éxitos o unen dos aspectos de éxitos.
4. Regla de probabilidad 4: la regla multiplicativa para eventos compuestos: la probabilidad de un evento compuesto es igual al múltiplo de las probabilidades de las partes separadas del evento.
5. Regla de probabilidad 5: toma en cuenta el reemplazo en eventos compuestos.

PREGUNTAS PARA EL CAPÍTULO 6

1. ¿Qué es teoría de probabilidad?
2. Menciona tres acciones recientes en tu vida cotidiana donde empleaste la teoría de las probabilidades (aunque no hayas calculado las probabilidades reales).
3. ¿Qué denota comúnmente el denominador de una fórmula de probabilidad?
4. ¿Qué denota comúnmente el numerador de una fórmula de probabilidad?
5. Si alguien reporta una probabilidad de 150%, ¿qué regla de probabilidad se ha violado?
6. Menciona dos eventos que tengan una probabilidad de ocurrencia de 100%.
7. Menciona dos eventos que tengan una probabilidad de ocurrencia de 0%.
8. Describe la regla de la adición de probabilidades y especifica cuándo se utiliza. Da un ejemplo.
9. Describe la regla multiplicativa de probabilidad y especifica cuándo se utiliza. Da un ejemplo.
10. Al calcular una probabilidad, un evento que cuenta doble un éxito o une dos aspectos de éxito se denomina __________.
11. Para una proporción de casos ajustando éxitos, ¿qué distingue una interpretación distribucional de una interpretación estadística? Ilustra con un ejemplo.
12. ¿Con variables de qué niveles de medida se utilizan más apropiadamente la media, la desviación estándar y la curva normal?
13. ¿Por qué es importante emplear el mismo símbolo, p, para proporción, probabilidad y área bajo una curva normal?
14. Cuando una puntuación de una variable normalmente distribuida está a la derecha de la media, se encuentra en la dirección __________.
15. Explica por qué no es adecuado emplear puntuaciones Z y la tabla de la curva normal para cualquier distribución de puntuaciones que no esté normalmente formada.
16. ¿Qué información proporciona el rango percentilar?
17. Explica qué significa tener buena o mala suerte.

EJERCICIOS PARA EL CAPÍTULO 6

Conjunto de problemas 6A

6A-1. Calcula las siguientes probabilidades para un dado de juego:

a) $p\,[6]$

b) $p\,[2 \text{ o } 4]$

c) $p\,[2 \text{ luego } 3 \text{ luego } 4]$

6A-2. Supón que tienes una caja con frijoles secos bien mezclados: 150 color rojo, 70 color blanco y 80 color negro. Calcula las probabilidades de sacar al azar de la caja lo siguiente.

a) *p* [blanco luego rojo luego negro] sin reemplazamiento

b) *p* [rojo luego rojo luego negro] con reemplazamiento

c) *p* [blanco luego negro luego blanco] sólo con reemplazamiento de los negros

6A-3. Para el lanzamiento de una moneda (C = cara, X = cruz), calcula lo siguiente.

a) *p* [*C*]

b) *p* [*X* luego *X*]

c) *p* [*X* luego *C* luego *C*]

6A-4. Calcula las siguientes probabilidades para sacar cartas de un mazo estándar con 52 cartas.

a) *p* [10]

b) *p* [7 o rey]

c) *p* [sota o diamante]

d) *p* [rey luego rey o as luego as] sin reemplazamiento

6A-5. Steelman, Powell y Carini (2000) exploraron la relación entre sindicatos de maestros y el desempeño educacional estudiantil medida según pruebas estandarizadas, como el examen de la American College Testing (ACT). Utiliza los estadísticos de la ACT que se muestran aquí para responder las preguntas siguientes. La distribución está normalmente formada. Traza la curva normal y marca todas las áreas objetivo. Considera que X = puntuación ACT.

$$\bar{X} = 22 \text{ ACT puntos}$$
$$s_X = 2 \text{ ACT puntos}$$
$$n = 441\,574 \text{ estudiantes}$$

a) ¿Qué proporción de estudiantes obtuvo una puntuación *arriba* de 26?

b) ¿Qué número de estudiantes que presentaron el examen ACT tuvieron una puntuación *entre* 17 y 19?

c) ¿Qué proporción de las puntuaciones cayó *entre* 18 y 23?

d) Determina la puntuación debajo de la cual caen 90% de las puntuaciones.

e) Si un aspirante tiene que estar al menos en el 90o. rango percentilar para ingresar a un programa universitario, ¿qué puntuación necesita obtener (respuesta breve)?

6A.6. Lynch, Maciejewski y Potenza (2004) examinaron la relación entre varias condiciones psiquiátricas y el comportamiento apostador en adolescentes y adultos jóvenes. Para duplicar los resultados, tú debes obtener datos de una muestra de adultos jóvenes con una edad media de 22 años y una desviación estándar de 2 años. Las edades en esta población están normalmente distribuidas. Tú seleccionarás al azar un individuo de esta población. Calcula las probabilidades siguientes. Traza la curva normal para cada problema. Considera que X = edad.

a) *p* [de seleccionar al azar a alguien entre las edades de 20 y 24 años]

b) *p* [de seleccionar al azar a alguien con "19 años o menor", o "25 años o mayor"]

c) Si al 10% más joven de la población de adultos jóvenes se le enviará una carta, ¿a qué edad deben dirigirse las cartas?

6A-7. Bastiaens (2004) examinó las respuestas de pacientes a un tratamiento antidepresivo en una clínica de salud mental comunitaria. La Escala de Depresión del Centro de Estudios Epidemiológicos (CESD) es usada para evaluar la severidad de los síntomas depresivos. Supongamos que tiene puntuaciones CESD normalmente distribuidas para algunos pacientes de salud mental. La puntuación CESD media es 27.2 puntos en la escala y la desviación estándar es 3.2. Su interés es en aislar puntuaciones CESD extremadamente bajas y altas para estos pacientes, aquellos que caen fuera del 95% medio (o 0.95) de las puntuaciones. Estas áreas caerán en las colas de la curva normal con 2.5% (o 0.025) en cada cola. Considera que X = puntuación CESD.

a) Traza y marca la curva normal. Utiliza la tabla de la curva normal para identificar la puntuación Z que separa una proporción de 0.025 del área en cada cola.

b) Determina las puntuaciones CESD (es decir, puntuaciones X) que definen los extremos fuera del área media del 95%. Interpreta tu respuesta en lenguaje común.

6A-8. Una prueba estandarizada de uso común es la Escala de Inteligencia de Stanford-Binet (Hollinger y Baldwin 1990). Esta prueba está diseñada para evaluar el desarrollo cognitivo en niños. La prueba está "normalizada" y diseñada tal que las puntuaciones caen en una distribución normal respecto a una media de 100 puntos en la escala con una desviación estándar de 13 puntos en la escala. Considera que X = puntuación de inteligencia Stanford-Binet.

a) Supongamos que una persona que se sometió a la prueba, Jack McGinley, tuvo una puntuación de 130 en la escala de inteligencia de Stanford-Binet. ¿Cuál es el rango percentilar de Jack en la escala? Es decir, ¿la puntuación de Jack fue igual a o mayor a qué porcentaje de quienes tomaron la prueba?

b) Bob Harris obtuvo 89 en la escala de inteligencia de Stanford-Binet. ¿Cuál es su rango percentilar?

Conjunto de problemas 6B

6B-1. Calcula las probabilidades siguientes al lanzar un dado de juego:

a) *p* [5]

b) *p* [5 luego 6]

c) *p* [1 o 3 o 6]

6B-2. Supongamos que se tiene una caja con 100 canicas color rojo, 50 color azul y 50 color verde. Calcula las probabilidades de sacar al azar de la caja lo siguiente:

a) *p* [roja luego roja luego verde] sin reemplazamiento

b) *p* [roja luego roja luego verde] con reemplazamiento

c) *p* [azul luego roja luego verde] sólo con reemplazamiento de las rojas

6B-3. Para el lanzamiento de una moneda (C = cara, X = cruz), calcula lo siguiente:

a) *p* [*C*]

b) *p* [*C* luego *X*]

c) *p* [*X* luego *X* luego *X*]

6B-4. Calcula las probabilidades siguientes al sacar cartas de un mazo estándar con 52 cartas.

a) *p* [as]

b) *p* [rey o sota]

c) *p* [reina o espadas]

d) *p* [as luego as o rey luego rey] sin reemplazamiento

6B-5. Gardner, Van Dyne y Pierce (2004) examinaron los efectos motivacionales del nivel de pago sobre el desempeño de los empleados. Supongamos que tienes los estadísticos descriptivos siguientes para las puntuaciones del desempeño en el trabajo, donde una puntuación alta indica buen trabajo. Utiliza estos datos para responder las preguntas siguientes. La distribución es normal. Traza la curva normal y marca todas las áreas objetivo. Considera que Y = puntuación de desempeño en el trabajo.

$$\bar{Y} = 78 \text{ puntos}$$
$$s_Y = 8 \text{ puntos}$$
$$n = 473 \text{ empleados}$$

a) ¿Qué proporción de empleados obtuvieron una puntuación *arriba* de 90 en el desempeño en el trabajo?

b) ¿Cuántos empleados obtuvieron una puntuación *entre* 88 y 98?

c) ¿Qué proporción de las puntuaciones cayó *entre* 70 y 90?

d) Determina la puntuación debajo de la cual cayeron 95% de las puntuaciones.

e) Si un aspirante tiene que estar al menos en el rango percentilar de 95% para obtener un bono en su pago, ¿qué puntuación debe obtener (respuesta breve)?

6B-6. Para una población grande de personas sin hogar, Wong y Piliavin (2001) examinaron factores de estrés, recursos y agotamiento psicológico empleando la Escala de Depresión del Centro de Estudios Epidemiológicos (CESD), un cuestionario de evaluación comunitario. Entre las personas sin hogar, la puntuación media CESD es 23.5 con una desviación estándar de 7.5 y la distribución es normal. Como trabajador en el área de admisiones en un refugio para personas sin hogar, tú deseas realizar tu investigación. Cuando llegan nuevos clientes, tú aplicas el CESD. Responde las preguntas siguientes. Traza una curva normal con cada solución. Considera que X = puntuación CESD.

a) Cualquier cliente con puntuación 16 o mayor se enviará a ver un doctor. ¿Cuál es la probabilidad de que tu próximo cliente será enviado a ver un doctor?

b) ¿Cuál es la probabilidad de que tu próximo cliente tenga una puntuación de 10 o menor?

c) Si las personas sin hogar con puntuación en el 15% superior en el CESD serán enviadas a los servicios de prevención de suicidios, ¿qué puntuación hace calificar a un cliente para estos servicios?

6B-7. He y Sutton (2004) evaluaron un método propuesto para dar seguimiento a la prevalencia de obesidad infantil en Canadá. Supongamos que tú tienes una muestra de niños a la que quieres dar seguimiento con base en el índice de masa corporal (IMC), que mide el peso relativo a la altura. El IMC medio de tu muestra es 26.8 kilogramos (kg) por metros al cuadrado y la desviación estándar de la muestra es 1.9 kg por metros al cuadrado. Los niños que se encuentran en el 5% superior del IMC se enviarán a un programa especial. Considera que X = puntuación IMC.

a) Traza y marca la curva normal. Utiliza la tabla de la curva normal para identificar la puntuación Z que aísla 0.05 del área en el extremo positivo de la curva.

b) Determina qué puntuación de IMC (es decir, puntuación X) aísla el 5% superior de los niños. Interpreta tu respuesta en lenguaje común.

6B-8. El examen de la American College Testing (ACT) es el examen de admisión universitario de uso más común. Este examen evalúa las habilidades de los estudiantes en cuatro áreas de habilidades. Para este ejercicio, supongamos que esta distribución de puntuaciones ACT es normal con una media de 22 puntos y una desviación estándar de 4 puntos. Traza la curva normal. Considera que X = puntuación ACT.

a) Jennifer O'Neal tuvo una puntuación de 31 en el examen. ¿Cuál es su rango percentilar? Es decir, ¿la puntuación de Jennifer es igual a o mayor a qué porcentaje de quienes presentaron la prueba?

b) Carl Lane tuvo una puntuación de 19 en la prueba ACT. ¿Cuál es su rango percentilar?

Conjunto de problemas 6C

6C-1. Calcula las probabilidades siguientes para el lanzamiento de un dado de juego.

a) *p* [2]

b) *p* [1 o 5]

c) *p* [2 luego 5 luego 6]

6C-2. Supongamos que se tiene un barril grande y bien mezclado con 70 pelotas color negro, 200 pelotas color azul y 120 pelotas color rojo. Calcula las probabilidades de sacar al azar del barril lo siguiente:

a) *p* [negra luego negra luego roja] sin reemplazamiento

b) *p* [azul luego roja luego negra] con reemplazamiento

c) *p* [azul luego roja luego azul] sólo con reemplazamiento de las rojas

6C-3. Para el lanzamiento de una moneda (C = cara, X = cruz), calcula lo siguiente.

a) *p* [*C*]

b) *p* [*C* luego *C*]

c) *p* [*C* luego *X* luego *X*]

6C-4. Calcula las probabilidades siguientes para el evento de sacar cartas de un mazo estándar de 52 cartas.

a) *p* [4]

b) *p* [9 o sota]

c) *p* [reina o tréboles]

d) *p* [sota luego sota u 8 luego 8] sin reemplazamiento

6C-5. La prueba de aptitud escolástica (SAT) es un examen de admisión universitario. Aunque Freedle (2003) argumenta que la prueba está sesgada cultural y estadísticamente contra algunos grupos minoritarios, aún se emplea mucho. Supón que los estadísticos descriptivos siguientes corresponden a una muestra de estudiantes que tomaron el SAT. La distribución está formada normalmente. Traza la curva normal y marca todas las áreas objetivo. Considera que Y = puntuación SAT.

$$\bar{Y} = 1\,100 \text{ SAT puntos}$$
$$s_Y = 100 \text{ SAT puntos}$$
$$n = 322\,763 \text{ estudiantes}$$

a) ¿Qué proporción de estudiantes tuvo una puntuación *arriba* de 1 300?

b) ¿Cuántos estudiantes tuvieron una puntuación *debajo* de 1 080?

c) ¿Qué proporción de las puntuaciones cayó *entre* 900 y 1 150?

d) Determina la puntuación debajo de la cual caen 85% de las puntuaciones.

e) Si un aspirante tenía que estar al menos en el rango percentilar de 85% para ingresar a un programa universitario, ¿qué puntuación necesita para lograrlo (respuesta breve)?

6C-6. Browning, Leventhal y Brooks-Gunn (2004) examinaron el impacto del contexto del vecindario y la raza en la iniciación de la actividad sexual entre adolescentes jóvenes. Supón que tiene una población de adolescentes jóvenes con una edad media de 13 años y una desviación estándar de 1 año. Las edades en esta población están normalmente distribuidas. Tú seleccionarás al azar individuos de esta población. Calcula las probabilidades siguientes y traza la curva normal para cada problema. Considera que X = edad.

a) *p* [de seleccionar al azar a alguien entre las edades de 12 y 14 años]

b) *p* [de seleccionar al azar a alguien "11.5 años o más joven", o "14.5 años o mayor"]

c) El 10% más joven de los adolescentes se seleccionarán para entrevistas de seguimiento. ¿A qué edad y debajo de ésta calificará un sujeto para esta entrevista?

6C-7. Egan y Kadushin (2004) examinaron la satisfacción en el trabajo entre trabajadores sociales de salud en casa dentro del entorno del sistema de pago interino de Medicare. Tú empleas una escala similar de satisfacción en el trabajo (variando de 0 a 36) para otra muestra de trabajadores de salud en casa. La media de tu muestra es 23.6 puntos en la escala de satisfacción en el trabajo y la desviación estándar es 2.3. La distribución es normal. Tú aislarás un total de 5% de las puntuaciones, el 2.5% que es extremadamente bajo y 2.5% que es extremadamente alto. Considera que X = escala de satisfacción en el trabajo.

a) Traza y marca la curva normal. Utiliza la tabla de la curva normal para identificar las puntuaciones Z que aíslan la proporción de las personas que respondieron con puntuaciones en el 2.5% superior e inferior.

b) Determina las dos puntuaciones de la escala de satisfacción (es decir, puntuaciones X) para el 2.5% superior y 2.5% inferior de trabajadores. Interpreta tu respuesta en lenguaje común.

6C-8. La prueba de aptitud escolástica (SAT) es un examen de admisión universitaria que evalúa habilidades verbales y numéricas. Ram (2004) utilizó el SAT para estudiar los efectos de gastos escolares sobre el logro estudiantil en Estados Unidos. Las puntuaciones SAT están normalmente distribuidas con una media de 1 100 puntos y una desviación estándar de 150. Considera que X = puntuación SAT.

a) Brian Fitzsimmons obtuvo una puntuación de 1 420 en la prueba. ¿Cuál es su rango percentilar? Es decir, ¿la puntuación de Brian es igual a o mayor a qué porcentaje de los que presentaron la prueba?

b) Marie Larelle obtuvo una puntuación de 974 en la prueba ACT. ¿Cuál es su rango percentilar?

Conjunto de problemas 6D

6D-1. Calcula las probabilidades siguientes para el lanzamiento de un dado de juego.

a) *p* [4]

b) *p* [2 luego 4]

c) *p* [1 o 4 o 5]

6D-2. Supón que tienes un recipiente grande con bebidas gaseosas bien mezclado: 100 son regulares; 50, sin cafeína, y 70, dietéticas. Calcula las probabilidades de seleccionar al azar lo siguiente del recipiente. No debes abrirlas. Simplemente estás tratando de impresionar a tus amigos.

a) *p* [regular luego dietéticas luego sin cafeína] sin reemplazamiento

b) *p* [dietéticas luego dietéticas luego regular] con reemplazamiento

c) *p* [regular luego dietéticas luego regular] sólo con reemplazamiento de las bebidas dietéticas

6D-3. Para el lanzamiento de una moneda (C = cara, X = cruz), calcula lo siguiente.

a) *p* [*C*]

b) *p* [*X* luego *C*]

c) *p* [*C* luego *X* luego *X*]

6D-4. Calcula las probabilidades siguientes al sacar cartas de un mazo estándar con 52 cartas.

a) *p* [rey]

b) *p* [reina o as]

c) *p* [reina o espadas]

d) *p* [cinco luego cinco o reina luego reina] sin reemplazamiento

6D-5. Klem y Connell (2004) examinaron dimensiones de apoyos a maestros y su relación con el logro estudiantil según la medición mediante la puntuación en el examen semestral. Utiliza estos estadísticos para responder las preguntas siguientes. La distribución es normal. Traza la curva normal y marca todas las áreas objetivo. Considera que X = puntuación en el examen semestral.

$$\bar{X} = 81 \text{ puntos}$$
$$s_X = 4 \text{ puntos}$$
$$n = 212 \text{ estudiantes}$$

a) ¿Qué proporción de estudiantes obtuvieron una puntuación *arriba* de 90?

b) ¿Qué número de estudiantes obtuvieron una puntuación *entre* 86 y 91?

c) ¿Qué proporción de las puntuaciones cayó *entre* 77 y 87?

d) Determina la puntuación en el examen debajo de la cual cayeron 95% de las puntuaciones.

e) Si un estudiante tenía que estar al menos en el 95o. rango percentilar en este grupo de estudiantes, ¿qué puntuación necesita obtener (respuesta breve)?

6D-6. Greiner y otros (2004) examinaron el impacto de los factores de estrés sobre la hipertensión entre operadores de tránsito urbano. Tú deseas validar este estudio empleando una muestra de 200 trabajadores de tránsito en otra ciudad. Tú administras una escala de estrés y determinas una calificación media de 18.5 puntos en la escala de estrés con una desviación estándar de 4.5. La distribución es normal. Considera que X = puntuación en la escala de estrés.

a) Cualquier trabajador que obtenga una puntuación de 14 o mayor en la escala de estrés es elegible para participar en una parte extensiva de la entrevista de su estudio. Dados sus estadísticos, ¿a cuántos de los 200 trabajadores en tu muestra les harás una entrevista profunda?

b) ¿Cuál es la probabilidad de que el siguiente trabajador examinado tenga una puntuación de 10 o menor?

c) A los trabajadores con puntuaciones en el 15% superior en la escala de estrés se les harán pruebas cardiacas extensivas. ¿Qué puntuación en la escala de estrés los califica como participantes para estos servicios?

6D-7. Riebschleger (2004) estudió las experiencias de niños viviendo con un miembro familiar que habían sido diagnosticados con una discapacidad psiquiátrica. Supón que tú realizas un estudio similar. La edad media de tu muestra de niños es 11.7 años y la desviación estándar es 1.4 años. La distribución es normal. Tú aislarás el 5% de los niños con mayor edad y realizarás una entrevista de seguimiento a ellos. Identifica este rango de edad. Considera que X = edad.

a) Traza y marca la curva normal. Utiliza la tabla de la curva normal para identificar la puntuación Z que aísla el 0.05 superior del área bajo la curva.

b) Determina la edad (es decir, puntuación X) que corta el 5% superior de los niños. En otras palabras, ¿arriba de qué edad cae el 5% de los niños en esta muestra? Interpreta tu respuesta en lenguaje común.

6D-8. El Graduate Record Exam (GRE) es una prueba de admisión para estudios de posgrado en Estados Unidos. Goldberg y Pedulla (2002) evaluaron diferencias de desempeño para varios métodos de tomar el examen. La prueba está "normalizada" (es decir, diseñada para ajustarse a la curva normal). Tú tienes una muestra de 897 estudiantes que presentaron el examen de manera electrónica. La media de la muestra es 1 000 puntos y la desviación estándar es 140 puntos. Utiliza la curva normal para responder las preguntas siguientes. Considera que X = puntuación GRE.

a) Una estudiante, Caroline van Nostren, obtuvo una puntuación de 1 340 en la prueba. ¿Cuál es su rango percentilar? Es decir, ¿la puntuación de Caroline es igual a o mayor a qué porcentaje de los que presentaron el examen?

b) John Riley obtuvo 843 puntos. ¿Cuál es su rango percentilar?

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 6

Para las clases donde se utilizan computadoras, visita el sitio web *The Statistical Imagination* en www.mhhe.com/richey2 y abre Computer Application Exercises del capítulo 6. Estos ejercicios se enfocan en el uso de distribuciones de frecuencias de puntuaciones como distribuciones de probabilidad y en el uso de puntuaciones Z para calcular probabilidades con variables de intervalo/razón normalmente distribuidas. Además, el apéndice D de este texto proporciona un repaso breve de secuencias de comandos *SPSS* para procedimientos estudiados en este capítulo.

CAPÍTULO

7

Uso de la teoría de la probabilidad para producir distribuciones muestrales

RESUMEN DEL CAPÍTULO

Introducción: estimación de parámetros

Como un repaso breve, una población es un conjunto grande de personas respecto de quienes deseamos obtener información. En general, para ahorrar tiempo y dinero, obtenemos muestras en lugar de observar un grupo tan grande. Los estadísticos de una muestra proporcionan estimaciones de los parámetros del la población total. Supongamos que nuestra población de interés es de 16 000 estudiantes de una universidad. De este campus universitario seleccionamos una muestra aleatoria de 200 estudiantes. Nos interesan parámetros como los siguientes: ¿Cuál es el promedio general de calificaciones? ¿Qué porcentaje de los estudiantes apoya la apertura de la biblioteca del campus las 24 horas del día? ¿Cuál es la edad media? Sin embargo, nuestro interés no se centran en las medias o proporciones de los 200 estudiantes en la muestra. Buscamos respuestas para todo el cuerpo estudiantil de 16 000 estudiantes. La muestra sólo es una herramienta para obtener información acerca de los parámetros de esta población total del campus.

No obstante, los estadísticos calculados en una muestra sólo proporcionan estimaciones. ¿Cómo podemos reconocer y tratar este hecho? ¿Existen herramientas que nos permitan refinar estas estimaciones enunciándolas con un grado de confianza y niveles conocidos del error de muestreo? Las respuestas a estas preguntas estriban en una buena comprensión de lo que denominamos distribuciones muestrales.

Estimaciones puntuales

Supongamos que designamos X como el PG y para una muestra de 200 estudiantes determinaos una media de 2.46 "puntos PG" (es decir, créditos obtenidos por hora-crédito tomada). ¿Nos asegura esto que la media de la población también es 2.46? Por supuesto que no. Existe un error de muestreo que debemos considerar. El **error de muestreo** es *la diferencia entre el valor calculado de un estadístico de la muestra y el valor real de un parámetro de la población*, que por lo general se desconoce. Por definición, los estadísticos de la muestra sólo son estimaciones de parámetros. Si reportamos esta cifra individual de 2.46 puntos PG, estamos proporcionando lo que se llama una **estimación puntual**, que es *un estadístico proporcionado sin indicar un rango de error*. Esto no es mucho mejor que una buena suposición. ¿Por qué? Porque si tomamos una segunda, una tercera y una cuarta muestra, es probable que obtengamos medias calculadas ligeramente diferentes para cada una. En otras palabras, hay una variabilidad en los resultados estadísticos de una muestra a otra.

> **Error de muestreo** Diferencia entre el valor calculado de un estadístico de la muestra y el valor real de un parámetro de la población.
>
> **Estimación puntual** Estadístico proporcionado sin indicar un rango de error.

Predicción del error de muestreo

Fue el descubrimiento de la variabilidad de la muestra, el reconocimiento que cada estadístico de la muestra difiere ligeramente del siguiente, lo que es el fundamento de la comprensión básica del error de muestreo. Al igual que los estadísticos antiguos que lanzaban dados repetidamente, los estadísticos posteriores aprendieron sobre el error de muestreo mediante el **muestreo repetido**, *tomando una muestra y calculando sus estadísticos y luego tomando una segunda muestra, una tercera, una cuarta, y así sucesivamente*. Estos estadísticos "cuenta frijoles" aprendieron dos hechos naturales importantes acerca del muestreo repetido de una población. Primero, los resultados calculados serán distintos de una muestra a otra. Segundo, los cálculos realizados en una muestra, un grupo que es menor que toda la población, sólo son estimaciones. Es decir, los estadísticos de una muestra estarán ligeramente errados de los valores reales de los parámetros de la población.

> **Muestreo repetido** Tomar una muestra y calcular sus estadísticos y luego tomar una segunda muestra, una tercera, una cuarta, y así sucesivamente. El muestreo repetido revela la naturaleza del error de muestreo.

El muestreo aleatorio repetido y la variabilidad consiguiente en los resultados estadísticos se ilustran en la figura 7-1, la cual presenta una población de niños cuyas edades varían de cero (bebés menores de un año) a nueve años. Para los estadísticos muestrales, cálculos realizados en datos de la muestra, por lo común empleamos símbolos de letras como $\bar{X}$ y s_X (con los que ya estamos familiarizados). Cuando es factible, empleamos letras griegas para los parámetros de la población. En específico, utilizamos los símbolos siguientes para representar parámetros de la población para variables de intervalo/razón:

Para la variable de intervalo/razón X,

μ_X = media de una población (pronunciada *mu* subíndice X)
σ_X = desviación estándar de una población (se pronuncia *sigma* subíndice X)

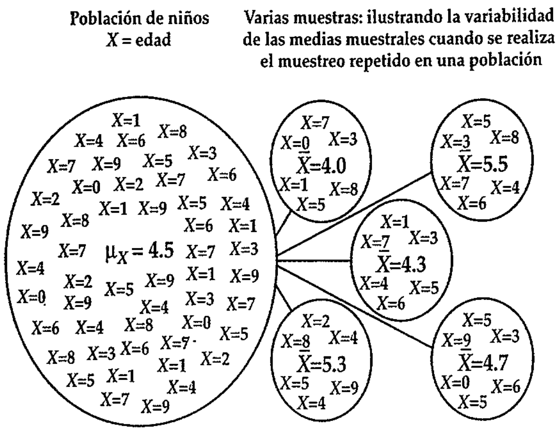


FIGURA 7-1
Variabilidad muestral con muestreo repetido: X = edades de niños, cero a nueve años

> **Símbolos matemáticos de uso común para distinguir poblaciones y muestras**
>
> Para estadísticos muestrales: letras latinas
> Para parámetros poblacionales: letras griegas

En la figura 7-1, X = la edad, y la edad media *en la población* de niños es $\mu_X = 4.5$ años. Observa que las medias muestrales representadas $\bar{X}$ en los círculos más pequeños varían respecto de esta media poblacional de 4.5 años. Cada media muestral es ligeramente mayor o menor que 4.5, reflejando la variabilidad muestral causada por la ocurrencia natural del error en el muestreo aleatorio. Por ejemplo, la muestra arriba a la derecha de la figura anterior donde $\bar{X} = 5.5$ años tiene un error de 1.0 en el parámetro real de 4.5; es decir, el error de muestreo se calcula como sigue:

$$\bar{X} - \mu_X = 5.5 - 4.5 = 1.0 \text{ año}$$

Por tanto, una buena manera de demostrar que los estadísticos muestrales no son valores exactos de parámetros poblaciones es muestrear repetidamente. Si no estás convencido de esto, toma un par de muestras aleatorias de la población (es decir, del círculo grande) de las edades de los niños en la figura 7-1. Son muchas las posibilidades de que obtenga medias muestrales ligeramente diferentes. Cada muestra es una parte muy pequeña de la población mayor, y cada una está compuesta de un conjunto distinto de seis niños. En una muestra —sólo por casualidad— pueden aparecer más niños mayores que menores, lo que resulta en una media muestral ligeramente mayor que 4.5 años. En una segunda muestra —sólo por casualidad— pueden aparecer más niños menores que mayores, lo que resulta en una media muestral menor. El muestreo repetido genera resultados estadísticos variados.

Hace más de 200 años los teóricos de la probabilidad reconocieron algunas "malas noticias": un estadístico de una muestra única sólo es una estimación de un parámetro poblacional. Pero mediante el muestreo repetido —muchas horas empleadas tomando una muestra tras otra— estos teóricos descubrieron algunas buenas noticias: el error de muestreo tiene patrones y es sistemático, y por lo tanto es predecible.

El primer punto predecible encontrado a partir del muestreo repetido fue que las medias muestrales resultantes eran similares en valor y tendían a agruparse alrededor de un valor particular. Los teóricos de la probabilidad sospecharon que este valor central era el valor real del parámetro de la población, la media de la población en sí misma (μ_X). Empleando modelos similares al de la figura 7-1, compararon resultados muestrales con parámetros conocidos y determinaron que, en efecto, una distribución de estadísticos muestrales se centra en el parámetro poblacional real. Esto tenía sentido. Si la edad promedio de una población de niños es 4.5 años, la media calculada en una muestra verdaderamente aleatoria debería estar cercana a este valor. Segundo, los teóricos de la probabilidad descubrieron que la variabilidad en el muestreo se podía predecir de forma matemática, a partir de curvas de probabilidad. Tomaron los resultados de muestras repetidas y trazaron histogramas. La mayoría de sus medias calculadas caían muy cerca del valor del parámetro de la población, y conforme uno se alejaba de este parámetro en cualquier dirección, había cada vez menos resultados. En otras palabras, descubrieron que los resultados estadísticos ocurren de acuerdo con las curvas de probabilidad como la curva normal. Por último, cuando compararon muestras de tamaños diferentes, estos teóricos determinaron que entre mayor era el tamaño de la muestra, menor era el rango de los errores en muestras repetidas.

Distribuciones muestrales

Cuando se trazan en histogramas las distribuciones de los estadísticos de muestras tomadas repetidamente, obtenemos una imagen representativa de la previsibilidad del error en el muestreo. A esa distribución la denominamos distribución muestral. *A partir del muestreo repetido*, una **distribución muestral** es *una descripción matemática de todos los resultados posibles y la probabilidad de cada uno.*

> **Distribución muestral** A partir del muestreo repetido, una descripción matemática de todos los resultados posibles del muestreo y la probabilidad de cada uno.

Distribuciones muestrales para variables de intervalo/razón

Para ilustrar las singularidades de una distribución muestral de medias, veamos qué sucede si muestreamos repetidamente a partir de una población *con una media conocida*. Supongamos que determinamos de los registros de titulados de una universidad que la edad media *de la población* de todos los médicos practicantes titulados en Estados Unidos es 48 años. Como estos datos son para toda la población, esta media es un parámetro conocido, simbolizado como μ_X, donde X = edad del médico. Supongamos también que la desviación estándar de esta población es seis años, simbolizada como σ_X. La distribución de frecuencias de las *puntuaciones brutas* de esta *población* de edades se presenta en la figura 7-2. Observa que esta distribución no es una curva normal con forma de campana. Es importante tener en cuenta que en el eje horizontal de la figura 7-2 trazamos puntuaciones brutas (puntuaciones X), las edades reales de los médicos.

FIGURA 7-2
Distribución de frecuencias de las puntuaciones brutas de edades para toda la población de médicos practicantes activos en Estados Unidos (datos ficticios)

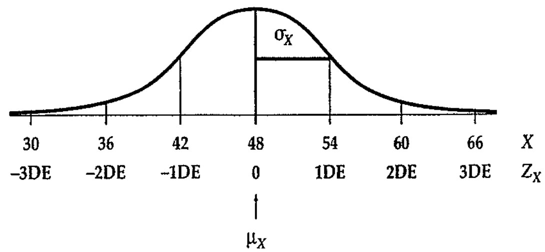
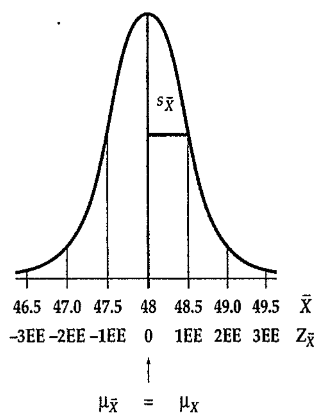


FIGURA 7-3
Distribución muestral de la edad media de médicos en Estados Unidos



Ahora enfoquemos nuestra atención lejos de la distribución de puntuaciones brutas de la figura 7-2 y consideremos una distribución muestral de medias. Para determinar todos los resultados posibles de la muestra, debemos imaginar que tomamos repetidamente muestras de esta población. Digamos que tomamos 10 000 muestras de 144 médicos. Para cada muestra, calculamos la edad media de la muestra, $\bar{X}$. Un pensamiento momentáneo nos convencerá que la mayoría de las medias de la muestra se calcularán alrededor de 48 años. Pero debido al error de muestreo, *no* nos sorprendería si cada media muestral estuviera ligeramente errada, digamos 47.9 años o 48.2 años.

Ahora imaginemos que trazamos los valores de estas 10 000 medias muestrales en un histograma. Es decir, para cada muestra tratamos la edad media calculada como una observación individual. Por tanto, trazamos las $\bar{X}$ en lugar de las X en el eje horizontal. ¿Adivina qué forma tendrá este histograma? En efecto, la de una distribución normal. Resulta que, cuando el tamaño muestral, n, es mayor que 121 casos, *una distribución muestral de medias tiene forma normal.* Esta normalidad se ilustra en la curva con forma de campana de la figura 7-3.

La mayoría de las medias muestrales cae en o alrededor de 48 años. Conforme nos alejamos de 48 años en cualquier dirección, la curva tiene pendiente hacia abajo, indicando cada vez menos resultados. Además, si sumamos los valores de todas las 10 000 medias muestrales y dividimos el resultado entre 10 000, esta *media de medias muestrales* es 48 años, la edad media de los médicos en la población. *La media de una distribución muestral de medias* se simboliza como $\mu_{\bar{X}}$ y siempre será igual a la media poblacional (μ_X). Es más, al igual que con cualquier curva normal, la desviación estándar es la distancia hasta el punto de inflexión de la curva. En resumen, la curva en la figura 7-3 es una distribución muestral de medias (X barra, no X). Esto describe de manera matemática *todos los resultados muestrales posibles y la probabilidad de cada resultado.* (El motivo por el cual las puntuaciones estandarizadas aparecen en unidades de error estándar "EE" en lugar de desviaciones estándar "DE" lo explicaremos en breve.)

¿Qué nos indica esta distribución muestral? Primero, cualquier distribución muestral (por definición) representa todos los resultados posibles del muestreo. La figura 7-3 revela todos los resultados estadísticos que ocurren si tomamos repetidamente muestras de tamaño 144 de la población de médicos y calculamos la media de cada muestra. Segundo, puesto que una distribución muestral de medias adopta una forma normal cuando $n > 121$ casos, se utiliza la tabla de la curva normal para calcular la probabilidad de ocurrencia de cualquier resultado muestral. Por tanto, con esta distribución normal, casi 68% de las observaciones caen dentro de 1 desviación estándar a ambos lados de la media. En específico, casi 68% de las veces nuestras medias muestrales ($\bar{X}$) estarán entre 47.5 y 48.5 años; aproximadamente 95% de las veces, entre 47.0 y 49.0 años; y casi 100% de las veces, entre 46.5 y 49.5 años. En resumen, esta distribución proporciona una descripción de todos los resultados posibles del muestreo cuando $n = 144$ y la media de la población es 48 años. Una distribución muestral nos indica con qué frecuencia un estadístico muestral tiene la probabilidad de fallar respecto al valor real del parámetro de la población y por cuánto.

El error estándar

La desviación estándar de una distribución muestral tiene un nombre especial, error estándar, debido a que es una medida de los errores predecibles en el muestreo. **El error estándar** es *la desviación estándar de una distribución muestral.* Observa que para la distribución muestral de la figura 7-3, las unidades de medida están marcadas como EE (errores estándar) en lugar de DE (desviaciones estándar). El error estándar mide la dispersión del error de muestreo que ocurre cuando se muestrea una población repetidamente.

> **Error estándar** Desviación estándar de una distribución muestral. El error estándar mide la dispersión del error de muestreo que ocurre cuando se muestrea repetidamente una población.

Los matemáticos han determinado que una buena estimación del *error estándar de una distribución muestral de medias* es la desviación estándar de la muestra divida entre la raíz cuadrada del tamaño de la muestra (n). Observa el símbolo que indica que este error estándar estimado se basa en datos muestrales.

Cálculo del error estándar de una distribución muestral de medias cuando se desconoce σ_X (para una variable de intervalo/razón)

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}}$$

donde

$s_{\bar{X}}$ = error estándar de medias estimado para la variable X

s_X = desviación estándar de una muestra

n = tamaño de la muestra

Para la distribución muestral de las edades medias de médicos representadas en la figura 7-3, el error estándar es medio año:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{6}{\sqrt{144}} = \frac{6}{12} = .5 \text{ años}$$

Regresemos a la figura 7-3 y observemos el eje horizontal. Podemos considerar este eje como una regla para medir el error de muestreo. El error estándar nos indica dónde hacer las marcas en la regla.

Ley de los números grandes

Un análisis con detenimiento de la fórmula para el error estándar de una distribución muestral de medias revela un punto importante acerca de la dispersión del error de muestreo. *Entre mayor sea el tamaño de la muestra, menor será el error estándar.* Este principio, denominado **ley de los números grandes,** tiene sentido (Sheynin 1970). Una muestra grande funciona mejor que una pequeña. Esto es aparente en la composición de la fórmula para el error estándar. Cuando n se reemplaza con valores cada vez mayores, esto incrementa el tamaño del denominador y reduce el tamaño del cociente. Para muestras de médicos, reemplace n con valores cada vez mayores y observe cómo el error estándar calculado disminuye.

Ley de los números grandes Entre mayor sea el tamaño de la muestra, menor será el error estándar (es decir, entre menor sea el rango de error en la distribución muestral).

Teorema del límite central

Para ayudarnos a distinguir entre puntuaciones brutas y distribuciones muestrales, observa las similitudes y diferencias entre las curvas de las figuras 7-2 y 7-3. En ambas distribuciones las medias son iguales a la media de la población, μ_X. Sin embargo, en la distribución muestral de la figura 7-3, en el eje horizontal trazamos *medias muestrales* (X-barra, no puntuaciones X brutas), y la desviación estándar es el error estándar. Además, el error estándar en la figura 7-3 tiene una fórmula y un símbolo diferentes a los de la desviación estándar en la figura 7-2. (Repase el capítulo 5.)

También observa las diferencias en el tamaño de la dispersión de estas distribuciones. En la distribución de las puntuaciones brutas de la figura 7-2, las edades reales de los médicos varían de casi 30 a 66 años. En la distribución muestral de la figura 7-3, *las medias muestrales calculadas* tienen un rango mucho más angosto, desde una media muestral de casi 46 años hasta una de casi 50 años. Esto se destaca en la figura 7-4, en la cual se superponen las distribuciones. Tiene sentido intuitivo que la distribución muestral tenga un rango menor. Como un estadístico de la tendencia central, la media para cualquier muestra es probable que se calcule dentro de un área central de una distribución de puntuaciones brutas. Un promedio es probable que sea un promedio. Por tanto, cuando trazamos un número grande de medias, éstas se agrupan *estrechamente* respecto a un valor central que sucede que es la media de la población, μ_X. Matemáticamente, la dispersión más estrecha de una distribución muestral es aparente en la fórmula para el error estándar. La desviación estándar está en el numerador y por tanto se divide en partes menores. El error estándar siempre se calculará con un valor menor que el de la desviación estándar.

La tendencia es fuerte para que una distribución muestral tenga un rango pequeño de valores dentro del centro de la distribución de las puntuaciones brutas. De hecho, esta tendencia es tan fuerte que ocurre aún cuando la propia distribución de las puntuaciones brutas no esté

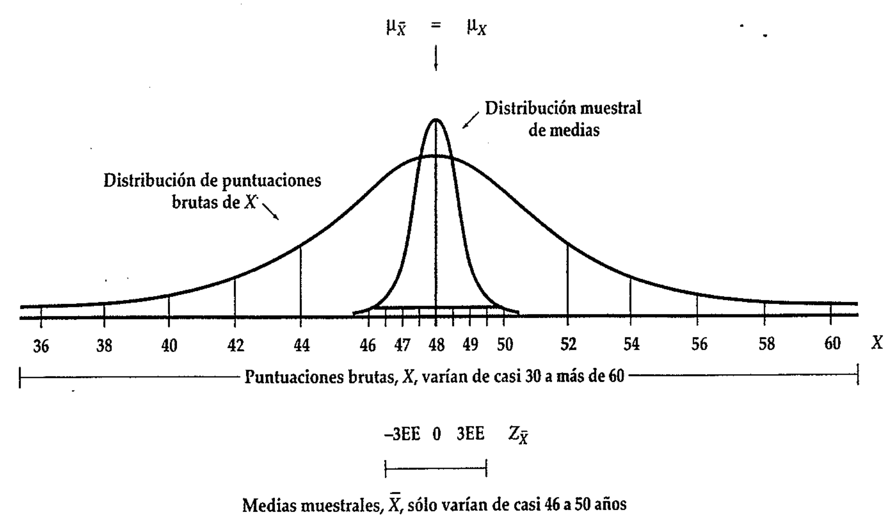


FIGURA 7-4

Comparación de las dispersiones de una distribución de puntuaciones brutas con su distribución muestral: edades de médicos en Estados Unidos normalmente formada. *Sin importar la forma de una distribución de puntuaciones brutas, su distribución muestral será normal cuando el tamaño de la muestra, n, sea mayor que 121 casos y se centrará en la media de la población real.* Entre los matemáticos a este descubrimiento se le refiere como **teorema del límite central,** que fue concebido primero por Pierre Laplace (Fischer 2000). (Como explicaremos en el capítulo 10, aun cuando el tamaño de la muestra sea menor que 121, la forma de la distribución se aproximará a la curva normal.)

Teorema del límite central Sin importar la forma de una puntuación bruta de una variable de intervalo/razón, su distribución muestral será normal cuando el tamaño de la muestra, *n*, sea mayor que 121 casos y se centrará en la media de la población verdadera.

Para ilustrar el teorema del límite central, analicemos la tabla de números aleatorios (tabla estadística A del apéndice B). Cada número individual en la tabla se denomina dígito. Estos "dígitos aleatorios individuales" varían de 0 a 9. Cada dígito se puede considerar como una puntuación de la variable *X*, como las edades de niños de nueve años y menores en la figura 7-1. ¿Cómo se generó esta tabla de números aleatorios? Imaginemos que la elaboramos a mano. Iniciamos escribiendo cada uno de los dígitos (0 a 9) en tiras de papel separadas. Colocamos estas tiras en un sombrero y las revolvemos, aleatorizamos, tal que cada una tenga una posibilidad igual de ser seleccionada. Sacamos una tira, registramos el resultado, la reemplazamos en el sombrero, y repetimos el proceso un número infinito de veces. Como cada dígito tiene una posibilidad igual de selección en cada evento, a la larga registraremos tantos ceros como unos, doces, y así sucesivamente, hasta el nueve. La distribución de frecuencias de estas puntuaciones brutas (es decir, puntuaciones *X*) aparecerían como en la figura 7-5. Esta distribución tiene forma "rectangular"; cada columna del histograma tiene una altura igual a las otras. Éste es el caso debido a que todos los dígitos tienen una posibilidad de selección igual y por lo tanto cada uno ocurrirá con la misma frecuencia. Por tanto, esta distribución ni siquiera es remotamente normal, no tiene "colas".

FIGURA 7-5
Distribución de frecuencias de las puntuaciones brutas de un conjunto infinito de dígitos individuales aleatorios

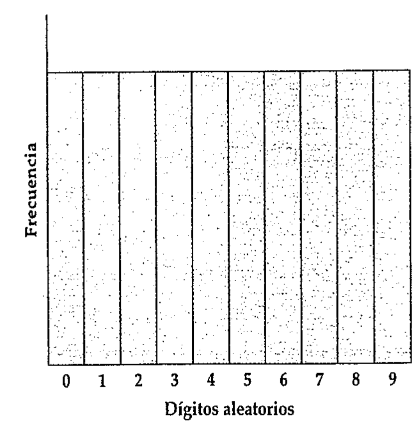
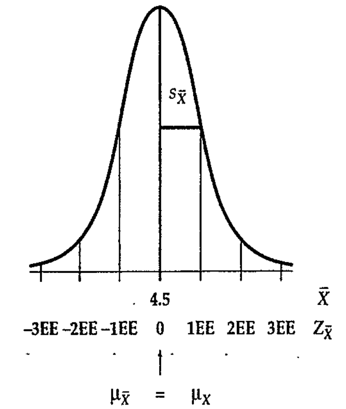


FIGURA 7-6
Distribución muestral de medias de muestras de dígitos aleatorios individuales



Ahora imaginemos que estamos tomando *muestras* de esta tabla de números aleatorios con $n > 121$. Seleccionamos una muestra. Calculamos la media muestral ($\bar{X}$), y repetimos este proceso muchas veces. Demostrando el teorema del límite central, la forma de un histograma de la distribución muestral resultante es normal aunque la distribución de las puntuaciones brutas no tenga ni remotamente forma normal. Esto se ilustra en la figura 7-6.

Para tener un sentido de este fenómeno resulta una buena idea considerar la tabla de números aleatorios (tabla estadística A del apéndice B) como si fuera una población de dígitos aleatorios y tomar muestras repetidas de ella. (Vea los ejercicios al final del capítulo.)

Por fortuna, para determinar la forma de una distribución muestral, no tenemos que tomar muestras repetidamente. Los teóricos de la probabilidad, los contadores de frijoles del pasado, emplearon su tiempo haciendo esto y proporcionaron las fórmulas del error estándar. Sólo necesitamos tomar una muestra y emplear su desviación estándar para estimar el error estándar de una distribución muestral.

Distribuciones muestrales para variables nominales

Con variables normales contamos la frecuencia de categorías y calculamos proporciones. Con frecuencia nuestro objetivo es una categoría de "éxito" particular y deseamos determinar su parámetro en la población. Para obtener una distribución muestral de la proporción de éxito, hacemos la pregunta: ¿qué sucede si tomamos una muestra, calculamos la proporción para esta categoría, y luego repetimos estos procedimientos una y otra vez? ¿Qué forma toma la distribución de los resultados de la muestra? Como resultado, una distribución muestral de proporciones adopta la forma de una distribución normal cuando *n* es lo suficientemente grande, como se analiza a continuación.

Para ilustrar una distribución muestral de proporciones, analicemos la proporción de médicos que son mujeres entre todos los médicos en práctica activa en Estados Unidos. En específico, los datos del reporte *Características y distribución de médicos en Estados Unidos, 1996-1997* de la American Medical Association (1997) revelan que en 1995 había 720 325 médicos en práctica activa, de los cuales 149 404 eran mujeres. Es decir, el parámetro poblacional conocido de la proporción de mujeres era .2074 (casi 21%). Al pensarlo un momento deberá convencernos que si tomamos, por ejemplo, 10 000 muestras aleatorias de, digamos, 300 médicos, la proporción de mujeres en cada muestra deberá ser de alrededor de

.21. Sin embargo, debido al error de muestreo esperado, para una muestra dada no nos sorprenderíamos si la proporción calculada fuera ligeramente distinta —digamos, .20 o .22—, debido al error de muestreo. Tendríamos que la mayoría de las 10 000 proporciones muestrales caen alrededor de .21, y conforme nos movemos en cualquier dirección, obtendremos menos y menos resultados. Un histograma de estos resultados es normal en forma. Además, la media de *estas 10 000 proporciones muestrales* es .21; es decir, si sumamos todas las 10 000 proporciones en conjunto y las dividimos entre 10 000, el resultado será .21, la proporción de médicos mujeres en la población. La desviación estándar de esta distribución muestral de proporciones se denomina error estándar de proporciones.

Al igual que con las distribuciones muestrales de medias, las distribuciones muestrales de proporciones tienen sus propios símbolos. Como nuestro interés es en la proporción de médicos mujeres (una variable nominal), denotaremos *la proporción de mujeres entre médicos en Estados Unidos* como P (es decir, éxito) y *la proporción de hombres entre los médicos de Estados Unidos* como Q (es decir, fracaso). Por tanto,

$P = p$ [de médicos en Estados Unidos que son mujeres]
$Q = p$ [de médicos en Estados Unidos que son hombres]

Utilizaremos los símbolos con subíndice siguientes para representar parámetros de la población. Para evitar anidar la letra p en los símbolos, emplearemos el subíndice u para *universo*, otro término para población. Por tanto,

$P_u = p$ [de la población de médicos en Estados Unidos que son mujeres] $= .21$
$Q_u = p$ [de la población de médicos en Estados Unidos que son hombres] $= 1 - P_u$
$= 1 - .21 = .79$

Utilizaremos un subíndice s para representar los estadísticos muestrales. Por tanto,

$P_s = p$ [de los médicos muestreados en Estados Unidos que son mujeres]
$Q_s = p$ [de los médicos muestreados en Estados Unidos que son hombres]

Queremos saber, con base en las proporciones relativas de médicos hombres y mujeres en la población, cuánto error se espera en el muestreo repetido. Al igual que con las medias, el tamaño del error está relacionado al tamaño de la muestra: entre mayor sea la muestra, menor será el rango de error. Para determinar el error de muestreo, calculamos la desviación estándar de esta distribución muestral —su error estándar— para un tamaño de la muestra de 300 médicos. Si los valores de P y Q se conocen, como es el caso con los médicos, el error estándar de proporciones se simboliza con sigma subíndice P subíndice s (σ_{P_s}) con la fórmula siguiente:

Cálculo del error estándar de una distribución muestral de proporciones cuando se conocen P_u y Q_u (para una variable nominal)

$$\sigma_{P_s} = \sqrt{\frac{P_u Q_u}{n}}$$

donde

σ_{P_s} = error estándar de proporciones para una variable nominal con $P = p$ [de la categoría de éxito]

$P_u = p$ [de la categoría de éxito en la población]

$Q_u = p$ [de la categoría de fracaso en la población]

n = tamaño de la muestra

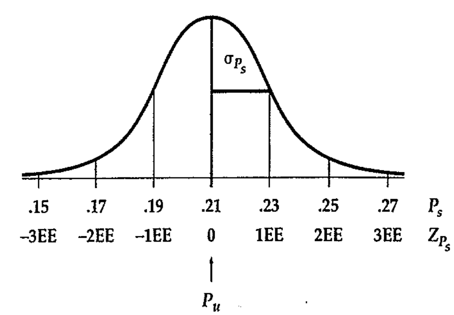
Para el error estándar de la proporción de médicos mujeres con muestras de tamaño 300,

$$\sigma_{P_s} = \sqrt{\frac{P_u Q_u}{n}} = \sqrt{\frac{(.21)(.79)}{300}} = .02$$

La figura 7-7 muestra la distribución de estas muestras, con P_s, la proporción de mujeres en una muestra, trazada a lo largo del eje horizontal. Al igual que con cualquier curva normal, ésta nos dice qué esperar si tomamos muestras repetidamente de la población de médicos: casi 68 de las veces nuestros resultados de la muestra, P_s, estarán entre .19 y .23 (es decir, $-21 \pm \sigma_{P_s}$; casi 95% de las veces, P_s estará entre .17 y .25 (es decir, $.12 \pm 2\sigma_{P_s}$; y sólo un porcentaje muy pequeño de las veces cualquier P_s será menor que .15 o mayor que .27 (es decir, $.21 \pm 3\sigma_{P_s}$). Entonces, esta distribución muestral nos indica con qué frecuencia esperar que un resultado muestral no acierte el parámetro real (P_u) y por cuánto. Además, como esta distribución de resultados muestrales es normal, podemos describir de forma matemática su media y un error estándar (como en la figura 7-7). Por último, al calcular puntuaciones Z, podemos dividir la curva y calcular la probabilidad de la ocurrencia de cualquier resultado muestral individual o cualquier rango de resultados de muestreo. Por tanto, esta distribución normal derivada de manera matemática es una distribución muestral, una descripción de todos los resultados muestrales posibles y la probabilidad de cada uno.

Esta ilustración es inusual en que se conocen los parámetros poblacionales reales de las proporciones de médicos mujeres y hombres. En muchos casos de investigación la muestra

FIGURA 7-7
Distribución muestral de la proporción de médicos mujeres en Estados Unidos en 1995



es la única fuente de datos disponibles y se emplean estadísticos muestrales para estimar el error estándar de las proporciones. Por tanto, si P_u y Q_u no se conocen, σ_{P_s} se estima como sigue, empleando proporciones muestrales P_s y Q_s.

Cálculo del error estándar de una distribución muestral de proporciones cuando no se conocen P_u y Q_u (para categorías de una variable nominal)

$$s_{P_S} = \sqrt{\frac{P_S Q_S}{n}}$$

donde

s_{P_S} = error estándar de proporciones estimado para una variable nominal, con $P = p$ [de la categoría de éxito]

$P_s = p$ [de la categoría de éxito en la muestra]

$Q_s = p$ [de la categoría de fracaso en la muestra]

n = tamaño de la muestra

Reglas respecto a una distribución muestral de proporciones

El muestreo repetido y el cálculo de las proporciones (P_s) revelan las características siguientes respecto de una distribución muestral de proporciones.

1. La media de una distribución muestral de proporciones es igual al parámetro de la población, P_u.
2. Una distribución muestral de proporciones tendrá forma normal cuando el parámetro menor (P_u o Q_u) multiplicado por n resulte ≥ 5. (Si no se conoce el parámetro, las estimaciones de la muestra, P_s y Q_s, se aplican al tomar esta decisión). En el ejemplo anterior P_u (la proporción de médicos mujeres = .21) es menor que Q_u (la proporción de hombres = 0.79). Por tanto, establecemos p_{menor} en .21. Para verificar si una distribución normal es apropiada, debemos determinar que p_{menor} por n sea mayor que o igual a 5:

$$(p_{menor})(n) = (.21)(300) = 63 \qquad 63 \geq 5$$

Por tanto, la distribución muestral de la proporción de médicos mujeres donde $n = 300$ es normal. En general, para cualquier valor de P_u y Q_u (cualquiera que sea menor), el tamaño mínimo de la muestra necesario para suponer una distribución normal se determina empleando la fórmula siguiente:

Cálculo del tamaño mínimo de la muestra para suponer que la distribución muestral de proporciones tiene forma normal

$$\text{Mínimo } n = \frac{5}{p_{menor}}$$

donde

n mínima = tamaño mínimo de la muestra necesario para suponer normalidad

p_{menor} = el menor de P_u o Q_u (si se conocen estos parámetros) o el menor de P_s o Q_s (si no se conocen P_u y Q_u)

Por ejemplo, cuando $P_s = .1$, la n mínima = 50; cuando $P_s = .3$, la n mínima = 17; y cuando $P_s = .5$, la n mínima = 10. Estos ejemplos ilustran que cuando P_s es moderada (es decir, alrededor de .5) en lugar de extrema, una muestra menor será suficiente para asegurar que la distribución muestral es normal. Por último, cuando n mínima es menor que 5, la distribución muestral apropiada se denomina distribución binomial, la cual se analiza en el capítulo 13.

Por último, observa que en nuestra ilustración de la distribución muestral de médicos mujeres, empleamos la variable dicotómica *género del médico.* Recuerda que una variable dicotómica sólo tiene dos categorías, como mujer u hombre. Si tienes una variable nominal con más de dos categorías, la categoría éxito siempre se define como P. Sin embargo, la categoría fracaso Q, es la proporción de todas las otras categorías combinadas y esta proporción es igual a $1 - P$. Por ejemplo, si $P = p$ [de estudiantes en la State University en su último año], entonces $Q = 1 - P = p$ [de estudiantes en la State University en segundo, tercer y primer año, y otras clasificaciones]. En resumen, la distribución muestral para una variable nominal se enfoca en la proporción de éxito. Las categorías restantes de "fracasos" se consideran como "no P"; por tanto, $Q = 1 - P$.

El conteo de frijoles como una forma para desarrollar la imaginación estadística

Las distribuciones muestrales son representaciones de lo que sucede si tomamos repetidamente muestras de una población y calculamos estadísticos de una variable. Mediante este muestreo repetido descubrimos todos los resultados estadísticos posibles y la probabilidad de cada uno. El error estándar, cuyo tamaño depende del tamaño de la muestra, nos proporciona un dispositivo de medición para calcular estas probabilidades.

Las distribuciones muestrales son una característica esencial del análisis estadístico debido a que son útiles como curvas de probabilidad. Al descubrir estas tendencias predecibles en los resultados muestrales, los teóricos de la probabilidad empezaron a formular preguntas: ¿podemos emplear el conocimiento acerca de la previsibilidad de los resultados muestrales de una manera tal que nos permita evitar tener que "muestrear" toda la población para determinar la media real de una variable? Por ejemplo, ¿podemos evitar el gasto de entrevistar a todos los 16 000 estudiantes en un campus universitario? ¿Podemos examinar una sola muestra y simplemente estimar los errores? Por ejemplo, si una muestra de un solo estudiante tiene un promedio general de calificaciones de 2.44, ¿podemos usar nuestro conocimiento acerca de la previsibilidad de las muestras repetidas para inferir un rango de error, digamos, más o menos .1? ¿Existe una forma para calcular el grado de confianza que tenemos en la exactitud de un estadístico a partir de una sola muestra? Sabemos que un estadístico de la muestra es una estimación, pero no 100% correcta. Pero, ¿podemos decir que estamos 90, 95 o 99% seguros de nuestros resultados?

Quizá te des cuenta ahora de que una distribución muestral es una distribución de probabilidad. Para un tamaño dado de la muestra, una distribución muestral nos indica con qué frecuencia esperar cualquiera y todos los resultados muestrales cuando tomamos muestras aleatorias. Esto no es diferente del cálculo, digamos, de la probabilidad de obtener cara y luego cara al lanzar dos monedas, como lo hicimos en el capítulo anterior (figura 6-2). De hecho, algunas de las simples ilustraciones de distribuciones de probabilidad en ese capítulo son útiles como distribuciones muestrales. Y nuestro arduo trabajo en el capítulo 6, dividiendo áreas de la curva normal, lo hicimos para prepararnos para tratar la distribución muestral como una curva normal.

Producir distribuciones muestrales a mano, mediante el muestreo repetido, es importante para comprender verdaderamente el concepto. Esto es probablemente lo que los primeros estadísticos hicieron. Primero, se propusieron determinar qué sucede cuando una variable nominal se muestrea repetidamente. Para representar una categoría de una variable como el género, ellos pudieron haber sustituido una caja con frijoles para la población, haciendo que los frijoles color blanco representaran hombres y los frijoles color rojo representaran mujeres. Luego de manera repetitiva tomaron muestras de la caja con un tamaño de éstas, digamos, de 100 frijoles. Para cada muestra calcularon la proporción de mujeres (es decir, P_s, donde $P = p$ [de frijoles que son color rojo]). Trazaron estas proporciones muestrales en un histograma y descubrieron la curva normal, y la fórmula para el error estándar. Trataron varios tamaños de la muestra ($n = 90, 80, 70$, etc.) y determinaron que siempre y cuando el tamaño de la muestra fuera lo suficientemente grande, una distribución de muestreo de proporciones adoptaba la forma de campana de una curva normal.

Para variables de intervalo/razón, los primeros estadistas utilizaron tablas de números aleatorios, como la tabla estadística A del apéndice B, tratando los dígitos como si cada uno representara un caso para una variable como la edad. Después de tomar muestras repetidamente, calcular las medias y trazarlas en histogramas, descubrieron nuevamente el fenómeno natural de la normalidad. Por último, estos matemáticos "cuenta frijoles" desarrollaron fórmulas para el error estándar y descubrieron la ley de los números grandes y el teorema del límite central. Los estadistas modernos están agradecidos por los esfuerzos incansables de estos matemáticos y estadistas pioneros. En la actualidad sólo necesitamos tomar una muestra aleatoria y emplear fórmulas para estimar la forma de una distribución muestral.

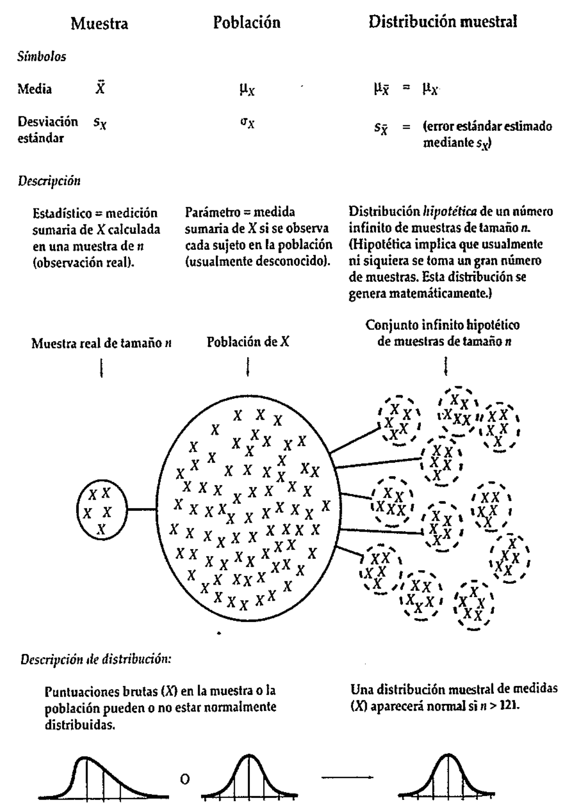
El teorema del límite central en esencia estipula que el muestreo aleatorio resulta en curvas normales: distribuciones simétricas que se agrupan en medio y se extienden hacia los lados. La normalidad en el muestreo aleatorio es un fenómeno natural que existió mucho antes que los estadísticos lo midieran y formularan, al igual que la gravedad existía mucho antes que Isaac Newton la midiera y explicara.

Una cosa es que hablemos respecto de cómo los primeros estadísticos aprendieron acerca de las distribuciones muestrales y otra es que experimentemos el proceso de generarlas para nosotros mismos. Los ejercicios y aplicaciones en computadora para este capítulo proporcionan formas simples para producir distribuciones muestrales a la manera antigua, contando frijoles realmente. *Contador de frijoles* es un término peyorativo que con frecuencia se aplica a burócratas tacaños que se aferran a los centavos, pero siendo un contador de frijoles literal no es una cosa mala para adquirir una visión de los procesos naturales detrás de un error muestral aleatorio. Si deseas que el resto de este curso se desarrolle de manera fluida, genera distribuciones muestrales suficientes para obtener un sentido de dominio sobre la idea. Vuélvete un contador de frijoles.

Distinción entre poblaciones, muestras y distribuciones muestrales

En este punto es útil repasar algunos de los símbolos y fórmulas que hemos empleado hasta aquí. Tenemos que utilizar mucho nuestra imaginación al tratar con los estadísticos, debido a que la única cosa que tocamos es la muestra. Mientras tanto, tenemos que imaginar cómo luce la población total. Además, debemos imaginar una distribución muestral y describir su forma de manera matemática y gráfica.

FIGURA 7-8
Distinción entre mediciones en muestras, poblaciones y distribuciones muestrales (para la variable *X* de intervalo/razón)



En la figura 7-8 se presenta un resumen de los símbolos empleados para estadísticos muestrales, parámetros de la población y estadísticos de la distribución muestral para medias. A partir de este punto será muy importante tener claras estas ideas y símbolos.

Insensatez y falacias estadísticas: tratar una estimación puntual como si fuera absolutamente cierta

Como hemos aprendido en este capítulo, ningún estadístico, una medida sumaria con base en datos de la muestra, es la última palabra al estimar un parámetro poblacional. El muestreo repetido indica que es probable que la siguiente muestra de la misma población resulte en un resultado estadístico ligeramente diferente. Aunque no es poco común, en especial en los medios de comunicación masiva, que las estimaciones puntuales se acepten rápidamente y luego se traten como si fueran absolutamente verdaderas.

Entre los encuestadores políticos puede haber ventajas al tener una estimación que no sea perfecta. Por ejemplo, una sola encuesta puede mostrar que poco más de 50% de los encuestados —una mayoría simple— apoya una propuesta de ley. Es probable que el encuestador y los partidarios de la legislación difundan este resultado y después hablen de él como si fuera un hecho absoluto. Una segunda encuesta puede indicar que considerablemente menos de 50% apoyan la propuesta. Si los políticos no tienen interés en la verdad absoluta sino que simplemente quieren apoyar sus posiciones, no es probable que realicen un debate de las conclusiones indefinidas hechas con estimaciones puntuales. Un estadístico entrenado o un ciudadano escéptico tal vez no caigan en esta artimaña, pero muchos ciudadanos quizá sí. Estos encuestadores no sólo nos engañan, insultan la inteligencia de las masas. En el capítulo 8 veremos que existe una manera para especificar el error en una estimación puntual y expresar el grado de confianza que tenemos, que es una estimación exacta de un parámetro de la población.

RESUMEN

1. Comprender las distribuciones muestrales es un elemento clave al realizar un análisis estadístico. Las distribuciones muestrales son curvas de probabilidad que nos permiten calcular el rango de error y el nivel de confianza que podemos evaluar al utilizar estadísticos de la muestra para sacar conclusiones acerca de parámetros de la población.
2. Los estadísticos de la muestra por lo general se denotan con letras latinas. Cuando es posible, los parámetros de la población se denotan con letras griegas.
3. Para determinar el valor de un parámetro de la población con certeza absoluta, se tendrían que hacer cálculos para cada sujeto en una población. En general, hacer esto es demasiado costoso y tardado, por lo que nos basamos en el muestreo aleatorio. El conocimiento de las distribuciones muestrales nos permite calcular la probabilidad del error de muestreo.
4. Una estimación puntual es un estadístico de la muestra proporcionado sin indicar un rango de error y se debe considerar con precaución. Esto se hace aparente mediante el muestreo repetido, que es el procedimiento de tomar una muestra y calcular sus estadísticos y después tomar una segunda muestra, una tercera, una cuarta, etcétera.

5. El muestreo repetido revela varios puntos acerca de la naturaleza del error de muestreo: *a*) Los estadísticos calculados diferirán ligeramente de una muestra a otra; *b*) un estadístico muestral dado estará ligeramente errado del valor real de su parámetro de la población; *c*) el error de muestreo tiene un patrón y es sistemático, y por lo tanto se puede predecir de forma matemática a partir de curvas de probabilidad denominadas distribuciones muestrales.
6. El teorema del límite central estipula que sin importar la forma de la distribución de las puntuaciones brutas de una variable de intervalo/razón, la distribución muestral de medias tendrá una forma normal cuando el tamaño de la muestra, *n*, sea mayor que 121. Esta distribución normal estará centrada en la media real de la población. Aún si una distribución de puntuaciones brutas en una población está sesgada, no obstante una distribución muestral de medias de esta población adoptará la forma de una curva normal.
7. El error estándar es la desviación estándar de una distribución muestral. Hay fórmulas que nos permiten estimar rápidamente un error estándar empleando estadísticos de una sola muestra.
8. La ley de los números grandes estipula que entre mayor sea el tamaño de la muestra, menor será el error estándar de una distribución muestral. Muestras grandes con sus errores estándar pequeños proporcionan estimaciones más precisas de parámetros de la población que las muestras pequeñas.
9. Un sentido de proporción acerca de la relación entre el tamaño de la muestra y el error muestral es aparente en la fórmula para el error estándar de la media, donde la desviación estándar de una muestra se divide entre la raíz cuadrada de *n*. Una *n* grande en el denominador de la ecuación produce un cociente pequeño, indicando que entre mayor sea la muestra, menor será el rango del error de muestreo.
10. Para una variable nominal/ordinal, las proporciones calculadas para una población muestreada repetidamente se calculan con valores similares que se agrupan alrededor del valor del parámetro de la población, la proporción de la población. Una distribución muestral de proporciones será normal cuando el parámetro menor (la probabilidad de éxito o fracaso en la población) multiplicado por el tamaño de la muestra es mayor que o igual a 5 (es decir, cuando $p_{menor} \geq 5$).

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 7 del material del texto disponibles en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2 incluyen sugerencias adicionales sobre la comprensión de las distribuciones muestrales.

FÓRMULAS PARA EL CAPÍTULO 7

Cálculo del error estándar de una distribución muestral de medias:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}}$$

Cálculo del error estándar de una distribución muestral de proporciones (para un variable nominal):

Cuando se conocen P_u y Q_u | Cuando no se conocen P_u y Q_u

$$\sigma_{P_s} = \sqrt{\frac{P_u Q_u}{n}} \qquad S_{P_s} = \sqrt{\frac{P_s Q_s}{n}}$$

Cálculo del tamaño mínimo necesario para suponer que una distribución muestral de proporciones tiene forma normal:

$$\text{Mínimo } n = \frac{5}{p_{menor}}$$

Por tanto, supongamos una distribución normal cuando $(p_{menor})(n) > 5$.

PREGUNTAS PARA EL CAPÍTULO 7

1. ¿Cuál es la diferencia entre un estadístico y un parámetro? En general ¿Cuál de los dos es la incógnita para la variable? Ilustra los símbolos que empleamos para los estadísticos y parámetros de variables de intervalo/razón y de variables nominales.
2. Define una distribución muestral. Distínguela de una distribución de puntuaciones brutas de una población.
3. ¿Cómo podemos demostrar que un estadístico calculado en una muestra única sólo proporciona *una estimación* de un parámetro?
4. Si trazamos un histograma para representar una distribución de *puntuaciones brutas*, digamos, la distribución de edades para una muestra de 200 estudiantes, ¿qué puntos se trazan en el eje horizontal del histograma?
5. Si trazamos un histograma para representar la distribución de *edades medias* para 1 000 *muestras* de 200 estudiantes, ¿qué puntos se trazan en el eje horizontal? ¿Cómo se denomina esta distribución?
6. Para una variable de intervalo/razón, proporciona los símbolos para la desviación estándar y el error estándar. ¿Qué mide cada estadístico de dispersión? ¿Cómo se relacionan matemáticamente?
7. Enuncia y explica la ley de los números grandes.
8. ¿En qué circunstancias una distribución muestral de proporciones se ajusta a la distribución normal?
9. Relaciona los símbolos de la izquierda con las definiciones de la derecha.
   - *a*) X _____ La desviación estándar para una *muestra* de puntuaciones brutas X
   - *b*) μ_X _____ La desviación estándar para una *población* de puntuaciones en bruto X
   - *c*) $\bar{X}$ _____ El símbolo para una variable de intervalo/razón y sus puntuaciones brutas
   - *d*) s_X _____ El error estándar de una *distribución muestral* de medias para la variable X, estimada a partir de la desviación estándar de la muestra
   - *e*) σ_X _____ La media de una *muestra* de puntuaciones brutas de la variable X
   - *f*) $s_{\bar{X}}$ _____ La media de una *población* de puntuaciones brutas de la variable X
10. ¿A variables de qué niveles de medición se aplican los símbolos en la pregunta 9?
11. Relaciona los símbolos de la izquierda con las definiciones de la derecha.
    - *a*) P _____ La proporción en la categoría "éxito" en una *población* de sujetos
    - *b*) Q _____ p [de la categoría de éxito]
    - *c*) P_u _____ La proporción en la categoría "éxito" en una *muestra* de sujetos
    - *d*) Q_u _____ El error estándar de una distribución de proporciones muestrales calculado con los valores conocidos de P_u y Q_u
    - *e*) P_s _____ El error estándar de una distribución de proporciones muestrales estimado a partir de los estadísticos de la muestra P_s y Q_s
    - *f*) σ_{P_s} _____ La proporción de la categoría "fracaso" en una *población* de sujetos
    - *g*) s_{P_s} _____ p [de fracaso], donde "fracaso" es la ausencia de una categoría o característica definida de una variable
12. ¿A variables de qué niveles de medición se aplican los símbolos en la pregunta 11?
13. Explica e ilustra con fórmulas por qué un error estándar de medias siempre será menor que la desviación estándar de esa variable.

EJERCICIOS PARA EL CAPÍTULO 7

Conjunto de problemas 7A

7A-1. Elder y otros (2004) analizaron la prevalencia y las características de las visitas a hospitales relacionadas con el alcohol entre una muestra de departamentos de emergencia y determinaron que las visitas frecuentes estaban relacionadas con agresión. Supongamos que tú reúnes una muestra aleatoria de 190 registros de los expedientes de adultos jóvenes acusados de agresión en los últimos seis meses y determinas que la edad media de los acusados es 20.8 años con una desviación estándar de 3.1 años. Utiliza estos estadísticos para calcular el error estándar de una distribución muestral de edades.

7A-2. Los siguientes datos ficticios son de una muestra de 437 empleados de una corporación transnacional. Completa la tabla siguiente calculando los errores estándar.

Variable	Desviación estándar o *P*	Error estándar
a) Salario mensual	$ 1 200	
b) Edad	4 años	
c) Proporción de mujeres	.39	
d) Años de servicio	2.7 años	
e) Proporción de trabajadores en divisiones de manufactura	.57	

7A-3. Elabora una distribución muestral de la proporción de caras en el lanzamiento repetido de 10 monedas. Lanza las monedas todas a la vez. Haz esto 100 veces. (Funciona mejor cuando se lanzan sobre una cama.)

a) En cada lanzamiento, cuenta el número de "caras" y registra el resultado en la tabla siguiente en la columna A con un tallo (es decir, /). A esto se le llama diagrama de tallo y hojas.

b) Después de los 100 lanzamientos, cuenta el número de tallos y registra la frecuencia de cada combinación de caras en la columna B (por ejemplo, /// // = siete veces).

c) Traza la distribución resultante de lanzamientos muestrales en papel milimétrico como un histograma de frecuencias.

d) Calcula la probabilidad de cada resultado muestral y regístralo en la columna C como "*p* del resultado".

e) Exhibe tu imaginación estadística describiendo en términos comunes por qué la distribución muestral adoptó esa forma.

	(A)	(B)	(C)
Núm. de caras	Diagrama de tallos de la frecuencia	Frecuencia de ocurrencia registrada	*p* del resultado
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

7A-4. Elabora una distribución muestral de medias para una muestra de tamaño 60. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar 60 dígitos aleatorios *individuales*; es decir, X = un dígito aleatorio individual. Calcula la media de esta muestra y regístrala con un lugar decimal. Haz esto 100 veces para obtener 100 medias muestrales ($\bar{X}$) de $n = 60$.

b) En papel milimétrico, traza un histograma de esta distribución muestral.

c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de dígitos aleatorios (μ_X) de una tabla de números aleatorios.

d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\bar{X}}$) de esta distribución muestral.

e) Utiliza tu conocimiento básico de la curva normal para aproximar con qué frecuencia ocurren los resultados muestrales dentro de 1, 2 y 3 errores estándar a ambos lados.

f) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica del muestreo de variables de intervalo/razón?

7A-5. El objetivo de este ejercicio es producir una distribución muestral de medias para un tamaño de la muestra de 7.

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar siete dígitos aleatorios *individuales*; es decir, X = un dígito aleatorio individual. Calcula la media de esta muestra y regístrala con un lugar decimal. Haz esto 120 veces para obtener 120 medias muestrales ($\bar{X}$) de $n = 7$.

b) En papel milimétrico, traza un histograma de esta distribución muestral.

c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de dígitos aleatorios (μ_X) de una tabla de números aleatorios.

d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\bar{X}}$) de esta distribución muestral.

e) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica de muestreo de variables de intervalo/razón?

f) Si se te asignara resolver el ejercicio 7A-4, compara los resultados de los ejercicios 7A-4 y 7A-5 y comenta con referencia a la ley de los números grandes.

7A-6. Elabora una distribución muestral de proporciones. En una caja (o recipiente grande), vacía una libra (453 gramos) de frijoles color rojo (secos, sin cocinar) y una libra de frijoles color blanco; mezcla bien (es decir, aleatoriza). Ésta es una población de frijoles. Con una cuchara extrae *dos* cucharadas para obtener una muestra de frijoles de esta población. Con dos lugares decimales, calcula P_s, la proporción de frijoles color rojo en la muestra, donde $P = p$ [de los frijoles que son color rojo]. Reemplaza los frijoles y mezcla bien. Haz esto 100 veces y traza las distribuciones muestrales resultantes de P como un histograma. Observa la distribución muestral y responde las preguntas siguientes *sin hacer cálculos*.

a) Proporciona una estimación de la proporción de frijoles color rojo en la población (es decir, el parámetro para toda la caja, P_u).

b) Proporciona una estimación del error estándar de esta distribución muestral (es decir, s_{P_s}).

c) Utiliza tu conocimiento básico de la curva normal para describir con qué frecuencia ocurren los resultados muestrales dentro de 1, 2 y 3 errores estándar a ambos lados.

d) A partir de esta experiencia de contar frijoles, ¿qué aprendiste acerca de la dinámica del muestreo de variables nominales?

7A-7. Harmelink y VanDenburgh (2003) analizaron la función de los contadores públicos certificados (CP) en la protección de las inversiones de sus clientes. Supongamos que tú quieres describir la distribución muestral de la proporción de personas satisfechas con los servicios recibidos de los contadores. Tú encuestas una muestra aleatoria de clientes de 40 contadores y determinas que 36 de los clientes están satisfechos.

a) ¿Sería apropiado emplear una distribución normal para describir la distribución muestral? ¿Por qué sí o por qué no?

b) Suponiendo que esta proporción muestral de los que están satisfechos es una buena estimación para la población de clientes de contadores, ¿qué tamaño de la muestra se necesita para emplear una curva normal como una descripción de esta distribución muestral?

Conjunto de problemas 7B

7B-1. Guo (2004) analiza la relación entre investigación y prácticas de marketing. Como un investigador de marketing en la Yeasty Feásty Bakery, tú realizas encuestas de compra de productos, implementando una variedad de sugerencias de Guo. En tu área de marketing determinas que el número medio de hogazas de pan consumidas por mes por hogar es 5.3 hogazas con una desviación estándar de 1.5 hogazas. Estos datos se basan en una muestra de 200 hogares. Utiliza estos estadísticos para calcular el error estándar de una distribución muestral de las hogazas medias consumidas por mes.

7B-2. Una compañía de marketing ha encuestado 395 hogares para evaluar los hábitos de ver televisión. Completa la tabla siguiente calculando errores estándar.

Variable	Desviación estándar o *P*	Error estándar
a) Edad del jefe del hogar	5 años	
b) Horas que la TV está encendida después de las 5:00 a.m.	1.5 horas	
c) Proporción de poseedores de la casa	.59	
d) Años de educación	1.9 años	
e) Proporción de hogares con más de dos televisores	.32	

7B-3. Elabora una distribución muestral de la proporción de caras en el lanzamiento repetido de ocho monedas. Toma las ocho monedas y lánzalas al mismo tiempo. Haz esto 100 veces. (Funciona mejor cuando se lanzan sobre una cama.)

a) En cada uno de los lanzamientos, cuenta el número de "caras" y registra el resultado en la tabla de la página 229 en la columna A con un tallo (es decir, /). Éste se denomina diagrama de tallo y hojas.

b) Después de los 100 lanzamientos, cuenta el número de tallos y registra la frecuencia de cada combinación de caras en la columna B (por ejemplo, *////* *//* = siete veces).

c) Traza la distribución resultante de lanzamientos muestrales en papel milimétrico como un histograma de frecuencias.

d) Calcula la probabilidad de cada resultado muestral y regístralo en la columna C como "*p* de resultado".

e) Exhibe tu imaginación estadística describiendo en términos comunes por qué la distribución muestral adoptó esa forma.

	(A)	(B)	(C)
Núm. de caras	Diagrama de tallos de la frecuencia	Frecuencia registrada de ocurrencias	*p* de resultado
0			
1			
2			
3			
4			
5			
6			
7			
8			

7B-4. Elabora una distribución muestral de medias para un tamaño de muestra de 50. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar 50 dígitos aleatorios *individuales*; es decir, X = un dígito aleatorio individual. Calcula la media de esta muestra y regístrala con un lugar decimal. Haz esto 100 veces para obtener 100 medias muestrales ($\overline{X}$) de $n = 50$.

b) En papel milimétrico, traza un histograma de esta distribución muestral.

c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de los dígitos aleatorios (μ_X) de una tabla de números aleatorios.

d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\overline{X}}$) de esta distribución muestral.

e) Utiliza tu conocimiento básico de la curva normal para describir con qué frecuencia los resultados del muestreo ocurren entre 1, 2 y 3 errores estándar para ambos lados.

f) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica del muestreo de variables de intervalo/razón?

7B-5. Elabora una distribución muestral de medias para una muestra de tamaño 6. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar seis dígitos aleatorios *individuales*; es decir, X = un dígito aleatorio individual. Calcula la media de esta muestra y regístrala con un lugar decimal. Haz esto 120 veces para obtener 120 medias muestrales ($\overline{X}$) de $n = 6$.

b) En papel milimétrico, traza un histograma de esta distribución muestral.

c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de los dígitos aleatorios (μ_X) de una tabla de números aleatorios.

d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\overline{X}}$) de esta distribución muestral.

e) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica del muestreo de variables de intervalo/razón?

f) Se te asignó el ejercicio 7B-4, compara los resultados de los ejercicios 7B-4 y 7B-5 y comenta con referencia a la ley de los números grandes.

7B-6. Elabora una distribución muestral de proporciones. En una caja (o recipiente grande), vacía una libra (453 gramos) de frijoles color rojo (secos, sin cocinar) y una libra de frijoles color blanco (secos, sin cocinar); mézclalos bien (aleatoriza). Esta es una población de frijoles. Con una cuchara saca *dos* cucharadas para obtener una muestra de frijoles de esta población. Con dos lugares decimales, calcula P_s, la proporción de frijoles color blanco en la muestra, donde $P = p$ [de los frijoles que son color blanco]. Reemplaza los frijoles y mezcla bien. Haz esto 100 veces y traza la distribución muestral resultante de P como un histograma. Observa la distribución muestral y responde las preguntas siguientes *sin hacer cálculos*.

a) Proporciona una estimación de la proporción de frijoles color blanco en la población (es decir, el parámetro para toda la caja, P_u).

b) Proporciona una estimación del error estándar de esta distribución muestral (es decir, s_{P_s}).

c) Utiliza tu conocimiento básico de la curva normal para describir con qué frecuencia los resultados muestrales ocurren dentro de 1, 2 y 3 errores estándar a ambos lados.

d) A partir de esta experiencia del conteo de frijoles, ¿qué aprendiste acerca de la dinámica del muestreo de variables nominales?

7B-7. Spoge y Trewin (2003) analizan la creciente popularidad de la declaración de impuestos sobre el ingreso mediante medios electrónicos, lo que proporciona a las personas un tiempo de respuesta más rápido de sus reembolsos de impuestos en caso de haber declarado un pago de impuestos mayor al debido. Tú debes describir la distribución muestral de la proporción de personas satisfechas con la prontitud de regreso del reembolso de impuestos del Internal Revenue Service. Tú obtienes una muestra de 20 personas que recibieron reembolso y determinas que 16 están satisfechas.

a) ¿Sería apropiado emplear una distribución normal para describir la distribución muestral? ¿Por qué sí o por qué no?

b) Suponiendo que esta proporción muestral de los contribuyentes satisfechos es una buena estimación para la población de los que recibieron reembolso, ¿qué tamaño de muestra se necesita para emplear la curva normal como una descripción de esta distribución muestral?

Conjunto de problemas 7C

7C-1. Wee y otros (2005) estimaron los gastos del cuidado de la salud asociados con la obesidad en Estados Unidos, examinando la influencia de la edad, la raza y el género. Supón que tú reúnes una muestra aleatoria de 275 registros médicos de un centro de salud de una comunidad pequeña del suroeste, calculando el índice de masa corporal (IMC) más reciente de cada paciente en kg/m^2. Tú descubres que el IMC para esta muestra es 30.4 kg/m^2, con una desviación estándar de 3.2 kg/m^2. Utiliza estos estadísticos para calcular el error estándar de una distribución muestral de índices de masa corporal.

7C-2. Supongamos que los datos siguientes son de una muestra aleatoria de 511 estudiantes de tiempo completo y parcial en una universidad urbana importante. Completa la tabla siguiente calculando los errores estándar.

Variable	Desviación estándar o *P*	Error estándar
a) Ayuda financiera mensual	$300	
b) Edad	5 años	
c) Proporción de hombres	.41	
d) Horas de trabajo semanal	3.7 horas	
e) Proporción de estudiantes actualmente empleados	.71	

7C-3. Elabora una distribución muestral de la proporción de colas en el lanzamiento repetido de 10 monedas. Toma 10 monedas de un peso y lánzalas al mismo tiempo. Haz esto 100 veces. (Es más fácil si las lanza sobre una cama.)

a) En cada lanzamiento, cuenta el número de "cruces" y registra el resultado en la tabla siguiente en la columna A con un tallo (es decir, /). Este se denomina diagrama de tallo y hojas.

b) Después de 100 lanzamientos, cuenta el número de tallos y registra la frecuencia de cada combinación de cruces en la columna B (por ejemplo, *### //* = siete veces).

c) Traza la distribución resultante de los lanzamientos muestrales en papel milimétrico como un histograma de frecuencias.

d) Calcula la probabilidad de cada resultado muestral y regístralo en la columna C como "*p* del resultado".

e) Exhibe tu imaginación estadística describiendo en términos comunes por qué la distribución muestral adoptó esa forma.

	(A)	(B)	(C)
Núm. de cruces	Diagrama de tallos y hojas de la frecuencia	Frecuencia de ocurrencia registrada	*p* del resultado
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

7C-4. Elabora una distribución muestral de medias para una muestra de tamaño 60. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar 60 dígitos aleatorios *individuales*; es decir, X = un dígito aleatorio simple. Calcula la media de esta muestra y regístrela con un lugar decimal. Haz esto 100 veces para obtener 100 medias muestrales ($\bar{X}$) de $n = 60$.

b) En papel milimétrico, traza un histograma de esta distribución muestral.

c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de los dígitos aleatorios (μ_X) de una tabla de números aleatorios.

d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\bar{X}}$) de esta distribución muestral.

e) Utiliza tu conocimiento básico de la curva normal para aproximar con qué frecuencia los resultados muestrales ocurren dentro de 1, 2 y 3 errores estándar a ambos lados.

f) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica del muestreo de variables de intervalo/razón?

7C-5. Elabora una distribución muestral de medias para una muestra de tamaño 5. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar cinco dígitos aleatorios *individuales*; es decir, X = un dígito aleatorio individual. Calcula la media de esta muestra y regístrala con un lugar decimal. Haz esto 120 veces para obtener 120 medias muestrales ($\bar{X}$) de $n = 5$.

b) En papel milimétrico, traza un histograma de esta distribución muestral.

c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de dígitos aleatorios (μ_X) de una tabla de números aleatorios.

d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\bar{X}}$) de esta distribución muestral.

e) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica del muestreo de variables de intervalo/razón?

f) Si se te asignara el ejercicio 7C-4, compara los resultados de los ejercicios 7C-4 y 7C-5 y comenta con referencia a la ley de los números grandes.

7C-6. Elabora una distribución muestral de proporciones. En una caja (o recipiente grande), vacía una libra (453 gramos) de frijoles color rojo (secos, sin cocinar) y una libra de frijoles color blanco; mezcla muy bien (es decir, aleatoriza). Ésta es una población de frijoles. Con una cuchara saca *dos* cucharadas para obtener una muestra de frijoles de esta población, calcula P_s, la proporción de frijoles color rojo en la muestra, donde $P = p$ [de los frijoles que son color rojo]. Reemplaza los frijoles y mezcla bien. Haz esto 100 veces y traza la distribución muestral resultante de P como un histograma. Observa la distribución muestral y responde las preguntas siguientes *sin hacer cálculos*.

a) Proporciona una estimación de la proporción de frijoles color rojo en la población (es decir, el parámetro para toda la caja, P_u).

b) Proporciona una estimación del error estándar de esta distribución muestral (es decir, s_{P_s}).

c) Utiliza tu conocimiento básico de la curva normal para describir con qué frecuencia ocurren los resultados muestrales dentro de 1, 2 y 3 errores estándar a ambos lados.

d) A partir de esta experiencia de conteo de frijoles, ¿qué aprendiste acerca de la dinámica del muestreo de variables nominales?

7C-7. Martin (2003) analiza el impacto de premios por votar (por ejemplo, disposición de fondos, beneficios presupuestales, etc.) sobre la participación política y los resultados de los votantes. Supongamos que tú tienes interés en describir la distribución muestral de la proporción de estudiantes que votaron en la elección presidencial de 2004, para lo cual obtienes una muestra aleatoria de 30 estudiantes en el campus de una universidad, los encuestas y descubres que 18 de ellos votaron.

a) ¿Sería apropiado emplear una distribución normal para describir la distribución muestral? ¿Por qué si o por qué no?

b) Suponiendo que esta proporción muestral de votantes estudiantes es una buena estimación para la población de estudiantes que decidieron votar, ¿cuál es la muestra menor necesaria para utilizar la curva normal como una descripción de esta distribución muestral?

Conjunto de problemas 7D

7D-1. En el Reino Unido, Prosser y Walley (2005) exploraron el alcance hasta el cual los médicos generales consideran los costos financieros cuando prescriben medicamentos controlados. Suponga que tú tienes interés en el mismo fenómeno. Tú descubres que el costo medio de 650 prescripciones de pacientes británicos es 32 libras esterlinas (£). Utiliza estos estadísticos para calcular el error estándar de una distribución muestral de costos de prescripción.

7D-2. Supongamos que los datos siguientes provienen de una muestra aleatoria de 298 madres de niños inscritos en un programa de campo educacional de un día. Completa la tabla siguiente calculando los errores estándar.

Variable	Desviación estándar o *P*	Error estándar
a) Edad del niño mayor	1.9 años	
b) Horas trabajadas por semana	1.5 horas	
c) Proporción rentando casa	.36	
d) Años de educación	2.8 años	
e) Proporción de madres con más de un niño	.44	

7D-3. Elabora una distribución muestral de la proporción de cruces en el lanzamiento repetido de 10 monedas. Lánzalas al mismo tiempo. Haz esto 100 veces. (Funciona mejor cuando se lanzan sobre una cama.)

a) En cada lanzamiento, cuenta el número de "cruces" y registra el resultado en la tabla siguiente en la columna A con un tallo (es decir, /). A éste se le denomina diagrama de tallo y hojas.

b) Después de 100 lanzamientos, cuenta los tallos y registra la frecuencia de cada combinación de cruces en la columna B (por ejemplo, ~~////~~ // = siete veces).
c) Traza la distribución resultante de los lanzamientos muestrales en papel milimétrico como un histograma de frecuencias.
d) Calcula la probabilidad de cada resultado muestral y regístralo en la columna C como "*p* de resultado".
e) Exhibe tu imaginación estadística describiendo en términos comunes por qué la distribución muestral adoptó esa forma.

	(A)	(B)	(C)
Núm. de cruces	Diagrama de tallos y hojas de la frecuencia	Frecuencia de ocurrencia registrada	*p* del resultado
0			
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

7D-4. Elabora una distribución muestral de medias para una muestra de tamaño 60. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)
a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar 60 números aleatorios *individuales*; es decir, X = un número aleatorio individual. Calcula la media de esta muestra y regístrela con un lugar decimal. Haz esto 100 veces para obtener 100 medias muestrales ($\bar{X}$) de $n = 60$.
b) En papel milimétrico, traza un histograma de esta distribución muestral.
c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de dígitos aleatorios (μ_X) de una tabla de números aleatorios.
d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\bar{X}}$) de esta distribución muestral.
e) Utiliza su conocimiento básico de la curva normal para describir con qué frecuencia ocurren los resultados muestrales dentro de 1, 2 y 3 errores estándar a ambos lados.
f) A partir de esta experiencia de muestreo repetido, ¿qué aprendiste acerca de la dinámica del muestreo repetido de variables de intervalo/razón?

7D-5. Elabora una distribución muestral de medias para una muestra de tamaño 8. (*Nota:* este problema es menos incómodo si se hace como proyecto de grupo en el salón de clases o en el laboratorio.)

a) Utilizando la tabla de números aleatorios (tabla estadística A del apéndice B), selecciona al azar ocho números aleatorios *individuales*; es decir, X = un número aleatorio individual. Calcula la media de esta muestra y regístrala con un lugar decimal. Haz esto 120 veces para obtener 120 medias muestrales ($\bar{X}$) de $n = 8$.
b) En papel milimétrico, traza un histograma de esta distribución muestral.
c) Mediante la observación (sin hacer cálculos) proporciona una estimación de la media de la población de dígitos aleatorios (μ_X) de una tabla de números aleatorios.
d) Mediante la observación (sin hacer cálculos) proporciona una estimación del error estándar ($s_{\bar{X}}$) de esta distribución muestral.
e) A partir de esta experiencia de muestreo repetido, ¿qué aprendió acerca de la dinámica del muestreo de variables de intervalo/razón?
f) Si se te asignara el ejercicio 7D-4, compara los resultados de los ejercicios 7D-4 y 7D-5 y comenta con referencia a la ley de los números grandes.

7D-6. Elabora una distribución muestral de proporciones. En una caja (o tazón grande), vacía una libra (453 gramos) de frijoles color rojo (secos, sin cocinar) y una libra de frijoles color blanco; mézclelos bien (es decir, aleatorice). Ésta es una población de frijoles. Con una cuchara saca *dos* cucharadas para obtener una muestra de frijoles de esta población. Con dos lugares decimales, calcula P_s, la proporción de frijoles color blanco en la muestra, donde $P = p$ [de los frijoles color blanco]. Reemplaza los frijoles y mezcle bien. Haz esto 100 veces y traza la distribución muestral resultante de P como un histograma. Observa la distribución muestral y responde las preguntas siguientes *sin hacer cálculos*.
a) Proporciona una estimación de la proporción de frijoles color blanco en la población (es decir, el parámetro para toda la caja, P_u).
b) Proporciona una estimación del error estándar de esta distribución muestral (es decir, s_{P_s}).
c) Utiliza tu conocimiento básico de la curva normal para describir con qué frecuencia los resultados muestrales ocurren dentro de 1, 2 y 3 errores estándar a ambos lados.
d) A partir de esta experiencia de conteo de frijoles, ¿qué aprendiste acerca de la dinámica del muestreo de variables nominales?

7D-7. La *adherencia médica* es un término empleado para referirse a cuando un paciente sigue las ordenes del doctor. Kim, Kaplowitz y Johnston (2004) examinaron los efectos de la empatía médica en la adherencia. A lo largo de líneas similares, tú describirás la distribución muestral de la proporción de pacientes que se apegan a las recomendaciones de sus doctores, para lo cual obtienes una muestra aleatoria de 57 pacientes de la oficina de un médico y determinas que 33 de ellos se apegaron a sus recomendaciones.
a) ¿Sería apropiado emplear una distribución normal para describir la distribución muestral? ¿Por qué si o por qué no?
b) Suponiendo que esta proporción muestral de pacientes que se apegaron a las recomendaciones es una buena estimación para la población de pacientes, ¿cuál es la muestra menor necesaria para utilizar una curva normal como una descripción de esta distribución muestral?

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 7

Si tu grupo de clase utiliza las aplicaciones computacionales opcionales de este texto, abre los ejercicios del capítulo 7 en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2. Estos ejercicios reforzarán tu comprensión de la relación entre el tamaño de la muestra y el error muestral. Además, el apéndice D de este texto proporciona un repaso breve de las secuencias de comandos *SPSS* para procedimientos estudiados en este capítulo.

CAPÍTULO 8

Estimación de parámetros empleando intervalos de confianza

RESUMEN DEL CAPÍTULO

Introducción

Anoche Kristi asistió a un concierto de rock en el estadio del campus universitario. Al regresar a su residencia universitaria, se dio cuenta que en el alboroto del evento había perdido un anillo de bajo costo pero con valor sentimental, heredado de su abuela. Ha pasado la mayor parte del día buscando en el campo de juego del estadio y empieza a perder las esperanzas de encontrarlo. Hasta que recuerda que su amiga Sarah tiene un detector de metales y la llama. Resulta que el detector de metales de Sarah no es muy preciso para señalar objetos metálicos, pero es muy confiable dentro de un margen de error. En específico, el detector emite una señal sonora cuando está dentro de 2 yardas de un objeto metálico. Sarah llega y camina por el campo con su detector de metales, y éste emite una señal sonora. Pero entonces ella dice que debe apresurarse para reunirse con otro amigo para cenar. Kristi pregunta, "¿Dónde está mi anillo"? Sarah le dice que busque alrededor de dos yardas en cualquier dirección de la línea central del campo, cerca de la marca más lejana. Kristi le pregunta: "¿Estás segura de que lo encontraré?" Sarah responde que su detector tiene una precisión de hasta 4 yardas 95% de las veces. Sarah está bastante segura —95%— por tanto le apuesta una cena a Kristi que encontrará el anillo.

Sarah no puede señalar la ubicación exacta del anillo, pero tiene un grado de confianza de que esté dentro del área de 4 yardas que describió. (Por cierto, Kristi encontró su anillo en el transcurso de algunos minutos e invitó a cenar a Sarah al día siguiente.)

La búsqueda de la ubicación de un objeto es similar a estimar el valor del parámetro poblacional empleando los estadísticos de una muestra. Como aprendiste en el capítulo 7, los estadísticos de una muestra son estimaciones —cálculos que sólo caen cerca del valor del parámetro poblacional real— al igual que el detector de metales de Sarah sólo llevó a Kristi cerca de la ubicación de su anillo. ¿Podemos hacer lo que hizo Sarah y no sólo señalar un punto sino también dar un margen confiable de área dentro de la cual buscar un parámetro? Por ejemplo, ¿podemos tomar una muestra de alumnos de décimo grado y estimar la estatura media de todos los alumnos de décimo grado hasta dentro de una pulgada —una estimación puntual más menos media pulgada (digamos, 67.5 pulgadas ± .5 pulgadas)? Nuestra conclusión sería que la estatura media está entre 67 y 68 pulgadas, no es exacta pero es cercana. Y ¿podemos, como Sarah, declarar que estábamos 95% seguros de esta estimación del intervalo? Como observamos en el capítulo 7, el muestreo repetido revela que cualquier estimación puntual individual sólo nos acerca cuando se estima un parámetro poblacional, al igual que Sarah señaló un punto en el campo. En este capítulo aprenderemos a decir con seguridad qué tan cercana se encuentra esta estimación puntual individual del parámetro real *dentro de un rango de error*, al igual que Sarah le indicó a Kristi buscar dentro de dos yardas en cualquier dirección del punto señalado por el detector. A ese tipo de estimación se le denomina intervalo de confianza.

Un **intervalo de confianza** es *un rango de valores posibles de un* **parámetro** *expresado con un grado específico de confianza.* Con los intervalos de confianza tomamos una estimación puntual y la acoplamos con el conocimiento acerca de las distribuciones muestrales. Proyectamos un rango conocido y calculable —o "intervalo"— de error respecto a la estimación puntual. Por ejemplo, donde X = promedio de calificaciones (PC), supongamos que tomamos una muestra de 300 estudiantes de la Crosstown University y calculamos una media muestral PC ($\overline{X}$) de 2.46. Nuestro conocimiento del muestreo repetido y de las distribuciones muestrales nos indica que este estadístico muestral deberá estar cerca del parámetro poblacional real. ¿Qué tan cerca? En el capítulo 7 vimos que una distribución muestral de medias, producida muestreando repetidamente una población, adopta la forma de una curva normal cuando $n > 121$. Con una muestra de 300 estudiantes, podemos reportar los resultados del muestreo repetido de una manera distribucional y decir, por ejemplo, que 95% de las muestras caen dentro de casi 2 errores estándar del parámetro real. Podemos interpretar este porcentaje de una forma probabilística y afirmar que *si sólo tomamos una muestra*, hay una probabilidad de 95% que esta media muestral única cae dentro de casi 2 errores estándar del parámetro, cualquiera que éste sea. Este error predecible, el producto de aproximadamente 2 veces el error estándar, se denomina término del error de un intervalo de confianza de 95% para esta situación cuando n es mayor que 121. Este término del error nos proporciona una interpretación probabilística de un cálculo de una muestra única.

Al calcular intervalos de confianza, *no* muestreamos, trazamos y calculamos áreas repetidamente bajo una curva de distribución muestral. En lugar de eso, sólo trazamos una

> **Intervalo de confianza** Rango de valores posibles de un parámetro expresado con un grado de confianza específico.

muestra y calculamos una estimación puntual como la media. Luego calculamos un error estándar y lo multiplicamos por una puntuación Z elegida para un nivel de confianza deseado. El resultado es un rango de error con base en el conocimiento acerca de la previsibilidad del error a partir del muestreo repetido. Después sumamos y restamos esta cantidad de la estimación puntual para obtener un intervalo dentro del cual es probable que se encuentre el parámetro. Este término del error es una cantidad "más o menos algún error" (al igual que Sarah recomendó buscar dentro de un par de yardas a cualquier lado de la línea de la yarda 50). Por ejemplo, si calculamos el intervalo de confianza de 95% del PC medio de los estudiantes de la Crosstown University, nuestra respuesta puede adoptar la forma de un intervalo de valores, digamos, 2.16 a 2.76 puntos del PC, la media muestral de 2.46 (una estimación puntual) más o menos .30 (un término del error). El resultado es una estimación del intervalo de la media real del PC (μ_X), un rango de valores del PC en el cual es probable que se encuentren las medias reales del campus. Aunque no decimos que sabemos el valor exacto de la CP media de todo el cuerpo estudiantil, estamos 95% seguros de que este parámetro está entre 2.16 y 2.76. El valor calculado de 2.16 determina el límite de confianza inferior (LCI), el valor menor que consideramos μ_X podría tener. De manera similar, 2.76 determina el límite de confianza superior (LCS), el valor mayor que consideramos μ_X podría tener. Reconocemos que el PC medio de la población podría ser tan bajo como 2.16 o tan alto como 2.76 o en algún punto intermedio. Es decir, μ_X podría ser 2.16, 2.17, 2.18, 2.28, 2.34 o cualquier valor hasta 2.76. No insistimos que hemos encontrado el valor exacto más de lo que Sarah insistió que encontraría el punto exacto donde se encontraba el anillo de Kristi. Pero igual que Sarah, podemos apostar con una confianza de 95% que el intervalo calculado contiene al valor poblacional real. Entonces el objetivo de calcular un intervalo de confianza, es estimar un parámetro poblacional dentro de un rango o "intervalo" específico de valores.

Los intervalos de confianza se emplean con frecuencia en estudios de exploración. Recuerde del capítulo 1 que los estudios de exploración buscan información acerca de fenómenos nuevos para los cuales se conoce muy poco que formular una teoría es imposible. Los intervalos de confianza hacen la primera pregunta básica: ¿cuál es el valor de un parámetro desconocido? Si en tu curso se utiliza software estadístico *SPSS*, observarás que los intervalos de confianza se calculan bajo el menú "Explore".

Calcular un intervalo de confianza es como lanzar una red hacia un estanque en donde sólo hay un pez. La ubicación del pez en un momento dado representa el parámetro desconocido. ¿Está a 10 pies de la orilla, a 20 pies o a 30 pies, etc.? Tenemos una oportunidad para lanzar la red y queremos sentirnos 95% confiados de atrapar al pez. Una estimación puntual de la ubicación proporciona alguna información aproximada, indicándonos en qué parte del estanque lanzar la red, digamos, cerca de un tocón por la orilla. Calcular el intervalo de confianza nos indica qué tan lejos lanzar la red. Nuestro nivel de confianza estipulado nos dice nuestra tasa de éxito, con qué frecuencia atraparemos el pez si empleamos una red con cierto ancho: el ancho del intervalo de confianza calculado. El **nivel de confianza** es *un grado de confianza calculado que un procedimiento estadístico realizado con datos muestrales producirá un resultado correcto para la población muestreada.*

> **Nivel de confianza** Grado de confianza calculado que un procedimiento estadístico realizado con datos muestrales producirá un resultado correcto para la población muestreada.

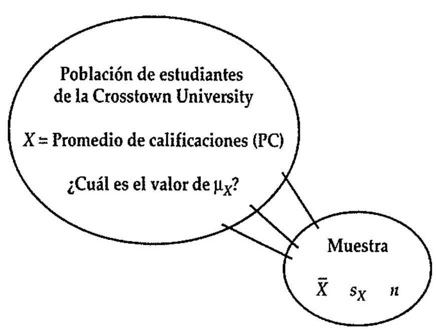
Intervalo de confianza de una media poblacional

Para cualquier variable de intervalo/razón, como el PC, nos proponemos estimar la media de una población. La pregunta que queremos responder es: ¿Cuál es el valor de μ_X? Los estadísticos muestrales son las herramientas que utilizamos para obtener esta estimación. Esto se representa en la figura 8-1.

Supongamos, por ejemplo, que estamos estudiando la estructura salarial de una planta industrial que emplea varios miles de ensambladores de computadoras pero no tenemos acceso a todos los expedientes de la compañía. Obtenemos una muestra aleatoria de 129 expedientes del personal con datos sobre los salarios por hora, una variable de razón X. Nuestro objetivo es emplear estos datos muestrales para hacer generalizaciones acerca de toda la población de ensambladores de computadoras. Así, calculamos un intervalo de confianza para el salario medio, μ_X, de todos los ensambladores. Nuestra pregunta de investigación es: dentro de un margen específico de cantidades en dólares, ¿cuál es el parámetro μ_X, el salario por hora medio de *la población* de ensambladores de computadoras? ¿Está entre, digamos, \$9 y \$10, o entre \$14 y \$15, o dónde? Con un intervalo de confianza de 95%, estaremos 95% confiados que el salario medio está dentro del margen de cantidades en dólares que calculamos.

Al confiar en una muestra, sabemos que hay un error en nuestra conclusión debido a que sabemos acerca del error de muestreo. De hecho, la única forma de estar completamente seguros o 100% seguros, es eliminar cualquier error de muestreo reuniendo datos sobre toda la población y calcular el parámetro correcto μ_X. Esto es demasiado costoso y tardado. Por tanto, nos conformamos con utilizar una muestra, sabiendo que tendremos algún grado de error en nuestra conclusión. Por fortuna, la cantidad de este error esperado es conocida. El **nivel del error esperado** es *la diferencia entre el nivel de confianza declarado y la "confianza perfecta" de 100%*. En otras palabras, si estamos 95% seguros acerca de nuestra conclusión, estamos 5% inseguros acerca de ella. Por tanto, tenemos un nivel de error esperado de 5%.

FIGURA 8-1
Uso de los estadísticos muestrales para obtener una estimación del intervalo de un parámetro poblacional para una variable de intervalo/razón $X = PC$



Las conclusiones acerca de μ_X con base en observar $\bar{X}$: estamos 95% seguros de que el PC medio de los estudiantes en la Crosstown University está entre 2.16 y 2.76.

Al calcular un intervalo de confianza, utilizamos la letra griega *alfa* (α) para representar el nivel de error esperado. Ese nivel de error esperado también se denomina *nivel de significación*, un término analizado de manera minuciosa en el capítulo 9. Al calcular el intervalo de confianza de 95%, nuestro nivel de significación o error esperado es 5%:

$$\text{Nivel de confianza} = 95\%$$

$$\text{Nivel de significación (error esperado)} = \alpha = 100\% - \text{nivel de confianza}$$
$$= 100\% - 95\% = 5\%$$

Entonces, en general, el nivel de confianza y el nivel de significación están inversamente relacionados; cuando uno aumenta, el otro disminuye. Juntos, el nivel de confianza y el nivel de significación suman 100%. Por tanto:

Cálculo del nivel de confianza y del nivel de significación

$$\text{Nivel de confianza} = 100\% - \alpha$$

Por tanto,

$$\alpha = 100\% - \text{nivel de confianza}$$

donde

$$\alpha = \text{nivel de significación (o error esperado)}$$

Para calcular un intervalo de confianza, calculamos el error estándar. Luego, a partir de la tabla de la curva normal (tabla estadística B del apéndice B), obtenemos una puntuación Z que corresponda a los niveles de confianza y significación elegidos (α). A ésta la denominamos puntuación *Z crítica*, simbolizada Z_α. Multiplicamos Z_α por el error estándar para obtener un "término del error". El término del error es la cantidad de más y menos el error, como ±3% que reportamos con nuestra estimación del intervalo. Para calcular el intervalo de confianza, sumamos y restamos este término del error a la media muestral. La dispersión de valores resultantes es una estimación del intervalo de confianza de la media poblacional:

$$\text{Intervalo de confianza} = \text{una estimación puntual} \pm \text{un término de error}$$

Cálculo del error estándar para un intervalo de confianza de una media poblacional

El objetivo de un intervalo de confianza es determinar una aproximación del parámetro poblacional. El parámetro, entonces, es una incógnita. Para una variable de intervalo/razón, tanto la media como la desviación estándar de la población son incógnitas. Por tanto, debemos utilizar la desviación estándar de la media para estimar el error estándar de la media. Recuerda del capítulo 7 que este error estándar estimado es como sigue:

Cálculo del error estándar (estimado) de un intervalo de confianza de una media poblacional

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}}$$

donde

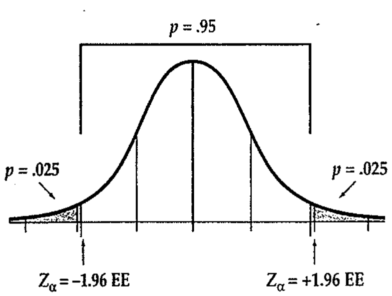
$s_{\bar{X}}$ = error estándar estimado de medias para una variable de intervalo/razón X

s_X = desviación estándar de una muestra

n = tamaño de la muestra

Selección de la puntuación Z crítica, Z_α

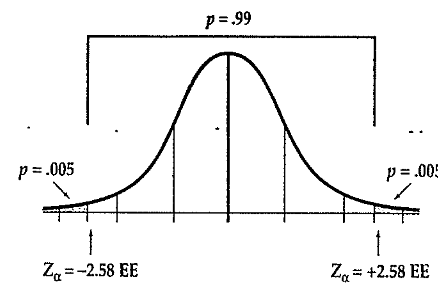
Con intervalos de confianza utilizamos nuestro conocimiento de las distribuciones muestrales para determinar los niveles de significación y confianza. Los intervalos de confianza por tradición se estipulan para una confianza de 95 y 99%. Recuerda del capítulo 7 que una distribución muestral de medias adopta la forma de una curva normal cuando $n > 121$. Además, la desviación estándar de una distribución muestral se denomina error estándar. Como observamos antes, en el muestreo repetido casi 95% de las medias muestrales estarán dentro de 2 errores estándar de la media de la población. Para ser más exactos, 95% de las medias muestrales están precisamente dentro de 1.96 errores estándar y 5% está en las dos colas. En la tabla de la curva normal, observa una puntuación Z de 1.96 en la columna A. En la columna C tenemos que .0250 (o 2.5%) de los casos caen fuera de esta puntuación a cada lado de la distribución.



A esta puntuación Z de ±1.96 se le refiere como puntuación *Z crítica* para el nivel de confianza de 95%. Empleamos Z_α para simbolizarla, donde α es el nivel de significación, que, de nuevo, es 1, el nivel de confianza. En el muestreo repetido estamos confiados que

95% de los resultados muestrales caerán dentro de este rango y 5% fuera de él en las dos colas de la curva.

Para el nivel de confianza al 99%, el nivel de significación es 1% o .01. Dividida en dos colas, tenemos .005 (la mitad de 1%) del área de la curva en cada cola. La puntuación Z crítica (Z_α) que aísla estas áreas de la curva es ±2.58.



Cálculo del término del error

Una vez calculado el error estándar, éste se multiplica por Z_α para obtener el término del error.

Cálculo del término del error de un intervalo de confianza de una media poblacional (cuando $n \geq 121$)

$$\text{Término del error} = (Z_\alpha)\,(s_{\bar{X}})$$

donde

α = nivel de significación (o error esperado)

Z_α = puntuación Z crítica que corresponde a los niveles estipulados de significación y confianza

$s_{\bar{X}}$ = error estándar estimado de un intervalo de confianza de la media

Cálculo del intervalo de confianza

Teniendo en cuenta que un intervalo de confianza de una media poblacional es una media muestral más y menos un término del error, la fórmula general para calcular el intervalo de confianza de una media poblacional es como sigue:

Cálculo de un intervalo de confianza (IC) de una media poblacional (cuando $n > 121$)

$$(100\% - \alpha)\ IC \text{ de } \mu_X = \bar{X} \pm (Z_\alpha)\,(s_{\bar{X}})$$

donde

α = nivel de significación (o error esperado, expresado como un porcentaje)

$(100\% - \alpha)$ = nivel de confianza

IC de μ_X = "intervalo de confianza de una media poblacional"

$\bar{X}$ = media muestral

Z_α = puntuación Z crítica que corresponde al nivel estipulado de significación y confianza

$s_{\bar{X}}$ = error estándar (estimado) de un intervalo de confianza de la media

Una vez más, dos intervalos de confianza muy comúnmente reportados son los IC al 95 y 99%. En la situación común de un tamaño de la muestra de n mayor que 121, se utilizan las fórmulas siguientes:

Cálculo de los intervalos de confianza al 95 y 99% de una media poblacional para la situación común de $n > 121$

$$IC \text{ de } 95\% \text{ de } \mu_X = \bar{X} \pm (1.96)\,(s_{\bar{X}})$$

y

$$IC \text{ al } 99\% \text{ de } \mu_X = \bar{X} \pm (2.58)\,(s_{\bar{X}})$$

donde

X = una variable de intervalo/razón

IC de 95% de μ_X = "intervalo de confianza de 95% de la media poblacional de X"

IC al 99% de μ_X = "intervalo de confianza al 99% de la media poblacional de X"

$\bar{X}$ = media muestral

$s_{\bar{X}}$ = error estándar estimado de la media

Cuándo calcular un intervalo de confianza de una media poblacional (cuando $n > 121$)

1. La pregunta de investigación requiere estimar un parámetro poblacional.
2. La variable de interés (X) es de nivel de medición de intervalo/razón. Por tanto, debemos proporcionar una estimación del intervalo del valor de un parámetro de la población μ_X.
3. Estamos trabajando con una muestra única representativa de una población.
4. El tamaño de la muestra es mayor que 121.

Los cinco pasos para calcular un intervalo de confianza de una media poblacional, μ_X

Calcularemos intervalos de confianza siguiendo estos cinco pasos: 1) Enuncia la pregunta de investigación, identifica el nivel de medición de la variable, enumera los "datos" y traza un diagrama (como la figura 8-1) representando la población objetivo, el parámetro que se estimará, la muestra y sus estadísticos; 2) calcula el error estándar y el término del error; 3) utilizando la fórmula general para intervalos de confianza, calcula el LCI y LCS; 4) proporciona una interpretación de las averiguaciones en lenguaje común dirigido a individuos y grupos que no estén familiarizados con la estadística (por ejemplo, administradores universitarios y de compañías, funcionarios del ayuntamiento, reporteros de periódicos y público en general); y 5) proporciona una interpretación estadística ilustrando la noción de "confianza en el procedimiento". La lista de verificación siguiente es útil para recordar estos pasos. Después resolveremos un problema de ejemplo.

Lista de verificación breve de los cinco pasos para calcular intervalos de confianza

Paso 1. Enuncia la pregunta de investigación. Traza diagramas conceptuales representando los datos, incluyendo la población y la muestra en estudio, la variable (por ejemplo, X = ...) y su nivel de medición, y los estadísticos dados o calculados.

Paso 2. Calcula el error estándar y el término del error.

Paso 3. Calcula el LCI y el LCS del intervalo de confianza.

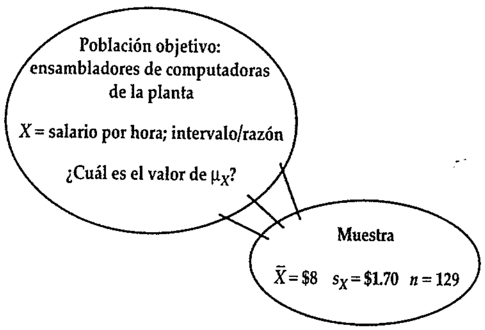
Paso 4. Proporciona una interpretación en lenguaje común.

Paso 5. Proporciona una interpretación estadística ilustrando la noción de "confianza en el procedimiento".

Cómo calcular un intervalo de confianza de una media poblacional

Problema: estamos realizando un estudio de la estructura salarial de una planta industrial que emplea varios miles de ensambladores de computadoras. Necesitamos obtener una idea aproximada del salario por hora medio de esta población de ensambladores. Seleccionamos al azar 129 expedientes del personal y registramos los salarios por hora. En este ejemplo encontramos una media de \$8.00 y una desviación estándar de \$1.70. Calcula el intervalo de confianza de 95% para el salario por hora medio de los ensambladores de la planta. (Al resolver un problema, no es necesario enunciar las instrucciones proporcionadas entre paréntesis.)

Paso 1. Pregunta de investigación: con un rango especificado de cantidades en dólares, ¿cuál es el parámetro μ_X, el salario por hora medio de la población de ensambladores de computadoras? *Datos:*



Paso 2. (error estándar, puntuación Z crítica y término del error)

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{1.70}{\sqrt{129}} = \$.15$$

Para un nivel de confianza de 95%, $Z_\alpha = 1.96$.

$$\text{Término del error} = (Z_\alpha)\,(s_{\bar{X}}) = (1.96)\,(\$.15) = \$.29$$

Paso 3. (LCI y LCS)

$$IC \text{ de } 95\% \text{ de } \mu_X = \bar{X} \pm (1.96)\,(s_{\bar{X}})$$

$$= \text{media muestral} \pm \text{término del error}$$

$$= \$8.00 \pm (1.96)\,(\$.15) = \$8.00 \pm \$.29$$

$$LCI = \$8.00 - \$.29 = \$7.71$$

$$LCS = \$8.00 + \$.29 = \$8.29$$

Paso 4. (interpretación en lenguaje común): "estoy 95% seguro de que el salario por hora medio de los ensambladores de computadoras de la plana está entre \$7.71 y \$8.29".

Paso 5. (interpretación estadística ilustrando la noción de "confianza en el procedimiento"): "si se realizan los mismos procedimientos muestrales y estadísticos 100 veces, 95 veces el parámetro poblacional μ_X estará comprendido en los intervalos calculados y 5 veces no. Por tanto, tengo una confianza de 95% que este intervalo de confianza individual que calculé incluye al parámetro real".

Interpretación apropiada de los intervalos de confianza

Cuando empleamos el enunciado "estoy 95% seguro", *en realidad estamos expresando confianza en nuestro método.* Para el problema de ejemplo anterior, esto se enuncia como la **interpretación estadística** de nuestros resultados. Para un intervalo de confianza de la media de 95%, nuestra interpretación estadística inicia: si los mismos procedimientos muestrales y estadísticos se realizan 100 veces, 95 veces la media poblacional real μ_X estará comprendida en los intervalos calculados. Recuerda que, como no reunimos datos para todos y cada uno de los miembros de la población, no podemos declarar una valor exacto, real de la media poblacional (el parámetro). Por tanto, hay una posibilidad que el intervalo de confianza calculado no incluya al parámetro real. Para regresar a nuestra analogía de pescar, no sabemos exactamente dónde se ubica el pez. Podemos lanzar la red y no atraparlo. Con un intervalo de confianza de 95%, esta posibilidad de fracaso es 5% (100% – 95% = 5%), el nivel de significación (o error esperado). Aunque vamos a lanzar la red sólo una vez, nuestro conocimiento del error de muestreo y su previsibilidad con una curva normal nos asegura que si lanzamos la red 100 veces, atraparemos al pez 95 veces. A la larga, un intervalo de confianza de 95% basado en una muestra única es correcto 95% de las veces.

Para comprender de manera apropiada los intervalos debemos utilizar la imaginación estadística y emplear lo que sabemos acerca del muestreo repetido, de las distribuciones muestrales y de la teoría de las probabilidades. La figura 8-2 representa la noción de muestrear de manera repetida y calcular intervalos de confianza para un tamaño muestral mayor que 121. Es decir, 95 de cada 100 medias muestrales se calcularán dentro de 1.96 errores estándar de la media poblacional real. Esto se debe a que en el muestreo repetido 95 de 100 muestras caen así de cerca. La figura 8-2 sugiere que el procedimiento estadístico de calcular de manera repetida intervalos de confianza, resulta en la media poblacional real cayendo dentro del intervalo un predecible 95% de las veces (19 de 20). Esto significa, por supuesto, que el intervalo de confianza calculado errará el parámetro correcto un predecible 5% de las veces (como es el caso para la muestra número 7). ¿Qué muestras aciertan y fallan? *Para las 95 medias dentro de 1.96 errores estándar*, los intervalos de confianza calculados incluirán el parámetro poblacional real, su media real. Estas medias muestrales que están lo suficientemente cerca para que sus intervalos de confianza abarquen hasta μ_X. Para las cinco medias que resultan fuera de 1.96 errores estándar, los intervalos de confianza calculados no tendrán el parámetro real. En la vida real, sólo tomamos una muestra y calculamos su intervalo de confianza. Estamos contando en la posibilidad de que esta muestra única es una de las 95 que

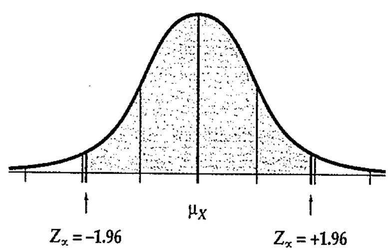
FIGURA 8-2

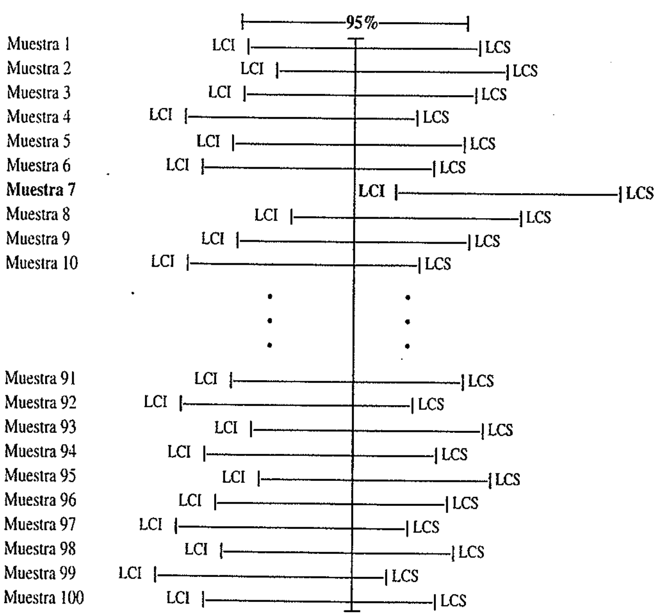
Tasa de éxito de un intervalo de confianza de 95% al proporcionar una estimación del intervalo que comprende el valor real del parámetro poblacional

μ_X = media **desconocida** de X en la población (es decir, el parámetro)

LCI = límite de confianza inferior
LCS = límite de confianza superior

Imaginemos que tomamos 100 muestras con un tamaño muestral de, digamos, 122. Calculamos $\bar{X}$ para cada muestra y trazamos los resultados como una distribución muestral. Esta distribución tendrá forma normal, con medias muestrales centradas en el valor real del parámetro poblacional, μ_X (cualquiera que sea su valor). Como $n > 121$, una puntuación Z crítica de ± 1.96 deja 5% del área en las colas de la curva y 95% en medio. Por tanto, 95% de estas medias caen dentro de 1.96 errores estándar del parámetro real como se ilustra en la curva normal siguiente. Imaginemos también que calculamos intervalos de confianza de 95% para cada una de esas 100 medias muestrales. Para las 95 muestras que resultan dentro de 1.96 errores estándar, sus intervalos de confianza se dispersarán lo suficientemente amplios para comprender (o "atrapar") al parámetro real, μ_X. Para las cinco medias muestrales fuera de 1.96 errores estándar, sus intervalos de confianza calculados *no* se dispersarán lo suficientemente amplios para atrapar al parámetro real. En otras palabras, el procedimiento para calcular un intervalo de confianza de 95% funciona 95% de las veces. El diagrama siguiente presenta 20 de 100 muestras con intervalos de confianza calculados. 95% (19 de 20) comprenden el parámetro poblacional (μ_X). La muestra número 7 representa la única de 20 (es decir, 5%) que falla en incluir la media poblacional.





caen dentro de 1.96 errores estándar del parámetro real. Si es una de las 95, cuando lanzamos 1.96 errores estándar atraparemos el μ_X real cualquiera que sea su valor. Por supuesto, aún no conocemos el parámetro real y nuestro intervalo de confianza sólo es una estimación; algunas veces esta estimación es muy amplia. Nunca conoceremos el parámetro real, μ_X, a menos que tengamos dinero y tiempo suficientes para obtener datos de cada miembro de la población, pero 95 posibilidades de 100 son muy buenas.

Malinterpretaciones comunes de los intervalos de confianza

Un intervalo de confianza trata acerca del tamaño de parámetros, no de puntuaciones individuales. Pensar en términos de puntuaciones individuales es una malinterpretación común de un intervalo de confianza. En el ejemplo de un intervalo de confianza del salario por hora medio de los ensambladores de computadoras de una planta, declaramos: "Estoy 95% seguro de que *el salario por hora medio* de los ensambladores de computadoras de la planta está entre \$7.71 y \$8.29". ¡*No estamos* diciendo que 95% de los ensambladores de computadoras ganan salarios por hora entre esas cifras! Si nuestro propósito hubiera sido describir un rango de puntuación en el cual caen 95% de los ensambladores, hubiéramos empleado la desviación estándar de la muestra —no el error estándar— para hacer la proyección (como en el capítulo 6). El intervalo de confianza aborda cuestiones de estadísticos sumarios, no de puntuaciones individuales.

También debemos tener cuidado de no empezar a tratar nuestras medias como si fueran la media población misma. En el capítulo 7 aprendimos que el muestreo repetido produce una distribución muestral con medias muestrales centradas en la media de la población, μ_X. Pero estaría mal tomar la media muestral individual, $\bar{X}$, de nuestro estudio y tratarla como si todas las otras medias muestrales se centraran en ella. En otras palabras, con un intervalo de confianza, *no* decimos que 95% de las muestras repetidas tendrán medias entre los límites de confianza superior e inferior calculados *a partir de esta media muestral individual*. Es la media poblacional desconocida respecto a la cual caen estas otras muestras. La interpretación del intervalo de confianza se basa en nuestra muestra individual, cuya media no es probable que sea igual a la media poblacional. En resumen, el intervalo de confianza simplemente nos proporciona un rango de valores posibles para el parámetro poblacional desconocido.

El nivel de confianza seleccionado y la precisión del intervalo de confianza

Para la muestra de 129 ensambladores de computadoras, seleccionamos el nivel de confianza de 95% y utilizamos la puntuación Z crítica de 1.96 para calcular el término del error. El uso de puntuaciones Z al calcular intervalos de confianza está relacionado a nuestro conocimiento de distribuciones muestrales a partir del muestreo repetido. En el capítulo 7 aprendimos que si tomamos muchas muestras y calculamos sus medias (como hicieron nuestros ancestros contadores de frijoles), la distribución muestral será una distribución normal cuando $n > 121$. Las puntuaciones Z miden qué tan alejada está una media muestral de la media poblacional real. Con la ayuda de la tabla de probabilidad de la distribución normal, estas puntuaciones determinan la probabilidad de ocurrencia de resultados muestrales.

Al calcular un intervalo de confianza, *una vez que se extraído una muestra*, su media, su desviación estándar y su tamaño muestral son "datos". Es decir, no podemos deshacernos de ellos. Estos datos determinan el error estándar del intervalo de confianza, y por tanto, tienen una gran influencia en el ancho del intervalo de confianza calculado. Si la desviación estándar es mayor o el tamaño de la muestra es pequeño, el intervalo de confianza resultará amplio; no será muy preciso. Sin embargo, después que se "conformó" la muestra aún podemos influenciar la precisión de un intervalo de confianza mediante nuestra selección del nivel de confianza. El nivel de confianza elegido determina el tamaño de la puntuación Z crítica (Z_α). Por tanto, en el cálculo de los límites de confianza, una Z_α produce un término del error grande y un intervalo de confianza menos preciso (o más amplio).

Por ejemplo, sustituyamos una Z_α de 2.58 en lugar de 1.96 en el problema de ejemplo del intervalo de confianza del salario medio de ensambladores de computadora. En la tabla de la distribución normal, ésta es la Z_α que corresponde a un intervalo de confianza de 99%:

$$IC\ 99\%\ \text{de}\ \mu_X = \bar{X} \pm (2.58)\,(s_{\bar{X}}) = \$8.00 \pm (2.58)\,(\$.15)$$

$$= \$8.00 \pm \$.39 = \$7.61 \text{ a } \$8.39$$

Comparando los dos intervalos de confianza, podemos ver que tenemos una mayor seguridad en el nivel de confianza de 99%, pero nuestra estimación es menos precisa:

IC 95% de μ_X = \$7.71 a \$8.29; este intervalo tiene una amplitud de \$.58

IC 99% de μ_X = \$7.61 a \$8.39; este intervalo tiene una amplitud de \$.78

La relación entre el nivel de confianza y el grado de precisión

Entre mayor sea el nivel de confianza estipulado, mayor será el término del error y por tanto será menos preciso el intervalo de confianza.

Esto tiene sentido. Si vamos a tener mucha fe (o confianza) en una respuesta, debemos tener precaución permitiendo bastante error. Por ejemplo, podríamos decir que estamos 99.9999% seguros (y estaríamos dispuestos a apostar \$100 a favor de nuestra respuesta) que el salario medio de los ensambladores de computadoras está entre \$3 y \$100 por hora. En esta situación absurda estamos seguros, pero el grado de precisión es tan bajo que no tiene sentido. Por otro lado, si proporcionamos una estimación con un alto grado de precisión, digamos 10 centavos —\$7.95 a \$8.05—, no apostaríamos demasiado. Para regresar una vez más a la analogía de pescar; alguien podría decir que está 100% seguro que el pez está entre una orilla y la otra, pero esta "ayuda" es tan imprecisa que es inútil. Por otra parte, si preguntáramos si el pez estaba entre nosotros y un muelle alejado 20 pasos, alguien podría responder, "no estoy seguro".

El tamaño de la muestra y la precisión del intervalo de confianza

Existe una manera para obtener un alto grado de precisión y mantener un alto nivel de confianza: *asegúrate antes de recolectar datos* que el tamaño de la muestra es lo suficientemente grande para producir errores estándar pequeños e intervalos de confianza precisos. Veamos cómo afecta el tamaño de la muestra al ancho de un intervalo de confianza. Volvamos a calcular el intervalo de confianza de 95% para la población de ensambladores de computadora, pero utilicemos un tamaño de muestra de 1 000 en lugar de 129. Supongamos que la media

muestral y la desviación estándar permanecen iguales, y volvamos a calcular el error estándar, el término del error y el intervalo de confianza:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{1.70}{\sqrt{1\,000}} = \$.05$$

$$\text{Término del error} = (Z_\alpha)\,(s_{\bar{X}}) = (1.96)\,(\$.05) = \$.10$$

$$IC\ \text{de}\ 95\%\ \text{de}\ \mu_X = \bar{X} \pm (Z_\alpha)\,(s_{\bar{X}})$$

$$= \$8.00 \pm (1.96)\,(\$.05)$$

$$= \$8.00 \pm \$.10 = \$7.90 \text{ a } \$8.10$$

Comparando esta muestra de 1 000 con la muestra de 129, observamos que la estimación de la muestra más grande es más precisa:

Con n = 129: IC de 95% de μ_X = \$7.71 a \$8.29; este intervalo tiene una amplitud de \$.58.

Con n = 1 000: IC de 95% de μ_X = \$7.90 a \$8.10; este intervalo tiene una amplitud de \$.20.

El intervalo de confianza más preciso para la muestra de n = 1 000 tiene sentido intuitivo y se deriva de la ley de los números grandes (capítulo 7). Entre mayor sea la muestra, menor será el error de muestreo y por tanto mayor será la precisión del intervalo de confianza.

Relación entre el tamaño de la muestra y el grado de precisión

Entre mayor sea el tamaño de la muestra, más preciso será el intervalo de confianza.

Intervalos de confianza de las medias para muestras pequeñas Para un intervalo de confianza de la media cuando el tamaño de la muestra (n) es menor que o igual a 121, las puntuaciones críticas de ±1.96 y ±2.58 no son apropiadas. Estas puntuaciones se basan en el conocimiento que las muestras mayores que 121 casos producen distribuciones muestrales que se ajustan a la curva normal. Como la media es susceptible a la distorsión por puntuaciones extremas, muestras "pequeñas" de $n \leq 121$ producen distribuciones muestrales que están más planas que la forma de campana de una curva normal. Estas distribuciones se denominan distribuciones aproximadamente normales y sus puntuaciones críticas se denominan puntuaciones t en lugar de puntuaciones Z. En la fórmula para el intervalo de confianza para muestras pequeñas, las puntuaciones t se sustituyen por puntuaciones Z. Esta modificación la analizaremos para muestras pequeñas en el capítulo 10.

Intervalo de confianza de una proporción poblacional calculado a partir de una muestra grande

Con variables de nivel nominal/ordinal, los intervalos de confianza proporcionan una estimación de la proporción de una población que cae en la categoría de "éxito" de la variable. Supongamos que realizamos una encuesta de elección para una candidata política, Chantrise Jones. Deseamos obtener una estimación de intervalo de su apoyo, realizando una encuesta telefónica de votantes probables dos días antes de la elección. Definimos $P = p$ [de votantes probables apoyando a Chantrise]. Por supuesto, no podemos darnos el lujo de encuestar a todos los votantes probables; por tanto, tomamos una muestra. La proporción muestral, P_s, se utiliza para estimar el parámetro poblacional, P_u, dentro de un intervalo con un error de muestreo calculado. Al igual que el caso de los intervalos de confianza de la media, utilizamos un estadístico muestral, P_s, como una estimación puntual de P_u, y sumamos y restamos un término del error. La fórmula completa para calcular el intervalo de confianza de la proporción poblacional es:

$$(100\% - \alpha)\ IC \text{ de } P_u = P_s \pm (Z_\alpha)\,(s_{P_s})$$

$$= \text{proporción muestral} \pm \text{término del error}$$

Aquí $P = p$ [de la categoría de éxito] de una variable nominal/ordinal, α = nivel de significación (o error esperado) $(100\% - \alpha)$ = el nivel de confianza, IC de P_u se lee "el intervalo de confianza de una proporción poblacional," P_s = la proporción muestral, Z_α = la puntuación Z crítica (de la tabla de la distribución normal) que corresponde al nivel de confianza y significación estipulados, y s_{P_s} = error estándar estimado de un intervalo de confianza de una proporción.

Las siguientes son las circunstancias en las que es apropiado calcular un intervalo de confianza de una proporción de la población:

Cuándo calcular un intervalo de confianza de una proporción de la población (para una variable nominal/ordinal)

1. Cuando debemos proporcionar una estimación de un intervalo del valor de un parámetro de la población, P_u, donde $P = p$ [de la categoría de éxito] de una variable nominal/ordinal.
2. Cuando tenemos una sola muestra representativa de una población.
3. Cuando el tamaño de la muestra (n) es lo suficientemente grande que $(p_{menor})(n) \geq 5$, resultando en una distribución muestral que es normal.

El requerimiento de que el tamaño de la muestra (n) sea lo suficientemente grande tal que $(p_{menor})(n) \geq 5$ es la única restricción sobre el tamaño de la muestra. La Z_α crítica para un

intervalo de confianza de 95% siempre será ±1.96 y para el intervalo de confianza de 99% será ±2.58.

Un error estándar estimado lo calculamos con base en los datos muestrales (como en el capítulo 7) y el término del error como sigue:

Cálculo del error estándar de un intervalo de confianza de una proporción de la población (para una variable nominal/ordinal)

$$s_{P_s} = \sqrt{\frac{P_s Q_s}{n}}$$

donde

s_{P_s} = error estándar estimado de proporciones para una variable nominal con $P = p$ [de la categoría de éxito]

$P_s = p$ [de la categoría de éxito en la muestra]

$Q_s = p$ [de la categoría de fracaso en la muestra] $= 1 - P_s$

n = tamaño de la muestra

Cálculo del término del error de un intervalo de confianza de una proporción de la población

$$\text{Término del error} = (Z_\alpha)\,(s_{P_s})$$

donde

α = nivel de significación (o error esperado)

Z_α = puntuación Z crítica que corresponde al nivel de confianza y significación estipulados

s_{P_s} = error estándar estimado de proporciones para una variable nominal/ordinal donde $P = p$ [de la categoría de éxito]

Para los niveles de confianza de 95 y 99% tradicionales utilizamos las dos ecuaciones siguientes:

Cálculo de intervalos de confianza de 95 y 99% de una proporción de la población cuando $(p_{menor})(n) \geq 5$ (para una variable nominal/ordinal)

$$IC \text{ de } 95\% \text{ de } P_u = P_s \pm (1.96)\,(s_{P_s})$$

y

$$IC \text{ de } 99\% \text{ de } P_u = P_s \pm (2.58)\,(s_{P_s})$$

donde

$P = p$ [de la categoría de éxito] de una variable nominal/ordinal

IC 95% de P_u = intervalo de confianza de 95% de una proporción de una población

IC 99% de P_u = intervalo de confianza al 99% de una proporción de la población

P_s = proporción de la muestra

s_{P_s} = error estándar estimado de un intervalo de confianza de una proporción

Como observamos en el capítulo 7, una distribución muestral de proporciones está normalmente distribuida sólo cuando el valor menor de P_s y Q_s por n es mayor que o igual a 5. Si $(p_{menor})(n) < 5$, la mejor solución es aumentar el tamaño de la muestra.

Ahora calculemos el intervalo de confianza de 95% de la proporción que apoya la candidatura al Senado de Chantrise Jones. Seguimos la lista de verificación de cinco pasos para calcular intervalos de confianza.

Cómo calcular un intervalo de confianza de una proporción de la población

Problema: trabajamos para Chantrise Jones, que es candidata al Senado de Estados Unidos. Faltan dos días para la elección. Con una seguridad de 95%, ella quiere saber si es probable que gane. ¿Cuál es su nivel de apoyo entre los votantes probables? En una encuesta telefónica de 1 393 votantes probables, 752 indican que tienen pensado votar por ella.

Paso 1. Pregunta de investigación: con una seguridad de 95%, ¿podemos concluir es probable que Chantrise Jones que gane la elección? Es decir, ¿parece probable que obtenga más de .50 (50%) de los votos? Dentro de un rango especificado del porcentaje de apoyo, ¿cuál es el parámetro P_u, la proporción de la población de votantes probables con intención de votar por Chantrise Jones? *Datos:*

$P = p$ [de los votantes probables apoyando a Chantrise]

$Q = p$ [de los votantes probables apoyando a otro candidato]

Muestra: $n = 1\,393$ probables votantes # apoyando a Chantrise = 752

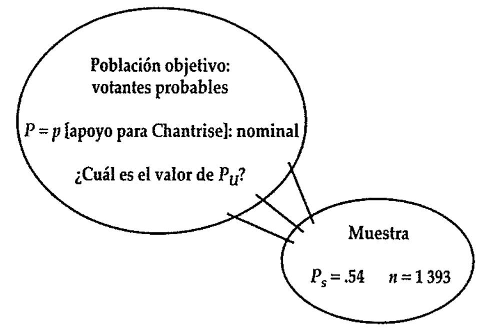
$$P_s = \frac{\#\text{ apoyando a Chantrise}}{\text{número total encuestado}} = \frac{752}{1\,393} = .54$$

$$Q_s = 1 - P_s = 1 - .54 = .46$$

[Verifique si n es lo suficientemente grande]. Compruebe si $(p_{menor})\,(n) > 5$].

$$(p_{menor})(n) = (.46)(1\,393) = 640.78, \quad 640.78 > 5$$

(Por tanto, podemos continuar con los cálculos)



Paso 2. (error estándar y término del error)

$$s_{P_s} = \sqrt{\frac{P_s Q_s}{n}} = \sqrt{\frac{(.54)\,(.46)}{1\,393}} = .0134$$

(para una seguridad de 95%, $Z_\alpha = 1.96$)

Término del error = $(Z_\alpha)\,(s_{P_s}) = (1.96)\,(.0134) = .0263$

Paso 3. (el LCI y el LCS del intervalo de confianza)

$$\begin{aligned} IC \text{ de } 95\% \text{ de } P_u &= P_s \pm (1.96)\,(s_{P_s}) \\ &= .54 \pm (1.96)\,(.0134) \\ &= .54 \pm .0263 \\ &= \text{proporción de la muestra} \pm \text{término del error} \end{aligned}$$

$$LCI = .54 - .0263 = .5137 = 51.37\%$$

$$LCS = .54 + .0263 = .5663 = 56.63\%$$

Paso 4. (interpretación en lenguaje común): "estoy 95% seguro de que el porcentaje de votantes probables apoyando a Chartrise Jones está entre 51 y 57%". Las posibilidades de ganar de Chartrise son buenas. Si la elección se llevara a cabo hoy, obtendría al menos 51% de los votos. (*Nota:* redondeamos a un porcentaje exacto para beneficio de una audiencia pública.)

Paso 5. (interpretación estadística ilustrando la noción de "seguridad en el procedimiento"): "si se realizan los mismos procedimientos de muestreo y el estadístico 100 veces, 95 veces el parámetro poblacional real, P_u, estará comprendido en los intervalos calculados y 5 veces no. Por tanto, estoy 95% seguro de que el intervalo de confianza individual que calculé incluye al parámetro real".

Selección de un tamaño de la muestra para elecciones, encuestas y estudios de investigación

Tamaño de la muestra para un intervalo de confianza de una proporción de la población

Una pregunta que cada investigador enfrenta es: ¿qué tamaño de muestra necesito? Como recién aprendimos, el tamaño de la muestra es un componente importante en el tamaño de un error estándar. En las ecuaciones del cálculo del error estándar tanto para medias como para proporciones, el tamaño de la muestra (n) está en el denominador de las ecuaciones. Por tanto, en un tamaño grande la muestra es mejor debido a que producirá un error estándar pequeño. Sin embargo, debido a factores de costo no podemos simplemente seleccionar un tamaño de la muestra enorme. No obstante, podemos seleccionar un tamaño de la muestra *adecuado* para el grado de precisión que deseamos para los resultados reportados. El grado de precisión depende de los objetivos de la investigación, de la cantidad de tiempo y dinero disponibles para la investigación, y de otras consideraciones. Por ejemplo, una compañía de encuestas políticas quizás utilice muestras pequeñas al inicio pero aumente su tamaño para mejorar la precisión cuando se aproxima la elección. Dependiendo de estos puntos, podemos elegir reportar los resultados con un error de más o menos 1%, 3%, 5%, etc. La precisión elegida depende del tamaño del término del error de la ecuación del intervalo de confianza.

Demostremos la elección del tamaño de la muestra para un intervalo de confianza de proporciones. Las variables de nivel nominal/ordinal, como la proporción de votantes probables a favor de un candidato o apoyando una causa, se emplean comúnmente en las encuestas políticas. Un estándar tradicional en las encuestas políticas así como en las investigaciones de marketing es reportar resultados con una seguridad de 95% y un rango de error de ±3%. Una vez que se elige el tamaño del término del error, se determina el tamaño de la muestra requerido para alcanzar ese nivel de error, despejando para n en la ecuación del término del error. El término del error para un intervalo de confianza de proporciones se puede desarrollar como sigue:

$$\text{Término del error} = (Z_\alpha)(s_{P_s}) = (Z_\alpha))\sqrt{\frac{P_sQ_s}{n}}$$

Despejando para n resulta la ecuación siguiente para calcular el tamaño muestral necesario:

Cálculo del tamaño muestral para el intervalo de confianza de una proporción poblacional (para una variable nominal/ordinal)

$$n = \frac{(P_sQ_s)(Z_\alpha)^2}{\text{término del error}^2}$$

donde

n = tamaño muestral necesario

Z_α = puntuación Z que corresponde al nivel de confianza y significación estipulados (por ejemplo, $Z_\alpha = 1.96$ para un nivel de confianza de 95%)

$P_s = p$ [de la categoría de éxito en la muestra]

$Q_s = p$ [de la categoría de falla en la muestra]

Término del error = precisión deseada en los resultados que se reportarán

Para despejar para n, se deben conocer todos los otros términos en la ecuación o de lo contrario deben estimarse. *Seleccionamos* el nivel de confianza, que determina Z_α. Si seleccionamos el nivel de 95%, $Z_\alpha = 1.96$. También seleccionamos el grado de precisión, qué tan grande queremos que sea el término del error. Por ejemplo, podríamos seleccionar el ±3% tradicional (es decir, ±.03). Puesto que aún no hemos reunido datos, debemos estimar P_s y Q_s para las variables importantes en el estudio, como el porcentaje apoyando a un candidato. Estas cifras se pueden estimar con base en una investigación previa. Si no se dispone de esa investigación, una estimación conservadora es establecer P_s en .5. Como $Q_s = 1 - P_s$, entonces Q_s también se estimará en 0.5. (Estas estimaciones son conservadoras en el sentido que estarán erradas en el tamaño grande; es decir, el error de 3% reportado será un escenario en el peor de los casos donde el error se reporta en exceso en lugar de reportarlo con un valor menor.) Con todos los términos previamente desconocidos ahora especificados, despejemos para el tamaño muestral necesario cuando queremos un error de ±3% en el nivel de confianza de 95%:

$$n = \frac{(P_sQ_s)(Z_\alpha)^2}{\text{término del error}^2} = \frac{(.5)(.5)(1.96)^2}{.03^{\ 2}} = 1\ 067 \text{ encuestados}$$

Observamos que es necesario un tamaño muestral considerable para un 3% de error reportado en el nivel de confianza de 95%. Debido al alto costo del muestreo, muchas compañías que realizan encuestas han empezado a conformarse con muestras más pequeñas y por tanto con un error mayor (como ±5%). Esto es cierto en especial en encuestas telefónicas nocturnas, cuyos costos se han incrementado como resultado de los retardos de tiempo ocasionados por las máquinas contestadoras y teléfonos celulares. Para realizar el procedimiento de cálculo del tamaño muestral necesario para un intervalo de confianza de una media poblacional, ve las extensiones del capítulo 8 del material de este texto disponibles en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2.

Insensatez y falacias estadísticas: es más y menos el término del error

En los reportes de medios populares de estudios de encuestas, un error común es tratar al término del error como igual al ancho del intervalo de confianza. Por ejemplo, en una contienda presidencial, una cadena importante reportó resultados de una encuesta de votantes probables una semana antes de la elección. Se reportó que el candidato republicano tenía 42% de los votos contra 36% para el candidato democrático con el resto indeciso. El rango de error se reportó como más menos 3.5%. Como la diferencia en apoyo para el candidato fue 6% y esto es mayor que 3.5%, los noticieros nocturnos iniciaron con la historia que el candidato republicano tenía "una ventaja sustancial". Esta declaración fue incorrecta. El límite de confianza inferior para el candidato republicano fue $42 - 3.5 = 38.5\%$; en la población de votantes probables su apoyo podría haber sido así de bajo. El límite de confianza superior para el candidato democrático fue $36 + 3.5 = 39.5\%$; su apoyo podría haber sido así alto. En otras palabras, los dos intervalos de confianza se sobrepusieron y era posible que el candidato democrático mantuviera la ventaja. Una interpretación correcta de la encuesta hubiera sido que la elección podría resultar en un empate. No había datos suficientes para concluir que cualquiera de los candidatos tenía la ventaja.

La persona que escribió el guión del noticiero probablemente cometió el error común de simplemente hacer notar que las dos estimaciones puntuales de 42 y 36% estaban separadas 6 puntos porcentuales y que esto era mayor que 3.5%. Sin embargo, para mostrar la separación entre los niveles de apoyo de los candidatos, las dos estimaciones puntuales deben estar separadas por más de 7.0%, 2 por el término del error de 3.5%. Recuerda que los estadísticos muestrales sólo son herramientas para sacar conclusiones acerca de los parámetros poblacionales. Es esencial que el error de muestreo se evalúe de manera adecuada. Una diferencia *muestral* de 6% puede parecer impresionante, pero para esa muestra no fue suficiente para establecer una diferencia en la *población* de votantes probables.

El reporte erróneo de una cadena de noticias importantes de los términos del error y de los intervalos de confianza no es inusual. De hecho, muchas veces los reporteros de los medios populares simplemente ignoran el término del error y tratan una estimación puntual como si fuera verdadera. Como hemos aprendido, ningún estadístico individual —una medida sumaria basada en datos muestrales— es la última palabra en la estimación de un parámetro de la población. El muestreo repetido demuestra que la siguiente muestra de la misma población es probable que resulte en un resultado estadístico ligeramente distinto. No obstante, no es poco común que las estimaciones puntuales se acepten rápidamente y después se traten como si fueran absolutamente ciertas. Una atención adecuada a un término del error, al ancho de un intervalo de confianza y al nivel de confianza es esencial para comprender por completo e interpretar adecuadamente las averiguaciones estadísticas.

RESUMEN

1. Con base en el análisis de los estadísticos en una muestra, un intervalo de confianza es un rango posible de valores de un parámetro poblacional expresado con un grado de confianza específico. El objetivo de un intervalo de confianza es proporcionar una estimación intervalar del valor de un parámetro poblacional desconocido y expresar la seguridad que tenemos de que el parámetro cae dentro de ese intervalo. Con un intervalo de confianza respondemos la pregunta: ¿Cuál es el valor de un parámetro poblacional, más y menos un poco de error de muestreo conocido?

2. Para elaborar un intervalo de confianza, tomamos una estimación puntual y utilizamos el conocimiento acerca de las distribuciones muestrales para proyectar un intervalo de error respecto a ésta. La fórmula se establece como sigue: intervalo de confianza = estimación puntual ± término del error.
3. Un intervalo de confianza se calcula empleando una media o proporción muestral, un error estándar y una puntuación Z crítica de la tabla de la curva normal. Esta puntuación crítica corresponde a un nivel de significación (es decir, el error esperado) y su nivel de confianza. Para muestras grandes, se utilizan puntuaciones críticas de 1.96 y 2.58 para los niveles de confianza de 95 y 99%, respectivamente. El término del error se calcula multiplicando un error estándar por la puntuación crítica.
4. El nivel de confianza y el nivel de significación, alfa (α), están inversamente relacionados, cuando uno aumenta, el otro disminuye. El nivel de confianza más el nivel de significación dan una proporción de 1.00 o un porcentaje de 100%. Así, por ejemplo, un nivel de confianza de 95% tiene un nivel de significación de 5% (es decir, $\alpha = .05$).
5. Un error estándar grande ocurre cuando n es pequeña o la desviación estándar de la muestra es grande.
6. Un término del error grande resulta cuando el error estándar o la puntuación Z crítica es grande.
7. Un intervalo de confianza de una media poblacional se calcula cuando la variable de interés (X) es de nivel de medición de intervalo/razón.
8. La interpretación en "lenguaje común" del intervalo de confianza de una media poblacional aborda la pregunta: ¿cuál es la estimación del parámetro, μ_X (mu subíndice X)? El intervalo reportado es una estimación de *la media poblacional*, no una estimación del rango de puntuaciones X.
9. La interpretación estadística del intervalo de confianza describe el proceso lógico mediante el cual podemos estipular un nivel de confianza. Con un nivel de confianza de 95%, estamos asegurando que el intervalo calculado es una buena estimación 95% de las veces. (Consulte la figura 8-2.)
10. Con variables nominales/ordinales, los intervalos de confianza proporcionan una estimación dentro de un rango de error de la proporción de una población que cae en la categoría de "éxito" de la variable. Para un intervalo de confianza de una proporción poblacional, el tamaño de la muestra debe ser lo suficientemente grande tal que $(p_{menor})(n) \geq 5$, resultando en una distribución muestral que es normal en forma.
11. La precisión de un intervalo de confianza (es decir, su ancho) es la diferencia entre el LCS y LCI. Entre mayor sea el nivel de confianza, menos preciso será el intervalo de confianza. Entre mayor sea la muestra, más preciso será el intervalo de confianza.
12. Para obtener un alto grado de precisión *y* mantener un alto nivel de confianza, un investigador debe utilizar una muestra suficientemente grande, tal que produzca un error

estándar pequeño, y por tanto, un intervalo de confianza más preciso. El tamaño de la muestra se puede elegir para ajustarse a un nivel de confianza y rango de error deseados.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 8 del material de este texto en el sitio en la red *The Statistical Imagination* en www.mhhe.ritchey2 incluyen el cálculo del tamaño mínimo de la muestra para un intervalo de confianza de la media.

FÓRMULAS PARA EL CAPÍTULO 8

Cálculo de un intervalo de confianza de una media poblacional cuando $n > 121$

Datos: una variable X de intervalo/razón

Pregunta de investigación: ¿cuál es el valor de la media poblacional, μ_X?

$$IC \text{ de } 95\% \text{ de } \mu_X = \overline{X} \pm (1.96)\,(s_{\overline{X}})$$

$$IC \text{ de } 99\% \text{ de } \mu_X = \overline{X} \pm (2.58)\,(s_{\overline{X}})$$

$$s_{\overline{X}} = \frac{s_X}{\sqrt{n}}$$

Cálculo de un intervalo de confianza de una proporción de la población (cuando $[(p_{menor})(n)] \geq 5$)

Datos: una variable nominal/ordinal con $P = p$ [de la categoría de éxito]

Pregunta de investigación: ¿cuál es el valor de la proporción de la población, P_u?

$$IC \text{ de } 95\% \text{ de } P_u = P_s \pm (1.96)\,(s_{P_s})$$

$$IC \text{ de } 99\% \text{ de } P_u = P_s \pm (2.58)\,(s_{P_s})$$

$$s_{P_s} = \sqrt{\frac{P_s Q_s}{n}}$$

La precisión o ancho de un intervalo de confianza = LCS – LCI

El tamaño muestral para el intervalo de confianza de una proporción poblacional (para una variable nominal/ordinal):

$$n = \frac{(P_s Q_s)\,(Z_\alpha)^2}{\text{término del error}^2}$$

PREGUNTAS PARA EL CAPÍTULO 8

1. En *lenguaje común*, ¿qué es un intervalo de confianza?
2. ¿Cuál es el objetivo de calcular un intervalo de confianza?
3. Al calcular un intervalo de confianza, ¿cuáles son los dos factores que afectan el tamaño del intervalo?
4. Al calcular intervalos de confianza, ¿cuál es la relación entre el tamaño muestral y el tamaño del error estándar?
5. Al calcular intervalos de confianza, ¿cuál es la relación entre el tamaño del intervalo de confianza y la puntuación Z crítica empleada para su cálculo?
6. Escribe la interpretación estadística de cualquier intervalo de confianza para el cual tienes una seguridad de 99%.
7. Enumera los cinco pasos al calcular un intervalo de confianza.
8. El nivel de significación (o error esperado) en el cálculo de un intervalo de confianza se simboliza por la letra griega α. De forma matemática, ¿cuál es la relación entre α y el nivel de confianza?

EJERCICIOS PARA EL CAPÍTULO 8

Conjunto de problemas 8A

8A-1. Siguiendo los cinco pasos para calcular un intervalo de confianza, calcula e interpreta el intervalo de confianza de 95% para los datos siguientes:

X = edad	n = 189 ejecutivos corporativos
Media = 57 años	Desviación estándar = 9 años

8A-2. Vuelve a realizar los últimos tres pasos para calcular un intervalo de confianza para calcular el intervalo de confianza de 99% con los datos del ejercicio 8A-1. Compara los resultados con los del ejercicio 8A-1 y comenta sobre esta comparación.

8A-3. Tú deseas calcular una estimación del intervalo del ingreso medio de planeadores urbanos en 150 Áreas Estadísticas Metropolitanas (AEM) en Sun Belt. Para esto obtienes una muestra aleatoria de 214 planeadores urbanos y determinas un ingreso medio de \$43 571 con una desviación estándar de \$4 792. Siguiendo el procedimiento de cinco pasos, elabora el intervalo de confianza de 99%. Como parte del paso 4, explica tus resultados al director del Departamento de Estudios Urbanos de la universidad local.

8A-4. Es el año 2010 y tú trabajas para la presidenta Shirley D. Fendus como encuestador. Ella desea saber qué proporción de sus 8 469 funcionarios de partido apoyan su proyecto legislativo para aumentar el gasto en defensa. Tú encuestas a 306 funcionarios seleccionados al azar y determinas que 108 apoyan su proyecto. Para informarle, calcula e interpreta el intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos.

8A-5. Tú has obtenido un cargo como asistente de un representante en la Casa de Representantes de Estados Unidos. Tu candidato, y su partido, buscan la aprobación del congreso de una propuesta importante sobre seguridad. Por consiguiente, tú deseas saber cuántos de los otros 434 representantes piensan emitir un voto de apoyo para la propuesta. Tú seleccionas al azar 137 miembros de la casa y descubres que 76% apoyan la legislación propuesta. Calcula e interpreta el intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos.

8A-6. Tú realizarás una encuesta para determinar el porcentaje de votantes registrados que actualmente apoyan al candidato A. Los resultados se reportarán con una seguridad de 95% y un término del error de 2 puntos porcentuales. ¿Qué tamaño de la muestra debes obtener? (*Sugerencia:* en este punto se desconoce P_s, pero supongamos que será .5.)

8A-7. Completa la tabla siguiente, donde n = 1 000 y P_s = .5. Responde las preguntas que siguen.

Nivel de confianza	Z_α	s_{P_s}	Término del error	LCI	LCS	Ancho del intervalo de confianza
95%	1.96	.0158	.0310	.4690	.5310	.0620
99%	____	____	____	____	____	____
99.9%	____	____	.0520	____	____	____

a) ¿Cuál es la relación entre el tamaño del nivel de confianza y el ancho del intervalo de confianza? Explica.

b) ¿Cómo se calcula el ancho del intervalo de confianza?

c) ¿Cuál es la relación entre el tamaño del nivel de confianza y el tamaño del término del error? Explica.

Conjunto de problemas 8B

8B-1. Siguiendo el procedimiento de los cinco pasos para calcular un intervalo de confianza, calcula e interpreta el intervalo de confianza de 95% para los datos siguientes (ficticios) de una muestra aleatoria de hombres en un estudio de nutrición nacional:

X = peso n = 147 hombres

Media = 174 libras   Desviación estándar = 6 libras

8B-2. Vuelve a realizar los últimos tres pasos del procedimiento descrito para calcular el intervalo de confianza de 99% para los datos del ejercicio 8B-1. Compara los resultados con los del ejercicio 8B-1 y comenta sobre esta comparación.

8B-3. La Dra. Latisia Latham, una consejera matrimonial, aplica la Escala de Angustia Global (EAG) que mide la discordia matrimonial global. Esta escala consiste de 43 preguntas de tipo verdadero/falso con una puntuación total combinada para los dos cónyuges (Snyder, Willis y Grady-Fletcher, 1991). Ella te pide que le proporcione una estimación aproximada de la puntuación promedio de su clientela. Tú obtienes una muestra aleatoria de 125 parejas y determinas una puntuación EAG media de 59 puntos en la escala EAG con una desviación estándar de 5.2. Siguiendo el procedimiento de cinco pasos, proporciona un intervalo de confianza de 95% de la puntuación EAG media de los clientes de la Dra. Latham.

8B-4. El senador Daniel "Dandy" Barker considera postularse para la presidencia. Pide a una persona que encueste a 90 votantes registrados seleccionados al azar y determina que 51% lo apoyan contra su oponente. Si la elección fuera hoy, ¿podría asegurar la victoria el senador Barker? Explica empleando un intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos.

8B-5. El presidente electo de la Asociación de Estudiantes de la Escuela de Ciencias Sociales y del Comportamiento (CSC) ha decidido postularse para otro periodo. Una encuesta aleatoria entre 219 estudiantes universitarios y graduados indica que casi 63% apoyan su próximo periodo como presidente. Si la elección fuera hoy, ¿podría asegurar su victoria el candidato? Explica empleando un intervalo de confianza de 95%.

8B-6. Tú realizarás una encuesta para determinar el porcentaje de pacientes de una organización de cuidado de la salud que está satisfecho con sus médicos generales. Tú deseas reportar los resultados con una seguridad de 99% y un término del error de 3 puntos porcentuales. ¿Qué tamaño de la muestra deberás tomar? (*Sugerencia:* en este punto se desconoce P_s, pero supongamos que será .5.)

8B-7. Completa la tabla siguiente, donde n = 225 y P_s = .6. Responde las preguntas que siguen.

Nivel de confianza	Z_α	s_{P_s}	Término del error	LCI	LCS	Ancho del intervalo de confianza
95%	1.96	.0327	.0641	.5359	.6641	.1282
99%	____	____	____	____	____	____
99.9%	____	____	.1076	____	____	____

a) ¿Cuál es la relación entre el tamaño del nivel de confianza y el ancho del intervalo de confianza? Explica.

b) ¿Cómo se calcula el ancho del intervalo de confianza?

c) ¿Cuál es la relación entre el tamaño del nivel de confianza y el tamaño del término del error? Explica.

Conjunto de problemas 8C

8C-1. Siguiendo el procedimiento de cinco pasos para calcular un intervalo de confianza, calcula e interpreta el intervalo de confianza de 95% para los datos siguientes:

X = años de educación n = 226 solicitudes de trabajo

Media = 15 años   Desviación estándar = 1.5 años

8C-2. Vuelve a realizar los últimos tres pasos para calcular un intervalo de confianza para calcular el intervalo de confianza de 99% con los datos del ejercicio 8C-1. Compara los resultados con la respuesta del ejercicio 8C-1 y haz un comentario sobre esta comparación.

8C-3. Tú necesitas calcular un intervalo de las edades medias de adultos con edad para votar en un distrito electoral, para esto obtienes una muestra aleatoria de 272 adultos con edad para votar y determina una edad media de 36.3 años con una desviación estándar de 9.7 años. Siguiendo el procedimiento de cinco pasos, elabora un intervalo de confianza de 99%.

8C-4. Tú has sido elegido por la mayoría de sus compañeros de clase como representante de la University Congress. Mientras intentas obtener una propuesta orientada a una reforma en las votaciones para el congreso, deseas saber cuántos de los 516 compañeros representantes tienen pensado apoyar tu propuesta. Tú seleccionas al azar 112 representantes, descubriendo que 68% apoyan tus reformas. Calcula e interpreta el intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos. Con base en tus resultados, ¿puedes estar 95% seguro de que tu propuesta tendrá un voto mayoritario en el congreso?

8C-5. Tú has sido contratado por William Burns Presidente del Comité de Adquisiciones del Senado. Él quiere saber qué proporción de sus 6 421 funcionarios del partido piensan apoyar su nueva resolución del gasto del presupuesto. Tú seleccionas al azar 279 funcionarios del partido y descubres que 101 apoyan la resolución. A fin de mantener informado al senador Burns de las últimas cifras de las encuestas, calcula e interpreta el intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos.

8C-6. Tú realizarás una encuesta para determinar el porcentaje de empleados de un negocio que apoya ampliar las horas hábiles del negocio durante la temporada de vacaciones próxima. Los resultados se reportarán con una seguridad de 95% y un término del error de 3 puntos porcentuales. ¿Qué tamaño de la muestra de empleados debes obtener? (*Sugerencia:* en este punto P_s se desconoce, pero supongamos que será .5.)

8C-7. Completa la tabla siguiente, donde $n = 550$ y $P_s = .4$. Responde las preguntas que siguen.

Nivel de confianza	Z_α	s_{P_s}	Término del error	LCI	LCS	Ancho del intervalo de confianza
95%	1.96	.0209	.0410	.3590	.4410	.0820
99%	____	____	____	____	____	____
99.9%	____	____	.0690	____	____	____

a) ¿Cuál es la relación entre el tamaño del nivel de confianza y el ancho del intervalo de confianza? Explica.

b) ¿Cómo se calcula el ancho del intervalo de confianza?

c) ¿Cuál es la relación entre el tamaño del nivel de confianza y el tamaño del término del error? Explica.

Conjunto de problemas 8D

8D-1. Siguiendo el procedimiento de cinco pasos para calcular un intervalo de confianza, calcula e interpreta el intervalo de confianza de 95% para los datos siguientes de una muestra aleatoria de estudiantes en un estudio en una universidad.

X = promedio general de calificaciones $\quad$ n = 314 estudiantes

Media = 2.9 puntos $\quad$ Desviación estándar = .3 puntos

8D-2. Vuelve a realizar los últimos tres pasos del procedimiento descrito para calcular el intervalo de confianza de 99% con los datos del ejercicio 8D-1. Compara los resultados con la respuesta del ejercicio 8D-1 y comenta sobre esta comparación.

8D-3. Tú deseas calcular una estimación de un intervalo de las puntuaciones medias del Graduate Record Examination (GRE) entre estudiantes graduados de nuevo ingreso en una universidad urbana grande. Tú haz reunido una muestra aleatoria de 129 estudiantes y encuentras una puntuación GRE media de 1 200 puntos con una desviación estándar de 60 puntos. Siguiendo el procedimiento de cinco pasos, determina el intervalo de confianza de 99%.

8D-4. El gobernador Jackson Fitzpatrick busca un segundo periodo gubernamental. Las encuestas entre 324 votantes registrados indican que casi 61% apoyan su reelección. Si la elección fuera hoy, ¿podría asegurar su victoria el gobernador Fitzpatrick? Explica empleando un intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos.

8D-5. El profesor William Carr considera postularse para el Senado de Estados Unidos. Él te ha comisionado para encuestar 126 votantes al azar, y tú descubres que 54% lo apoyan en este momento. Si la elección fuera hoy, ¿podrías asegurar la victoria del profesor Carr? Explica empleando un intervalo de confianza de 95%. Sigue el procedimiento de cinco pasos.

8D-6. Como estadístico, en una compañía de correduría, tú realizarás una encuesta para determinar el porcentaje de inversionistas satisfechos con sus consejeros financieros. Tú deseas reportar los resultados con una seguridad de 99% y un término de error de 2 puntos porcentuales. ¿Qué tamaño de la muestra necesitas seleccionar? (*Sugerencia:* en este momento P_s se desconoce, pero supongamos que será .5.)

8D-7. Completa la tabla siguiente, donde $n = 185$ y $P_s = .7$. Responde las preguntas que siguen.

Nivel de confianza	Z_α	s_{P_s}	Término del error	LCI	LCS	Ancho del intervalo de confianza
95%	1.96	.0337	.0661	.6339	.7661	.1322
99%	____	____	____	____	____	____
99.9%	____	____	.1112	____	____	____

a) ¿Cuál es la relación entre el tamaño del nivel de confianza y el ancho del intervalo de confianza? Explica.

b) ¿Cómo se calcula el ancho del intervalo de confianza?

c) ¿Cuál es la relación entre el tamaño del nivel de confianza y el tamaño del término del error? Explica.

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 8

Si en tu clase se utilizan las aplicaciones computacionales opcionales de este texto, abre los ejercicios del capítulo 8 en el sitio en la red de *The Statistical Imagination* en www.mhhe.com/richey2. Los ejercicios son sobre el cálculo de intervalos de confianza en *SPSS para Windows* con énfasis sobre la importancia de examinar los efectos del sesgo en los cálculos. Además, el apéndice D de este texto proporciona un repaso breve de las secuencias de comandos en *SPSS* para procedimientos analizados en este capítulo.

CAPÍTULO 9

Prueba de hipótesis I: los seis pasos de la inferencia estadística

RESUMEN DEL CAPÍTULO

Introducción: teoría científica y desarrollo de hipótesis comprobables

Como vimos en el capítulo 1, los estadísticos inferenciales se calculan para demostrar relaciones de causa efecto, y para probar hipótesis y teorías científicas. Una teoría se prueba haciendo predicciones específicas acerca de datos. Por ejemplo, al estudiar desórdenes civiles como disturbios, podríamos proponer una "teoría de protesta", que indique que el comportamiento alborotador se estimula por la práctica de una autoridad represiva, como los incidentes de brutalidad policiaca. Esta teoría proporciona conceptos (como desorden civil, disturbios, represión y protesta) así como ideas acerca de cómo funciona el mundo social (como la idea que una autoridad estatal abusiva conduce a la protesta). Más importante aún, una teoría dirige nuestros pensamientos tal que podamos concebir un conjunto de proposiciones acerca de relaciones entre variables medidas. Por ejemplo, si la autoridad represiva conduce a la protesta, las medidas de protesta deben ser altas (como alta incidencia de desordenes civiles y disturbios) en situaciones donde la represión estatal también es alta (como mucha brutalidad policiaca). En resumen, una teoría es un conjunto de ideas interrelacionadas y organizadas de manera lógica que explica un fenómeno de interés y permite probar la solidez de estas ideas contra hechos observables. El proceso de determinar qué hechos son válidos y cuáles no lo son se denomina prueba de hipótesis, que es el tema de este capítulo.

Relacionar las ideas bien organizadas de una teoría con eventos reales constituye el lado creativo de la imaginación estadística. Para esto, se requiere "ver al futuro", al menos con respecto a cómo se comportará un fenómenos de interés. Una teoría "motiva" hipótesis al incitarnos a pensar en términos de demostrar nuestras afirmaciones. Una **hipótesis** es *una predicción acerca de la relación entre dos variables, que afirma que las diferencias entre las mediciones de una variable independiente corresponderán a diferencias entre las mediciones de una variable dependiente.* Una hipótesis es una predicción que necesita corroboración mediante observación y análisis de datos. (La palabra *hipótesis* comparte su raíz con la palabra *hipotético*, que significa "imaginemos por un momento".) Las hipótesis ponen las ideas teóricas en práctica al estipular que dada la lógica de la teoría, deberán aparecer hechos observables de una cierta manera. Si los datos resultan como sugiere la teoría, esa teoría puede ser una explicación útil del fenómeno de interés. "En cambio, si las predicciones no se ven respaldadas por los hechos, esa teoría no es sólida y se debe rechazar o modificar de manera importante. En un sentido amplio, las pruebas de hipótesis tienen como propósito corroborar una teoría."

Hipótesis Predicción acerca de la relación entre dos variables, que afirma que las diferencias entre las mediciones de una variable independiente corresponderán a diferencias entre las mediciones de una variable dependiente.

El propósito teórico de una prueba de hipótesis es corroborar la teoría probando ideas contra hechos.

Realización de predicciones empíricas

Un desafío de la prueba científica de hipótesis es averiguar cómo hacer predicciones empíricas. En realidad, hacer ese tipo de predicciones no es difícil. Lo hacemos con mucha frecuencia en nuestras vidas cotidianas. Las siguientes son algunas preguntas que quizá formulas y algunas hipótesis que puedes probar hoy, junto con las observaciones empíricas que puedes hacer para probarlas.

Pregunta 1. ¿Debo llevar un paraguas hoy?

Hipótesis: un aumento en nubosidad (variable independiente) está asociado con un aumento en la lluvia (variable dependiente).

Observación: mira por la ventana.

Pregunta 2. ¿Obtendré mejores calificaciones en mis exámenes de estadística si estudio más?

Hipótesis: entre mayor sea el tiempo de estudio (variable independiente), mejor será la calificación (variable dependiente).

Observación: estudia más y ve qué pasa.

Pregunta 3. ¿Me lastimaré si me paro frente a un autobús en movimiento?

1. *Una hipótesis de la ciencia física:* dadas las leyes físicas conocidas que dos objetos no pueden ocupar el mismo espacio al mismo tiempo y que los objetos con mayor masa moviéndose a velocidades mayores desplazarán a los objetos de menor masa y velocidad, el autobús masivo arrollará a la persona menos masiva.

Observación: párate frente a un autobús.

2. *Una hipótesis de la ciencia del comportamiento humano:* habrá una mayor incidencia de lesiones de cabeza y cuerpo (variable dependiente) en las víctimas de choques de autobuses con peatones que en personas que no han experimentado esos choques (variable independiente).

Observación: analiza registros de hospitales. (A esto se le llamaría estudio retrospectivo, debido a que estaríamos viendo experiencias pasadas.)

Sugerencia: prueba esto con los datos retrospectivos o con maniquíes de choques.

La última pregunta ilustra el hecho que en realidad empleamos cantidades enormes de información corroborada de manera empírica cuando vamos de un sitio a otro en nuestras vidas cotidianas. No tenemos que pararnos frente a un autobús en movimiento para saber que ocurrirán lesiones.

Inferencia estadística

Inferir es sacar una conclusión acerca de algo. La **inferencia estadística** implica *sacar conclusiones acerca de una población con base en estadísticos de una muestra.* Como estudiamos en el capítulo 2, las inferencias estadísticas deben tomar en cuenta el error de muestreo. Y como aprendimos en el capítulo 7, se puede esperar que una medición hecha en una muestra esté ligeramente errada del parámetro real de la población.

Inferencia estadística Sacar conclusiones acerca de una población con base en estadísticos de una muestra.

Saber cómo varía el error de muestreo y cómo se puede predecir con distribuciones muestrales es la clave para comprender la inferencia estadística. Ilustremos esto con un ejemplo simple. Supongamos que estamos sentados en clase un día cuando un desconocido bien parecido entra al salón vestido como vaquero de rodeo o músico de country and western. Él lleva puesto un sombrero emplumado, una camisa con lentejuelas, corbata de lazo con nudo con turquesas, un cinturón ancho con hebilla con forma del estado de Texas y las botas puntiagudas más brillantes que hayamos visto. Se presenta como Billy "Tex" Cooper de Dallas, Texas.

Tex dice que escuchó que los cursos de estadística tratan de la predicción del futuro y que le gustaría aprender más acerca de esto debido a que quiere ser apostador profesional. Saca un par de dados de su bolsillo y propone un juego. Él dice, "Les propongo algo. Lanzaré los dados. Si sale 2, entonces todos ustedes me pagarán un dólar. Si sale cualquier otra combinación —3, 4, 5, etc.—, entonces les pagaré a cada uno de ustedes un dólar". Le pedimos ver sus dados y parecen legítimos.

Parece un buen trato. ¿Por qué? Porque nuestro conocimiento de las probabilidades nos indica que tendremos una buena ventaja. La teoría de la probabilidad nos permite averiguar *qué pasa a largo plazo*, es decir, cuando se lanzan los dados una y otra vez. Si podemos proyectar de manera matemática la frecuencia de cada lado de un dado, podemos determinar si algunas combinaciones de dos dados aparecerán con más frecuencia que otras. En otras palabras, podemos producir la distribución muestral para el evento de lanzar dos dados.

Como aprendimos en el capítulo 7, la distribución muestral es una descripción matemática de todos los resultados posibles y la probabilidad (p) de cada uno. La figura 9-1 es una matriz que ilustra la distribución muestral para el lanzamiento de dos dados, es decir, todas las combinaciones posibles y la probabilidad de cada una. La matriz revela que hay 36 resultados posibles. Cuando en el primer dado sale 1, en el segundo dado puede salir 1 o 2 o 3 o 4 o 5 o 6, y así sucesivamente. Después, observando la esquina superior izquierda de la matriz, tenemos que cuando en los dos dados sale 1, la combinación es 2, el menor resultado posible. Esta es la única manera posible de obtener un resultado de 2, y hay un total de 36 resultados posibles; por tanto, la probabilidad de sacar un 2 con dos dados es 1 de 36. La combinación de 3 puede ocurrir de dos formas; el primer dado 1 y el segundo 2 y el primer dado 2 y segundo dado 1; por tanto, la probabilidad de sacar un 3 es 2 de 36. El resto de la matriz proporciona las combinaciones para todos los otros lanzamientos de los dados, indicando cada resultado posible (de 2 a 12) en el lado derecho de cada combinación.

La matriz de la figura 9-1 se resume en la tabla 9-1, que proporciona una descripción clara de la distribución muestral para un lanzamiento individual de dos dados. Es una lista de todos los resultados y la probabilidad (p) de cada uno. Observa la forma familiar casi normal de la gráfica de asteriscos o "curva" de estos resultados. Al igual que las medias y proporciones que determinamos en el muestreo repetido en el capítulo 7, esta distribución tiene un máximo (en este caso una puntuación de 7) con frecuencias gradualmente menores a medida que nos movemos hacia las "colas" de la curva. Tomando nota que 7 ocurre 6/36 de las veces, podemos calcular rápidamente las probabilidades de otros resultados alejándonos del 7 y restando 1. Por tanto, la probabilidad de sacar 6 es 5/36, la de sacar 8 es 5/36, la de sacar 5 es 4/36, y así sucesivamente. (¡La figura 9-1 y la tabla 9-1 valen el costo de la colegiatura de este curso! Serán invaluables si te gusta jugar a los dados o al Monopolio, backgammon o cualquier otro juego de azar que use dados.)

FIGURA 9-1
Matriz de todas las combinaciones posibles en el lanzamiento de dos dados

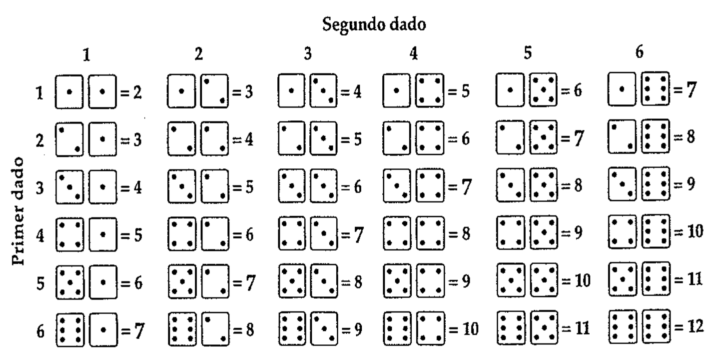


TABLA 9-1 | Distribución muestral del lanzamiento de dos dados

Resultado posible	Fracción p	Diagrama de asteriscos
2	1/36 = .0278	*
3	2/36 = .0556	**
4	3/36 = .0833	***
5	4/36 = .1111	****
6	5/36 = .1389	*****
7	**6/36 = .1667**	******
8	5/36 = .1389	*****
9	4/36 = .1111	****
10	3/36 = .0833	***
11	2/36 = .0556	**
12	1/36 = .0278	*
Totales	36/36 = 1.0001*	

*El total no suma 1 debido al error de redondeo.

Volviendo al caso de Tex, debemos tener una ventaja. La probabilidad de que gane con un solo lanzamiento con un 2 es 1/36 o .0278. Utilizando la regla de la adición para eventos independientes, la probabilidad de que ganemos en un solo lanzamiento es 35/36 o .9722, las probabilidades de todos los otros resultados combinados, o simplemente, 1 − .0278. Deberíamos sentirnos muy bien en este momento. Aquí tenemos a este encantador muchacho, vestido de manera elaborada, que dice que es un tanto ignorante acerca de las probabilidades, que son la esencia del juego y las apuestas. (¡Y ahora nosotros estamos pensando que somos muy inteligentes debido a que sobrevivimos el curso de estadística hasta este punto!)

Pero supongamos que él lanza los dados y sale 2. ¡Espera un momento! Algo no está bien aquí. En este punto quizá sospechemos que Tex es un estafador y que los dados están cargados (cargados de tal forma que sólo sale 2) o que está utilizando un juego de manos para cambiar a un par cargado antes de lanzarlos. ¿Por qué surgen nuestras sospechas? En parte debido a que su aparición en clase es inesperada y tenemos prejuicios contra la gente que viste de manera diferente y nos anima a apostar nuestro dinero. Pero también sospechamos debido a que la teoría de la probabilidad nos indica qué esperar del lanzamiento de dos dados aún antes de su lanzamiento. Hemos determinado que *un par legítimo de dados* saldrá 2 sólo .0278 de las veces. Esto es menor que 3/100. En otras palabras, si 100 apostadores hubieran aparecido y lanzado sus dados *y todos los dados estuvieran marcados y pesados legítimamente*, menos de tres hubieran sacado un 2. ¡Tex aparece de la nada y lo hace! Difícil de creer. No podemos estar absolutamente seguros (quizá Tex en realidad sea afortunado), pero sospechamos que "nos estafaron" y que Tex es un truhán. Es probable que concluyamos que el resultado fue *el efecto* de los dados cargados, no el efecto de muy buena suerte en la caída aleatoria de dados legítimos.

Aunque no seguimos procedimientos estrictos, probamos la hipótesis de que Tex es un tramposo. Todas las pruebas de hipótesis tienen los elementos de este evento. Primero, se formula una pregunta: ¿Es honesto o un tramposo Tex? Segundo, las predicciones de resultados muestrales normales se hacen con una distribución muestral: los dados legítimos previsiblemente caen como se indica en la tabla 9-1. Tercero, observamos los estadísticos de una muestra individual y calculamos la probabilidad de ocurrencia contra la distribución muestral: Tex saca un 2. Con dados legítimos la probabilidad de este resultado es inusualmente pequeña, sólo .0278. Cuarto, decidimos si este resultado es el efecto del error muestral normal o el efecto de los dados cargados. Los lanzamientos de Tex parecen muy inusuales para dados normales. Por tanto, rechazamos la suposición que los dados son legítimos y concluimos que sus lanzamientos son el efecto de dados cargados. Por último, enunciamos una conclusión: ¡Tex es un tramposo! En términos simples, los dados no se comportan como lo hacen los dados legítimos por tanto no pueden ser legítimos.

Además del propósito de corroborar una teoría, todas las pruebas de hipótesis tienen un fin estadístico. Responden las preguntas generales: si un resultado muestral observado parece normal, ¿es el efecto del error de muestreo? O ¿es muy inusual el resultado y por tanto el efecto de algo más? La distribución muestral proporciona un patrón de medida contra el cual se comparará un estadístico muestral individual observado para determinar si es inusual. Cuando la probabilidad de un resultado muestral es baja, como .0278, rechazamos la noción que este resultado se debe simplemente al error de muestreo. La única razón por la que podemos llamar tramposo a Tex es que sabemos cómo caen los dados *legítimos*. Este conocimiento se describe con certeza matemática en la distribución muestral de la tabla 9-1.

Objetivo estadístico de una prueba de hipótesis Determinar si los resultados muestrales indican: 1) efectos reales en la población o 2) un error de muestreo.

Cuando decidiste si llevar una sombrilla hoy, tú seguiste el mismo procedimiento lógico:

Pregunta de investigación: ¿va a llover?

Hipótesis: si hay un cielo obscuro y nuboso, hay una buena posibilidad de que llueva.

Observación: el cielo es color azul claro. Esto no se ajusta a las predicciones empíricas de la hipótesis. La probabilidad de que llueva sin nubes es muy baja.

Conclusión: rechaza la noción de que va a llover. Deja la sombrilla en la casa.

En esencia, la lógica de la prueba de hipótesis implica decidir si aceptar o rechazar una afirmación con base en observaciones de datos. Comparamos predicciones con evidencia empírica.

La importancia de las distribuciones muestrales para pruebas de hipótesis

Ahora tenemos una idea general de la lógica de las pruebas de hipótesis. Tomemos un ejemplo empleando los estadísticos que aprendimos en capítulos anteriores e iniciemos la integración de los detalles de las pruebas de hipótesis. Supongamos que realizamos un estudio de atletas de preparatoria. Deseamos examinar si el estereotipo común de que los atletas son "cabezas huecas" —sólo músculos sin cerebro— tiene algún mérito. De manera más específica: en promedio, ¿son los atletas menos inteligentes que otros estudiantes de preparatoria?

Utilizando nuestra imaginación estadística, buscamos una forma para comparar atletas con estudiantes en general. Obtenemos una prueba del cociente intelectual (CI) que tiene un promedio nacional de 100 para todos los estudiantes de preparatoria. Éste es un valor de referencia al cual los atletas de preparatoria se pueden comparar. Nuestra pregunta de investigación es: ¿es menor el CI medio de atletas de preparatoria que el CI medio de todos los estudiantes de preparatoria?

El contar con este valor numérico específico de 100 puntos de CI nos permite probar una hipótesis. Para hacer esto, tomamos una muestra única de 144 atletas de preparatoria y encontramos una media muestral de 99 puntos de CI con una desviación estándar de 12 puntos de CI. Nuestra media muestral está 1 punto debajo del CI promedio. Si no tuviéramos un conocimiento adecuado, de inmediato concluiríamos que, en efecto, los atletas tienen un CI menor y esto apoya el estereotipo de que son "cabezas huecas". Sin embargo, estamos indecisos de extraer esta conclusión, debido a que no calculamos el CI medio *de la población* de todos los atletas de preparatoria. Simplemente empleamos una muestra de 144 de ellos. Sabemos por experiencia propia al muestrear repetidamente una población y calcular medias a partir de ella (capítulo 7) que una segunda muestra de una población es probable que proporcione un resultado ligeramente diferente. Por ejemplo, sería igual de probable obtener una media muestral de 101, que está 1 punto arriba del promedio. Y una tercera muestra proporcionaría incluso otro resultado. En otras palabras, aún si la puntuación media de la *población* de atletas es 100 puntos de CI y no difiere de la de todos los estudiantes de preparatoria, *una muestra* puede resultar ligeramente distinta a 100 debido al error de muestreo esperado. Por tanto, ¿indica algo significativo esta diferencia de –1 punto entre la observación empírica y el valor anticipado o simplemente es un error de muestreo?

Por fortuna, disponemos de una manera para calcular qué tan erradas pueden estar las medias muestrales. En el capítulo 7 aprendimos que si una población tiene una media conocida, μ_X, cuando se toman muestra repetidamente de la población, las medias muestrales, $\bar{X}$, estarán centradas alrededor del valor de μ_X. Para el caso en cuestión, si el CI medio de atletas no es diferente del de todos los estudiantes, las medias muestrales se centrarán alrededor de 100. Además, el error de muestreo es predecible. Cuando el tamaño de la muestra es mayor que 121, un gráfico de las medias de un muestreo repetido tomará la forma de una curva normal. Para describir con precisión la distribución muestral, calculamos el error estándar de las medias empleando el tamaño de la muestra y la desviación estándar de nuestra muestra:

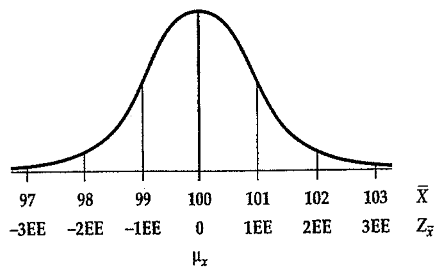
$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{12}{\sqrt{144}} = 1.00 \text{ punto de CI}$$

Ahora trazamos la curva normal para esta distribución muestral (figura 9-2).

¿Qué nos indica la curva de la figura 9-2? Podemos observar que *si el CI medio de la población de atletas estudiantes es en efecto 100*, en un muestreo repetido casi 68% de la veces podemos esperar obtener una media muestral, $\bar{X}$, entre 99 y 101 puntos de CI. Casi 95% de las veces podemos esperar obtener una media muestral entre 98 y 102 puntos de CI.

Ahora sabemos que los resultados muestrales ocurren con la suposición que *no hay diferencia* entre los CI medios de los atletas y de todos los atletas. La distribución muestral proporciona un patrón de medición contra el cual podemos comparar nuestro resultado muestral actual de $\bar{X} = 99$ puntos de CI. Formulamos las preguntas: ¿Es un resultado usual una media muestral de 99 puntos de CI —simplemente error muestral— cuando la media poblacional $\mu_X = 100$ puntos de CI?

FIGURA 9-2
Distribución muestral del CI medio de atletas de preparatoria suponiendo que la media normal es 100 puntos de CI



O, como una explicación alternativa, ¿es una media muestral de 99 muy inusual cuando se muestrea repetidamente de una población con una media de 100? Al igual que hicimos con Tex, podemos calcular la probabilidad de la "extrañeza" de un resultado muestral. En breve haremos el cálculo preciso de la probabilidad de este resultado muestral. Sin embargo, en la figura 9-2 ya podemos observar que una media muestral de 99 no es poco común proviniendo de una población cuyo CI medio es 100. Una diferencia de −1 punto de CI está dentro de 1 error estándar del valor anticipado de 100 puntos de CI. Casi 68% de las medias se calcularán en este rango. Por tanto, concluimos: el CI promedio de atletas no parece diferente del de todos los estudiantes; por tanto, la noción de "cabezas huecas" es un estereotipo falso. Sacamos esta conclusión debido a que la media muestral observada de 99 podría haber resultado fácilmente del error de muestreo.

Con la pregunta de Tex el apostador obtuvimos un resultado muestral poco común de baja probabilidad, por que rechazamos "dados normales". Con nuestra pregunta de si son "cabezas huecas" los atletas de preparatoria obtuvimos un resultado de alta probabilidad conduciéndonos a no rechazar la inteligencia normal. La idea general tras la prueba de hipótesis es la misma para todas las disciplinas científicas, y para todas las aplicaciones estadísticas. Identifica un enunciado de hipótesis que permita predicciones de todos los resultados muestrales posibles en la forma de una distribución muestral. Haz una observación empírica de un solo estadístico muestral y observa cómo se compara contra las predicciones. La habilidad para concebir distribuciones muestrales es la característica central de las pruebas de hipótesis.

Hay muchos tipos de pruebas de hipótesis. Una prueba particular se forma respecto a la distribución muestral que es apropiada para ella. Los detalles de una distribución muestral dependen de cuántas variables se impliquen, de sus niveles de medición, del tamaño de la muestra y de otras características.

Los seis pasos de la inferencia estadística para una prueba de medias de una muestra única grande

Procedamos a definir los detalles de las pruebas de hipótesis utilizando nuestro ejemplo del estereotipo de "cabezas huecas". Para esta situación, empleamos lo que se denomina prueba de medias de una muestra única grande. En este caso, *grande* se define como una muestra con más de 121 casos, lo que nos permite emplear la curva normal como nuestra distribución muestral. En específico, utilizamos una prueba de medias de una muestra única grande cuando se cumplen los criterios siguientes:

Cuando utilizar una prueba de medias de una muestra única grande (distribución Z, n = 121 casos)

En general: resulta útil para probar una hipótesis que establezca que la media de una población es igual a un valor objetivo.

1. Sólo hay una variable.
2. El nivel de medición es de intervalo/razón.
3. Sólo hay una medición y una población.
4. El tamaño de la muestra es mayor que 121 casos.
5. Hay un valor objetivo de la variable para la cual podemos calcular la media de la muestra.

Para el ejemplo de "cabezas huecas", tenemos una sola variable: la puntuación de CI. Es de nivel de medición intervalo/razón. Tenemos una única población de interés: atletas de preparatoria. Nuestra muestra de 144 atletas es mayor que 121. Tenemos un valor de referencia que proporciona una predicción cómo caen los resultados muestrales: un CI normal es una puntuación de 100. Por tanto, podemos emplear una prueba de medias de una muestra única grande.

Cada prueba de hipótesis sigue un mismo proceso lógico que se compone de seis partes. Nos referimos a este proceso como "los seis pasos de la inferencia estadística" o "los seis pasos de prueba de hipótesis". Antes de la prueba misma, debemos prepararnos para ella. El siguiente es un resumen del proceso lógico.

Lista de verificación breve para los seis pasos de la inferencia estadística

PREPARACIÓN DE LA PRUEBA

Enuncia la pregunta de investigación. Traza diagramas conceptuales representando los datos, incluyendo la o las poblaciones y la o las muestras en estudio, las variables (por ejemplo, X = ..., Y = ...,) y sus niveles de medición, y los estadísticos y los parámetros dados o calculados. Formula el procedimiento apropiado de la prueba estadística.

SEIS PASOS

Utilizando el símbolo H para *hipótesis*:

1. Formula H_0 y la H_A y estipula la dirección de la prueba.
2. Describe la distribución muestral.
3. Declara el nivel de significación (α) y la dirección de la prueba y especifica el valor crítico de la prueba.
4. Observa los resultados muestrales reales y calcula los efectos de la prueba, el estadístico de la prueba y el valor p.

5. Toma la decisión de rechazo.
6. Interpreta y aplica los resultados y proporciona las mejores estimaciones en lenguaje común.

Preparación de la prueba

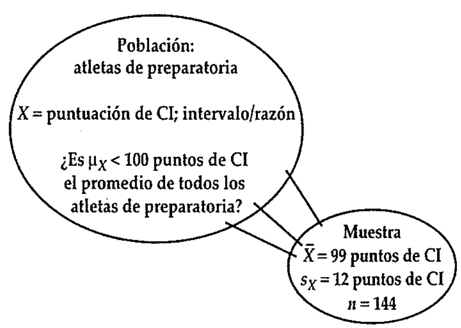
Iniciamos una prueba de hipótesis con su preparación. Esto es similar a reunir los ingredientes de una receta antes de mezclarlos. Primero, identificamos y formulamos una pregunta de investigación, "un objetivo que se pueda enunciar en términos de una hipótesis" (Bailey 1978:10). Como destacamos, las preguntas de investigación se formulan para resolver puntos prácticos o responder preguntas que surgen a partir de la teoría. Para el ejemplo de "cabezas huecas", nuestra pregunta de investigación es: ¿es menor el CI medio de los atletas de preparatoria que el CI medio de todos los estudiantes, el cual es 100 puntos de CI? Luego, identificamos los "datos", incluyendo las variables implicadas, la población, el tamaño de la muestra, los parámetros proporcionados y los estadísticos proporcionados o calculados. Para ayudar a distinguir estos elementos, los organizamos en un diagrama que distinga la población de la muestra. Esto se presenta en la figura 9-3. Observa que los símbolos en el óvalo de la población pertenecen a parámetros y los del óvalo de la muestra pertenecen a estadísticos.

Nuestra prueba de hipótesis es acerca de la población y sus parámetros. Los estadísticos de la muestra sólo son estimaciones de los parámetros de la población y la muestra solamente es una herramienta para hacer inferencias estadísticas acerca de la población. El paso final en la preparación de la prueba es declarar qué prueba estadística se empleará. En este caso estamos utilizando la prueba de medias de una muestra única grande.

Los seis pasos

Paso 1: La hipótesis nula En una prueba de la hipótesis de "cabezas huecas", al igual que hicimos con el caso de Tex el apostador, debemos poner nuestras observaciones estadísticas en un contexto mayor que tome en cuenta el error de muestreo. Debemos encontrar una "hipótesis estadística", un enunciado que proporcione un valor numérico y proyecte una distribución muestral alrededor de él. A esa hipótesis se le denomina **hipótesis nula**, *una hipótesis enunciada de tal manera que sabremos qué resultados estadísticos ocurrirán en el muestreo repetido si esta hipótesis es cierta.*

FIGURA 9-3
Representación de los "datos" para una prueba de hipótesis



Simbolizamos la hipótesis nula como H_0. Para nuestra hipótesis de "cabezas huecas", la hipótesis nula se enuncia como:

H_0: $\mu_{X\,(atletas\ de\ preparatoria)}$ = 100 puntos de CI (la media conocida, μ_X, de todos los estudiantes de preparatoria)

Es decir, el CI medio de los atletas de preparatoria no es diferente del de todos los estudiantes de preparatoria.

Observa la forma de presentación, que es consistente para todas las pruebas de hipótesis. Las hipótesis nulas siempre se relacionan con parámetros de la población, no con estadísticos de una muestra. La población a la que aplica el parámetro, en este caso atletas de preparatoria, se escribe como subíndice del símbolo del parámetro.

No podemos realizar una prueba de hipótesis a menos que podamos identificar una hipótesis nula relacionada con la pregunta de investigación. Para hacer eso preguntamos: ¿Existe una forma para predecir resultados muestrales suponiendo ningún efecto, cero efecto o ninguna diferencia? La definición del diccionario de la palabra *nula* es "*ninguna*". De hecho, esa es la razón por la que empleamos el símbolo H_0 (*H* subíndice cero). Aunque nuestra pregunta de investigación es si los atletas tienen un CI medio menor, nuestra hipótesis nula postula un CI medio igual (es decir, ninguna diferencia). Tenemos que enunciarla de esta manera a fin de elaborar una distribución muestral, la cual describimos en el paso 2.

Otra forma de darle sentido a la palabra *nula* es observar con qué frecuencia probamos una hipótesis examinado un enunciado que "nulifique" la pregunta de investigación, invirtiendo o negando sus palabras. Queremos saber si el CI medio de los atletas de preparatoria es diferente, por que probamos la hipótesis nula de que no es diferente. En este sentido, estamos haciendo una evaluación indirecta de la pregunta de investigación. Tratamos de demostrar algo rechazando otra cosa, al igual que buscamos demostrar que Tex es un tramposo desmintiendo que sea honesto. Es más fácil desmentir algo que demostrarlo. Si deseamos demostrar que todas las aves son color blanco, el enfoque simple es desmentirlo señalando a las aves coloreadas. Con frecuencia, la hipótesis nula se determina invirtiendo las palabras de la pregunta de investigación.

Hipótesis nula (o hipótesis estadística): H_0 Una hipótesis enunciada de tal manera que sabremos qué resultados estadísticos ocurrirán en un muestreo repetido si esta hipótesis es cierta.

Una hipótesis nula es un enunciado de "ningún efecto" o "ninguna diferencia". Nos guía al seleccionar una distribución muestral para una prueba de hipótesis. A menudo, la hipótesis nula es la negación o inversión de la pregunta de investigación, la cual se demuestra rechazando la hipótesis nula.

En la investigación de laboratorio, se establece una línea base de "ningún efecto" con un grupo de control. Por ejemplo, suponga que queremos probar la efectividad de un medicamento para una alergia provocada por el polen. A un grupo experimental de sujetos le damos la medicina real, en tanto que un grupo control recibe una píldora similar de *placebo* o producto inocuo. Los sujetos en los dos grupos se exponen al polen y calculamos los números medios de síntomas de alergia para los dos grupos. El número medio de síntomas entre los sujetos en el grupo control proporciona una medida de la línea base de "ningún efecto", debido a que la píldora de placebo no contiene medicamento. Probaríamos la hipótesis nula de que el medicamento no tiene "ningún efecto", que quiere decir que no hay una diferencia en el número medio de síntomas con o sin el medicamento. Sin embargo, si el número medio de síntomas del grupo experimental es significativamente menor, rechazaríamos "ningún efecto" y concluiríamos que el medicamento *es* efectivo. Para probar una pregunta de investigación, rechazamos la hipótesis nula.

Sin embargo, lo importante acerca de una hipótesis nula es que debe ser una *hipótesis estadística.* Es un enunciado que proporciona una distribución muestral, predicciones de resultados estadísticos como si sacáramos un número infinito de muestras para determinar la naturaleza del error de muestreo. La distribución muestral proporciona una forma de medición para calcular la probabilidad del estadístico, calculada para una muestra que en realidad tomamos.

Retornando al ejemplo de los "cabezas huecas", ¿por qué no elaborar la distribución muestral respecto a la pregunta de investigación misma? Porque la afirmación que el CI medio *de la población* de estudiantes atletas es menor que 100 no proporciona un valor de referencia para la distribución muestral. Si la media de la población es menor que 100, entonces ¿cuál es su valor? Esto aún se tiene que determinar. Además, sería un error terrible usar nuestra media muestral observada de 99 puntos de CI para elaborar la distribución muestral. Si cada uno de varios investigadores toma muestras de 144 atletas, cada investigador obtendrá una media muestral diferente —debido al error de muestreo— y por tanto cada uno tendría una distribución muestral desviada, centrada en un valor aleatorio distinto. Por otro lado, centrar nuestra distribución muestral en 100 puntos de CI proporciona una predicción firme. Al final, podemos o no aferrarnos a esta predicción. O rechazamos esta hipótesis nula o fracasamos en rechazarla.

La hipótesis alternativa, H_A En cada prueba de hipótesis se requiere la hipótesis nula o "sin efecto" a fin de proyectar resultados muestrales. También debemos decidir con anticipación qué concluiremos si rechazamos la hipótesis nula. Este enunciado se denomina **hipótesis alternativa (H_A)**, *la hipótesis que aceptaremos si se rechaza la hipótesis nula.* En general, la hipótesis alternativa es la que aborda directamente la pregunta de investigación, y este es el caso para nuestra pregunta de investigación de los "cabezas huecas":

$$H_A: \mu_{X\,(atletas\ de\ preparatoria)} < 100 \text{ puntos de CI}$$

Es decir, el CI medio de los atletas de preparatoria es menor que el de todos los estudiantes de preparatoria, una cola.

> **Hipótesis alternativa: H_A** Hipótesis que aceptaremos si se rechaza la hipótesis nula. Con frecuencia, la hipótesis nula es un enunciado directo de la pregunta de investigación.

Hipótesis alternativas posibles Para cualquier prueba de hipótesis, existe una sola hipótesis nula y una sola hipótesis alternativa. No obstante, hay tres hipótesis alternativas *posibles* y para distinguirlas utilizamos el término *dirección.* Cuando anticipamos una dirección, estamos afirmando que tenemos una razón para creer que la media muestral caerá arriba o abajo de la media hipotética de 100. Por razones que analizaremos en breve, también empleamos los términos *de una cola* y *de dos colas* para referirnos a las colas en la curva de la distribución muestral. Éstas son las tres hipótesis alternativas posibles para nuestra hipótesis de "cabezas huecas".

Opción 1: hipótesis alternativa de una cola, en la dirección positiva:

$$H_A: \mu_{X\,(atletas\ de\ preparatoria)} > 100 \text{ puntos de CI}$$

Es decir el CI medio de los atletas de preparatoria es *mayor que* el de todos los atletas estudiantes de preparatoria, una cola.

Aquí, *positiva* significa en el lado superior del CI medio. Utilizaremos una curva de distribución muestral para calcular la probabilidad de nuestro resultado muestral. Cuando predecimos la dirección positiva, calcularemos puntuaciones Z positivas en la cola de la curva a la derecha arriba de la media.

Opción 2: hipótesis alternativa de una cola, en la dirección negativa:

$$H_A: \mu_{X\,(atletas\ de\ preparatoria)} < 100 \text{ puntos de CI}$$

Es decir, el CI medio de los atletas de preparatoria es *menor que* el de todos los estudiantes de preparatoria, una cola.

Aquí, *negativa* significa en el lado inferior del CI medio. Cuando utilizamos una curva de distribución muestral para calcular la probabilidad de nuestro resultado muestral, calcularemos puntuaciones Z negativas en el lado izquierdo o cola izquierda de la curva.

Opción 3: hipótesis alternativa de dos colas, no direccional:

$$H_A: \mu_{X\,(atletas\ de\ preparatoria)} \neq 100 \text{ puntos de CI}$$

Es decir, el CI medio de los atletas de preparatoria *no es igual al* de todos los estudiantes de preparatoria, dos colas.

La tercera opción es no direccional. No propone que el CI medio de los atletas sea mayor o menor, sólo diferente. En el cálculo de la probabilidad de los resultados emplearemos los dos lados o colas de la curva de la distribución muestral.

Cuando probamos una hipótesis, debemos decidir cuál de estas tres hipótesis alternativas aplica. Probamos sólo una de ellas. Esa determinación se hace con base en teoría o en consideraciones prácticas. El estereotipo común acerca de "cabezas huecas" propone que son menos inteligentes, por tanto elegimos la opción 2. Pero otro investigador quizá proponga que los atletas de preparatoria tienen un CI medio mayor, debido a que se requiere ser inteligente para cumplir con los estudios y practicar un deporte. Este investigador estipula H_A en la dirección positiva como en la opción 1. Un tercer investigador puede estar al tanto de las dos teorías y quiera resolver el debate proponiendo una H_A no direccional. Este investigador estipula la H_A como la opción 3. Aunque hay tres hipótesis alternativas opcionales para cualquier prueba de hipótesis, debemos elegir sólo una. Además, por razones que aclararemos más adelante, esa opción se elige antes de observar los datos muestrales en el paso 4. Por último, al afirmar que no se puede predecir una dirección con anticipación, una prueba de dos colas es más conservadora que una prueba de una cola. Es más difícil rechazar H_0 y aceptar H_A, cuando se realiza una prueba de dos colas. Las razones de esto se presentan en el capítulo 10.

Para establecer la dirección de una prueba estadística, examina con detenimiento la pregunta de investigación. Si hay palabras que sugieran una direccionalidad positiva (*mayor que, más que, aumento, más pesado que, más largo que, más rápido, ganancia*), se debe realizar una prueba de una cola. Si hay palabras que sugieran una direccionalidad negativa (*menor que, disminuye, perder, más bajo que, más lento que*), se debe realizar una prueba de una cola. Por supuesto, cuando no se estipula una dirección, utilizamos una prueba de dos colas.

Decisión sobre la dirección de una prueba de hipótesis: Formula la hipótesis alternativa en una de tres formas

1. De una cola en la dirección positiva

El contenido de la pregunta de investigación incluye términos como *mayor que, más, aumento, más rápido, más pesado* y *ganancia*.	→	Utiliza una prueba de una cola, positiva, en la hipótesis alternativa y un signo >.

2. De una cola en la dirección negativa

El contenido de la pregunta de investigación incluye términos como *menor que, menos, disminuye, más lento, más ligero* y *pérdida*.	→	Utiliza una prueba de una cola, negativa, en la hipótesis alternativa y un signo <.

3. De dos colas, no direccional

El contenido de la pregunta de investigación incluye enunciados acerca de dirección o simplemente afirma desigualdad.	→	Utiliza una prueba de dos colas, neutral, en la hipótesis alternativa y un signo ≠.

Paso 2: Describe la distribución muestral El segundo paso en una prueba de hipótesis es describir la distribución muestral. Para una prueba de hipótesis, la distribución muestral es una descripción de todos los resultados posibles y la probabilidad de cada resultado *cuando H_0 es cierta*. La distribución muestral se elabora respecto al parámetro hipotético de la hipótesis nula. Si es cierto que el CI medio de la población de atletas es igual a 100, entonces el muestreo repetido de esta población y una gráfica de las $\overline{X}$ produce una curva de distribución normal cuando $n > 121$. La descripción completa de la distribución muestral para la hipótesis de "cabezas huecas" se formula así: si H_0 es cierta y se toman muestras repetidas de tamaño 144 de la población de atletas de preparatoria, las medias muestrales ($\overline{X}$) estarán centradas en 100 como una distribución normal con un error estándar:

$$s_{\overline{X}} = \frac{s_X}{\sqrt{n}} = \frac{12}{\sqrt{144}} = 1.00 \text{ punto de CI}$$

Al terminar un ejercicio, trazamos la curva de la distribución muestral como en la figura 9-2. La transición lógica entre los pasos 1 y 2 es como sigue: el paso 1 proporciona un enunciado, H_0, que permite predicciones precisas de resultados muestrales. El paso 2 imagina (es decir, hipotetiza) que este enunciado es cierto y describe estas predicciones muestrales. Además, esta descripción de resultados muestrales es exhaustiva. Se presenta cada resultado muestral. Y empleando las habilidades de partición que aprendimos en el capítulo 6, podemos calcular la probabilidad de cualquier combinación de resultados. Una vez más, la distribución muestral es la esencia de la prueba de hipótesis y ni siquiera tenemos que contar frijoles para obtenerla. Los matemáticos nos proporcionan el conocimiento de las formas de las curvas de las distribuciones muestrales y las fórmulas del error estándar requerido para trazarlas con precisión. La fórmula del error estándar toma en cuenta el tamaño de la muestra, que es un factor importante en el tamaño de los errores de muestreo.

Paso 3: El nivel de significación En el paso 3, establecemos un nivel de significación (simbolizado por la letra griega alfa, α). Recuerda del capítulo 8, que nos referimos al nivel de significación como el nivel del error esperado. Regresaremos a la noción de error más adelante en este capítulo. Por ahora, enfoquémonos en una función importante del nivel de significación en los seis pasos: ayuda a determinar si rechazar la hipótesis nula (H_0) o fallar en rechazarla. Para la prueba de hipótesis, el nivel de significación es la cantidad de probabilidad crítica que define qué tan inusual debe ser un resultado muestral para rechazar el valor del parámetro proyectado en H_0. Por ejemplo, ¿qué tan alejado del CI normal de 100 debe caer el CI medio muestral de los atletas, antes que concluyamos que su CI medio es significativamente menor? ¿Será suficiente una diferencia de 1 punto debajo de 100, o 2 puntos o 3 puntos, etcétera? Sin embargo, el nivel de significación se presenta como una probabilidad en una curva de distribución muestral. Esto nos permite usar tablas estándar, como la tabla de la curva normal, para calcular probabilidades. No tenemos una tabla para diferencias en puntos de CI.

Es muy común en la investigación de encuestas de ciencias sociales establecer $\alpha = .05$. Hacemos esto en el paso 3. En el paso 4 calcularemos la probabilidad de obtener una media muestral de 99 puntos de CI. Si esta probabilidad es baja, menor que nuestra α de .05, rechazaremos H_0 que la media es 100 y aceptaremos la hipótesis alternativa (H_A) que es menor. Si la probabilidad de obtener una media de 99 puntos de CI resulta ser alta (es decir, mayor que .05), concluiremos que no es inusual proviniendo de una población con una media de 100. "Fracasaremos en rechazar" H_0. Como no tenemos una tabla para puntos de CI, debemos calcular una puntuación Z a fin de utilizar la tabla de la curva normal para obtener una probabilidad de ocurrencia.

Paso 4: Observa la muestra real: calcula los efectos de la prueba, el estadístico de prueba y el valor *p* En el paso 4 finalmente observamos nuestra muestra. Observamos la media muestral y la comparamos con el valor hipotético de 100. Para determinar la probabilidad de ocurrencia, calculamos una puntuación Z para transformar puntos de CI en errores estándar. Luego llevamos la puntuación Z a la tabla de la curva normal para obtener la probabilidad de ocurrencia del resultado muestral.

Para calcular la puntuación Z, tomamos la diferencia entre el valor del estadístico de la muestra y el valor del parámetro proyectado por H_0. Esta diferencia se denomina **"efecto de la prueba"**. Con nuestra hipótesis de "cabezas huecas", el efecto de la prueba es –1 punto de CI:

$$\overline{X} - \mu_X = 99 - 100 = -1 \text{ punto de CI}$$

Efecto de la prueba de una prueba de hipótesis Diferencia entre el valor del estadístico de la muestra y el valor del parámetro proyectado por la hipótesis nula.

El efecto de la prueba es una puntuación de desviación. Aborda la pregunta: ¿cuánto se desvía el valor del estadístico de la muestra del valor del parámetro de la población proyectado por la hipótesis nula?

Un efecto de prueba es una puntuación de desviación. Recuerda del capítulo 5 que una puntuación de desviación es la diferencia (o distancia) entre la media en el centro de una curva normal y algún punto (es decir, puntuación) en el eje horizontal o X. También recuerda que las puntuaciones de desviación se expresan en la unidad original de medición de la puntuación bruta. En el ejemplo anterior, esto sería puntos de CI. Para calcular la probabilidad de un efecto de prueba, debemos estandarizar la puntuación —transformarla en unidades estándar de desviación— tal que podamos utilizar tablas de probabilidad como la tabla de la distribución normal. (¡No tenemos una tabla con puntos de IC en ella!) Para pruebas de hipótesis, estas puntuaciones estandarizadas se expresan en unidades de error estándar. Un **estadístico de prueba** que se empleará en conjunto con curvas de probabilidad y tablas estadísticas de probabilidad es *una fórmula para medir efectos estadísticos de prueba en unidades de error estándar.*

Estadístico de prueba Fórmula para medir efectos estadísticos de una prueba en unidades de error estándar. Estas fórmulas se utilizan en conjunto con las curvas de probabilidad y las tablas estadísticas del apéndice B.

La distribución muestral para nuestra hipótesis de "cabezas huecas" es la curva normal y nuestro estadístico es una puntuación Z. Los efectos estadísticos de prueba por lo general se miden en el numerador del estadístico de prueba y luego se estandarizan dividiéndolos entre el error estándar. Este es el caso para una prueba de medias de una muestra única grande.

Cálculo del estadístico de prueba para una prueba de medias de una muestra única

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}}$$

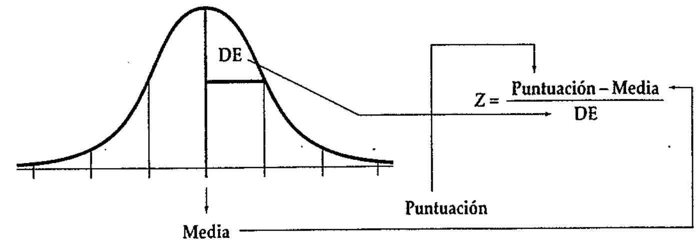
$$= \frac{\text{el efecto de prueba}}{\text{error estándar}} = \begin{array}{l}\text{número de la diferencia de errores} \\ \text{estándar (EE) entre lo que se observó} \\ \text{y lo que se hipotetiza}\end{array}$$

donde

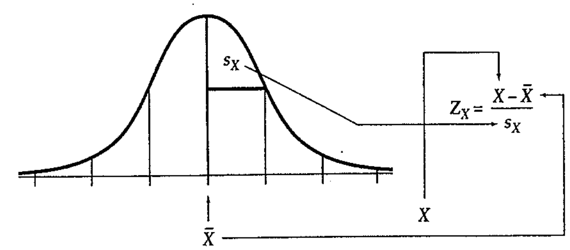
$Z_{\bar{X}}$ = número de errores estándar que se desvía una media muestral ($\bar{X}$) de la media poblacional hipotética (μ_X)

$s_{\bar{X}}$ = error estándar estimado de la distribución muestral de medias

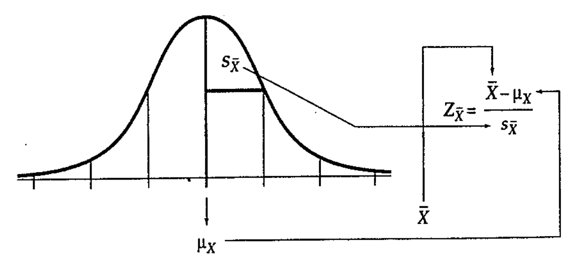
Una mirada a los términos de esta fórmula en relación con la curva de distribución es informativa. Cualquier puntuación Z (repasa el capítulo 6) es una medida de la desviación ("que tan alejado" cae el estadístico muestral observado) de un valor esperado. Cualquier curva de la puntuación Z tiene una media y una desviación estándar (DE) y una puntuación de intervalo/razón medida a lo largo de su eje horizontal.



Por ejemplo, como analizamos en el capítulo 6, cuando la medición de intervalo/razón es una puntuación bruta X, entonces Z_X se calcula como sigue:



Los elementos de una distribución normal de medias muestrales son como sigue:



Ten cuidado en distinguir el error estándar de esta distribución de medias, $s_{\bar{X}}$, de la desviación estándar de la distribución de la puntuación bruta, s_X.

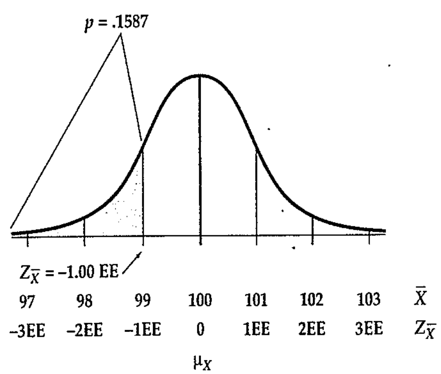
Antes de calcular el estadístico de prueba, estimemos su valor a ojo. Observa la figura 9-2 en la página 274. Podemos observar que un resultado muestral de 99 puntos de CI está 1 error estándar debajo de la media hipotética de 100; por tanto,

$$\text{Estadístico de prueba} = Z_{\overline{X}} = \frac{\overline{X} - \mu_X}{s_{\overline{X}}} = \frac{99 - 100}{1.00} = -1.00 \text{ EE}$$

El valor p Para determinar si un efecto de prueba es lo suficientemente grande para conducirnos a rechazar una hipótesis nula, debemos calcular la probabilidad de su ocurrencia. Hicimos esto con la hipótesis de Tex el apostador. Determinamos que la probabilidad de sacar *unos* es .0278. Esta probabilidad de ocurrencia *inusualmente baja* para dados legítimos nos condujo a rechazar su honestidad. De manera similar, para nuestra hipótesis del estereotipo de "cabezas huecas", podemos determinar con precisión la probabilidad del efecto observado. Esta probabilidad se denomina valor *p*. El **valor *p* de la prueba de hipótesis** es *una medida de la rareza de un resultado muestral cuando la hipótesis nula es cierta*. En general, el valor *p* es el cálculo siguiente:

> *p* [un resultado muestral tan inusual como o más inusual que el observado cuando la hipótesis nula es cierta]

Para distribuciones muestrales que se ajustan a curvas de probabilidad, como la curva normal, el valor *p* se calcula como un área en una o en las dos colas de la curva. Si la hipótesis alternativa (H_A) es una prueba de una cola en la dirección negativa "menor que", como nuestra hipótesis de "cabezas huecas", el valor *p* se calcula como el área en la curva del valor medio muestral observado de 99 puntos de CI y más allá hacia la izquierda. Para evitar confusión, sombreemos el área de referencia del valor *p*:



El valor *p* de una prueba de hipótesis Medición de la rareza de un resultado muestral cuando la hipótesis nula es cierta.

En curvas de probabilidad, como la curva normal, el valor *p* será un área en la cola de la curva.

Utilizamos la tabla de la curva normal (tabla estadística B del apéndice B) para obtener su valor numérico. El área sombreada es un área de columna tipo C. En la columna A de la tabla ubicamos una puntuación Z de 1.00 e imaginamos un signo negativo puesto que nuestro valor del estadístico de prueba es –1 EE. En la columna C encontramos que la proporción del área en la cola es .1587. Por tanto, declaramos nuestro valor *p* como sigue:

> Valor *p*: *p* [sacar una muestra con una media ($\overline{X}$) tan inusual como o más inusual que 99 cuando la media poblacional verdadera (μ_X) es 100] = .1587.

¿Qué sugiere la frase "tan inusual como o más inusual qué? Al calcular el valor *p* en la cola de la curva, la palabra *o* es una clave para utilizar la regla de la adición de la probabilidad y lo que hacemos es sumar las probabilidades. Nuestro valor *p* es la probabilidad de obtener una muestra con una media tan inusual como 99 más la probabilidad de cualquier resultado más inusual, como un CI medio de 98 o 97 o 96 o 95, etc. La hipótesis nula (H_0) se rechaza cuando el valor *p* es pequeño. Si hubiéramos tomado sólo la probabilidad de un resultado muestral de 99, sería un área diminuta en la curva, arriba de esa puntuación individual. Hubiéramos rechazado H_0 muy rápido, aunque 99 no es un resultado inusual. Así, una razón para tomar el área más allá del resultado observado es para evitar esta equivocación. La razón real para incluir "o más inusual que" es esta: si consideráramos una media muestral de 99 puntos de CI lo suficientemente inusual para rechazar H_0, con seguridad rechazaríamos H_0 para resultados muestrales más extremos, como 98, 97 o 96. La curva se precipita hacia abajo en la cola. Resultados más allá de 99 puntos de CI tienen probabilidades aún menores.

¿Qué sucede si utilizamos una hipótesis alternativa distinta (H_A)? SI H_A es una prueba de una cola en la dirección positiva "mayor que", el valor *p* se calcula como el área en la curva desde la media muestral observada y más allá hacia la derecha. Si H_A es una prueba no direccional de dos colas, el valor *p* se calcula para incluir las áreas de las dos colas de la curva. Recuerda, el cálculo del valor *p* supone que H_0 es cierta. Si no se estipula una dirección de antemano en H_A, en el muestreo repetido sería tan inusual errar el parámetro hipotético de 100 puntos de CI en el lado alto como en el lado bajo. Puesto que la curva normal es simétrica (es decir, una imagen de espejo en cada lado), para una prueba de dos colas simplemente multiplicamos el área de la columna C por 2. Por ejemplo, si H_A para la hipótesis de "cabezas huecas" hubiera sido simplemente "no igual a 100", nuestro valor *p* hubiera sido (.1587) (2) = .3174, el área más allá de 1 EE en ambos lados de la curva.

Paso 5: La decisión de rechazo Una vez que determinamos el valor *p* en el paso 4, lo comparamos con el nivel de significación (α) del paso 3, que es .05. Si el valor *p* es menor que o igual a α, entonces rechazamos H_0 y aceptamos H_A. Por otro lado, si el valor *p* es mayor que α, "fracasamos en rechazar" H_0. Este paso de la prueba de hipótesis se denomina decisión de rechazo. Para nuestra pregunta de "cabezas huecas", nuestro valor de .1587 es mayor que .05, por tanto fracasamos en rechazar H_0.

Paso 6: Interpreta y aplica los resultados y proporciona las mejores estimaciones en términos comunes El paso final en la prueba de hipótesis es proporcionar una interpretación de los resultados. Tomamos en cuenta la audiencia a la cual se reportarán los resultados. Para una audiencia pública, que no está muy versada en términos estadísticos, la interpretación se restringe a lenguaje común. Para una presentación profesional o entrega a una revista científica, se proporciona información estadística adicional.

Para nuestra hipótesis del estereotipo de "cabezas huecas", imaginemos que estamos presentando nuestros resultados a una audiencia pública. En esencia, abordamos tres puntos en la interpretación:

1. Vuelve a estipular H_0 o H_A, cualquiera que hayamos decidido en el paso 5.
2. Proporciona las mejores estimaciones con base en los datos de la muestra. Observa el estadístico muestral. Cuando se rechaza H_0 y se acepta H_A, presenta el resultado muestral como una mejor estimación del parámetro. Cuando no se rechaza H_0, presenta el resultado muestral y el efecto de la prueba como simplemente debido al error de muestreo.
3. Proporciona una respuesta directa para concluir con el procedimiento de la prueba de hipótesis.

Con nuestra hipótesis de "cabezas huecas" fracasamos en rechazar H_0. Por tanto, descartamos H_A y, en su lugar, volvemos a estipular e interpretar H_0 en lenguaje directo: la puntuación CI media de atletas de preparatoria no parece diferente de la puntuación de CI media de 100. *Mejor estimación:* estimamos que el CI medio de atletas es 100, el mismo que el de otros estudiantes. La diferencia de –1 punto entre la puntuación CI media de la muestra de atletas de preparatoria y el CI medio normal de 100 se debe al error de muestreo esperado. *Respuesta*: el estereotipo de "cabezas huecas" está equivocado. Los atletas de preparatoria son, en promedio, tan inteligentes como otros estudiantes.

Observa los elementos del paso 6. Primero, regresamos al paso 1 y volvemos a estipular H_0. Por tanto, en el paso 1, H_0 y H_A se deben enunciar de manera apropiada y clara para evitar confusión. Segundo, al volver a enunciar H_0, utilizamos el lenguaje conservador "*no parece* diferente" en lugar de "*es* igual". Puesto que nos basamos en datos muestrales, no podemos declarar una conclusión acerca de la población con 100% de seguridad. Tercero, al proporcionar las estimaciones mejores, como fracasamos en rechazar H_0, al menos debemos mencionar que el efecto de prueba se debe al error de muestreo. Una audiencia pública puede desear saber por qué nos apegamos a un CI medio de 100 cuando los datos muestrales parecen demostrar que el promedio de los atletas es 1 punto de CI menor. Una audiencia pública puede preguntar por qué elegimos ignorar el efecto y quizá nos acusen de preferencia o "rectitud política". Para responder, debemos explicar con claridad la noción de error del muestreo aleatorio. Por ejemplo, mencionamos que una segunda muestra podría haber tenido una media de 101 —1 punto más que la media general— y hubiéramos sacado la misma conclusión. Al proporcionar mejores estimaciones, también existe la opción de calcular un intervalo de confianza (capítulo 8). Por ejemplo, para nuestra muestra de "cabezas huecas", el intervalo de confianza de 95% se leería: estamos 95% seguros de que el CI medio (μ_X) de la población de atletas de preparatoria está entre 97 y 101 puntos (no se muestran los cálculos). Observa que una μ_X de 100 puntos de CI cae dentro de este intervalo de confianza. Es por eso que concluimos que el CI de los atletas no parece diferente del de todos los estudiantes.

Por último, recuerda responder la pregunta, aún si ésta es un nuevo planteamiento de otras partes de la interpretación. Las hipótesis nula y alternativa se plantean en lenguaje matemático y estadístico. La respuesta a la pregunta de investigación puede abordar preguntas

conceptuales más amplias. En este caso probamos el CI medio, en tanto que nuestra pregunta más amplia fue si el estereotipo de "cabezas huecas" es cierto.

Nota especial sobre los símbolos

Por razones que se explican en el capítulo 10, los programas de software de cómputo como *SPSS* se refieren a cualquier prueba de medias muestrales únicas, sin importar el tamaño de la muestra, como una prueba *t*. Esto se aplica a la prueba Z que se presenta en este capítulo. En su software de cómputo, busca una prueba *t* de medias para correr el procedimiento de la prueba Z.

Comprensión del lugar de la teoría de la probabilidad en la prueba de hipótesis

Calcular probabilidades es la operación matemática esencial en la prueba de hipótesis. Observa que en los seis pasos de la inferencia estadística analizamos "probabilidad" muchas veces. Primero, en el paso 2, la distribución muestral es una predicción de todos los resultados muestrales posibles y la probabilidad de cada resultado cuando se imagina que la hipótesis nula es cierta (del paso 1). Segundo, en el paso 3, determinamos el nivel de significación (α), el nivel de probabilidad que quizá nos conduzca a rechazar la hipótesis nula. Tercero, en el paso 4, calculamos la probabilidad de nuestro resultado muestral real, aún imaginando que la hipótesis nula es cierta. En el paso 5, comparamos dos probabilidades: la calculada en el paso 4 y la formulada en el paso 3. Esta comparación nos condujo a una conclusión: o rechazamos la hipótesis nula y aceptamos la hipótesis alternativa, o "fracasamos en rechazar" la hipótesis nula y la hacemos valer. *Si empieza a perder sentido de lo que es prueba de hipótesis, recuerda que se basa en comparar dos probabilidades: la que ocurre en realidad en nuestra única muestra observada y la de lo que esperamos que ocurra en el muestreo repetido.*

Un enfoque sobre valores p

Comprender el valor *p* es esencial para dominar la lógica de la prueba de hipótesis. Apliquemos el pensamiento proporcional haciendo estas preguntas: ¿qué sucede cuando aumenta la magnitud de un valor *p*? ¿Qué significa cuando el valor *p* se calcula pequeño o grande?

Cuando el valor *p* es grande relativo a alfa, es decir, cuando $p > \alpha$, fracasamos en rechazar la hipótesis nula Un valor *p* grande nos indica que nuestro resultado muestral observado no es muy distinto o no está muy "alejado" del resultado anticipado por la hipótesis nula; en otras palabras, el efecto de la prueba es pequeño. Por ejemplo, supongamos que hubiéramos obtenido una media de 99.9 puntos de CI para nuestra muestra de "cabezas huecas". Nuestro efecto de prueba hubiera sido –0.10 puntos de CI. La distribución muestral de la figura 9-2 sugiere que este efecto de prueba pequeño bien pudiera haber resultado del error muestral normal. Tiene una alta probabilidad de ocurrencia y eso es lo que mide el valor *p*. Cuando obtenemos un efecto de prueba pequeño y su valor *p* grande, en el argot científico decimos: no hay "una diferencia estadísticamente significativa" entre lo que se observa y lo que se hipotetiza. La diferencia podría haber resultado fácilmente del error muestral normal.

Cuando el valor p es pequeño en relación a alfa, es decir, cuando $p \leq \alpha$, rechazamos la hipótesis nula Un valor *p* pequeño nos indica que *suponiendo* que la hipótesis nula sea cierta, nuestro resultado muestral es inusual. Supongamos, por ejemplo, que nuestro CI medio de la muestra hubiera sido 97 puntos de CI. El efecto de prueba hubiera sido −3 puntos de CI. Consultando la figura 9-2, podemos ver que este efecto grande está 3 errores estándar debajo de la media hipotética de 100 y no hay mucha área más allá de ese punto en la cola de la curva. De hecho, nuestro valor *p* hubiera sido .0013 (de la columna C de la tabla de la curva normal). Esto es tan poco común que tal vez la media poblacional no es 100. Valores *p* pequeños ocurren cuando el resultado muestral no ajusta razonablemente el parámetro hipotético. Recuerda la ilustración simple: cuando no se observan nubes, rechaza la hipótesis de que va a llover pronto. La probabilidad (o valor *p*) de tener lluvia de un cielo sin nubes con seguridad que es pequeña, por tanto la hipótesis de lluvia se rechaza. Cuando obtenemos un efecto de prueba grande y un valor *p* pequeño, en argot científico decimos: hay "una diferencia estadísticamente diferente" entre lo que se observa y lo que se hipotetiza. En resumen, existe una relación inversa entre el tamaño del efecto de prueba y su valor *p* calculado, como se resume en el cuadro siguiente.

Relaciones entre el tamaño del efecto, valores *p* y decisiones de rechazo

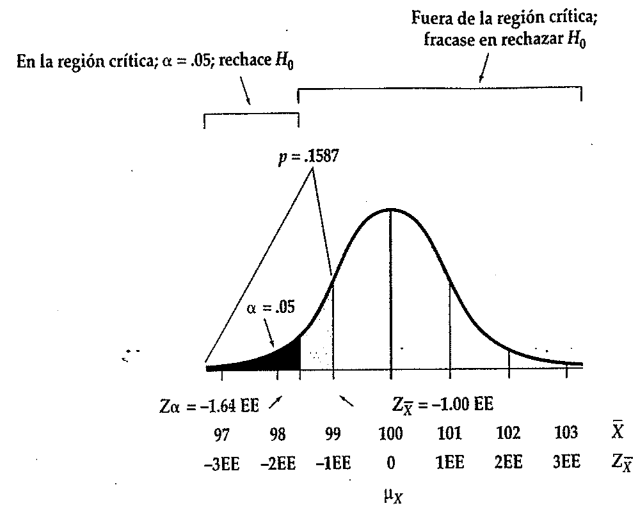
Un efecto de prueba pequeño → un valor *p* grande → "fracase en rechazar" H_0

Un efecto de prueba grande → un valor *p* pequeño → "rechace" H_0

El nivel de significación y las regiones críticas de la curva de la distribución de muestreo

Como demostramos, el valor *p* es la probabilidad de un resultado tan inusual como o más inusual que el observado. En el capítulo 6 destacamos que cuando se divide la curva normal, una probabilidad se puede representar de manera gráfica. Hicimos esto en nuestro problema de ejemplo sombreando la curva de la distribución muestral para indicar que el área del valor *p* es igual a .1587. También sombreamos el área que representa el nivel de significación, α = .05. Para hacer esto dividimos .05 del área en la cola negativa/izquierda de la curva, debido a que en la hipótesis alternativa (H_A) anticipamos una cola en esa dirección. Esta tarea es un problema de tipo 5 del capítulo 6. En la columna C de la tabla de la curva normal (tabla estadística B del apéndice B) buscamos el valor de .0500 o el más cercano a él. En la columna A encontramos 1.64 y agregamos un signo negativo para indicar que estamos trabajando con la mitad inferior de la curva. Esto indica que una puntuación Z de −1.64 corta el .05 inferior de la curva. A esta puntuación Z se le refiere como **puntuación crítica de la prueba,** *la que es lo suficientemente grande para indicar una diferencia significativa entre el estadístico muestral observado y el parámetro hipotetizado.* Se simboliza Z_α. *El área en la cola de la curva que está más allá de* Z_α se denomina **región crítica** de la curva. Superponiendo esta área α = .05 en la curva junto con el área del valor *p*.

La palabra *crítica* se utiliza debido a que la región crítica comprende puntuaciones Z que nos conducen a criticar la validez de la hipótesis nula. En una curva de distribución muestral



las ubicaciones relativas del valor crítico de prueba, Z_α, y el valor calculado del estadístico de prueba, $Z_{\bar{X}}$, proporciona una manera rápida para evaluar la decisión de rechazo en el paso 5 del procedimiento de seis pasos. Observa la curva y compara los *valores absolutos* (las puntuaciones ignorando los signos) de $Z_{\bar{X}}$ del paso 4 con el valor crítico de prueba (Z_α) del paso 3. Para el problema del estereotipo de "cabezas huecas", $Z_\alpha = -1.64$. Si el estadístico de prueba ($Z_{\bar{X}}$ del paso 4) cae en la región crítica de la curva, rechazamos H_0, debido a que eso significa que $p \leq \alpha$. En nuestro problema del estereotipo de "cabezas huecas" no caímos en esa región crítica y, por tanto, fracasamos en rechazar H_0.

Puntuación crítica de la prueba (Z_α) Puntuación estadística de prueba que es lo suficientemente grande para indicar una diferencia significativa entre el estadístico muestral observado y el parámetro hipotetizado.

Región crítica de la curva de la distribución muestral Área en la cola de la curva que está más allá de el valor crítico de prueba (Z_α) del nivel de significación estipulado (α).

En general, existe una relación inversa entre estas dos puntuaciones, como se indica en el cuadro siguiente.

Las relaciones entre el estadístico de prueba, el valor crítico de la prueba y el valor *p*

Si $Z_{\bar{X}} \geq Z_\alpha$, entonces $p \leq \alpha$; es decir, el área del valor *p* es menor que el área de la región crítica. Rechace H_0 y acepte H_A.

Si $Z_{\bar{X}} < Z_\alpha$, entonces $p > \alpha$; es decir, el área del valor *p* es mayor que el área de la región crítica. Fracase en rechazar H_0.

La puntuación Z crítica de uso más frecuente es ±1.96. Noventa y cinco por ciento del área bajo una curva normal cae entre +1.96 y −1.96, dejando 5% del área disponible en las dos colas (2.5% en cada cola). Es el área en las colas de la curva lo que constituye la región crítica o probabilidad α. Como el enfoque es en dos colas, a ésta se le denomina región crítica con dos colas. Entonces una puntuación Z crítica de ±1.96, corresponde la región crítica "α = .05, de dos colas".

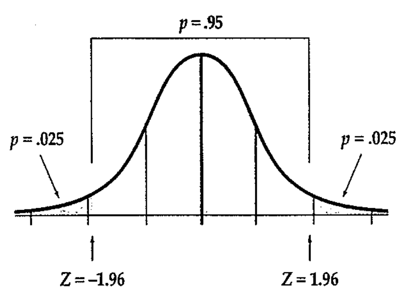
También podemos tener una región crítica concentrada en un lado de la curva una región crítica de una cola. Como se mostró antes, la puntuación Z crítica de 1.64 es una región crítica de una cola; 5% de la curva está más allá de 1.64 en un lado. Entonces una puntuación Z crítica de 1.64, corresponde a la región crítica "α = .05, una cola". Estas dos puntuaciones críticas y sus regiones críticas son tamaños "cómodos" (es decir, probabilidades α). Observe que estas regiones críticas son tamaños "cómodos" (es decir, 5, 1 y .1%). Por ejemplo, si se te pide calificar el desempeño de los miembros de un grupo de rock, tú podrías responder que el grupo califica en el 5 o 1% superior. Es probable que no utilices un porcentaje incómodo como 4%.

FIGURA 9-4

Puntuaciones Z para α = .05

Ilustración A: puntuación Z crítica de ±1.96 de colas; el total del área de la región crítica es .05 (5%) distribuida en las dos colas.

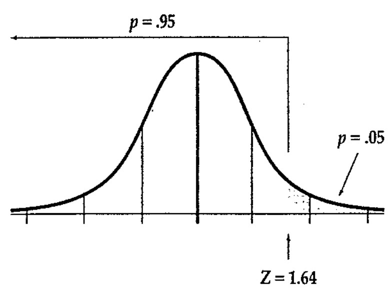
Designada: región crítica para α = .05, dos colas



α = (.025) + (.025) = .05

Ilustración B: puntuación Z crítica de 1.64 de una cola; el total del área de la región crítica es .05 (5%).

Designada: región crítica para α = .05, una cola.



α = .05

TABLA 9-2 | Puntuaciones Z críticas de uso más común y probabilidades α (p en la región crítica)

		Área en la región crítica				
		En una cola		En dos colas		
Región crítica (α)	Puntuación Z crítica (Z_α)	p	%	p	Σ% de ambos lados	(% en un lado)
α = .05, 1 cola	1.64	.05	5%			
α = .01, 1 cola	2.33	.01	1%			
α = .001, 1 cola	3.08	.001	.1%			
α = .05, 2 colas	1.96			.05	5%	(2.5%)
α = .01, 2 colas	2.58			.01	1%	(.5%)
α = .001, 2 colas	3.30			.001	.1%	(.05%)

La unidad de medida de la puntuación bruta para una región crítica La noción de región crítica nos permite considerar la decisión de rechazo aún de otra manera. Tanto el estadístico de prueba, $Z_{\bar{X}}$, como el valor crítico de prueba, Z_α, están en unidades de error estándar (EE). El estadístico de prueba, $Z_{\bar{X}} = -1.00$ EE es simplemente una versión estandarizada de $\bar{X} = 99$ puntos CI. También podemos calcular una puntuación crítica de prueba en las unidades de puntos CI de puntuaciones brutas. Una puntuación crítica de prueba de −1.64 errores estándar corresponde a una $\bar{X}$ de 98.3 puntos CI. Este cálculo es un problema tipo 7 del capítulo 6. Despejamos para $\bar{X}$ y sustituimos el valor crítico de prueba de $Z_\alpha = -1.64$ para $Z_{\bar{X}}$:

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}}$$

por tanto,

$$\bar{X} = \mu_X + (Z_{\bar{X}})(s_{\bar{X}})$$
$$= 100 + (-1.64)(1.00) = 98.36 \text{ puntos CI}$$

Este resultado nos indica que una media muestral menor que *98.36 puntos CI* tiene una probabilidad de ocurrencia menor que .05 de una población con una media de 100. Como nuestra media observada de 99 puntos CI es mayor que 98.36, fracasamos en rechazar H_0.

Obtener un buen sentido de las relaciones entre los resultados muestrales observados, su representación como puntuaciones de prueba, sus valores p, el nivel de significación elegido, las puntuaciones críticas de prueba y las regiones críticas se adquieren mejor observando la curva de la distribución muestral. El trazo de esta curva es una herramienta de aprendizaje importante para la prueba de hipótesis.

Selección del nivel de significación Es en el paso 3 del procedimiento de seis pasos de la inferencia estadística que establecemos el nivel de significación, α. En el paso 5 regresamos a α, donde tomamos la "decisión de rechazo" comparando el valor p con α. Cuando $p \leq \alpha$, rechazamos la hipótesis nula; cuando $p > \alpha$, fracasamos en rechazar la hipótesis nula. Como

TABLA 9-3 | Resultados posibles de decisiones de rechazo

	La verdad desconocida acerca de parámetros	
Nuestra decisión de rechazo	**Cuando H_0 en realidad es cierta**	**Cuando H_0 en realidad es falsa**
Rechazamos H_0	Error tipo I	Decisión correcta
Fracasamos en rechazar H_0	Decisión correcta	Error tipo II

destacamos antes, a menos que observemos toda la población, nuestros resultados sólo son estimaciones y la decisión de rechazo y las conclusiones hechas a partir de ella pueden estar equivocadas. Cualesquiera conclusiones basadas en el muestreo tienen un error esperado, como analizamos en el capítulo 8, donde nos referimos a α como el error esperado. A éste le llamamos error, en lugar de equivocación, debido a que somos capaces de estipular sus posibilidades de ocurrencia con precisión. La determinación del nivel de significación nos permite controlar las posibilidades de tomar una decisión equivocada o "error".

La tabla 9-3 ilustra la relación entre los resultados reales y la decisión de rechazo revela cuatro ocurrencias posibles. Ten en cuenta que nunca sabremos si la hipótesis nula es cierta o falsa *a menos que "muestremos" toda la población y obtengamos el parámetro real.* Realizamos una prueba estadística con el conocimiento que quizá obtendremos una conclusión equivocada.

Aunque nunca sabremos con seguridad cuando la hacemos, al rechazar la hipótesis nula cuando es falsa, hemos tomado la decisión correcta. De igual forma, cuando fracasamos en rechazar la hipótesis nula al ser cierta, hemos tomado la decisión correcta. Sin embargo, cuando *rechazamos una hipótesis nula cierta*, cometemos un **error tipo I.** En cualquier prueba donde rechazamos la hipótesis nula, existe una posibilidad que no la deberíamos haber rechazado. Por ejemplo, ¿podría ser que Tex el apostador simplemente tuvo suerte? De manera similar, en cualquier prueba de hipótesis donde fracasamos en rechazar la hipótesis nula, existe una posibilidad que la hubiéramos rechazado. Esto es un asunto de *fracasar en rechazar una hipótesis nula falsa* y a este tipo de error lo denominamos **error tipo II.** Éste hubiera sido el caso de concluir que Tex era honesto cuando de hecho no lo fue.

Nunca tendremos la seguridad si tomamos la decisión correcta o si cometimos un error. Sin embargo, podemos manejar y controlar la magnitud de esos errores en una variedad de formas. Primero, si rechazamos la hipótesis nula, no podríamos haber cometido un error tipo II debido a que este error implica *no* rechazar una hipótesis. De igual manera, cuando fracasamos en rechazar la hipótesis nula, sabemos que no podríamos haber cometido un error tipo I debido a que este error implica *rechazar* una hipótesis. Segundo, podemos controlar de manera fácil la cantidad de error tipo I que estamos dispuestos a permitir. Éste es el caso debido a que el nivel de significación, α, que determinamos a nuestra propia discreción, es la probabilidad de cometer un error tipo I. Por tanto,

$$\alpha = p \text{ [de cometer un error tipo I]}$$

Una vez más, se rechaza la hipótesis nula cuando el valor p del paso 4 es pequeño. Si hubiéramos elegido determinar α baja (digamos, .001), hubiéramos hecho difícil rechazar la hipótesis nula debido a que el valor p hubiera tenido que ser muy pequeño para estar "debajo" de .001. Al hacer la hipótesis nula difícil de rechazar, hicimos difícil rechazar un

error. Por tanto, cuando determinamos α baja, reducimos la posibilidad de un error tipo I, de rechazar la hipótesis nula cuando de hecho es cierta.

En contraste, si elegimos determinar un nivel α alto (digamos, .10), facilitamos rechazar la hipótesis nula debido a que el valor p del paso 4 no hubiera tenido que calcularse muy pequeño para ser menor que una α de .10. Al facilitar el rechazo de la hipótesis nula, reducimos la posibilidad de cometer la equivocación de no rechazarla cuando es falsa (es decir, reducimos la posibilidad de cometer un error tipo II).

Utilizamos la letra griega *beta* (β) para denotar la probabilidad de un error tipo II. Por tanto,

$$\beta = p \text{ [de cometer un error tipo II]}$$

Por desgracia, controlar β es muy difícil. Establecer α es posible debido a que se basa en la distribución esperada de los resultados descrita por la distribución muestral, cuando la hipótesis nula es cierta. Sin embargo, beta depende de que la hipótesis nula sea falsa. Como una hipótesis puede ser falsa en una variedad de formas, no disponemos de una base matemática fácil para calcular las probabilidades de estos resultados falsos. No obstante, podemos controlar β de manera indirecta cuando establecemos nuestro nivel alfa. Esto se debe a que α y β están inversamente relacionadas; es decir, cuando α aumenta, β por necesidad debe disminuir, y viceversa. Aunque en general no calculamos β, sabemos que cuando α se establece alta, esto facilita rechazar la hipótesis nula. Esto disminuye la posibilidad de fracasar en rechazarla y por tanto disminuye la posibilidad de fracasar en rechazarla *cuando es falsa.*

Error tipo I Inadvertidamente, tomar la decisión incorrecta de rechazar una hipótesis nula cierta

$$\alpha = p \text{ [de cometer un error tipo I]}$$

Error tipo II Inadvertidamente tomar la decisión incorrecta de fracasar en rechazar una hipótesis nula falsa

$$\beta = p \text{ [de cometer un error tipo II]}$$

De nuevo, es α, el nivel de significación, el que establecemos. Sin embargo, la decisión de su valor no es problemática debido a que los científicos en un campo particular siguen convenciones (tradiciones) que se basan en los tipos de preguntas siendo estudiadas y en lo que otros científicos aceptarán. Los cuatro niveles α convencionales se presentan en la tabla 9-4, donde se muestra la relación entre estos niveles y la probabilidad de rechazar una hipótesis nula. El nivel de significación (α) se debe establecer bajo cuando las consecuencias de un error tipo I sean serias. Por ejemplo, si nuestra hipótesis nula es que un nuevo medicamento controlado es tóxico (es decir, venenoso), no queremos rechazar esta hipótesis de manera prematura y cometer un error tipo I. Por tanto, estableceríamos α baja (digamos, .001). Esto requeriría de una evidencia contundente de que el medicamento sea seguro antes de rechazar su toxicidad. En la investigación de encuestas sociales el nivel α convencional es .05, un nivel moderado. A menos que tengas una buena razón para hacer lo contrario, sigue esta convención. Si te enfrentas con una situación con implicaciones éticas, ve las extensiones del capítulo 9 en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2 para ver el procedimiento sistemático para establecer el nivel de significación.

TABLA 9-4 | Niveles de significación convencionales y la posibilidad de rechazar la hipótesis nula (H_0)

Posibilidad de rechazar H_0	Nivel de significación* (α)	Usos comunes
Alta	.10	Investigación exploratoria, donde se conoce poco acerca de un tema
Moderada	.05 y .01	Niveles convencionales en investigación de encuestas e instrumentos de evaluación psicométrica y educacional
Baja	.01 y .001	Niveles convencionales en investigación biológica, de laboratorio y médica, en especial cuando un error tipo I es una amenaza a la vida (como la prueba de toxicidad de medicamentos)

*Estos niveles convencionales se aplican en el análisis estadístico bivariado. En el modelado estadístico multivariado como LISREL, el ajuste del modelo se puede probar con una α establecida tan alta como .5. Ese análisis está más allá del alcance de este texto.

Seleccione el nivel de significación antes de observar datos Es esencial que decidamos qué tan críticos vamos a ser respecto a la hipótesis nula (H_0) *antes* de hacer nuestra observación muestral. Es decir, debemos establecer α en el paso 3 del procedimiento de seis pasos, antes de observar los datos muestrales en el paso 4. ¿Por qué? Si esperamos hasta después de observar el resultado muestral en el paso 4, podríamos establecer α en un nivel ligeramente mayor que el valor p calculado y esto nos aseguraría rechazar H_0. En otras palabras, podríamos amañar la prueba de hipótesis para obtener el resultado que deseamos. Por ejemplo, supongamos que estamos obstinados en demostrar que los atletas son "cabezas huecas". Podríamos observar nuestro valor p de .1587 y luego establecer $\alpha = .20$. Después de "encontrar" $p < \alpha$, rechazaríamos H_0. Sin embargo, desde el punto de vista de la integridad científica esto sería hacer trampa. Permitiría que la preferencia personal entrara en el proceso científico. Además de la integridad científica, esto sería deshonesto. Si se hace de manera intencional, revelaría ignorancia acerca de la lógica de la prueba de hipótesis. Nos convertiríamos en "estadísticos tontos".

Al analizar datos, es tentador echar un vistazo a los resultados antes de establecer el nivel de significación. En el mundo de la investigación científica, obtener los resultados que deseamos quizás apoye los argumentos de nuestra teoría, conducir a publicaciones en revistas respetables y hacernos famosos. En el mundo de los negocios, obtener los resultados que deseamos puede destacar nuestra posición con el jefe (mostrando, por ejemplo, que hubo un aumento estadísticamente significativo en las ganancias de la compañía). En encuestas políticas, obtener los resultados que deseamos puede influenciar a votantes indecisos. En efecto, el análisis estadístico se puede manipular estableciendo un nivel α ventajoso. Pero ¡no caiga en la tentación! Los científicos profesionales con entrenamiento adecuado consideran el adulteramiento de los datos como poco ético. Además, como se analizó en el capítulo 1, el proceso de la investigación científica tiene verificaciones y comparaciones (como revisiones ciegas de entregas de artículos a revistas) para atrapar el comportamiento sin ética o sin cuidado. Estas verificaciones y comparaciones no sólo minimizan el error humano sino también la vanidad humana.

El nivel de confianza

Cuando rechazamos la hipótesis nula (H_0), digamos, al nivel de significación de .05, estamos tomando un riesgo de 5% de rechazar H_0 cuando de hecho es cierta. Por ejemplo, supongamos que examinamos la pregunta de investigación que Tex el apostador es un embustero. Nuestra H_0 es que él honesto. Establecemos α en .05, un nivel de significación convencional. Calculamos el valor p de sacar "unos" y es .0278; por tanto, $p < \alpha$ y rechazamos H_0 y llamamos embustero a Tex. Sin embargo, nunca sabremos con seguridad si en realidad lo es. (¡Tomó los dados y se fue!) Hubo una posibilidad de 5% que él simplemente fue muy afortunado y que lo acusamos sin fundamentos. Pero, en el mismo orden de ideas, había una probabilidad de 95% que tomamos la decisión correcta y no rechazamos sin fundamentos su honestidad. A esto lo denominamos **nivel de confianza,** *la seguridad que tenemos de que no cometimos un error tipo I*, y es igual a $1 - \alpha$.

Nivel de confianza

Nivel de confianza = 1 − nivel de significación = $1 - \alpha$

Un nivel de significación de .05 corresponde a un nivel de confianza de 95%. De igual manera, un nivel de significación de .01 corresponde a un nivel de confianza de 99%, y así sucesivamente.

En el capítulo 8 definimos estos términos con respecto a los intervalos de confianza. Aquí las propiedades matemáticas son las mismas. El nivel de confianza y el nivel de significación (o error muestral esperado) son enunciados acerca de la seguridad que tenemos en nuestros procedimientos muestrales y estadísticos. En el nivel de *significación* de .05 estamos asegurando que *si la hipótesis nula en realidad es cierta* y realizamos nuestros procedimientos 100 veces, tomaremos la decisión correcta 95 veces. Por tanto, estamos 95% seguros de la conclusión que sacamos de este procedimiento individual de prueba de hipótesis. Nuestro nivel de confianza está inversamente relacionado al riesgo de cometer un error tipo I. Entre menor sea la posibilidad de riesgo de rechazar la hipótesis nula (es decir, entre menor establezcamos α), mayor seguridad tendremos en nuestra conclusión *cuando suceda que la rechazamos*.

La única vez que tenemos una seguridad de 100% en una conclusión es cuando se observa cada sujeto en una población. En esta situación poco común, los cálculos resultantes no son estimaciones (es decir, estadísticos) sino parámetros reales. También, el error de muestreo no es un problema; es decir, tenemos una probabilidad cero de error de muestreo. Por ejemplo, los registros de la Crosstown University quizás utilicen registros computarizados para proporcionar un promedio de calificaciones exacto de su cuerpo estudiantil actual, el parámetro poblacional actual. En la mayoría de las investigaciones, como en una encuesta telefónica de hogares en Estados Unidos, no tenemos acceso a cada observación para una población. Por fortuna, nuestra habilidad para manejar y controlar el error muestral hace innecesario gastar grandes sumas de dinero encuestando poblaciones completas.

Sugerencias de estudio: organización de las soluciones de problemas

Los pormenores de la prueba de hipótesis se aprenden mejor con la práctica. La tarea es muy similar a aprender a leer música. Toma práctica y se aborda mejor de forma sistemática.

Todas las pruebas de hipótesis siguen la lógica del procedimiento de seis pasos de la inferencia estadística. El enunciado general de estos pasos se muestra en la tabla 9-5. Es buena idea ir adelante y aprender las palabras de los seis pasos aún si no estás completamente cómodo con el significado de todas ellas. De esta manera, no te perderás cuando tu maestro haga referencia a un paso particular en la clase.

TABLA 9-5 | Los seis pasos de la inferencia estadística o prueba de hipótesis

Preparación de la prueba

Enuncia la pregunta de investigación. Traza diagramas conceptuales representando los datos, incluyendo la o las poblaciones y la o las muestras en estudio, las variables (por ejemplo, X = ..., Y = ...) y sus niveles de medición, y los estadísticos dados o calculados. Utilizando los criterios para seleccionar una prueba estadística, selecciona e indica el procedimiento estadístico de prueba seleccionado.

Seis pasos

Utilizando el símbolo H para *hipótesis:*

1. Formula la hipótesis nula (H_0). Formula la hipótesis alternativa (H_A) y estipula la dirección de la prueba (si es de una cola o de dos colas).

   H_0 es una hipótesis estadística, enunciada de tal manera que tú sabrás qué resultados estadísticos ocurrirán en el muestreo repetido aleatorio *si esta hipótesis es cierta*. H_A se acepta si H_0 se rechaza.

2. Describe la distribución muestral y traza su curva.

   La distribución muestral es una proyección de resultados muestrales que es probable que ocurran en el muestreo repetido *cuando H_0 es cierta*. Una distribución muestral consiste de un listado de resultados muestrales posibles y una estipulación de la probabilidad de cada uno.

3. Indica el nivel de significación elegido (α). Indica de nuevo si la prueba es de una cola o de dos colas. Especifica el valor crítico de prueba y marca la región crítica en la curva en el paso 2.

   Alfa es la cantidad de error de muestreo que estamos dispuestos a tolerar al llegar a una conclusión. El valor crítico de prueba se obtiene de las tablas estadísticas del apéndice B.

4. Observa los resultados muestrales reales. Calcula los efectos de prueba, el estadístico de prueba y el valor p, y marca el valor p en la curva en el paso 2.

   El *efecto de prueba* es la diferencia entre lo que se observa en la muestra y lo que se hipotetizó para H_0 (en el paso 1). El *estadístico de prueba* es una fórmula para medir la posibilidad del efecto observado. El *valor p* es la probabilidad (p) de resultados muestrales tan inusuales o más inusuales que el resultado observado con la suposición que H_0 es cierta.

5. Toma la decisión de rechazo.

   Compara el valor p con α.

   Si $p \leq \alpha$ rechaza H_0 y acepta H_A al nivel de confianza $1 - \alpha$.

   Si $p > \alpha$, no rechace H_0.

   Para determinar si $p \leq \alpha$, compara el valor absoluto del estadístico de prueba con el valor absoluto de el valor crítico de prueba. Observa el valor p en relación a la región crítica en la curva en el paso 2.

6. Interpreta y aplica los resultados, y proporcione las mejores estimaciones en términos comunes.

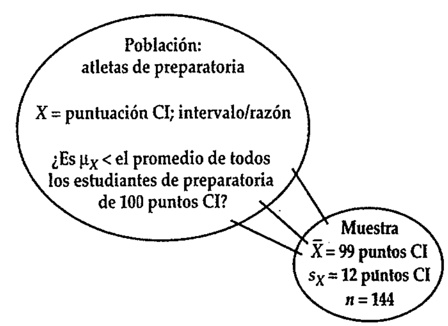
Cuadros de solución empleando los seis pasos

Cada procedimiento estadístico en los capítulos restantes incluye cuadros de solución para facilitar la resolución de los ejercicios del capítulo. Los dos cuadros siguientes comprenden presentaciones concisas de soluciones de pruebas de medias de una muestra única grande. El primer cuadro trata sobre nuestro problema del estereotipo de "cabezas huecas" donde fracasamos en rechazar la hipótesis nula, H_0. El segundo cuadro presenta una solución donde se rechaza H_0. En todas las pruebas de hipótesis, en los primeros cuatro pasos se supone que H_0 es cierta; por tanto, estos pasos son muy consistentes de una prueba a otra. Las distinciones principales entre las soluciones con decisiones de rechazo distintas ocurren en los pasos 5 y 6. Compara estas dos soluciones y pon mucha atención en estos dos pasos finales.

Solución para una prueba de medias de una muestra única grande (cuando $n > 121$): donde no se rechaza la hipótesis nula

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿es menor el CI medio de atletas de preparatoria que el CI medio de todos los estudiantes de preparatoria, el cual es 100 puntos CI? *Procedimiento estadístico:* prueba de medias de una muestra única grande. *Datos:*



SEIS PASOS

1. H_0: $\mu_{X\,(atletas\ de\ preparatoria)} = 100$ puntos CI (la media conocida (μ_X) de todos los estudiantes de preparatoria).

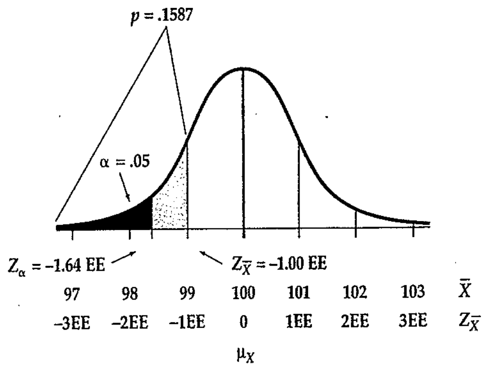
   Es decir, el CI medio de los atletas de preparatoria no es diferente del de todos los estudiantes de preparatoria.

   H_A: $\mu_{X\,(atletas\ de\ preparatoria)} < 100$ puntos CI.

   Es decir, el CI medio de los atletas de preparatoria es *menor que* el de todos los estudiantes de preparatoria. Una cola.

2. *Distribución muestral:* si H_0 es cierta y se toma una muestra de 144 de manera repetida de la población de atletas de preparatoria, las medias muestrales ($\bar{X}$) estarán centradas en 100 como una distribución normal con un error estándar:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{12}{\sqrt{144}} = 1.00 \text{ punto de CI}$$



(El sombreado de la curva se hace en los pasos 3 y 4.)

3. *Nivel de significación:* $\alpha = .05$. Una cola. Puntuación crítica $Z_\alpha = -1.64$ EE. (Sombrea y marca la región crítica como "$\alpha = .05$" en la curva en el paso 2.)
4. *Observación:*

$$\text{Efecto de prueba: } \bar{X} - \mu_X = 99 - 100 = -1 \text{ punto de CI}$$

$$\text{Estadístico de prueba } Z_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}} = \frac{99 - 100}{1.00} = -1.00 \text{ EE}$$

Valor *p*: *p* [sacar una muestra cuando ($\bar{X}$) tan inusual como o más inusual que 99 cuando la media poblacional verdadera (μ_X) es 100] = .1587.

(En la tabla de la curva normal, .1587 se encuentra en la columna C para $Z = -1.00$. Sombrea y marca el área del valor *p* en la curva en el paso 2.)

5. *Decisión de rechazo:* $|Z_{\bar{X}}| < |Z_\alpha|$ (es decir, $1.00 < 1.64$); por tanto, $p > \alpha$ (es decir, $.1587 > .05$); no rechaces H_0.
6. *Interpretación:* la puntuación CI media de atletas de preparatoria no parece diferente del CI medio de 100. *Mejor estimación:* estimamos que el CI medio de los atletas es 100, el mismo que el de los otros estudiantes. La diferencia de −1.00 punto ente la puntuación CI media de la muestra de atletas de preparatoria y el CI medio normal de 100 se debe al error de muestreo esperado.

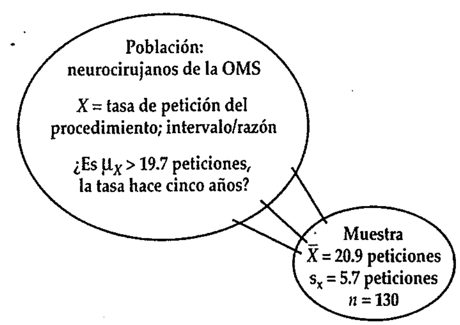
Respuesta: el estereotipo de "cabezas huecas" es erróneo. Los atletas de preparatoria son, en promedio, tan inteligentes como los otros estudiantes.

Solución para prueba de medias de una muestra grande única (cuando $n > 121$): cuando se rechaza la hipótesis nula

Problema. Tú eres un consultor en administración de calidad para una Organización de Mantenimiento de la Salud (OMS) nacional grande. Realizas un estudio de peticiones de neurocirujanos para un procedimiento de diagnóstico costoso. Se mide como "Tasa de peticiones" del procedimiento, el número de peticiones por cien pacientes examinados. Hace cinco años la tasa de peticiones media de todos los 1 567 neurocirujanos en la OMS fue 19.7 peticiones. Tú obtienes registros actuales en una muestra de 130 neurocirujanos en la OMS y obtienes una tasa de peticiones media de 20.9 peticiones con una desviación estándar de 5.7 peticiones. ¿Ha aumentado la tasa de peticiones media de neurocirujanos de la OMS para un procedimiento de diagnóstico costoso desde hace cinco años? Utiliza el procedimiento de seis pasos de la inferencia estadística y redondea el error estándar a dos lugares decimales.

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿ha aumentado la tasa de peticiones media de los neurocirujanos de la OMS para un procedimiento de diagnóstico costoso desde hace cinco años? *Procedimiento estadístico:* prueba de medias de una muestra grande única. *Datos:*



SEIS PASOS

1. H_0: $\mu_{X\,(neurocirujanos\ OMS\ hoy)} = 19.7$ peticiones (la media conocida de (μ_X) de neurocirujanos de la OMS hace cinco años).

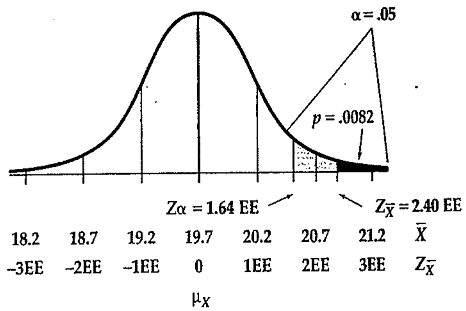
Es decir, la tasa de peticiones media de neurocirujanos de la OMS para el procedimiento no es diferente de la de hace cinco años.

H_A: $\mu_{X\,(neurocirujanos\ OMS\ hoy)} > 19.7$ peticiones.

Es decir, la tasa de peticiones media es mayor que hace cinco años. Una cola.

2. *Distribución muestral:* si H_0 es cierta y se toman muestras repetidas de tamaño 130 de la población de neurocirujanos de la OMS, las medias muestrales ($\bar{X}$) se centrarán en 19.7 como una distribución normal con un error estándar:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{5.7}{\sqrt{130}} = .50 \text{ peticiones}$$



(El sombreado de la curva se hace en los pasos 3 y 4.)

3. *Nivel de significación:* $\alpha = .05$. Una cola. Puntuación crítica de prueba Z_α = –1.64 EE.

(Sombrea y marca la región crítica como "$\alpha = .05$" en la curva en el paso 2.)

4. *Observación:*

Efecto de prueba: $\bar{X} - \mu_X = 20.9 - 19.7 = 1.20$ peticiones

$$\text{Efecto de prueba: } Z_{\bar{X}} = \frac{X - \mu_X}{s_{\bar{X}}} = \frac{20.9 - 19.7}{0.50} = 2.40 \text{ EE}$$

Valor p : p [sacar una muestra con una media ($\bar{X}$) tan inusual como o más inusual que 20.9 cuando la media poblacional verdadera (μ_X) es 19.7] = .0082. (En la tabla de la curva normal, .0082 se encuentra en la columna C para Z = 2.40. Sombrea y marca el área del valor p en la curva en el paso 2).

5. *Decisión de rechazo:* $|Z_{\bar{X}}| > |Z_\alpha|$ (es decir, 2.40 > 1.64); por tanto, $p < \alpha$ (es decir, .0082 < .05; rechaza H_0 y acepta H_A al nivel de confianza de 95%.

6. *Interpretación:* la tasa de peticiones media para el procedimiento parece ser mayor que 19.7 hoy. *Mejor estimación:* estimamos que la tasa de peticiones media de neurocirujanos de la OMS para un procedimiento de diagnóstico costoso ha aumentado desde hace cinco años.

Interpretación de resultados cuando se rechaza la hipótesis nula: la base hipotética de la prueba de hipótesis

Una vez más, la palabra *hipotético* significa "imaginemos por un momento". Para los dos cuadros de solución anteriores, en los pasos 1 a 4 del procedimiento de seis pasos de la inferencia estadística, hacemos enunciados hipotéticos o "si esto es cierto". Observa el segundo cuadro de solución. En el paso 1, aunque sospechamos que la tasa de peticiones media para un procedimiento costoso ha aumentado, formulamos la hipótesis nula (H_0) "imaginamos por un momento que no" y procedemos al paso 2 como si este fuera el caso. En el paso 2, continuamos imaginando que predecimos resultados muestrales para "cuando H_0 sea cierta"; es decir, suponemos que la tasa de peticiones no ha cambiado. La tasa de peticiones puede o no haber cambiado, pero nosotros describimos qué sucede en el muestreo repetido *si no ha cambiado*, debido a que esta es una forma para nosotros para hacer predicciones matemáticas sólidas acerca de resultados muestrales. En el paso 3, al establecer el nivel de significación y estipular el valor crítico, declaramos qué tan inusual debe ser un resultado muestral para que rechacemos H_0 que estamos *suponiendo ser cierta*. En el paso 4, cuando calculamos el valor p, esta probabilidad de nuestro resultado muestral se calcula *como si H_0 fuera cierta*. Los pasos 1 a 4 son imaginarios en el sentido que nosotros utilizamos nuestro conocimiento e imaginación estadística para predecir que esperar cuando observamos datos si la hipótesis nula es cierta. Sólo en el paso 5 concretamos una decisión y decidimos lo que en realidad creemos es cierto, la hipótesis nula o la alternativa. Al probar una hipótesis, decimos en cada uno de los pasos 1 a 4: "retenga este pensamiento. Si H_0 es cierta, entonces aquí está lo que sucede en el muestreo repetido". En el paso 5, decidimos si nuestros datos muestrales se ajustan a las predicciones.

Luego, es en los pasos 5 y 6 que aparecen diferencias distintas ente las pruebas de hipótesis donde H_0 se rechaza y no se rechaza. En el paso 5, si H_0 se rechaza, aceptamos la hipótesis alternativa (H_A). Al hacer esto, estamos corriendo un riesgo α de cometer un error tipo I, por tanto tenemos una confianza de $1 - \alpha$ en nuestra conclusión. Así, observa en el paso 5 en el cuadro de solución de la tasa de peticiones del procedimiento que, como $\alpha = .05$ = 5%, aceptamos H_A en el nivel de confianza de 95%. Ahora procedemos en el paso 6 a interpretar la hipótesis alternativa. El valor observado de la media muestral de 20.9 peticiones para el procedimiento costoso ahora se acepta como una estimación de valor del parámetro de la población.

Si en el paso 5 no se rechaza H_0, como fue el caso con la solución del estereotipo de "cabezas huecas", *no* aceptamos H_0 como cierta, simplemente fracasamos en rechazarla. Aquí somos conservadores al enunciar la decisión de rechazo debido a que hay una posibilidad de cometer un error tipo II. Es decir, la conclusión que los atletas tienen el mismo CI que los otros estudiantes puede ser falsa. A menos que nos tomemos la molestia considerable de calcular *beta*, la probabilidad de un error tipo II, no tenemos forma de saber cuánta seguridad colocar en nuestra conclusión. Después, en el paso 6 empleamos un lenguaje conservador como "parece diferente" o "no encontramos una diferencia significativa". Una herramienta de aprendizaje útil es comparar de manera continua las soluciones del procedimiento de seis pasos para cada nueva prueba de hipótesis en los capítulos restantes.

Selección de la prueba estadística a emplear

¿Cómo sabemos cuáles son las fórmulas correctas para un problema en particular? La parte más difícil de la prueba de hipótesis es elegir las distribuciones muestrales y las fórmulas

TABLA 9-6 | Criterios para la selección de una prueba estadística

1. **PREGUNTA:**	¿Cuántas variables observamos para esta prueba?
2. **PREGUNTA:**	¿Cuáles son los niveles de medición de las variables? Es decir, ¿son nominales/ordinales (para calcular conteos y proporciones) o de intervalo/razón (para calcular medias) las variables?
3. **PREGUNTA:**	¿Estamos tratando con una muestra representativa de una población única o más?
4. **PREGUNTA:**	¿Cuál es el tamaño de la muestra?
5. **PREGUNTA:**	¿Existen circunstancias peculiares que se deban considerar?

estadísticas apropiadas. Estas tareas se facilitan si se siguen criterios sistemáticos al tomar decisiones. Estos criterios se presentan en la tabla 9-6.

Aunque cada uno de estos criterios es importante al determinar el tipo de prueba que se debe emplear, el criterio 2 es muy útil. Quizá desees repasar los cuatro niveles de medición (capítulo 2). Un punto útil que se debe recordar es que la media, las puntuaciones de desviación, la varianza y la desviación estándar sólo se calculan para variables de intervalo/razón. Por tanto, las pruebas estadísticas para estas variables con frecuencia llevan el nombre *prueba de medias, diferencias de medias* o *análisis de la varianza*. En contraste, las variables de nivel nominal/ordinal por general implican contar frecuencias, porcentajes o proporciones de casos en categorías y a menudo llevan el nombre de *prueba de proporciones*. Para ayudarte en la selección del procedimiento apropiado, se proporcionan árboles de decisión en la segunda de forros del texto y al final de los capítulos.

Insensatez y falacias estadísticas: sentido común informado: más allá del sentido común observando datos

En el mundo social y en el físico se puede aprender mucho mediante el sentido común, aplicando un proceso de razonamiento claro a una situación. Pero los científicos permanecen ocupados debido a que muchos de los procesos de la naturaleza no son tan obvios. De hecho, los científicos sociales han establecido desde hace tiempo que como humanos, somos propensos a los prejuicios y falsedades simplistas que hemos llegado a creer debido a que el sentido común nos dice que son ciertas. Hay muchos mitos y supersticiones acerca de la realidad, en especial en la realidad social. La ciencia y la imaginación estadística con su procedimiento de prueba de hipótesis nos animan a cuestionar observaciones más cercanamente, ponderarlas contra resultados predecibles, y desafiar mitos y prejuicios.

Por ejemplo, el sentido común conduce a mucha gente a concluir que las mujeres son "obviamente" más débiles física y emocionalmente que los hombres. En promedio, es claro que los hombres tienen una fuerza mayor en el cuerpo superior. Pero la fuerza física tiene muchas dimensiones que desafían la afirmación de dominación de los varones. Por ejemplo, menos mujeres nacen muertas, y las niñas tienen un índice de mortandad infantil menor y mayor esperanza de vida. La fuerza emocional también es difícil de definir. Mucha gente supone que los hombres son emocionalmente más fuertes que las mujeres debido a que las mujeres lloran más rápido. Pero entonces, ¿por qué los hombres cometen más de 90% de todos los delitos violentos emocionalmente cargados como agresiones y homicidios? ¿Yace la confusión con restricciones culturales sobre cómo los hombres y las mujeres expresan las emociones? ¿Cómo se puede medir la fuerza emocional de manera confiable y con equidad?

Para comprender completamente la fuerza física y emocional, debemos iniciar con una definición clara de lo que en realidad es fuerza. Aunque el sentido común explica gran parte de la realidad, una mayor comprensión requiere de un razonamiento agudo, de una predicción significativa y de la observación y medición precisa. La observación metódica amplía y *moldea* el sentido común.

Esto no quiere decir que un reporte científico dado es la última palabra sobre un tema. Cualquier teoría científica siempre está abierta a una modificación posterior. Ni quiere decir que los científicos están arriba creando y apegándose a sus propios mitos. Por ejemplo, la mayoría de la investigación científica a finales de 1800 apoyaba la noción que las mujeres eran innatamente menos inteligentes que los hombres. Pero en la ciencia esos mitos tienden a no resistir el paso del tiempo. El proceso de la investigación científica tiene sistemas incorporados de verificaciones y comparaciones que aumentan las oportunidades para desmentir los mitos. La prueba de hipótesis es un proceso clave en separar hechos esenciales de hechos aparentes pero perjudiciales.

RESUMEN

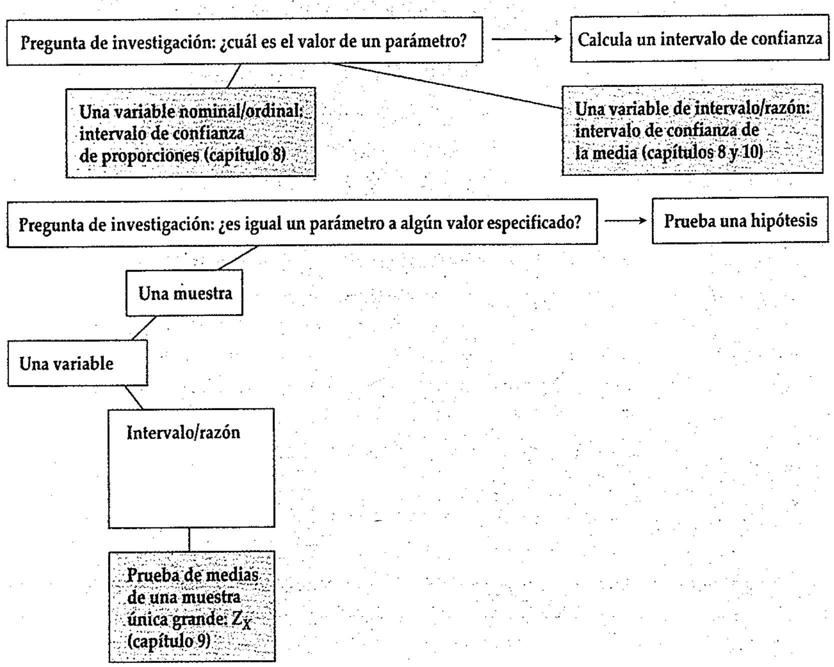
1. Una hipótesis es una predicción acerca de la relación entre dos variables que afirma que las diferencias entre las mediciones de una variable independiente corresponderán a diferencias entre las mediciones de una variable dependiente.
2. El propósito teórico de una prueba de hipótesis es corroborar la teoría probando ideas contra hechos.
3. El propósito estadístico de una prueba de hipótesis es determinar si los efectos estadísticos de una muestra indican: 1) efectos reales en la población o 2) error muestral.
4. Los estadísticos de una muestra sólo son herramientas para sacar conclusiones acerca de una población. Es la población acerca de la cual finalmente haremos declaraciones.
5. Una prueba de hipótesis se basa en predecir resultados muestrales para una hipótesis nula (H_0). Suponiendo que H_0 es cierta, predecimos todos los resultados muestrales posibles tomando en cuenta el error de muestreo. La distribución muestral resultante es un patrón de medición con el cual comparamos nuestro resultado muestral único real ver si es "significativamente diferente" del resultado anticipado por H_0.
6. El nivel de significación elegido (α), la dirección de la prueba y el valor crítico de una tabla estadística como la tabla de la curva normal establecen el punto en el cual rechazamos H_0 y la cantidad de error de muestreo que toleraremos al llegar a una conclusión. Esto se representa de manera gráfica como la región crítica bajo la curva de distribución muestral.
7. El efecto de prueba es la diferencia entre lo que se observó en la muestra y lo que se hipotetizó en el paso 1.
8. El estadístico de prueba transforma el efecto de prueba en errores estándar tal que se pueda emplear una tabla estadística para calcular el valor p.
9. La decisión de rechazo compara el valor p con α. Una valor p pequeño, menor que α, nos indica que nuestro resultado muestral es inusual cuando H_0 es cierta y justifica rechazar la verdad de H_0.

10. Los criterios listados en la tabla 9-6 nos ayudan al seleccionar el procedimiento apropiado de la prueba estadística. El nivel de medición de las variables y del tamaño de la muestra son criterios importantes.
11. Un error tipo I rechaza una H_0 cierta. El nivel de significación, α, es igual a la probabilidad de cometer un error tipo I. El error tipo I se controla fácilmente eligiendo el tamaño de α. Un error tipo II fracasa en rechazar una H_0 falsa. *Beta* es la probabilidad de cometer un error tipo II y es más difícil de controlar.
12. Una prueba de medias de una muestra única grande se utiliza con una variable única de intervalo/razón de una población con un tamaño muestral mayor que 121 casos. Además, debe haber un valor objetivo respecto al cual se enmarcan las predicciones muestrales para la prueba. La curva de la distribución Z normal es la distribución muestral.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 9 de material del texto disponibles en el sitio en la red *The Statistical Imagination* en www.mhhe.ritchey2 incluyen un procedimiento sistemático para seleccionar el nivel de significación para pruebas de hipótesis implicando situaciones éticas o situaciones donde las consecuencias de la prueba sean controversiales.

PROCEDIMIENTOS ESTADÍSTICOS ANALIZADOS HASTA ESTE PUNTO



FÓRMULAS PARA EL CAPÍTULO 9

Prueba de medias de una muestra única grande (prueba Z)

Datos: una variable de intervalo/razón X, una muestra y población únicas y $n > 121$ casos

Pregunta de investigación: ¿es igual μ_X (es decir, la media de X en la población) al valor objetivo?

H_0: μ_X = un valor objetivo

Distribución muestral: distribución normal Z; error estándar estimado utilizando la desviación estándar de la muestra.

Error estándar:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}}$$

Efecto de prueba $= \bar{X} - \mu_X$

Estadístico de prueba (para utilizar con la curva normal, tabla de la distribución Z, tabla estadística B, apéndice B):

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}}$$

PREGUNTAS PARA EL CAPÍTULO 9

1. Una teoría (un conjunto de ideas acerca de cómo funciona el mundo empírico) motiva hipótesis (predicciones específicas de las cuales se pueden esperar observaciones cuando una teoría es cierta). Supongamos que probamos una teoría de discriminación racial para explicar segregación residencial (es decir, la tendencia en un vecindario a ser ocupado por una sola raza). Con respecto al comportamiento de agentes de bienes raíces, ¿qué hipótesis se motiva por esta teoría?
2. Define y distingue los objetivos teóricos y estadísticos para probar una hipótesis. Ilustra con un ejemplo.
3. Al probar una hipótesis, determina si los efectos muestrales observados se deben a diferencias reales en parámetros poblacionales o simplemente se deben al error de muestreo. Matemáticamente, ¿cuáles dos cosas debemos predecir a fin de iniciar una prueba de hipótesis?
4. Rechazamos una hipótesis nula cuando el valor de p ¿es grande o pequeño?
5. ¿Cuál es la relación entre el tamaño del efecto de una prueba estadística y el valor p calculado para esa prueba? Ilustra con un ejemplo.
6. En lenguaje simple, ¿cuál es el nivel de significación de una prueba de hipótesis y cuál es su función en la prueba?

7. Relaciona lo siguiente:

a) Error tipo I ______ p [error tipo I]

b) Error tipo II ______ Rechazar H_0 cuando de hecho es cierta

c) Alfa (α) ______ p [error tipo II]

d) Beta (β) ______ Fracasar en rechazar H_0 cuando de hecho es falsa

8. Una distribución muestral es hipotética. ¿Qué significa esto?

9. ¿Con variables de qué niveles de medición utilizamos las pruebas de medias?

10. ¿Por qué debemos elegir el nivel de significación antes de observar los resultados estadísticos de nuestra muestra?

11. Enumera los seis pasos de la inferencia estadística.

12. Lista los criterios para la selección de una prueba estadística.

13. Ahora que conoces la distribución muestral para el lanzamiento de un par de dados (tabla 9-1), utiliza tu imaginación estadística para mejorar tu estrategia para el juego de tablero Monopolio. (Quizá quieras inspeccionar el juego real para responder estas las preguntas siguientes.)

a) Ganar el juego depende de poseer las propiedades más valiosas y recolectar frecuentemente la renta de éstas. Dado esto, si pudieras elegir poseer un color de propiedades (o calles) para iniciar el juego, ¿qué color elegirías? ¿Por qué?

b) ¿Cuál es la acción más tonta que un jugador recién enviado a la cárcel puede hacer en el siguiente turno si no posee las propiedades color púrpura o anaranjadas?

c) Los cuatro ferrocarriles no pagan mucha renta y por tanto con frecuencia no vale la pena comprarlos o retenerlos. Sin embargo, hay circunstancias cuando es ventajoso poseerlos. ¿Cuáles serían estas circunstancias? (*Sugerencia:* aquí es útil la regla de la adición de las probabilidades.)

EJERCICIOS PARA EL CAPÍTULO 9

Conjunto de problemas 9A

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se estipule lo contrario.

9A-1. Practica el arte de identificar las hipótesis nulas y concebir las distribuciones muestrales. En términos generales, anticipa qué resultados muestrales se pueden esperar que ocurran en el muestreo repetido cuando las hipótesis nulas (H_0) siguientes son ciertas. (Un repaso del capítulo 7 puede ser útil.)

a) H_0: la edad media de estudiantes en el campus es 21 años.

b) H_0: entre las corporaciones de *Fortune 500*, el porcentaje de miembros de la junta corporativa que son mujeres es 20%.

c) H_0: el peso medio de las barras de chocolate Lot-O-Candy es .75 onzas.

d) H_0: el maestro no tiene preferencia por hombres o mujeres al otorgar calificaciones de 10.

9A-2. Una pregunta de investigación es un objetivo de proyecto que se puede estipular en términos de una hipótesis. Practica el arte de determinar si cada una de las preguntas de investigación siguientes constituye la hipótesis nula (H_0) o la hipótesis alternativa (H_A). Explica tu respuesta.

a) En promedio, hay más de seis acciones de violencia por semana en cada una de las series de televisión estelar.

b) En una apuesta. Elbert acaba de lanzar 10 monedas al aire y todas salieron cara. ¿Parecen legítimas sus monedas?

c) ¿Es cierto el estereotipo que más de 90% de las personas indigentes son adictas al alcohol o drogas?

9A-3. La dirección y el signo de una prueba de hipótesis se especifican en la hipótesis alternativa. Decide si las hipótesis alternativas (H_A) siguientes son de una cola en la dirección positiva, de una cola en la dirección negativa o de dos colas no direccional. También, indica el signo matemático y explica tu elección.

a) H_A: más de 50% de las víctimas de cáncer pulmonar son o han sido fumadoras.

b) H_A: el promedio de calificaciones de estudiantes hombres y mujeres no es el mismo.

c) H_A: en el distrito escolar central de la ciudad menos de 60% de los graduados de preparatoria ingresan a una universidad.

9A-4. En un estudio de patrones de trabajo entre abogados, un investigador plantea la hipótesis de que los abogados que se especializan en leyes corporativas trabajan más horas por semana que los que se especializan en leyes estatales. En H_0, el investigador plantea la hipótesis de que el número medio de horas trabajadas por semana para los dos grupos es igual.

a) ¿Por qué el investigador enuncia la H_0 de esa manera en lugar de decir que la media para los abogados corporativos es mayor?

b) En H_A, ¿deberá utilizar una prueba de una cola o de dos colas? ¿Por qué?

c) En el paso 3 de la prueba, establece un nivel de significación (α) de .05. En el paso 4 de la prueba calcula un valor p de .23. En el paso 5, ¿rechazará H_0 o fracasará en rechazarla?

9A-5. Este ejercicio te familiarizará con las relaciones entre niveles de significación, valores p y decisiones de rechazo. Para los niveles de significación y valores p siguientes, indica si rechazarías o fracasarías en rechazar la hipótesis nula, H_0.

	Nivel de significación (α del paso 3 de los seis pasos)	Valor p (del paso 4 de los seis pasos)	Decisión de rechazo: rechaza H_0 o fracasa en rechazar H_0
a)	.001	.0007	
b)	.05	.0650	
c)	.01	.0099	
d)	.05	.0399	
e)	.001	.0110	
f)	.01	.0101	

9A-6. Alguien te pregunta si el promedio de calificaciones de los estudiantes de la Greene County High School es igual a B (es decir, un promedio de calificaciones (PC) de 3.0 en una escala de 4). Tú consideras que no es cierto y tienes razones para pensar que es menor que 3.0. Prueba la hipótesis con los datos muestrales siguientes:

$n = 155$ $\bar{X} = 2.91$ puntos PC $s_X = .9$ puntos PC

9A-7. Un banco regional grande busca ubicar sucursales en comunidades residenciales. El objetivo es enfocarse en préstamos para mejoras de viviendas, por que las casas de la comunidad deben tener una edad promedio mayor que 15 años. Tú seleccionas al azar 130 registros de propiedades de la ciudad de Clarksdale y determinas la edad media de las casas igual a 15.78 años con una desviación estándar de 3.1 años. ¿Es buena candidata para una sucursal Clarksdale?

9A-8. En el paso 3 del procedimiento de seis pasos de una prueba de hipótesis decidimos en un nivel de significación (α), la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazamos H_0. En la curva de la distribución muestral, ésta es la región crítica con una puntuación crítica expresada como un número de errores estándar.

a) Para el ejercicio 9A-6, ¿cuál es el valor crítico expresado en las unidades de la puntuación bruta de puntos PC?

b) Para el ejercicio 9A-7, ¿cuál es el valor crítico expresado en las unidades de la puntuación bruta de años?

Conjunto de problemas 9B

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos decimales. Utiliza α = .05 a menos que se indique lo contrario.

9B-1. Practica el arte de identificar la hipótesis nula y concebir distribuciones muestrales. En términos generales, anticipa qué resultados muestrales se puede esperar que ocurran en el muestreo repetido cuando las hipótesis nulas (H_0) siguientes son ciertas. (Un repaso del capítulo 7 puede ser útil.)

a) H_0: la mitad del público televidente ve un noticiero nocturno.

b) H_0: la velocidad media de automóviles en el tramo de la muerte en la carretera interestatal es 80 millas por hora.

c) H_0: 40% de los estudiantes de último año de preparatoria han consumido alcohol ilegalmente.

d) H_0: la edad media de los vicepresidentes corporativos es 49 años.

9B-2. Una pregunta de investigación es un objetivo del proyecto que se puede enunciar en términos de una hipótesis. Practica el arte de determinar si cada una de las preguntas de investigación siguientes constituye la hipótesis nula (H_0) o la hipótesis alternativa (H_A). Explica tu respuesta.

a) En promedio, ¿exceden el límite de velocidad de 70 millas por hora los conductores en el tramo de la muerte de la carretera interestatal?

b) Utilizando una muestra de 30 de los 125 jugadores, ¿es igual el peso promedio del equipo de fútbol de este año al del año pasado que fue de 224 libras?

c) ¿Utiliza dados cargados este casino?

9B-3. La dirección y el signo de una prueba de hipótesis se especifican en la hipótesis alternativa. Decide si las hipótesis alternativas (H_A) siguientes son de una cola en la dirección positiva, de una cola en la dirección negativa o de dos colas no direccional. También, indica el signo matemático y explica tu elección.

a) H_A: más de 80% de los presos en la cárcel del condado son encarcelados por cargos relacionados con las drogas.

b) H_A: para el nuevo medicamento Fixtall, la tasa de cura del grupo experimental que recibió el medicamento es mayor que la del grupo de control que recibió un placebo (es decir, una píldora de azúcar).

c) H_A: los porcentajes de bautistas y metodistas que creen que la Biblia no tiene errores no son iguales.

9B-4. La profesora Smith estudia la desigualdad de género en una compañía importante de comunicaciones. Con base en su experiencia pasada y en las teorías de la información especializada, tiene razones para creer que las mujeres que trabajan para la compañía tienen un ingreso medio menor que el de los hombres. En la hipótesis nula, H_0, ella hipotetiza que los ingresos medios de los hombres y las mujeres son iguales.

a) ¿Por qué enuncia H_0 de esa manera en lugar de decir que la media para los hombres es mayor?

b) En la hipótesis alternativa, H_A, ¿deberá utilizar una prueba de una cola o de dos colas? ¿Por qué?

c) En el paso 3 de la prueba, ella establece un nivel de significación (α) de .05. En el paso de la prueba calcula un valor p de .03. En el paso 5, ¿rechazará o aceptará H_0?

9B-5. Este ejercicio te familiarizará con las relaciones entre niveles de significación, valores p y decisiones de rechazo. Para los niveles de significación y valores p siguientes, indica si rechazarías o aceptarías la hipótesis nula, H_0.

	Nivel de significación (α del paso 3 de los seis pasos)	Valor p (del paso 4 de los seis pasos)	Decisión de rechazo: rechaza H_0 o fracasa en rechazar H_0
a)	.05	.0476	
b)	.05	.3297	
c)	.01	.0476	
d)	.001	.0028	
e)	.01	.0006	
f)	.05	.4996	

9B-6. Alguien te pregunta si el índice de masa corporal (IMC) promedio de los estudiantes de la Jackson Middle School es igual al IMC ideal sugerido por los Centers for Disease Control (es decir, un IMC de 25). Tú no estás de acuerdo y tienes razones para creer que es mayor que 25. Prueba la hipótesis con los datos muestrales siguientes:

$n = 170$ $\bar{X} = 27.6$ kilogramos por metro cuadrado $s_X = .9$ kilogramos por metro cuadrado

9B-7. Alguien te pregunta si el promedio de calificaciones (PC) de los estudiantes de la Highlands University es igual a B (es decir, un PC de 3.0 en una escala de 4). Tú no estás de acuerdo y tienes razones para creer que es menor que 3.0. Prueba la hipótesis con los datos muestrales siguientes:

$n = 210$ $\bar{X} = 2.92$ puntos PC $s_X = .9$ puntos PC

9B-8. En el paso 3 del procedimiento de los seis pasos de una prueba de hipótesis, decidimos sobre un nivel de significación (α), que es la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazamos H_0. En la curva de la distribución muestral, ésta es la región crítica con una puntuación crítica expresada como el número de errores estándar.

a) Para el ejercicio 9B-6, ¿cuál es su puntuación crítica expresada en las unidades de la puntuación bruta de metro al cuadrado?

b) Para el ejercicio 9B-7, ¿cuál es la puntuación crítica expresada en las unidades del valor crítico de puntos PC?

Conjunto de problemas 9C

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se indique lo contrario.

9C-1. Practica el arte de identificar hipótesis nulas y de concebir distribuciones muestrales. En términos generales, anticipa qué resultados muestrales se pueden esperar que ocurran en el muestreo repetido cuando las hipótesis nulas (H_0) son ciertas. (Un repaso del capítulo 7 puede ser útil.)

a) H_0: la edad media de los empleados de una compañía grande es 32 años.

b) H_0: el peso medio de los jugadores de un equipo de fútbol de la división 1-A es 207 libras.

c) H_0: el número medio de estudiantes inscritos en cursos de introducción a la psicología en una universidad urbana grande es 82.

d) H_0: las mujeres constituyen 65% de los médicos con licencia entre las especialidades pediátricas.

9C-2. Una pregunta de investigación es una meta de proyecto que se puede estimar en términos de una hipótesis. Practica el arte de determinar si cada una de las preguntas de investigación siguientes constituirá la hipótesis nula (H_0) o la hipótesis alternativa (H_A). Explica tu respuesta.

a) Utilizando una muestra de 120 estudiantes universitarios, ¿es la edad promedio de la clase de estudiantes de primer año la misma que en años pasados, que fue 18.4 años?

b) ¿Marca las cartas el jugador de poker a fin de aumentar su ventaja respecto a los otros jugadores?

c) ¿Es cierto que más de 60% de los nuevos doctores en sociología son mujeres?

9C-3. La dirección y el signo de una prueba de hipótesis se especifican en la hipótesis alternativa. Decide si las hipótesis alternativas (H_A) siguientes son de una cola en la dirección positiva, de una cola en la dirección negativa o de dos colas no direccional. También, incluye el signo matemático y explica tu elección.

a) H_A: más del 60% de los estudiantes universitarios se han involucrado en una borrachera en los últimos seis meses.

b) H_A: menos de 6% de la población de Estados Unidos tiene grado de doctor.

c) H_A: los porcentajes de sociólogos y psicólogos que atribuyen enfermedades mentales a atributos estructurales dentro de la sociedad no son iguales.

9C-4. En un estudio de patrones de logro académico entre estudiantes de preparatoria, un investigador hipotetizó que, en general, en los grados 10 a 12, las mujeres obtienen mejores calificaciones en las clases principales. En la hipótesis nula, H_0, el investigador hipotetizó que las calificaciones medias asignadas a los dos grupos de género son iguales.

a) ¿Por qué el investigador enunció la H_0 de esa manera en lugar de decir que la media de las mujeres es más alta?

b) En la hipótesis alternativa, H_A, ¿deberá utilizar el investigador una prueba de una cola o de dos colas? ¿Por qué?

c) En el paso 3 de la prueba, el investigador establece un nivel de significación (α) de .05. En el paso 4 de la prueba, compara un valor p de .41. En el paso 5, ¿rechazará o fracasará en rechazar H_0?

9C-5. Este ejercicio te familiarizará con las relaciones entre los niveles de significación, los valores p y las decisiones de rechazo. Para los niveles de significación y valores p siguientes, indica si rechazarías o fracasarías en rechazar la hipótesis nula, H_0.

	Nivel de significación (α del paso 3 de los seis pasos)	Valor p (del paso 4 de los seis pasos)	Decisión de rechazo: rechaza H_0 o fracasa en rechazar H_0
a)	.05	.3610	
b)	.01	.0031	
c)	.001	.0149	
d)	.01	.0067	
e)	.05	.0549	
f)	.05	.0476	

9C-6. Supongamos que alguien te pregunta si el número promedio de niños entre familias en una ciudad grande del oeste medio es 4. Tú no estás de acuerdo y crees que su promedio es menor que 4. Prueba la hipótesis con los datos muestrales siguientes:

$n = 195$ $\quad \overline{X} = 3.68$ niños $\quad s_X = 1.1$ niños

9C-7. Alguien te pregunta si el promedio de calificaciones (PC) promedio de los estudiantes de la State University es igual a B (es decir, un PC de 3.0 en una escala de 4). Tú no estás de acuerdo y tienes razones para creer que es menor que 3.0. Prueba la hipótesis con los datos muestrales siguientes:

$n = 140$ $\quad \overline{X} = 2.87$ puntos PC $\quad s_X = .8$ puntos PC

9C-8. En el paso 3 del procedimiento de los seis pasos de una prueba de hipótesis, decide si el nivel de significación (α), la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazar H_0. En la curva de la distribución muestral, ésta es la región crítica con una puntuación crítica expresada como un número de errores estándar.

a) Para el ejercicio 9C-6, ¿cuál es el valor crítico expresado en las unidades de la puntuación bruta del número de niños?

b) Para el ejercicio 9C-7, ¿cuál es el valor crítico expresado en las unidades de la puntuación bruta de puntos PC?

Conjunto de problemas 9D

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se indique lo contrario.

9D-1. Practica el arte de identificar hipótesis nulas y concebir distribuciones muestrales. En términos generales, anticipa qué resultados muestrales se pueden esperar que ocurran en el muestreo repetido cuando las hipótesis nulas (H_0) siguientes son ciertas. (Un repaso del capítulo 7 puede ser útil.)

a) H_0: la velocidad media de los conductores en un punto de control de la velocidad por parte de la policía es 78 millas por hora.

b) H_0: la edad media del cuerpo docente en una universidad pequeña de artes liberales es 42 años.

c) H_0: 70% de los estudiantes universitarios reportan haber participado en borracheras en los últimos seis meses.

d) H_0: el ingreso medio inicial entre graduados de una escuela de negocios es $47 000 por año.

9D-2. Una pregunta de investigación es una meta de proyecto que se puede enunciar en términos de una hipótesis. Practica el arte de determinar si cada una de las preguntas de investigación siguientes constituye la hipótesis nula (H_0) o la hipótesis alternativa (H_A). Explica tu respuesta.

a) Jane recién sacó un par de seises seguidos con un par de dados. ¿Está jugando con dados cargados?

b) En promedio, ¿abandonan su especialidad deseada de química los estudiantes de tercer año más que los estudiantes de primer año en otras disciplinas académicas?

c) En un área metropolitana grande, los oficiales de la policía promedian más de seis arrestos por semana implicando el manejar bajo la influencia del alcohol.

9D-3. La dirección y el signo de una prueba de hipótesis se especifican en la hipótesis alternativa. Decide si las hipótesis alternativas (H_A) siguientes son de una cola en la dirección positiva, de una cola en la dirección negativa o de dos colas no direccional. También, indica el signo matemático y explica tu elección.

a) H_A: ensayos de estudio indican que las tasas de absorción y rehidratación atribuidas al consumo de una bebida de desarrollo reciente son mayores que las tasas correspondientes entre las fórmulas existentes.

b) H_A: más de 10% de los maestros en un colegio comunitario local reportan haber tenido experiencia con mala conducta académica en sus grupos de clase el año pasado.

c) H_A: las edades medias entre hombres y mujeres en una muestra de estudio en todo el país no son iguales.

9D-4. En un estudio de profesiones en Estados Unidos, un investigador postula que las mujeres promedian niveles menores de estima ocupacional que los hombres en posiciones profesionales y administrativas idénticas. En la hipótesis H_0, el investigador hipotetiza que los niveles medios de estima profesional entre hombres y mujeres son iguales.

a) ¿Por qué el investigador enuncia H_0 de esa manera en lugar de decir que la media para las mujeres es menor?

b) En la hipótesis H_A, ¿deberá el investigador utilizar una prueba de una cola o de dos colas? ¿Por qué?

c) En el paso 3 de la prueba, el investigador establece un nivel de significación (α) de .05. En el paso 4 de la prueba, calcula un valor p de .024. En el paso 5, ¿rechazará H_0 o fracasará en rechazarla?

9D-5. Este ejercicio te familiarizará con las relaciones entre niveles de significación, valores p y decisiones de rechazo. Para los niveles de significación y valores p siguientes, indica si rechazarías o fracasarías en rechazar la hipótesis nula, H_0.

	Nivel de significación (α del paso 3 de los seis pasos)	Valor p (del paso 4 de los seis pasos)	Decisión de rechazo: rechaza H_0 o fracasa en rechazar H_0
a)	.01	.0124	
b)	.001	.0025	
c)	.05	.0059	
d)	.01	.0027	
e)	.05	.3110	
f)	.001	.0009	

9D-6. Estudios de personas indigentes de mediados del siglo XX determinaron de manera consistente la edad promedio de esta población como mayor de 40 años. Un estudio reciente de 150 personas indigentes en una ciudad del sur establecieron una edad media de 37.7 años con una desviación estándar de 9.6 años. ¿Es *menor que* 40 años la edad promedio de las personas indigentes en esa ciudad?

9D-7. Tú determinarás si los estudiantes que solicitan ingreso a un programa de posgrado tienen una puntuación CI media mayor que los otros estudiantes. Para hacer esto, tú aplicas una prueba del CI a una muestra aleatoria de 125 estudiantes que solicitan ingreso a un programa de posgrado. Se sabe que la puntuación CI promedio nacional de los estudiantes es 100. Su muestra produce una puntuación CI promedio de 105 con una desviación estándar de 10 puntos. ¿Cuál es tu conclusión?

9D-8. En el paso 3 del procedimiento de los seis pasos de una prueba de hipótesis, decidimos sobre un nivel de significación (α), la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazamos H_0. En la curva de la distribución muestral, ésta es la región crítica con una puntuación crítica expresada como el número de errores estándar.

a) Para el ejercicio 9D-6, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de años?

b) Para el ejercicio 9A-7, ¿cuál es el valor crítico expresado en las unidades de la puntuación bruta de puntos CI?

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 9

Si en tu clase se utilizan las aplicaciones computacionales opcionales que acompañan a este texto, abre los ejercicios en computadora del capítulo 9 en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2. Los ejercicios implican la prueba de medias de una muestra única grande y una orientación a los procedimientos estadísticos bivariados para *SPSS para Windows*. Además, el apéndice D de este texto proporciona un repaso breve de las secuencias de comandos *SPSS* para los procedimientos analizados en este capítulo.

CAPÍTULO 10

Prueba de hipótesis II: prueba de hipótesis de una muestra única: estableciendo la representatividad de las muestras

RESUMEN DEL CAPÍTULO

Introducción

En muchos aspectos el capítulo 10 es una extensión del capítulo 9, donde la prueba de medias de una muestra única grande se utiliza para introducir la lógica de las pruebas de hipótesis. En el capítulo 10, continuamos con pruebas de hipótesis de una muestra única e iniciamos con un análisis de sus aplicaciones comunes. Se presentan dos pruebas de hipótesis, la prueba de medias de una muestra única pequeña (cuando $n \leq 121$ casos) y la prueba de proporciones de una muestra única grande. Para esta última prueba de hipótesis, mostramos una aplicación importante: la prueba de la representatividad de la muestra. Por último, recordarás que el cálculo de un intervalo de confianza de una media poblacional (capítulo 8) tenía la restricción de que el tamaño de la muestra debía ser mayor que 121 casos. En este capítulo mostramos los ajustes mínimos necesarios para calcular un intervalo de confianza de una media poblacional para muestras con menos casos.

Como aprendimos en el capítulo 8, cuando queremos estimar un parámetro desconocido de una población, digamos, la edad media de los gerentes de restaurantes de comida rápida, calculamos un intervalo de confianza. Con los intervalos de confianza abordamos la pregunta: *¿Cuál es* el valor de la puntuación del parámetro? El intervalo de confianza proporciona una estimación de intervalo o rango. En este capítulo abordamos un tipo distinto de pregunta. En lugar de preguntar, *¿cuál es* el valor de la puntuación del parámetro? Preguntamos, *¿es igual el parámetro a algún valor objetivo elegido*?

¿De dónde provienen estos valores objetivo? Hay varias fuentes:

1. *Un parámetro poblacional conocido a partir de un grupo de comparación.* Por ejemplo, se sabe que la puntuación media del coeficiente intelectual (CI) para todas las personas es 100. Donde H_0 es la hipótesis nula y H_A es la hipótesis alternativa, podemos probar la hipótesis que el CI medio de un grupo específico, digamos, de inventores, es mayor:

   H_0: $\mu_{X(inventores)}$ = 100 puntos de CI (la media conocida, μ_X, de todos los adultos)

   H_A: $\mu_{X(inventores)}$ > 100 puntos de CI (la media conocida, μ_X, de todos los adultos). Una cola.

   Observa que en los lados izquierdos de estas ecuaciones anotamos como subíndice el nombre de la población en estudio y a la derecha indicamos la fuente del valor objetivo.

2. *Parámetros conocidos a partir de un periodo pasado.* Por ejemplo, los datos del censo de 2000 de Estados Unidos proporcionan parámetros reales, como la proporción de familias en Estados Unidos encabezadas por mujeres. Utilizando una muestra nacional actual de 1 500 familias, podemos probar la hipótesis que la proporción ha cambiado desde entonces:

   H_0: $P_{u\,(familias\ encabezadas\ por\ mujeres\ hoy)}$ = .19 (la proporción conocida, P_u, de familias encabezadas por mujeres en 2000).

   H_A: $P_{u\,(familias\ encabezadas\ por\ mujeres\ hoy)} \neq$ .19 (la proporción conocida, P_u, de familias encabezadas por mujeres en 2000). Dos colas.

3. *Una idea estadística.* Por ejemplo, como gerente de control de calidad de la Smellishou Perfume Company, tú seleccionas una muestra aleatoria de frascos de media onza de los anaqueles de la tienda y determinas si este volumen deseado se ha mantenido en todo el proceso de producción y entrega.

   H_0: $\mu_{X(frascos\ de\ Smellishou\ en\ los\ anaqueles\ de\ la\ tienda)}$ = .5 onzas (el valor objetivo deseado).

   H_A: $\mu_{X(frascos\ de\ Smellishou\ en\ los\ anaqueles\ de\ la\ tienda)}$ < .5 onzas (el valor objetivo deseado). Una cola.

4. *Se toma una muestra de una población y los estadísticos muestrales se comparan con algunos parámetros poblacionales conocidos para determinar si la muestra es representativa de la población.* Por ejemplo, supongamos que hemos muestreado 250 enfermeras en el estado. Todos los registros de las licencias estatales muestran que 64% de todas las enfermeras estatales son enfermeras registradas (ER). Éste es un parámetro conocido para nuestra población muestreada.

   Si nuestra muestra es representativa, casi 64% de las enfermeras en la muestra serán ER.

   H_0: $P_{u\,(población\ muestreada)}$ = .64 (el parámetro conocido (P_u) de la población objetivo de enfermeras del estado).

   Es decir, la muestra *es* representativa de las enfermeras del estado con respecto a la proporción de enfermeras con licencias de ER.

   H_A: $_{(población\ muestreada)} \neq$ .64 (el parámetro conocido (P_u) de la población objetivo de enfermeras del estado).

   Es decir, la muestra *no* es representativa de las enfermeras del estado con respecto a la proporción de enfermeras con licencias de ER.

Más adelante en este capítulo ilustraremos pruebas de hipótesis completas para algunas de estas fuentes de valores objetivo o meta.

Objetivo de una prueba de hipótesis de una muestra única Determinar si un parámetro de una población es igual a un valor "objetivo" especificado. Fuentes de valores objetivo: 1) grupos de comparación, 2) parámetros conocidos a partir de un periodo pasado, 3) ideales estadísticos y 4) parámetros conocidos de una población muestreada (para establecer la representatividad de la muestra).

La prueba de medias de una muestra única pequeña

En el capítulo 9, para ilustrar la lógica de la prueba de hipótesis aprendimos cómo hacer una prueba de medias de una muestra única *grande*. Al hacer esto, tuvimos precaución de advertir muchas veces que esta prueba, que utiliza la curva normal como distribución muestral, requiere un tamaño muestral mínimo de más de 121 casos. ¿Por qué es importante esto? Recuerda del capítulo 4 que la media tiene una limitación: el cálculo de una media se puede distorsionar por valores extremos o por una distribución sesgada.

Una distribución muestral de medias describe qué resultados muestrales ocurrirían si muestreamos y calculamos medias de manera repetida. La distribución muestral revela cómo cada media muestral diferirá ligeramente y con qué frecuencia esperar que una media muestral no dé con el parámetro real. Utilizamos la tabla de la curva normal y puntuaciones Z para calcular la probabilidad de resultados muestrales. Como resultado cuando el tamaño muestral, n, es 121 casos o menos, las posibilidades de que puntuaciones extremas distorsionen el cálculo de las medias muestrales son muchas. Si se incluye una puntuación extremadamente alta, la media tendrá un valor inflado. Si se incluye una puntuación extremadamente baja, la media estará desinflada. Cuando n es 121 o menos casos, los efectos de las puntuaciones extremas son tan grandes en el muestreo repetido que la curva de la distribución muestral tiene una dispersión grande en el error de muestreo y esto aplana la curva. A esta curva aplanada la denominamos distribución t.

La distribución muestral "t de Student"

Como se afirma en la ley de los números grandes, para una distribución muestral de medias, entre mayor sea el tamaño de la muestra, n, menor será el error estándar. Y lo contrario, en especial las muestras pequeñas tienen errores estándar tan grandes que la curva de probabilidad se alarga y se aplana. Este aplanamiento inicia su aparición en muestras con tamaños menores que 122 y es especialmente aparente cuando *n* es menor que 30.

La curva de la distribución muestral que utilizamos con muestras especialmente pequeñas se denomina *t* de Student o simplemente distribución *t*. Esta curva de la distribución es aproximadamente normal. Al igual que la curva normal, la curva de la distribución *t* es simétrica; es decir, la media, la mediana y la moda son iguales y una mitad de la curva es una imagen en espejo de la otra. Pero una curva de la distribución *t* es más plana que la curva normal (es decir, una distribución *t* es *platocúrtica* o aplanada como un plato).

Cuando se prueba una hipótesis, el estadístico de prueba —una medida de la distancia en el eje horizontal de esta curva de la distribución normal— se simboliza como *t* en lugar de *Z* para indicar que la curva sólo es aproximadamente normal. Para identificar puntuaciones críticas y evaluar valores *p*, utilizamos la tabla de la distribución *t* (tabla estadística C del apéndice B).

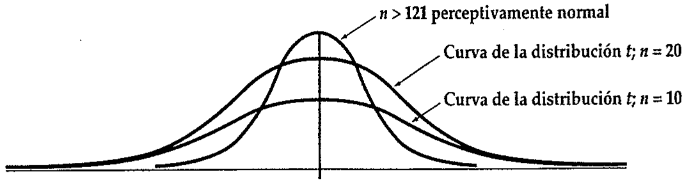
Distribución *t* aproximadamente normal Distribución muestral que es como la curva normal en que es simétrica, pero la curva está aplanada en vez de tener forma de campana. Para estadísticos de prueba, utiliza el símbolo *t* en lugar de *Z*. Usa la tabla de la distribución *t* en vez de la tabla de la curva normal para identificar puntuaciones críticas y evaluar valores *p*.

Para ilustrar el aplanamiento de una distribución muestral, enfoquémonos en el tamaño de la muestra. La figura 10-1 compara las distribuciones muestrales de tres tamaños muestrales de 10, 20 y 121 casos. Con un tamaño muestral mayor que 121, la curva será normal con su forma de campana distintiva. Un tamaño muestral de 121 o menos casos produce una curva de la distribución *t* que sólo es aproximadamente normal. Entre menor sea el tamaño de la muestra, más plana será la distribución *t*.

La forma de una curva de distribución *t* particular depende de un ajuste denominado grados de libertad (*gl*), que se definen y explican más adelante en este capítulo. Para una distribución muestral de medias, los grados de libertad se calculan como el tamaño de la muestra menos 1.

FIGURA 10-1

Aplanamiento de distribuciones muestrales de medias cuando *n* es menor que 121



Cálculo de los grados de libertad para una distribución muestral de medias

$$gl = n - 1$$

donde

gl = grados de libertad

n = tamaño de la muestra

Como se indica en la figura 10-1, entre menor sea la muestra, más plana será la curva. Para calcular probabilidades a partir de una curva de distribución *t*, necesitamos 120 tablas de distribución *t* al igual que la tabla de la curva normal, una para cada tamaño de muestra, de 121 hacia abajo hasta 2. Para ahorrar espacio, una sola tabla de distribución *t* (tabla estadística C del apéndice B) consolida la información de todas las 120 curvas. Sin embargo, esta tabla está diseñada de manera diferente de la tabla de la curva normal; estas diferencias se ilustran en la tabla 10-1.

La diferencia clave es que la tabla de la distribución *t* proporciona información sólo para las *regiones críticas* de $\alpha = .05$, $\alpha = .01$ y $\alpha = .001$ (capítulo 9). En la tabla de la distribución *t*, la columna izquierda proporciona los grados de libertad (*gl*) y la fila superior proporciona las áreas bajo la curva, pero sólo para estas regiones críticas. El cuerpo de la tabla contiene puntuaciones *t* críticas (t_α). En una curva de distribución *t*, estas puntuaciones *t* se utilizan de la misma manera que las puntuaciones *Z* críticas (Z_α).

Observa las distribuciones *t* de dos colas y de una cola en el lado derecho de la tabla 10-1. Para cualquier número de grados de libertad (*gl*), conforme se mueve hacia la derecha en la tabla, las puntuaciones *t* aumentan en tanto que los niveles de α disminuyen de .05 a .001. Esto simplemente refleja el hecho que entre mayor sea el valor de la puntuación crítica, menor será el área en la cola de la curva.

Observa que las puntuaciones t_α en la fila inferior de la tabla 10-1 son esencialmente equivalentes a puntuaciones Z_α. Las diferencias son tan pequeñas que no aparecen cuando las puntuaciones *t* se redondean. Es decir, las puntuaciones *Z* y las puntuaciones *t* son las mismas cuando el tamaño de la muestra (*n*) es 122 o más y, por tanto, los grados de libertad se calculan en 21 o más. Pero al subir por la columna α de tabla de la distribución la puntuación *t* siempre será mayor que la puntuación *Z* que aparece en la parte inferior de la tabla. Esto tiene sentido intuitivo debido a que las puntuaciones *Z* y las puntuaciones *t* miden la variabilidad en el error de muestreo, la distancia desde la media de la curva. Cuando el tamaño de la curva es pequeño, las medias calculadas tienen un error mayor debido a la sensibilidad de la media a puntuaciones extremas de *X*. No es inusual con una muestra pequeña obtener resultados muestrales ($\bar{X}$) más alejados de la cola de la curva, muy alejados del parámetro (μ_X). Este "empuje" de una región crítica más alejada de la media (μ_X) de la distribución muestral para una muestra pequeña se ilustra en la figura 10-2, que muestra los valores críticos *Z* y *t* al nivel .05, una cola cuando $n > 121$ y $n = 20$, respectivamente. Con una *n* de 20, sería tan inusual sacar una muestra cuya media esté a 1.725 errores estándar de μ_X como sería sacar una media a sólo 1.64 errores estándar de μ_X cuando *n* es mayor que 121.

TABLA 10-1 | Ubicaciones de información en las tablas de la distribución normal y de la distribución *t*

Tabla estadística B del apéndice B **Tabla de distribución normal**		**Tabla estadística C del apéndice B** **Tabla de la distribución *t***							
Columna A	Columna C		Dos colas				Una cola		
Puntuaciones Z	Área más allá de Z	*df*	.05	.01	.001	*df*	.05	.01	.001
1.64	0.0505 (≅ .05) una cola	•	•	•	•	•	•	•	•
•	•	5	2.571	4.032	6.869	5	2.015	3.365	5.893
2.33	0.0099 (≅ .01) una cola	•	•	•	•	•	•	•	•
•	•	10	2.228	3.169	4.587	10	1.812	2.764	4.144
3.08	0.0010 una cola	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•
1.96	0.0250 (× 2 = .05) dos colas	•	•	•	•	•	•	•	•
•	•	20	2.086	2.845	3.850	20	1.725	2.528	3.552
2.58	0.0049 (≃ .005 × 2 = .01) dos colas	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•
3.30	0.0005 (× 2 = .001) dos colas	•	•	•	•	•	•	•	•
•	•	30	2.042	2.750	3.646	30	1.697	2.457	3.385
•	•	120	1.980	2.617	3.373	120	1.658	2.358	3.160
•	•	∞	1.96	2.58	3.30	∞	1.64	2.33	3.08
↑			↑	↑	↑		↑	↑	↑

Las probabilidades α para regiones críticas se ubican aquí

Las Z críticas (Z_α) están en la columna A de la tabla de la distribución normal

Las t críticas (t_α) están en estas columnas de la tabla de la distribución t

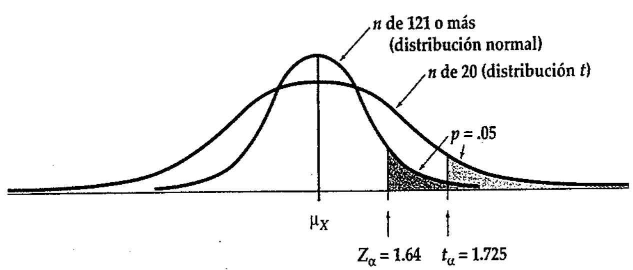


FIGURA 10-2
Comparación de las regiones críticas de una cola y puntuaciones críticas para tamaños de muestra de 121 o más (una distribución normal) y 20 (una distribución *t*)

No te confundas por esto. Una distribución t es simplemente una curva aproximadamente normal, una curva normal que está aplanada por una dispersión mayor del error de muestreo. Las puntuaciones t en la tabla de la distribución t son como puntuaciones Z excepto que son sólo para regiones críticas. Las puntuaciones t y Z calculadas miden lo mismo en estas curvas.

Selección de la puntuación crítica de probabilidad, t_α, a partir de la tabla de la distribución t

En el paso 3 del procedimiento de seis pasos de la inferencia estadística debemos identificar la puntuación crítica apropiada, t_α, a partir de la tabla de la distribución t (tabla estadística C del apéndice B). La puntuación t_α particular depende del nivel de significación elegido (α), de la dirección de la prueba (de dos colas/no direccional o de una cola/direccional) y del número de grados de libertad implicados por el tamaño de la muestra, con $gl = n - 1$. Los siguientes son algunos valores t_α de ejemplo tomados de la tabla de la distribución t:

Ejemplo 1: Queremos el nivel de significación de .05, la prueba es dos colas/no direccional $n = 15$. Calcula $gl = n - 1 = 14$. En la tabla estadística C, observa las cuatro columnas a la izquierda en "Prueba de dos colas o no direccional." Ubica 14 en la columna "*gl*". Desplázate hacia la puntuación crítica listada en $\alpha = .05$. La puntuación crítica $t_\alpha = 2.145$.

Ejemplo 2: Queremos el nivel de significación de .01, la prueba es de una cola/direccional y $n = 27$. Calcule $gl = n - 1 = 26$. En la tabla estadística C, observa las cuatro columnas a la derecha en "Prueba de una cola o direccional". Ubica 26 en la columna "*gl*". Desplázate hacia la puntuación crítica listada en $\alpha = .01$. La puntuación crítica, $t_\alpha = 2.479$.

Ejemplo 3: Queremos el nivel de significación de .05, la prueba es dos colas/no direccional y $n = 339$. Calcule $gl = n - 1 = 338$. En la tabla estadística C, observa las cuatro columnas a la izquierda en "Prueba de dos colas o no direccional". Encuentra "∞" en la columna "*gl*", significando $n > 121$ y, por tanto $gl > 120$. Desplázate hacia la puntuación crítica listada en $\alpha = .05$. La puntuación crítica, $t_\alpha = 1.96$. Esta puntuación crítica es la misma que la obtenida de la tabla de la curva normal (tabla estadística B del apéndice B). Se aplica a la prueba de una muestra grande que aprendimos en el capítulo 9 y a intervalos de confianza de una muestra grande en el capítulo 8. Como no es inusual tener muestras mayores que 121 y realizar pruebas de dos colas con un nivel de significación de .05, con frecuencia se utiliza t_α igual a 1.96.

Por último, el nombre t de Student tiene una historia interesante. El descubrimiento y la derivación matemática de la distribución t se hicieron a principios del siglo XX por un matemático llamado W.S. Gossett, quien trabajó para la Guinness Brewing Company en Dublín, Irlanda. Para proteger su ventaja competitiva en el negocio de la cerveza, la compañía no permitía que sus empleados publicaran su trabajo. Sin embargo, sus descubrimientos sobre tratar con muestras pequeñas fueron tan importantes para los estadísticos que la compañía hizo una excepción y permitió que Gossett publicara bajo el seudónimo Student. De aquí que firmara su trabajo de esa manera, y hasta la fecha los estadísticos se refieren a esta distribución muestral como t de Student.

Nota especial sobre los símbolos

Como mencionamos en el capítulo 9, los programas de software de cómputo como *SPSS* se refieren a cualquier prueba de medias de una muestra única, sin importar el tamaño de la muestra, como la prueba t. Ahora debes comprender por qué. Cuando $gl > 120$, una distribución muestral de medias se ajustará a la curva normal y Z es con certeza un símbolo apropiado para indicar que la distribución es normal. Pero para evitar complicar los programas de cómputo, sin considerar el tamaño de la muestra, utilizamos el símbolo t y nos referimos a todas las pruebas de medias de muestras únicas como pruebas t. Esto está bien. En cierto sentido, la curva Z normal se puede considerar como un caso especial de una curva t, el caso donde $n > 121$, proporcionando 121 o más grados de libertad. De nuevo, en tu software de cómputo, busca una prueba t de medias para correr cualquier prueba de medias de una muestra única para cualquier tamaño de la muestra. Además, en el resto de este texto nos referiremos a la prueba t aún cuando $n > 121$.

¿Qué son los grados de libertad?

Al hacer un análisis estadístico en una muestra, se debe tener cuidado de evitar procedimientos de investigación que conduzcan a conclusiones imprecisas acerca de la población. Cada instrumento de medición y técnica estadística tiene limitaciones que potencialmente distorsionan la interpretación de los datos. Por ejemplo, el famoso telescopio Hubble (que viaja en un satélite fuera de la atmósfera) proporciona imágenes distorsionadas debido a un desalineamiento microscópico en la curvatura de su lente. Como resultado, las imágenes fotográficas aparecen borrosas. Las estrellas no son borrosas; la borrosidad es una función de las limitaciones del instrumento de medición. El telescopio defectuoso se interpone en la evaluación precisa de las formas reales de las galaxias distantes. Ajustes en el lugar (por medio del transbordador espacial) compensan la curvatura del lente hasta cierto punto, pero las imágenes del telescopio Hubble aún no son puras. El telescopio tiene un grado de precisión pico, y este nivel es fijo. Las conclusiones extraídas acerca de la naturaleza de sus sujetos fotográficos (estrellas, galaxias, cuásares, etc.) están restringidas por las herramientas y los métodos utilizados para recolectar datos. Incluso con mejoras computacionales, los científicos a cargo del Hubble confrontan la falta de flexibilidad en la corrección de las distorsiones de su instrumento de medición. La fotografía que ellos ven sólo es una aproximación cercana de lo que realmente aparece allá.

De manera similar, los procedimientos estadísticos tienen limitaciones que potencialmente se interponen en una evaluación precisa de los parámetros de la población. Para estimar la dispersión de una distribución muestral de medias, debemos considerar los efectos de la limitación importante de la media (capítulo 4): el cálculo de la media se afecta por puntuaciones o valores extremos extremos. Este efecto distorsionante es problemático en especial con muestras pequeñas. Estando al tanto de esta limitación, ajustamos los cálculos para tomar en cuenta la sensibilidad de la media a valores extremos, al igual que los científicos del telescopio Hubble mejoran en computadora sus imágenes fotográficas. Todos los procedimientos estadísticos tienen límites: una falta de libertad total en cómo se utilizan. Empleamos el término *grados de libertad* para referirnos a qué tan flexible es un procedimiento estadístico. Entre más grados de libertad tengamos, mejor, debido a que los **grados de libertad** son el *número de oportunidades en el muestreo para compensar por las limitaciones, distorsiones y debilidades potenciales en los procedimientos estadísticos.*

¿Por qué se calculan los grados de libertad como $n - 1$ para la distribución t aproximadamente normal? Para una variable, una puntuación extrema en la muestra puede producir una media inflada o desinflada, una que no refleje el valor real del parámetro de la población de esa variable. Cuando el tamaño de la muestra es pequeño, esta distorsión puede ser un

tanto grande. Una vez que se saca de manera aleatoria una puntuación extrema con un valor muy alto en una muestra pequeña, no quedan muchas oportunidades para seleccionar un caso de valor bajo para sacar la media calculada más cerca del parámetro real de la población. Una puntuación extremadamente alta fija el cálculo de la media en el extremo del valor alto del rango de puntuaciones de la variable. La muestra pequeña es inflexible, no libre de la limitación de la media de la sensibilidad a puntuaciones extremas. Tiene pocos grados de libertad.

Para ilustrar estos principios, en el capítulo 7 (figuras 7-5 y 7-6) destacamos que un conjunto infinito de dígitos aleatorios individuales varía de 0 a 9 con una media de 4.5 (como en la tabla estadística A del apéndice B). Es decir, el parámetro poblacional real, μ_X, es 4.5. Sin embargo, suponga que no sabíamos esto, y para obtener una estimación de este parámetro poblacional, muestreamos estos dígitos y calculamos la media muestral, $\bar{X}$. Idealmente, esta estimación estaría cercana al parámetro real de 4.5. Esto se logra cuando en el proceso aleatorio de selección cada puntuación elegida en el lado alto (por ejemplo, 9, 8, 7 o 6) se equilibra por una puntuación en el lado bajo (por ejemplo, 0, 1, 2 o 3). Con un promedio real del parámetro de 4.5, una muestra aleatoria perfectamente precisa incluiría igual número de ceros como de nueves debido a que la media de estas dos puntuaciones es 4.5. De manera similar, esta muestra perfecta incluiría igual número de unos y ochos, doces y sietes, treces y seises, y cuatros y cincos. Pero supongamos que nuestro tamaño de muestra es pequeño, digamos, $n = 5$. Imagina aún más que el primer dígito obtenido para la muestra es un 9, una puntuación extrema. Cuando 9 se suma a ΣX al calcular la media muestral, es probable que cause que nuestra estimación del parámetro esté en el lado alto. Por ejemplo, la siguiente secuencia de extracciones muestrales podría ocurrir y resultar en una media muestral en el "lado alto" de 6.2:

Secuencia de muestreo: 9, 5, 3, 8, 6

$$\bar{X} = \frac{\Sigma X}{n} = \frac{31}{5} = 6.2$$

Con esta muestra pequeña, después de extraer el 9, es muy probable que nuestra estimación sea alta debido a que sólo nos quedan cuatro oportunidades de muestreo ($n - 1$) para obtener un 0 para equilibrar el cálculo de la media. Diríamos que sólo tenemos 4 grados de libertad. Extrayendo un 9 nos traba en una estimación en el lado alto. Una muestra de cinco no es muy flexible una vez que una puntuación extrema entra en la muestra.

Con una muestra grande, digamos, de tamaño 130, la extracción de un 9 al inicio no es un problema tan serio. Tenemos 129 más oportunidades de sacar un 0 para regresar el cálculo de la media al rango de 4.5. Con una muestra grande hay mayor libertad de ajuste en el procedimiento de muestreo.

Otra forma de ver el concepto de grados de libertad es considerar la "independencia de los eventos de muestreo". Por ejemplo, supongamos que alguien dice que reunió cinco dígitos aleatorios y calculó una media de 6.2, como en la ilustración anterior. Si este investigador nos dijo los valores de cuatro de los dígitos, podríamos determinar el quinto de manera matemática. Es decir, si los primeros cuatro dígitos son 9, 5, 3 y 8, para que la ΣX sea igual a 31 para producir una media de 6.2, el último dígito tiene que ser 6. En otras palabras, el último dígito no tiene libertad para variar; su valor depende de cómo se calcule la media. Por tanto, al calcular los grados de libertad para una media, restamos 1 al tamaño de la muestra. En este ejemplo esto nos deja 4 grados de libertad. Cuatro de los dígitos tienen "libertad para variar". Entonces, los grados de libertad se pueden considerar como el número de eventos muestrales independientes, eventos que son independientes de las limitaciones de la fórmula estadística utilizada.

Sólo con una distribución muestral de medias se calculan los grados de libertad como $n - 1$. Con otros procedimientos estadísticos, los ajustes para los grados de libertad dependen de las limitaciones particulares de un procedimiento. En los capítulos restantes, analizamos las limitaciones de los diversos procedimientos estadísticos y hacemos los ajustes de los grados de libertad. Vale la pena buscar en los capítulos restantes el símbolo *gl* para que te familiarices con la noción que los estadísticos de la muestra incluyen este ajuste.

> **Grados de libertad (*gl*)** Número de oportunidades en el muestreo para compensar por limitaciones, distorsiones y debilidades potenciales en los procedimientos estadísticos o el número de eventos muestrales independientes. Los grados de libertad se calculan de diversas maneras para varios procedimientos estadísticos.

Los cálculos de los grados de libertad representan el reconocimiento de las limitaciones de un procedimiento. Utilizaremos palabras como "*ajusta* para grados de libertad" y "*corrige* para grados de libertad". Con frecuencia decimos que una limitación particular nos ocasiona "*perder* grados de libertad". Por ejemplo, la sensibilidad de la media a valores extremos nos ocasiona perder 1 grado de libertad. Ajustar por los grados de libertad de un procedimiento es una parte esencial de la evaluación del error de muestreo. Debemos estar conscientes de manera constante que muestrear es mirar por un lente angosto. Debemos preguntar: ¿es lo mismo lo que vemos y lo que en realidad está allí? Si sabemos que nuestra lente está desenfocada, debemos tomarlo en cuenta, al igual que los científicos del telescopio Hubble "corrigieron" sus imágenes fotográficas digitales con mejoras computacionales. El cálculo de los grados de libertad es nuestro modo de corrección.

Los seis pasos de la inferencia estadística para una prueba de medias de una muestra única pequeña

En el capítulo 9 aprendimos los seis pasos de la inferencia estadística para una prueba de medias de una muestra única grande. Ahora que hemos determinado que se requieren consideraciones especiales para una prueba de medias cuando *n* es pequeña (121 o menos casos), apliquemos los seis pasos de la inferencia estadística a esta circunstancia.

En 2005, Estados Unidos sufrió un desastre natural histórico cuando el huracán Katrina devastó la Costa del Golfo de México desde el este de Texas hasta Alabama. Más de un cuarto de millón de personas fueron evacuadas y muchas perdieron sus casas y pertenencias. Las ciencias sociales tienen una gran tradición en la investigación de desastres, incluyendo el estudio de cómo esos eventos afectan la salud mental y física. Supongamos que en las semanas después del paso del huracán, encuestamos personas evacuadas de la Costa del Golfo quienes ahora viven en refugios públicos de emergencia en una ciudad algunos cientos de millas al norte. El estudio implica una variedad de variables relacionadas a servicios de emergencia, recursos sociales y capitales, prospectos futuros, y efectos en la salud por el desastre. Supongamos que entre nuestras medidas de salud está una serie de puntos de un cuestionario denominados Escala de Desorden de Estrés Agudo (ASDS; Bryant, Moulds y Guthrie, 2000). La escala mide el *estrés* agudo: las *reacciones inmediatas* a un evento traumático. Sus 19 enunciados en la encuesta requieren que los respondientes indiquen con qué frecuencia han

tenido síntomas desde el evento. La lista incluye síntomas de disociación, como entumecimiento, una sensación de irrealidad, y el sentimiento de estar en el exterior viendo cómo se desarrollan los eventos. Además, la escala mide el temor a volver a sufrir el trauma, evitar pensar y hablar acerca de éste, y excitación psicológica, como irritabilidad, nerviosismo, reaccionar exageradamente a estímulos, y sentimientos de angustia. La escala es especialmente útil al predecir qué poblaciones de personas estresadas son probables que sufran el desorden de estrés a largo plazo del Síndrome de Estrés Postraumático (PTSS).

La ASDS es de nivel de medición de intervalo, con puntuaciones variando de 19 a 95, donde una puntuación alta indica mayor estrés. Estudios anteriores utilizando la ASDS revelan que los individuos con una puntuación mayor que 56 tienen una buena probabilidad de desarrollar PTSS (Bryant, Moulds y Guthrie 2000:64-65). Podemos utilizar esta puntuación de corte como una indicación para saber si nuestra *población* de víctimas de desastre está en riesgo de PTSS a largo plazo. Si es así, es necesaria la intervención del departamento de salud pública, como protección de individuos en los refugios. Entonces se pueden proporcionar servicios de salud mental a los individuos con puntuaciones ASDS altas para disminuir el riesgo de PTSS.

Por tanto, nuestro interés se centra en la puntuación ASDS media de esta población de víctimas del huracán Katrina en refugios públicos. Si la puntuación ASDS media es mayor que 56, entonces este grupo está en riesgo de PTSS. Entrevistamos a 24 adultos seleccionados al azar en refugios. Determinamos una media de 64.54 puntos en la escala ASDS con una desviación estándar de 19.27 puntos en la escala.

Pensemos por un momento en nuestra pregunta de investigación y cómo podríamos probarla. Nuestra pregunta de investigación es: en promedio, ¿tienen puntuaciones ASDS medias arriba de la puntuación de corte de 56 las víctimas del huracán Katrina, poniéndolas en riesgo de PTSS? Para cualquier prueba de hipótesis de una muestra única, debemos tener un valor objetivo para formular la hipótesis nula (H_0) y éste lo proporciona la puntuación de corte de 56 puntos en la escala ASDS. Supondremos que este valor es verdadero y luego haremos predicciones como si lo fuera. En otras palabras, el valor objetivo para H_0 nos permite describir la distribución muestral, suponiendo que H_0 sea cierta. Nuestra hipótesis alternativa (H_A) será que el ASDS medio para esta población es mayor que 56. Podemos comparar nuestra media muestral de 64.54 con nuestra puntuación de corte de 56. La diferencia es el efecto de prueba:

$$\bar{X} - \mu_X = 64.54 - 56 = 8.54 \text{ puntos en la escala ASDS}$$

Observamos que en promedio nuestros sujetos muestrales tienen una puntuación 8.54 puntos arriba del punto de corte. Sin embargo, no podemos de manera rápida concluir que esta *población* de víctimas en refugios tiene una ASDS alta. ¿Por qué? Porque el efecto de la prueba podría ser el resultado del error del muestreo, en especial puesto que tenemos una muestra pequeña. En otras palabras, podríamos tomar una segunda muestra y determinar que su media está 8.54 puntos *debajo* de 56. Recuerda que el objetivo estadístico de cualquier prueba de hipótesis es determinar si un efecto estadístico calculado de una muestra indica un efecto real en la población o simplemente un error del muestreo. Nuestra prueba de hipótesis determinará si la media de 64.54 puntos ASDS para la muestra está tan arriba de 56 puntos ASDS como para ser estadísticamente diferente de 56.

Ahora examinaremos los criterios para la selección de la prueba *t* y presentaremos los seis pasos de la inferencia estadística.

Cuándo utilizar una prueba de medias de una muestra única pequeña ($n \leq 121$; distribución t, $gl = n - 1$)

En general: resulta útil para probar una hipótesis de que la media de una población, μ_X, es igual a un valor objetivo.

1. Existe sólo una variable.
2. El nivel de medición es de intervalo/razón.
3. Hay una sola variable y una población.
4. El tamaño de la muestra, n, es 121 o menos casos (aunque esta prueba se puede emplear para muestras mayores, se requiere cuando n es pequeña).
5. Hay un valor objetivo de la variable al cual podemos comparar la media de la muestra.

Examinemos nuestra pregunta de investigación y determinemos si se requiere esta prueba de hipótesis. Primero, tenemos la variable individual, ASDS. Segundo, es de nivel de medición de intervalo. Tercero, tenemos una muestra de la población de víctimas del huracán Katrina en refugios. Cuarto, nuestro tamaño de muestra de 24 es "pequeño" (es decir, 121 o menor). Por último, tenemos un valor objetivo respecto al cual podemos proyectar una distribución muestral.

Ahora apliquemos los seis pasos de la inferencia estadística a nuestra cuestión de si las víctimas del huracán Katrina en refugios públicos sufren niveles altos de ASDS. El siguiente cuadro presenta la solución. Observa que es compatible con los pasos de la tabla 9-5 de la página 296. Al igual que con todas las pruebas de hipótesis, estudia la solución y compárela con la tabla 9-5 así como con las pruebas de hipótesis que ha aprendido.

Lista de verificación breve de los seis pasos de la inferencia estadística

PREPARACIÓN DE LA PRUEBA

Enuncia la pregunta de investigación. Traza diagramas conceptuales representando los datos, incluyendo la o las poblaciones y la o las muestras en estudio, las variables (por ejemplo, $X = ...$, $Y = ...$,) y sus niveles de medición, y los estadísticos y parámetros dados o calculados. Formula el procedimiento estadístico de prueba apropiado.

SEIS PASOS

Utilizando el símbolo H para *hipótesis*:

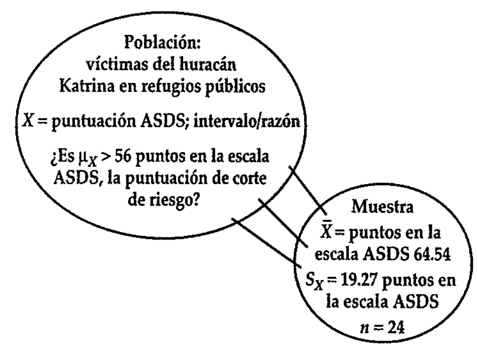
1. Formula H_0 y H_A y estipula la dirección de la prueba.
2. Describe la distribución de muestreo.

3. Estipula el nivel de significación (α) y la dirección de la prueba y especifique la puntuación crítica de prueba.
4. Observa los resultados muestrales actuales y calcula los efectos de prueba, el estadístico de prueba y el valor p.
5. Toma la decisión de rechazo.
6. Interpreta y aplica los resultados y proporciona las mejores estimaciones en términos comunes.

Solución para una prueba de medias de una muestra única pequeña (cuando $n \leq 121$)

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: en promedio, ¿tienen puntuaciones ASDS medias las víctimas del huracán Katrina arriba de la puntuación de corte de 56, poniéndolas en riesgo de PTSS? *Procedimiento estadístico*: prueba de medias de una muestra única pequeña. *Datos*:



SEIS PASOS

1. H_0: $\mu_{X(\text{víctimas de Katrina en refugios})} = 56$ puntos ASDS (la puntuación media de corte, μ_X, arriba de la cual una población está en riesgo de PTSS).

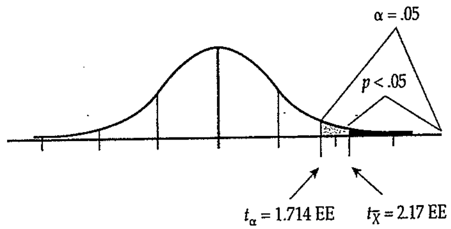
   Es decir, la ASDS media de víctimas del huracán Katrina en refugios no es diferente de la puntuación de corte.

   H_A: $\mu_{X(\text{víctimas de Katrina en refugios})} > 56$ puntos ASDS.

   Es decir, la ASDS media de víctimas del huracán Katrina está arriba de la puntuación de corte, indicando que está en riesgo de PTSS. Una cola.
2. *Distribución muestral*: si H_0 es cierta y se toman muestras repetidamente de tamaño 24 de la población de víctimas del huracán Katrina en refugios, las

medias muestrales ($\overline{X}$) estarán centradas en 56 como una distribución t normal $gl = n - 1 = 23$, con un error estándar:

$$s_{\overline{X}} = \frac{s_X}{\sqrt{n}} = \frac{19.27}{\sqrt{24}} = \frac{19.27}{4.90} = 3.93 \text{ puntos en la escala ASDS}$$



44.21	48.14	52.07	56	59.93	63.86	67.79	$\overline{X}$
−3EE	−2EE	−1EE	0	1EE	2EE	3EE	$t_{\overline{X}}$
			μ_X				

(El sombreado de la curva se hace en los pasos 3 y 4.)

3. *Nivel de significación*: $\alpha = .05$. Una cola. Puntuación crítica de prueba $t_\alpha = 1.714$ EE. (Sombrea y marca la región crítica en la curva en el paso 2.)

4. *Observación*:

   Efecto de la prueba: $\overline{X} - \mu_X = 64.54 - 56 = 8.54$ puntos en la escala ASDS

   $$\text{Estadístico de la prueba: } t_{\overline{X}} = \frac{\overline{X} - \mu_X}{s_{\overline{X}}} = \frac{64.54 - 56}{3.93} = 2.17 \text{ EE}$$

   Valor p: p [extraer una muestra con una media ($\overline{X}$) tan inusual o más inusual que 64.54 cuando la media poblacional verdadera (μ_X) es 56] $< .05$.

   (Sombrea y marca el área del valor p en la curva en el paso 2).

5. *Decisión de rechazo:* $|t_{\overline{X}}| > |t_\alpha|$ (es decir, $2.17 > 1.714$); por tanto, $p < \alpha$ (es decir, $p < .05$); rechaza H_0 y acepta H_A al nivel de confianza de 95%.

6. *Interpretación:* la puntuación media en la escala ASDS de víctimas de Katrina en refugios parece significantemente mayor que la puntuación de corte de 56. *Mejor estimación:* estimamos que la puntuación media en la escala ASDS de víctimas de Katrina es 64.54. *Respuesta:* las víctimas del huracán Katrina en refugios públicos parecen estar en un gran riesgo de sufrir un desorden de estrés agudo. Deberán ser el objetivo de los servicios de salud mental dirigidos a la detección y prevención de los efectos de este desorden así como del Desorden de Estrés Postraumático.

Algunos puntos a destacar acerca de esta prueba de hipótesis:

- En el paso 1, observa que la hipótesis es acerca de parámetros (μ_X), no de estadísticos ($\overline{X}$). Cada prueba de hipótesis es acerca de la población y sus parámetros. Los estadísticos muestrales sólo son estimaciones de parámetros poblacionales y la muestra sólo es una herramienta para hacer inferencias estadísticas acerca de la población.
- En el paso 1, enunciamos la hipótesis alternativa, H_A, como una prueba positiva de una cola y nos enfocamos en el lado derecho de la curva. Esta decisión *no* se hizo examinando la media muestral para ver si era mayor que 56. En lugar de eso, se eligió la dirección positiva con base en factibilidad. Puntuaciones *mayores que* 56 puntos en la escala ASDS son de interés para predecir el PTSS.
- En el paso 2, describimos la distribución muestral para el parámetro hipotético del paso 1. La distribución muestral nos indica que si esta población tiene una ASDS media de 56, no mayor o menor, y si muestreamos de manera repetida con $n = 24$, casi 68% de las veces las puntuaciones ASDS medias muestrales, $\overline{X}$, caerán entre 52.07 y 59.93; casi 95% de las veces entre 48.14 y 63.86, y así sucesivamente.
- En el paso 3, observamos la tabla de la distribución t para la puntuación t_α crítica para una prueba de una cola al nivel de significación .05 con $gl = 23$ y determinamos que es 1.714 errores estándar (EE). Esta puntuación crítica define la región crítica, el área sombreada y marcada en la curva en el paso 2 como $\alpha = .05$. Si, en efecto, la puntuación ASDS media de nuestra población de víctimas de Katrina es 56, sólo 5% de las veces en el muestreo repetido las medias muestrales caen más de 1.714 EE arriba de 56. Nuestra media muestral de 64.54 puntos ASDS caen en la región crítica, por tanto la consideramos significativamente diferente de 56 puntos y rechazamos H_0 en el paso 5.
- En el paso 4, en lugar de enfocarnos en la región crítica lo hacemos en el valor p. El cálculo del valor p nos indica qué tan inusual es el resultado muestral observado *cuando H_0 es verdadera*. Responde la pregunta: con muestreo repetido, ¿con qué frecuencia ocurre una media de 64.54 puntos ASDS o más cuando la media poblacional es 56? Como $p < .05$, consideramos la media muestral tan inusual como para rechazar 56 como la media poblacional verdadera.
- En el paso 4, regresamos al paso 2 y trazamos el área del valor p en la curva de la distribución muestral. Sin importar si H_0 es verdadera o no, en el paso 2 anticipamos qué ocurre en el muestreo "si H_0 es verdadera" y en el paso 4 el cálculo del valor p supone esto por el momento. Recuerda que los pasos 1 a 4 se basan en la suposición que H_0 es verdadera.
- En el paso 5, toma nota de las palabras cuidadosas de la decisión de rechazo. Cuando rechazamos H_0, aceptamos H_A, pero con precaución. Un resultado de 64.54 puntos ASDS no ocurre en el muestreo repetido cuando H_0 es 56, aunque sea menos del 5% de las veces. Como no muestreamos toda la población, no podemos estar 100% seguros de nuestro resultado. Pero estableciendo α en .05, sólo estamos corriendo un riesgo de 5% de rechazar H_0 cuando de hecho es verdadera. Es decir, estamos corriendo un riesgo de 5% de cometer un error tipo I y de sacar la conclusión errónea. Por otro lado, tenemos una posibilidad de 95% de no cometer ese error. Por tanto, estamos 95% seguros de nuestro resultado.

- En el paso 6, debido a que rechazamos la hipótesis nula, H_0, la interpretación se enfoca en la H_A aceptada. En este punto, podemos rechazar H_0. También observa que el tono del paso 6 es sustantivo. Es un análisis de la pregunta de investigación y aborda conceptos y variables. Los primeros pasos abordan aspectos técnicos del procedimiento estadístico y se enfocan en la teoría de las probabilidades.
- En el paso 5, rechazamos H_0 que $\mu_X = 56$ puntos ASDS. En el paso 6, una audiencia pública o profesional querrá saber qué aceptamos en lugar de 56. Por tanto, proporcionamos una *mejor estimación*. Como somos los únicos investigadores con una estimación estadística de ASDS para esta población, proporcionamos nuestro resultado muestral de 64.54 puntos ASDS. Esta es una estimación puntual. Para una audiencia profesional, se podría requerir un cálculo del intervalo de confianza (capítulo 8).

Adquiriendo un sentido de proporción acerca de la dinámica de una prueba de medias

Hasta este punto, hemos aprendido dos tipos de pruebas de hipótesis, las pruebas de medias de una muestra única grande y de una pequeña. En esencia, podríamos llamar a las dos pruebas t y utilizar la distribución t con el conocimiento que cuando $n > 121$, la distribución muestral es normal en lugar de solamente aproximadamente normal. Esto es obvio en la tabla 10-1 donde, cuando $n > 121$, las puntuaciones t de la tabla de la distribución t son iguales a las puntuaciones Z críticas de la tabla de la curva normal. Como hicimos notar, los programas de cómputo se refieren a todas las pruebas de medias de una muestra única como pruebas t.

Las pruebas de medias de una muestra única son buenas para aprender la idea básica detrás de las pruebas de hipótesis. El valor objetivo de la hipótesis nula de una prueba de medias es fácil de concebir. La lógica de la prueba de hipótesis es simple. Hipotetizamos una puntuación media de una variable de intervalo/razón X en H_0, digamos, por ejemplo, la hipótesis que el promedio de calificaciones en el campus, μ_X, es 3.0 (una B). Observamos una media muestral, $\bar{X}$. Si la media de la muestra está cercana al valor del parámetro hipotético, digamos, 2.9 puntos de PC, nos quedamos con el valor de X en H_0 y tratamos la diferencia (es decir, el efecto) como debida al error del muestreo. Si la media muestral no está en la vecindad del valor hipotético de X en H_0, digamos, 2.4 puntos de PC, entonces rechazamos H_0. Concluimos que el efecto es tan grande que es probable que no ocurriera debido al error del muestreo. Aunque este concepto es lo suficientemente simple, debemos determinar con exactitud qué es "cerca". Los seis pasos de la inferencia estadística definen "cerca" en relación al error de muestreo. Ahora examinemos con más detalle los elementos de la prueba de hipótesis para obtener un sentido de cómo se conectan las partes.

Relaciones entre parámetros hipotéticos, estadísticos muestrales observados, estadísticos de prueba calculados, valores p y niveles alfa

Es importante tener un sentido de proporción acerca de las relaciones entre el parámetro hipotético (μ_X), el estadístico muestral observado ($\bar{X}$), el estadístico de prueba calculado ($t_{\bar{X}}$), el nivel de significación y su puntuación crítica (α y t_α), y el valor p. Esta lista de conceptos puede parecer abrumadora, pero una vez que se comprendan sus interrelaciones, las cosas se empiezan a encajar y las pruebas de hipótesis parecen muy simples. Pensemos en términos de la posibilidad que la hipótesis nula (H_0) se rechace, después de lo cual se acepta la

hipótesis alternativa (H_A). En otras palabras, ¿en qué condiciones ya no estamos dispuestos a aceptar que la hipótesis nula es verdadera? ¿Qué observaciones muestrales conducen al rechazo de la hipótesis nula?

Regla 1: La hipótesis nula se rechaza cuando el efecto de la prueba es grande, lo suficientemente grande que el valor del estadístico de prueba sea mayor que la puntuación crítica de la prueba Una vez más, el efecto de la prueba es la diferencia entre el estadístico muestral observado y el parámetro hipotético. Es una puntuación de desviación en la curva. La hipótesis nula se rechazará cuando el efecto sea grande. Recuerda el ejemplo simple de hipotetizar que va a llover pronto. Esperamos ver nubes oscuras densas en el cielo. Si observamos un cielo azul claro, este efecto es tan diferente de lo que esperamos que rechazamos la hipótesis de lluvia.

Cuando el efecto de prueba es grande, el valor del estadístico de prueba también será grande. Para la prueba de medias de una muestra única, el estadístico de prueba es $t_{\bar{X}}$, que se calcula en el paso 4 de los seis pasos. Mide el efecto de prueba *como un número de errores estándar (EE)*. Cuando $|t_{\bar{X}}| > |t_\alpha|$, $p < \alpha$; rechazamos la hipótesis nula. Cuando $|t_{\bar{X}}| < |t_\alpha|$, $p > \alpha$ fracasamos en rechazar la hipótesis nula. Con una prueba de medias, por ejemplo, el efecto de prueba se calcula como $\bar{X} - \mu_X$ y este término aparece en el numerador del estadístico de prueba. En cualquier problema de división, cuando el numerador es grande, el cociente será grande. Imagina que hipotetizamos que el promedio de calificaciones (PC) de estudiantes es mayor que 2.6. Ésta sería una hipótesis alternativa, debido a que debemos enunciar la hipótesis nula como igual a 2.6 para producir una distribución muestral (es decir, H_0: μ_X = 2.6 puntos de PC). Comparemos muestras de estudiantes con efectos de prueba grandes y pequeños, una muestra de la State University y otra de la Crosstown University. Por simplicidad, supongamos que los errores estándar son los mismos e iguales a .2 puntos de PC y que los tamaños de las muestras son 500. Las medias muestrales de los dos campus difieren como sigue:

De la State University (un efecto de prueba y estadístico de prueba grandes):

Paso 1. H_0: $\mu_X = 2.6$ puntos de PC
H_A: $\mu_X = 2.6$ puntos de PC — Una cola

Paso 4. Media muestral observada:

$\bar{X} = 3.0$ puntos de PC

Efecto de la prueba $= \bar{X} - \mu_X = 3.0 - 2.6 = .4$ puntos de PC

$$\text{Estadístico de la prueba} = t_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}} = \frac{3.0 - 2.6}{.2} = \frac{.4}{.2} = 2.00 \text{ EE}$$

De la Crosstown University (un efecto de prueba y estadístico de prueba pequeños)

Paso 1. H_0: $\mu_X = 2.6$ puntos de PC
H_A: $\mu_X > 2.6$ puntos de PC — Una cola

Paso 4. Media muestral observada:

$$\bar{X} = 2.7 \text{ puntos de PC}$$

$$\text{Efecto de la prueba} = \bar{X} - \mu_X = 2.7 - 2.6 = .1 \text{ punto de PC}$$

$$\text{Estadístico de la prueba} = t_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}} = \frac{2.7 - 2.6}{.2} = \frac{.1}{.2} = .50 \text{ EE}$$

La media muestral de la State University falla en .4 puntos de PC, casi la mitad de una letra de calificación. Esto está 2.00 errores estándar alejado del PC de 2.6 que esperamos. Ésta es una diferencia grande entre el PC de 3.0 de la muestra observada y el parámetro hipotético de 2.6. Es tan grande que podemos decir con seguridad que *no* se debe al error aleatorio del muestreo. Rechazamos la hipótesis nula que el PC en la State University es 2.6 y concluimos que es mayor. Por otro lado, para la muestra de la Crosstown University, el efecto de la prueba fue pequeño —sólo .1 puntos de PC— y por tanto $t_{\bar{X}}$ es pequeña. La media muestral sólo es la mitad de un error estándar de la media esperada. Nuestra experiencia con distribuciones muestrales sugiere que esta diferencia pequeña podría resultar fácilmente del error del muestreo. Para resumir la regla 1, cuando el efecto de la prueba y por tanto el valor absoluto del estadístico de prueba $t_{\bar{X}}$ son grandes, la posibilidad de rechazar la hipótesis nula aumenta.

Regla 2: Entre mayor sea el efecto de la prueba y el estadístico de la prueba, menor será el valor p Efectos de la prueba y valores del estadístico de la prueba grandes son inusuales cuando la hipótesis nula es verdadera. El resultado muestral observado es muy inusual (es decir, tiene una probabilidad baja de ocurrencia) cuando el efecto de la prueba es grande. En la curva de la distribución *t*, la baja probabilidad de un efecto de la prueba es aparente en el área pequeña en la cola de la curva detrás de una $t_{\bar{X}}$.

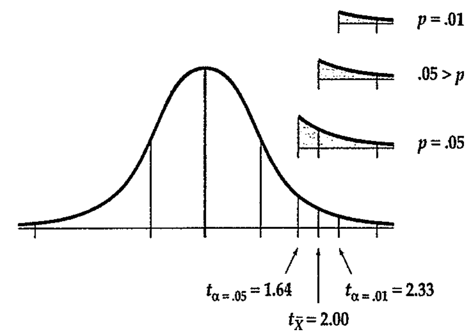
La curva de la distribución *t* aproximadamente normal es útil para obtener un sentido de proporción acerca de la relación entre el valor del estadístico de prueba y el valor *p*. Como la tabla de la distribución *t* proporciona áreas en la curva sólo para las regiones críticas, sólo podemos estimar el valor *p* de un valor estadístico de la prueba particular $t_{\bar{X}}$. Ilustremos esto examinando los ejemplos de la Crosstown University y de la State University utilizando una prueba de una cola. En las dos muestras, $gl = n - 1 = 499$. Los valores críticos t_α se encuentran en la fila "∞" de la tabla de la distribución *t*.

Para la State University:

$$H_0\text{: } \mu_X = 2.6 \text{ puntos de PC}$$
$$\bar{X} = 3.0 \text{ puntos de PC}$$
$$\bar{X} - \mu_X = .4 \text{ puntos de PC}$$
$$t_{\bar{X}} = 2 \text{ EE}$$

Estos valores del efecto de la prueba y del estadístico de la prueba son relativamente grandes. El valor *p* es relativamente pequeño. Para ver esto, observa la tabla de la distribución *t* para obtener el valor *p*. El valor *t* calculado, $t_{\bar{X}} = 2$ EE, cae entre los valores t_α críticos de 1.64 y 2.33; por tanto $p\ [t_{\bar{X}} = 2 \text{ EE}]$ es menor que .05 pero mayor que .01.

El valor *p* se estipula: *p* [sacar una muestra con una media ($\bar{X}$) tan inusual como o más inusual que 3.0 puntos de PC cuando la media de la población verdadera (μ_X) es 2.6 puntos de PC < .05.

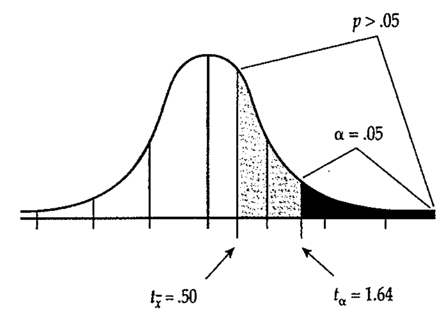


Para la Crosstown University:

$$H_0\text{: } \mu_X = 2.6 \text{ puntos de PC}$$
$$\bar{X} = 2.7 \text{ puntos de PC}$$
$$\bar{X} - \mu_X = .1 \text{ puntos de PC}$$
$$t_{\bar{X}} = .5 \text{ EE}$$

Estos valores del efecto de la prueba y del estadístico de la prueba son relativamente pequeños. El valor *p* es relativamente grande. Observando la tabla de la distribución *t*, el valor *t* calculado, $t_{\bar{X}} = .5$ EE, cae a la izquierda de la puntuación crítica t_α de 1.64; por tanto, $p\ [t_{\bar{X}} = .5]$ es mayor que .05.

El valor *p* se estipula como *p* [sacar una muestra con una media ($\bar{X}$) tan inusual como o más inusual que 2.7 puntos de PC cuando la media de la población verdadera (μ_X) es 2.6 puntos de PC] > .05.



En esta comparación, podemos ver que para la muestra de la Crosstown University *no* sería tan inusual sacar una media muestral de 2.7 puntos de PC de esta población si su media es 2.6 puntos de PC. De hecho, si el parámetro de la población es en efecto 2.6 puntos de PC, en el muestreo repetido, una gran proporción de las veces, sacaríamos medias muestrales que estarían cerca de la media hipotética. La probabilidad alta de este resultado es aparente en el área sombreada grande de la curva de la Crosstown University. Al probar al nivel de significación de .05, fracasaríamos en rechazar la hipótesis nula para la Crosstown University. En contraste, con la muestra de la State University, concluiríamos que *es* inusual sacar una media muestral de 3.0 puntos de PC de una población con una media igual a 2.6 puntos de PC. De hecho, este efecto de .4 puntos de PC raramente ocurre como un resultado del error de muestreo aleatorio, menos de 5 veces de 100 muestras. En la curva de la State University, esto es aparente en el área comparablemente pequeña más allá de $t_{\bar{X}} = 2.00$ EE. Al probar al nivel de significación de .05, rechazaríamos la hipótesis nula para el caso de la State University.

Para resumir esta segunda regla acerca de la relación de los estadísticos y los valores *p*, entre mayor sea el efecto de prueba, y por tanto entre mayor sea el estadístico de prueba, menor será la probabilidad de que el resultado muestral ocurrió como un resultado del error del muestreo aleatorio. Ahora analicemos los niveles de significación (niveles alfa) en relación con la dirección de la prueba.

Regla 3: Es más fácil rechazar con una prueba de una cola que con una prueba de dos colas Otro ejercicio que te ayudará a obtener un sentido de proporción acerca de las pruebas de hipótesis, implica examinar si es más fácil rechazar la hipótesis nula con una prueba de una cola o con una prueba de dos colas. Como mencionamos en el capítulo 9, la respuesta es una prueba de una cola, debido a que con una prueba de una cola el efecto de la prueba —la diferencia entre el resultado muestral observado y el parámetro hipotético— no tiene que ser tan grande para que el estadístico de la prueba $t_{\bar{X}}$ caiga más allá de la puntuación crítica del estadístico de la prueba, t_α. Cuando $t_{\bar{X}}$ cae más allá de t_α en la región crítica, el valor *p* caerá debajo de α, y la hipótesis nula se rechazará. Con una prueba de una cola, la región crítica está agrupada en un lado y t_α es menor (1.64 comparado con 1.96 para la prueba de dos colas), colocando la región crítica más cerca de la media hipotética. Por tanto, el efecto de prueba puede ser menor para alcanzar el área de la región crítica para una prueba de una cola.

Para ilustrar esto, supongamos que dos investigadores, Jerome y Charlotte, desconocidos entre sí, están probando la hipótesis que el PC medio de la State University es 2.6. Jerome no tiene un motivo para estipular una dirección en su hipótesis alternativa, y por tanto realiza una prueba de dos colas. Sin embargo, Charlotte está al tanto que la State University, en comparación con otras universidades en el área, tiene más becas para ofrecer a sus estudiantes y de esta manera atrae mejores estudiantes. Por tanto, su hipótesis alternativa estipula que el PC medio de la State University es mayor que 2.6, y utiliza una prueba de una cola positiva. En el paso 1 de la prueba estadística, ambos Jerome y Charlotte estipulan sus hipótesis nulas como sigue:

Paso 1. H_0: $\mu_{X(State\ University)} = 2.6$ puntos de PC

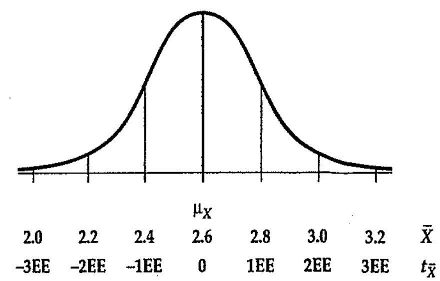
Sin embargo, sus hipótesis alternativas se estipularán de manera distinta:

Jerome: H_A: $\mu_{X(State\ University)} \neq 2.6$ puntos de PC. Dos colas.

Charlotte: H_A: $\mu_{X(State\ University)} < 2.6$ puntos de PC. Una cola.

Supongamos que los dos reúnen muestras de 122 y encuentran medias muestrales de 2.95 (es decir, $\bar{X} = 2.95$ puntos de PC). Sus desviaciones estándar muestrales (s_X) son ambas 2.2 puntos de PC y por tanto sus errores estándar calculados son .20 puntos de PC. Las dos distribuciones muestrales serían así:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{2.2}{\sqrt{122}} = .20 \text{ puntos PC}$$



En el paso 3, aunque los dos estipularon un nivel de significación (α) de .05, sus puntuaciones críticas (t_α) diferirán. Jerome determinará su t_α en 1.96 debajo de las columnas "prueba de dos colas o no direccional" en el lado izquierdo de la tabla de la distribución *t*. Charlotte determinará su t_α en 1.64 debajo de las columnas "prueba de una cola o direccional" en el lado derecho de la tabla de la distribución *t*. En el paso 4 de sus pruebas estadísticas, sus efectos de prueba y sus estadísticos de prueba calculados serán idénticos.

Paso 4. *Observación:*

$\bar{X} = 2.95$ puntos de PC $\quad s_X = 2.2$ puntos de PC $\quad n = 122$

Efecto de prueba $= \bar{X} - \mu_X = 2.95 - 2.6 = .35$ puntos de PC

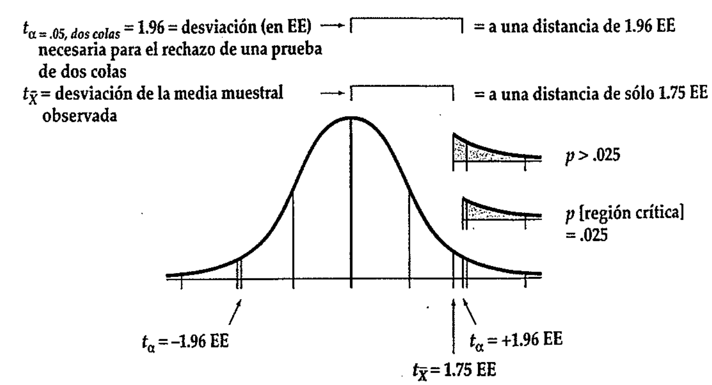
$$\text{Estadístico de prueba} = t_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}} = \frac{2.95 - 2.6}{.2} = \frac{.35}{2} = 1.75 \text{ EE}$$

Sin embargo, sus valores *p* (paso 4) diferirán debido a que se estimarán a partir de columnas diferentes de la tabla de la distribución *t*. Y en el paso 5 sus decisiones de rechazo serán distintas.

Valor *p* y la decisión de rechazo de Jerome:

p [sacar una muestra con una media ($\bar{X}$) tan inusual como o más inusual que 2.95 puntos de PC cuando la media de la población verdadera (μ_X) es 2.6 puntos de PC] > .05

La media muestral de Jerome debe estar a 1.96 EE del parámetro hipotético para rechazarla, pero el estadístico de prueba sólo es 1.75 EE y por tanto cae antes de la región crítica.

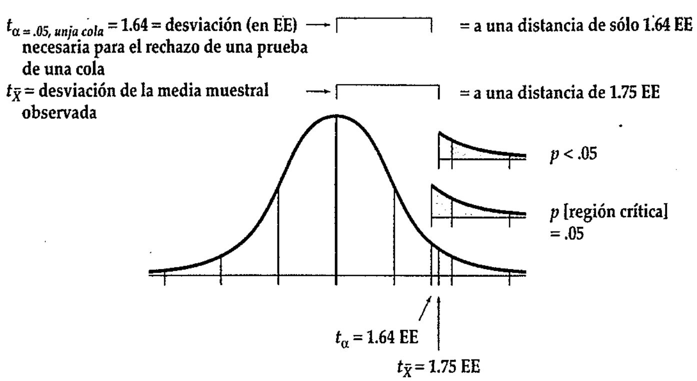


Paso 5. *Decisión de rechazo:* $|t_{\bar{X}}| < |t_\alpha|$ (es decir, 1.75 < 1.96); por tanto $p > \alpha$ (es decir, $p > .05$). Fracasa en rechazar H_0.

La media muestral de Charlotte sólo necesita estar a 1.64 EE del parámetro hipotético para rechazarlo. Como el estadístico de prueba es 1.75 EE, alcanza la región crítica para una prueba de una cola.

Valor p y decisión de rechazo de Charlotte:

p [sacar una muestra con una media ($\bar{X}$) tan inusual como o más inusual que 2.95 puntos de PC cuando la media de la población verdadera (μ_X) es 2.6 puntos de PC] < .05



Paso 5. *Decisión de rechazo:* $|t_{\bar{X}}| > |t_\alpha|$ (es decir, 1.75 > 1.64); por tanto $p < \alpha$ (es decir, $p < .05$). Rechaza H_0 y acepta H_A al nivel de confianza de 95%.

Tanto para Jerome como para Charlotte, el *efecto de la prueba* —la diferencia entre lo que se observa en las muestras y lo que se espera cuando la hipótesis nula es verdadera— es .35 puntos de PC. Sin embargo, en el paso 5 de los seis pasos de la prueba Jerome no rechazará la hipótesis nula en tanto que Charlotte sí. La región crítica de Charlotte es mayor con su 5% agrupado en un lado. Esto recorre su puntuación crítica (t_α = 1.64) hacia la media hipotética, reduciendo el tamaño del efecto de la prueba necesario para el rechazo. La región crítica de 5% de Jerome se divide a ambos lados para su prueba de dos colas, dejando su puntuación crítica (t_α = ±1.96) muy afuera en la cola, y esto requiere un efecto de la prueba mayor para alcanzar la región crítica. Lo que esto implica es que dividiendo el área de 5% en dos, el valor p de Jerome en realidad debe ser menor que .025 para que rechace la hipótesis nula. El valor p de una prueba de dos colas se divide a ambos lados, al igual que el área de la región crítica. En otras palabras, el efecto de la prueba debe ser el doble de inusual que el de Charlotte antes que se pueda rechazar.

La selección de la dirección de la prueba conduce a Jerome y Charlotte a conclusiones distintas. Jerome, al no rechazar la hipótesis nula de que el PC medio es 2.6, concluirá que el PC medio de la State University podría ser ese valor y atribuirá el efecto de la prueba de .35 puntos de PC al error del muestreo. Sin embargo, Charlotte al rechazar la hipótesis nula, concluirá que el efecto de la prueba de .35 puntos de PC es muy inusual proviniendo de una

población con un PC medio de 2.6; por tanto, el PC medio de la State University no debe ser 2.6. Ella declarará que es significativamente mayor que 2.6 puntos de PC. Charlotte aumentó sus posibilidades de rechazar la hipótesis nula al seleccionar una prueba de una cola.

Al seleccionar una prueba de una cola cargamos los dados a nuestro favor de rechazar la hipótesis nula y aceptar la hipótesis alternativa. Si éste es el resultado que deseamos para probar una teoría o impresionar al jefe, debemos tener buenas razones para utilizar una prueba de una cola. Además, la elección de la dirección no se toma observando estadísticos de la muestra. La dirección de la prueba se determina por la *pregunta* de investigación, que es respecto a los parámetros de la población. La dirección de la prueba *no* se determina por las respuestas de investigación que se encuentran en los estadísticos de la muestra.

Regla 4: Entre menor sea el nivel de significación, más difícil será rechazar la hipótesis nula Cuando el nivel de significación, α, es pequeño, la región crítica de la prueba será menor y su límite en la curva de probabilidad se encontrará más alejado del parámetro hipotético. Esto significa que un efecto de prueba debe ser muy grande para que el valor absoluto del estadístico de prueba alcance el valor de la puntuación t crítica (t_α). Por ejemplo, supongamos que un tercer investigador, Roger, prueba la hipótesis que el PC medio de la State University es 2.6. Sus datos son los mismos para Charlotte y Jerome —una media muestral de 2.95 puntos de PC y así sucesivamente— y el estadístico de prueba $t_{\bar{X}}$ se calcula en 1.75 errores estándar. Además, al igual que Charlotte, Roger realiza una prueba de una cola, *pero estipula su nivel de significación como .01*. Su puntuación crítica de la tabla de la distribución t es muy grande: t_α = 2.33. Por tanto, para rechazar la hipótesis nula, su efecto de prueba debe ser lo suficientemente grande para que su estadístico de prueba sea igual al menos a 2.33 errores estándar (EE). Veamos el valor p y la decisión de rechazo de Roger y comparémoslas con las anteriores.

Valor p y decisión de rechazo de Roger:

Paso 4. p [sacar una muestra con una media ($\bar{X}$) tan inusual como o más inusual que 2.95 puntos de PC cuando la media poblacional (μ_X) es 2.6 puntos de PC] < .05 pero > .01.

Paso 5. *Decisión de rechazo:* $|t_{\bar{X}}| < |t_\alpha|$ (es decir, 1.75 < 2.33); por tanto, $p > \alpha$ (es decir, $p > .01$). Fracase en rechazar H_0.

Tanto Roger como Charlotte tienen el mismo *efecto de la prueba*, .35 puntos de PC, que está a una distancia de 1.75 errores estándar de la media hipotética de 2.6. Sin embargo, en el paso 5 de la prueba de hipótesis Roger *fracasará* en rechazar la hipótesis nula con α = .01, en tanto que Charlotte la *rechazará* con α = .05. Roger concluirá que el PC medio de la State University podría ser 2.6, pero Charlotte concluirá que el PC medio de la State University es significativamente mayor que 2.6 puntos de PC. Aunque el valor del estadístico de prueba de 1.75 EE alcanzó la región crítica de Charlotte de .05, no llegó a la región crítica de Roger de .01. Es más difícil rechazar cuando el nivel de significación es bajo (digamos, α = .01 o .001). Y al contrario, es más fácil rechazar cuando es de moderado a alto (digamos, α = .05 o .10).

Regla 5. Cuando el resultado muestral observado está en la dirección opuesta de la anticipada por la hipótesis alternativa, de inmediato fracasamos en rechazar la hipótesis nula Debemos ser cuidadosos al calcular el valor p cuando realicemos una prueba de una cola donde el resultado muestral observado cae en la dirección opuesta a la anticipada por la hipótesis alternativa (H_A). Por ejemplo, supongamos que Shelia, otra investigadora en la State University, examina la pregunta de investigación que el PC medio es *menor que* 2.6, una predicción en la dirección negativa. Con $n = 500$ su media muestral se calcula ser 3.0, que está en la dirección *positiva*. Algunos pasos seleccionados en la prueba son como sigue:

Paso 1. H_0: $\mu_X = 2.6$ puntos de PC

H_A: $\mu_X < 2.6$ puntos de PC. Una cola

Paso 3. *Nivel de significación:* $\alpha = .05$. Una cola

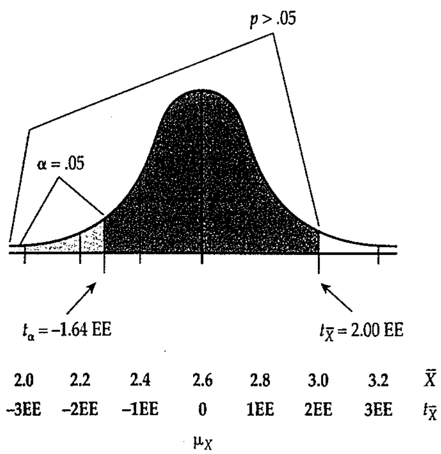
Puntuación crítica de prueba $t_\alpha = -1.64$ EE

Paso 4. *Media muestral observada:* $\bar{X} = 3.0$ puntos de PC

Efecto de la prueba $= \bar{X} - \mu_X = 3.0 - 2.6 = .4$ puntos de PC

$$\text{Estadístico de prueba: } t_{\bar{X}} = \frac{\bar{X} - \mu_X}{s_{\bar{X}}} = \frac{3.0 - 2.6}{.2} = \frac{.4}{.2} = 2.00 \text{ EE}$$

Si continuamos a ciegas con esta prueba observando los valores absolutos de la puntuación crítica de la prueba y del estadístico de la prueba, concluiríamos que la media muestral observada de 3.0 puntos de PC es significativamente diferente de 2.6 puntos de PC y rechazaríamos H_0. En efecto, el valor absoluto del estadístico de la prueba, $t_{\bar{X}}$, es mayor que el valor absoluto de la puntuación crítica t_α; es decir, $|2.00| > |-1.64|$. Sin embargo, el resultado no es en la dirección anticipada por H_A y no cae en la región crítica. ¿Cómo podemos justificar aceptar H_A que el PC medio es *menor que* 2.6 cuando la media muestral observada de 3.0 es *mayor que* 2.6? El cálculo apropiado del valor p es como sigue:



En la hipótesis H_A Shelia anticipó la dirección negativa. Por tanto, el área del valor p se calcula en esa dirección, a partir el valor del estadístico de prueba observado de 2.00 EE *a la izquierda* en la dirección negativa. Cuando el resultado está en la dirección equivocada, es obvio que el valor p debe ser mayor que .50, y por tanto, mayor que .05, debido a que al menos la mitad de la curva está comprendida por el área del valor p. Esto ilustra la importancia de trazar la curva de la distribución del muestreo.

Esta circunstancia también revela la importancia de justificar claramente la elección de la dirección de la prueba con una razón práctica o teórica. Si tú anticipas una dirección, te estás dando a ti mismo la ventaja de facilitar el rechazo de la H_0. Si el resultado muestral cae en la dirección opuesta, no puedes abandonar la predicción original. En lugar de eso, debes dar una explicación rápida. Por cierto, si Shelia no hubiera anticipado una dirección y hubiera utilizado una prueba de dos colas, podría haber rechazado de manera legítima H_0. Sin embargo, no puede correr una prueba de doble sentido, de una cola utilizando valores críticos de una cola sino esperar a ver la dirección en que ocurre el resultado.

Cuando utilices una computadora para probar hipótesis, pon mucha atención a si el resultado muestral de su prueba cae en la dirección anticipada. La computadora se programa para realizar una prueba de una cola o bien de dos colas, pero supone que tú anticipaste la dirección correcta. Si Shelia hubiera resuelto este problema en la computadora y simplemente hubiera observado el valor p proporcionado en la salida, ella hubiera visto que $p < .05$ en la salida y sacado la conclusión errónea. Este tipo de situaciones engañosas revelan el valor de aprender estadística resolviendo los ejercicios con lápiz y papel y no con base solamente en los caprichos de la programación de las computadoras.

Aunque una manera fácil para tomar la decisión de rechazo es observar el valor del estadístico de prueba, $t_{\bar{X}}$, y compararlo con la puntuación crítica, t_α, comprender el cálculo y el significado del valor p es importante. Éste es el caso especialmente si utiliza computadoras. La salida de la computadora reporta el valor del estadístico de prueba y el valor p, pero no la puntuación crítica. La comprensión deficiente de cualquiera de estos elementos de una prueba de hipótesis puede conducir a conclusiones erróneas.

Relaciones entre parámetros hipotéticos, estadísticos muestrales observados, estadísticos de prueba calculados, valores p y niveles alfa

Regla 1: La hipótesis estadística (H_0) se rechaza cuando el efecto de la prueba es lo suficientemente grande que el valor del estadístico de la prueba es mayor que el valor de la puntuación crítica de la prueba, por ejemplo, con una prueba de medias de una muestra única cuando

$$|t_{\bar{X}}| > |t_\alpha|$$

Regla 2: Entre mayores sean el efecto de la prueba y el estadístico de la prueba, menor será el valor p.

Regla 3: Es más fácil rechazar H_0 con una prueba de una cola que con una de dos colas.

Regla 4:	Entre menor sea el nivel de significación, más difícil será rechazar H_0.
Regla 5:	Cuando el resultado muestral observado es en la dirección opuesta a la anticipada, de inmediato fracase en rechazar H_0.

Uso de pruebas de hipótesis de una muestra única para establecer la representatividad de la muestra

Las pruebas de hipótesis de una muestra única son especialmente útiles al determinar si una muestra es representativa de la población de donde proviene. En el capítulo 2 analizamos la importancia de una muestra representativa, una en la que todos los segmentos de la población (como hombres, mujeres, blancos, afroamericanos, los jóvenes, los adultos mayores, los ricos y los pobres) se incluyan en proporción correcta respecto a su representación en la población. Por ejemplo, supongamos que un investigador en el ficticio condado Delaney realiza una encuesta telefónica para ver si los ciudadanos apoyan un aumento al impuesto predial. Su población de interés es todos los jefes de familia en el condado. Para obtener una muestra de 387 jefes de familia, el investigador utiliza un sistema de marcación aleatorio que asegura la inclusión de números telefónicos no listados. Sin embargo, es obvio que su encuesta excluye los hogares sin teléfonos. Como la mayoría de los hogares sin teléfono están habitados por gente pobre, debe determinar si al emplear una encuesta telefónica excluye injustamente a los pobres. Por tanto, ella quiere determinar si su muestra es representativa de los hogares del condado con respecto al índice de pobreza. El índice de pobreza de un condado es el porcentaje o proporción de hogares que están debajo del ingreso mínimo definido por el gobierno para sobrevivir, un punto denominado *línea de pobreza*.

¿Cómo una muestra no representativa puede conducir a una conclusión incorrecta acerca del apoyo para el aumento al impuesto predial? Supongamos que los adultos de hogares ricos son más probables que posean sus casas. Los propietarios de casas ven de manera directa las cantidades cargadas en sus recibos de impuestos y están menos inclinados a apoyar un aumento. Si los propietarios de casas están sobrerrepresentados en la muestra, sus respuestas contarán más que las de los miembros de hogares más pobres. Los resultados pueden indicar que la mayoría de los residentes del condado se oponen a un aumento de impuestos cuando de hecho los residentes más pobres que están más a favor no se les da una oportunidad justa de expresar sus opiniones.

Examinemos una población pequeña para ilustrar las consecuencias de una subrepresentación y una sobrerrepresentación. Supongamos que un condado tiene 10 familias: 7 con teléfono y 3 sin teléfono. De las 7 con teléfono, 3 apoyan el aumento de impuestos y 4 se oponen. Las 3 familias sin teléfono lo apoyan. Por tanto, el apoyo real de todo el condado es 6 *a favor* y 4 *en contra*. Entonces, una encuesta hecha de manera correcta deberá mostrar apoyo para el aumento de impuestos. Pero, ¿qué sucede si no se encuestan las familias sin teléfono? Los resultados de la encuesta harían aparecer de manera incorrecta que los residentes del condado estaban en contra del aumento 4 a 3. La representatividad de la muestra es un requisito esencial para hacer generalizaciones estadísticas: enunciados acerca de toda una población hechas con base en una muestra.

Valores objetivo para pruebas de hipótesis de la representatividad de la muestra

Para establecer la representatividad de una muestra, se deben reunir datos sobre algunos *parámetros conocidos* de la población. Si tenemos algunos parámetros conocidos, los podemos utilizar como valores objetivo hipotéticos en una serie de pruebas de hipótesis de una muestra única. Las variables demográficas como edad, género, estado civil, ingreso, índice de pobreza y raza son de uso común como parámetros conocidos para evaluar la representatividad de una muestra. Éste es el caso ya que el U.S. Bureau of the Census proporciona estos parámetros por código postal, número de distrito del censo, vecindario, área metropolitana, condado o estado. Para grupos y organizaciones como compañías, escuelas, clubes, y grupos voluntarios, los registros de organización son una buena fuente de esos datos de parámetros.

Gran parte de los datos de una oficina de censo del condado es de nivel de medición nominal/ordinal. La confiabilidad de los datos para variables nominales como género, raza e índice de pobreza es comúnmente mejor para la de las variables de intervalo/razón, como el ingreso en el hogar, que está muy sesgado. Una prueba de medias de una muestra única grande se puede utilizar para establecer la representatividad de la muestra con respecto a una variable de intervalo/razón, pero estas pruebas se deben realizar con mucho cuidado. Por tanto, las variables nominales comunes por lo general se eligen para establecer la representatividad de la muestra. Los parámetros conocidos para estas variables nominales/ordinales suministran las proporciones de la población ajustándose a una categoría, como la proporción de mujeres y hombres no caucásicos. Estas proporciones suministran valores objetivo para lo que se denomina prueba de proporciones de una muestra única grande.

Regresando al investigador en el condado de Delaney, para obtener un parámetro conocido sobre el índice de pobreza, investiga las cifras de la población del condado de la oficina del censo de Estados Unidos y determina un índice de pobreza de 22%, que es una proporción de .22. Si su muestra es representativa de los hogares del condado de Delaney, el índice de pobreza de los hogares muestreados en la encuesta también debe ser .22, más o menos un error de muestreo pequeño. Después de terminar todas las llamadas de la encuesta telefónica, determina que 66 de las 387 familias están debajo de la línea de pobreza para una proporción muestral de .17:

$$P_s = \frac{66}{387} = .17$$

El efecto de la prueba del procedimiento del muestreo para el índice de pobreza es la diferencia entre la proporción muestral observada, P_s, y la proporción real del condado, P_u, que es .22. Por tanto:

Efecto del procedimiento del muestreo al calibrar el índice de pobreza $= P_s - P_u = .17 - .22 = -.05$

Una prueba de proporciones de una muestra única grande se utiliza para determinar si este efecto de prueba se debe *a*) al error aleatorio del muestreo o *b*) a una falla en el muestreo y a representar de manera adecuada los hogares más pobres. A partir de su experiencia con el muestreo repetido, el investigador sabe que los estadísticos muestrales varían ligeramente de los parámetros poblacionales conocidos. ¿Es el efecto de $-.05$ tan pequeño como para ser un error del muestreo? La prueba de proporciones de una muestra única grande determina esto. Si $-.05$ no es estadísticamente significativo sino que simplemente se debe al error del muestreo, puede indicar que la muestra es representativa. Esto significa que el grupo poblacional que muestreó es el que quería muestrear, la población objetivo de adultos de todos los hogares del condado de Delaney incluyendo a los pobres. Significa que ella muestreó todos los segmentos de la población con sus proporciones correctas.

¿Cuál es la importancia de encontrar una representatividad con respecto a las variables demográficas? Si la muestra es representativa de la composición demográfica del condado, ella puede *suponer* con seguridad que es representativa de otras variables, como apoyo para el aumento de impuestos. Escrito de manera simple, si su procedimiento muestral muestrea de manera correcta el índice de pobreza, las posibilidades son que muestreó opiniones positivas, negativas y neutras de manera correcta sobre un aumento de impuestos. Un equilibrio correcto de la demografía sugiere que todo el procedimiento de muestreo está correctamente equilibrado. En contraste, si el perfil demográfico de la muestra no se ajusta al de la población, la muestra estará "sesgada" hacia un segmento de la población u otro. El sesgo en la respuesta introduce errores en los cálculos y conduce a conclusiones incorrectas.

Un reto al probar la representatividad de la muestra es concebir apropiadamente la población. Por supuesto, para probar una hipótesis debemos enunciarla de manera estadística, de tal manera que sabremos la forma de la distribución muestral *suponiendo que la hipótesis acerca de la población sea verdadera*. En el paso 1 de los seis pasos de la prueba de hipótesis es como sigue:

H_0: $P_{u(\text{población muestreada})} = .22$ (el parámetro conocido, P_u, de la población objetivo del condado de Delaney)

Es decir, la muestra *es* representativa de los hogares del condado de Delaney.

$$H_A: P_{u(\text{población muestreada})} \neq .22$$

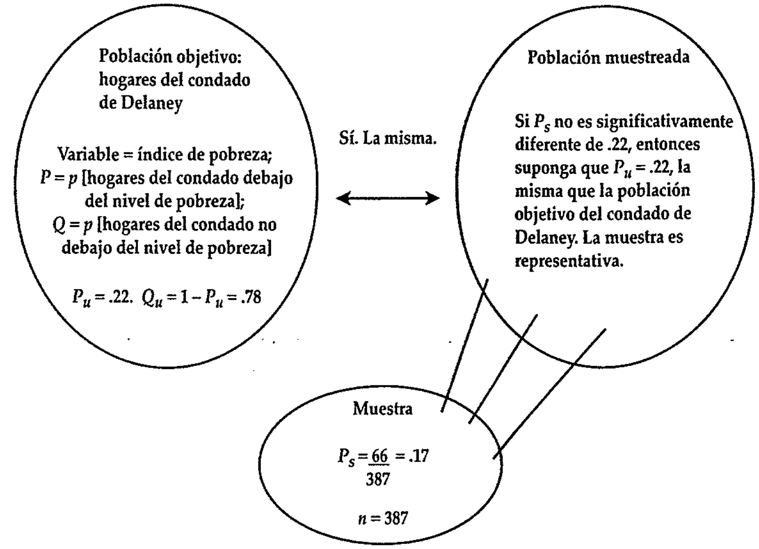
Es decir, la muestra *no es* representativa de los hogares del condado Delaney. Dos colas.

La representación de la noción de representatividad es un tanto difícil. Observa que *no* enunciamos la hipótesis con referencia al valor de la muestra (un artilugio utilizado con frecuencia en algunos libros de texto pero muy engañoso). Es decir, *no* hipotetizamos que la proporción de la muestra es igual a la proporción de la población (es decir, que $P_s = P_u$). Los *enunciados de hipótesis siempre se refieren a una población*. Los estadísticos muestrales observados nunca deben aparecer en el paso 1 de una prueba de hipótesis. Las muestras y sus estadísticos sólo son herramientas para abordar preguntas acerca de una población.

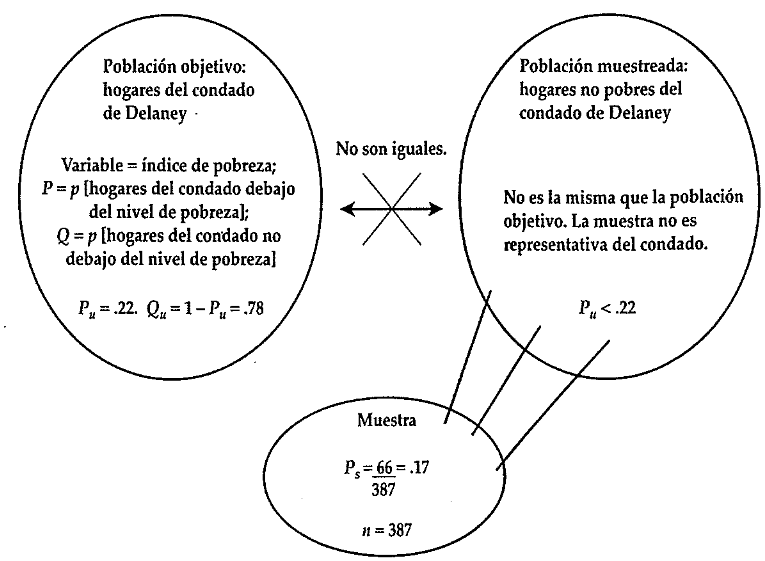
Para el ejemplo del condado de Delaney, es absurdo por al menos dos razones, enunciar la hipótesis como "el índice de pobreza de la muestra es igual al índice de pobreza de la población". Primero, podemos ver que no lo es; es obvio que, $.17 \neq .22$. Segundo, nuestra experiencia con distribuciones muestrales no indica que si tomamos otra muestra, su índice de pobreza probablemente diferirá del .17 de la muestra actual.

La pregunta real de la representatividad de la muestra es: ¿en realidad muestreamos la población objetivo o sin advertirlo obtuvimos demasiados (sobrerrepresentados) o muy pocos (subrepresentados) sujetos de muestra de algún segmento de esta población? Por ejemplo, en una encuesta telefónica en realidad podríamos muestrear la población de "hogares principalmente no pobres en el condado Delaney".

Observa que para esta prueba de hipótesis, queremos *fracasar en rechazar* la hipótesis nula (H_0). Cuando esto ocurre, afirmamos que la población objetivo y la población muestreada son iguales, hogares del condado de Delaney. Afirmamos también que nuestros resultados del estudio sobre las opiniones respecto a aumentos de impuestos se pueden aplicar, o *generalizar*, a todos los residentes del condado. Esta conclusión se concibe de manera gráfica como sigue:



Por otro lado, si rechazamos H_0 y aceptamos H_A, estamos afirmando que la muestra no es representativa y que un sesgo en nuestro procedimiento de muestreo nos ocasionó errar la población objetivo. Cuando esto ocurre, los resultados de nuestro estudio respecto a opiniones hacia el aumento del impuesto no se puede generalizar a todos los residentes del condado. Rechazar H_0 y aceptar H_A que la muestra no es representativa se puede concebir así:



En resumen, la representatividad muestral en realidad no es acerca de la muestra. Es una cuestión de a qué población representa la muestra y, por tanto, a qué población se pueden generalizar los resultados del estudio. Utilicemos este ejemplo para ilustrar los seis pasos de la inferencia estadística para una prueba de proporciones de una muestra única grande.

Prueba de proporciones de una muestra única grande

Una "prueba de proporciones de una muestra única grande", como la palabra *proporción* implica, se utiliza con una variable de nivel nominal/ordinal, donde $P = p$ [de la categoría de éxito] de la variable y $Q = p$ [de la o las categorías de fracaso]. Esta prueba se utiliza cuando la menor de P_u y Q_u por n es mayor que o igual a 5 (es decir, $[(p_{menor})\,(n)] \geq 5$)]. Si $[(p_{menor})\,(n)] < 5$, la prueba estadística apropiada se denomina distribución binomial, que se estudia en el capítulo 13.

Cuándo utilizar una prueba de proporciones de una muestra única grande (distribución t, $gl = \infty$)

En general: con una muestra lo suficientemente grande para que sea útil para probar una hipótesis que la proporción de una categoría de una variable nominal/ordinal en una población es igual a un valor objetivo. Especialmente útil para probar la representatividad de una muestra.

1. Sólo hay una variable.
2. La variable es de nivel de medición nominal/ordinal con $P = p$ [de la categoría de éxito].
3. Hay una muestra y una proporción.
4. El tamaño de la muestra es lo suficientemente grande para que $[(p_{menor})(n) \geq 5$, donde p_{menor} = la menor de P_u y Q_u.
5. Hay un valor objetivo de la variable con el cual podemos comparar la proporción de la muestra.

La distribución muestral de proporciones cuando $[(p_{menor})(n)] \geq 5$ es la distribución t aproximadamente normal. Por tanto, el estadístico de prueba es una puntuación t, y el valor p se estima con referencia a la tabla de la distribución t. Sin embargo, observa que si el tamaño de la muestra pasa como suficientemente grande, los grados de libertad son iguales al infinito (∞) y, por tanto, los valores críticos son iguales a los de la curva normal. De hecho, podríamos utilizar la tabla de la curva normal para esta prueba y llamarla prueba de la distribución Z. Sin embargo, seguiremos las convenciones del software de cómputo y la llamaremos prueba t.

Como destacamos en el capítulo 7, el error estándar de proporciones es como sigue:

Cálculo del error estándar de una distribución muestral de proporciones cuando se conocen P_u y Q_u (para una variable nominal)

$$\sigma_{P_s} = \sqrt{\frac{P_u Q_u}{n}}$$

donde

σ_{P_s} = error estándar de proporciones para una variable nominal/ordinal con $P = p$ [de la categoría de éxito], $Q = p$ [de la categoría de fracaso]

P_u = p hipotética [de la categoría de éxito en la población]

Q_u = p hipotética [de la categoría de fracaso en la población]

n = tamaño de la muestra

Observa que los símbolos en el numerador de esta ecuación indican parámetros de la población (P_u y Q_u). Esto difiere del cálculo del error estándar para intervalos de confianza, donde los estadísticos muestrales (P_s y Q_s) se emplean como estimaciones. Aunque P_u y Q_u con frecuencia no se conocen verdaderamente, en esta prueba de hipótesis el error estándar siempre se calcula de esta manera. Esto se hace debido a que la distribución muestral describe resultados muestrales suponiendo que los parámetros de la población sean iguales a los valores hipotéticos. P_u y Q_u se conocen en el sentido que sabemos qué esperar en el muestreo repetido cuando sus valores hipotéticos sean ciertos.

El estadístico de prueba se calcula como sigue:

Cálculo del estadístico de prueba para una prueba de proporciones de una muestra única grande

$$t_{P_s} = \frac{P_s - P_u}{\sigma_{P_s}}$$

$$= \frac{\text{el efecto}}{\text{el error estándar}} = \text{número de errores estándar (EE), diferencia entre lo que se observó y lo que se hipotetizó}$$

donde

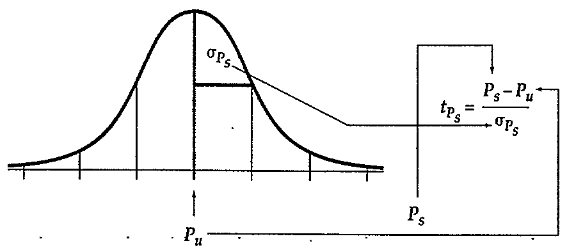
t_{P_s} = número de errores estándar que una proporción muestral (P_s) se desvía de la proporción hipotética de la población (P_u)

P_s = p [de la categoría de éxito en la muestra]

P_u = p [de la categoría de éxito hipotética para la población]

σ_{P_s} = error estándar de la distribución muestral de proporciones

La fórmula para t_{P_s} se deriva de las dimensiones de la curva de la distribución t aproximadamente normal como sigue. Observa la similitud con otros cálculos de puntuaciones t y Z.



Este estadístico de prueba, como el que calculamos para la prueba de medias, nos indica a cuantos errores estándar (EE) una proporción muestral observada cae de la proporción hipotética de la población. Ahora apliquemos los seis pasos de la inferencia estadística y analicemos los detalles de esta prueba de hipótesis para nuestro ejemplo del condado de Delaney.

Los seis pasos de la inferencia estadística para una prueba de proporciones de una muestra única grande

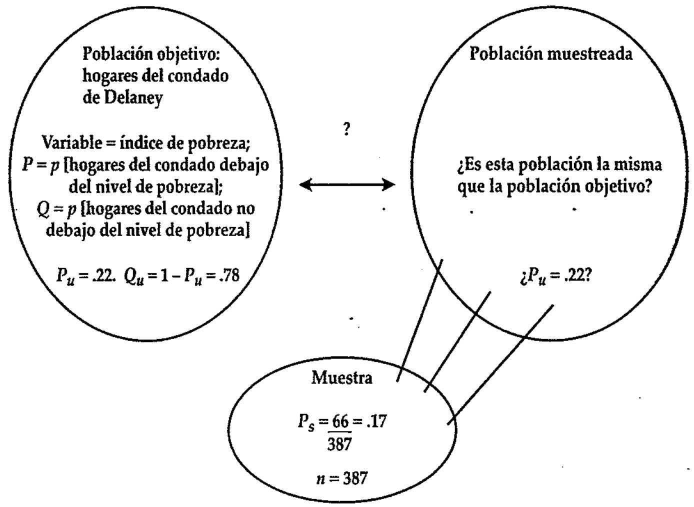
Problema: una investigadora en el ficticio condado de Delaney realiza una encuesta telefónica sobre el apoyo de los ciudadanos a un aumento del impuesto predial. Consciente del potencial para no subrepresentar a los hogares pobres, ella quiere determinar si su muestra es representativa de los hogares del condado con respecto al nivel de pobreza, el porcentaje de hogares con ingresos debajo del umbral de pobreza determinado por el gobierno de Estados Unidos. Los datos del censo de Estados Unidos revelan que 22% (.22) de los hogares del condado de Delaney están debajo del umbral de pobreza. Para la muestra de 387 hogares, 66 están debajo del umbral de pobreza.

Este problema cumple los criterios de una prueba de proporciones de una muestra única grande. Primero, tenemos una variable. Segundo, la variable es de nivel de medición nominal. Tercero, hay una muestra única de una población. Cuarto, el tamaño de la muestra, n, es lo suficientemente grande tal que $[(p_{menor})\,(n)] \geq 5$. Quinto, hay un parámetro conocido que proporciona un valor objetivo; es decir, los datos del censo revelan que el índice de pobreza del condado es .22.

Solución para una prueba de proporciones de una muestra única grande (distribución t, cuando $[(p_{menor})\,(n)] \geq 5$)

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿es representativa la muestra de 387 hogares encuestados del condado de Delaney de los hogares del condado con respecto al índice de pobreza? Procedimiento estadístico: prueba de proporciones de una muestra única grande. $[(p_{menor})\,(n)] \geq 5$; $(0.22)\,(387) = 85.12$; $85.12 > 5$.



1. H_0: $P_{u\,(población\ muestreada)} = .22$ (el índice de pobreza conocido, P_u, en el condado de Delaney objetivo)

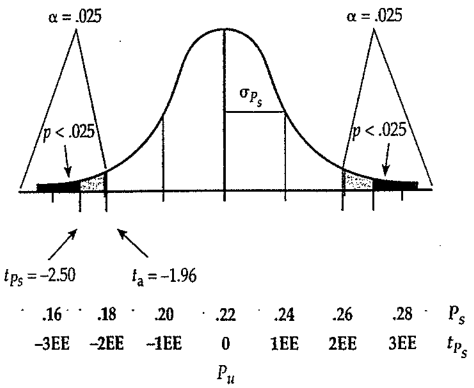
   Es decir, la muestra *es* representativa de los hogares del condado de Delaney con respecto al índice de pobreza.

   H_A: $P_{u\,(población\ muestreada)} \neq .22$. Dos colas.

   Es decir, hay un sesgo en nuestro procedimiento de muestreo resultando en que la muestra *no* es representativa con respecto al índice de pobreza.

2. Distribución muestral: si H_0 es verdadera y se toman muestras de manera repetitiva de tamaño 387 de los hogares del condado de Delaney, las proporciones muestrales (P_s) estarán centradas alrededor de .22 como una distribución t aproximadamente normal y $gl = \infty$, con un error estándar:

$$\sigma_{P_s} = \sqrt{\frac{P_u Q_u}{n}} = \sqrt{\frac{(.22)\,(.78)}{387}} = .02$$



(El sombreado de la curva se hace en los pasos 3 y 4.)

3. *Nivel de significación:* $\alpha = .05$. Dos colas. Puntuación crítica de prueba $t_\alpha = \pm 1.96$ EE.

   (Sombrea y marca la región crítica en la curva en el paso 2.)

4. *Observación:*

   $$\text{Efecto de la prueba: } P_s - P_u = .17 - .22 = -.05$$

   $$\text{Estadístico de la prueba: } t_{Ps} = \frac{P_sP_u}{\sigma_{Ps}} = \frac{.17 - .22}{.02} = -2.50 \text{ EE}$$

   Valor p: p [observar una proporción muestral, P_s, tan inusual como o más inusual que .17 cuando la proporción de la población verdadera, P_u, es .22] $< .05$.

   (Sombrea y marca el área del valor p en la curva en el paso 2.)

5. *Decisión de rechazo:* $|t_{P_s}| > |t_\alpha|$ (es decir, $2.50 > 1.96$); por tanto, $p < \alpha$ (es decir, $p < .05$). Rechaza H_0 y acepta H_A al nivel de confianza de 95%.

6. *Interpretación:* parece haber un sesgo en nuestro procedimiento de muestreo resultando en que la muestra no es representativa de los hogares del condado de Delaney con respecto al índice de pobreza. *Mejor estimación:* aunque 22% de los hogares del condado de Delaney están debajo del umbral de pobreza, sólo 17% de nuestra muestra lo está. Los hogares pobres están subrepresentados y los hogares no pobres están sobrerrepresentados. Nuestra población muestreada tiene demasiadas personas no pobres que respondieron.

Algunos puntos importantes a destacar acerca de esta prueba de hipótesis.

- En el paso 1, estipulamos la hipótesis alternativa como una prueba de dos colas. No sería apropiado examinar el estadístico de prueba, P_s, para determinar la dirección de la prue-

ba. En general, las pruebas de la representatividad de la muestra se hacen como pruebas de dos colas.

- En el paso 2, al igual que con cualquier prueba de hipótesis, ésta depende de describir la distribución muestral, la cual nos indica qué sucede con el muestreo repetido. Con este ejemplo, estamos haciendo una predicción para cualquier variable nominal/ordinal que tiene una proporción poblacional (es decir, un parámetro) de .22 cuando se toman muestras de manera repetitiva de tamaño 387: casi 68% de las veces la proporción muestral, P_s, se calculará entre .20 y .24, y casi 95% de las veces entre .18 y .26. Aunque la variable en sí es nominal y, por tanto, calculamos proporciones, las proporciones muestrales calculadas, P_s, constituyen una puntuación de nivel de razón y estas puntuaciones se trazan en el eje horizontal. La curva de la distribución tiene una proporción *media*, que en este caso es .22. En otras palabras, si tomaras, digamos, 10 000 muestras, sumaras las proporciones muestrales, P_s, y dividieras entre 10 000, el resultado sería .22.
- En el paso 4, la diferencia entre la proporción muestral observada y la proporción hipotética de los hogares pobres (es decir, $P_s - P_u = .05$ o 5%) es "el efecto" de la prueba. En esencia, el objetivo de la prueba de hipótesis es determinar si este efecto de la prueba se debe al error del muestreo aleatorio (como se afirma por la hipótesis nula) o debido a un sesgo en nuestro procedimiento de muestreo (como se afirma por la hipótesis alternativa).
- En el paso 4, determinamos que la proporción muestral observada ($P_s = .17$, tiene una baja probabilidad de ocurrencia en el muestreo repetido. Esto nos condujo a concluir en el paso 5 que la P_s observada de .17 no se debía al error del muestreo aleatorio normal, sino que se debía al hecho que la población de donde proviene nuestra muestra no era única y la misma que la población de hogares en el condado de Delaney. En otras palabras, la muestra no es representativa de la población.
- En el paso 4, al calcular el valor p, estamos estipulando "qué tan inusual" es el resultado muestral observado "si la hipótesis nula es verdadera". Por tanto, si la proporción de la población, P_u, es en efecto .22, y no tenemos una razón de antemano para predecir una dirección, entonces sería tan inusual sacar una proporción muestral 2.50 EE arriba de la proporción media de .22 como una proporción muestral 2.50 EE debajo de ella. Ésta es la razón porque identificamos áreas en *ambas colas* de la curva del paso 2 para una prueba de dos colas.
- En el paso 6, nos enfocamos en la hipótesis alternativa e ignoramos la hipótesis nula debido a que se rechazó en el paso 5. Con una audiencia pública, el concepto de representatividad quizá no tenga eco. Por tanto, para darle sabor a la respuesta, proporcionamos una mejor estimación —algunos números concretos— con la que se pueda relacionar una audiencia. Destacamos que los hogares pobres no están bien representados en 5% [es decir, $(.17 - .22)(100) = 5\%$].

¿Qué hacer si se determina que una muestra no es representativa?

Con nuestra muestra de 387 hogares del condado de Delaney, ahora hemos concluido que esta muestra no es representativa con respecto al índice de pobreza. Cuando ocurre esta situación, se deben solucionar tres puntos. Primero, ¿qué defecto en el diseño del muestreo condujo al muestreo inferior de los hogares pobres? El investigador sospecha que una encuesta telefónica excluyó muchos de los hogares pobres que no podían costear una línea telefónica.

Nuestra segunda pregunta es: ¿De qué maneras cambia nuestras conclusiones esta muestra sesgada? Es obvio que un sesgo de "clase media" —un sobremuestreo de ella— puede conducir a las conclusiones erróneas acerca de las opiniones relacionadas a puntos económicos como aumentos de impuestos. Por ejemplo, las personas en hogares más ricos en municipalidades suburbanas son más probables que posean casas, y el aumento del impuesto predial los afecta de manera directa. Ellos pueden apoyar un aumento al impuesto de la municipalidad que envíe ingresos a las escuelas locales a las que asisten sus hijos. Sin embargo, se pueden oponer a un aumento general en el país puesto que estos ingresos se dispersan ampliamente para incluir grandes partes del centro y más pobres del país. Como los hogares más ricos se sobremuestrearon, sus opiniones están sobrerrepresentadas. Esto podría conducir a la conclusión errónea que hay una oposición muy difundida al aumento del impuesto. Si la proporción correcta de hogares más pobres tuviera una oportunidad de responder a la encuesta sobre este punto, sus opiniones podrían inclinar la balanza y revelar que una mayoría de los residentes del país en realidad apoyan el aumento al impuesto. La estimación muestral obtenida puede ser menor que la proporción real de apoyo para el aumento al impuesto (es decir, el parámetro). Una muestra no representativa puede ser una herramienta peligrosa.

Nuestra tercera pregunta es: ¿qué ajustes son pertinentes cuando una muestra no es representativa? Se pueden hacer varias cosas para compensar el sesgo de la muestra. Primero, se pueden seleccionar sujetos adicionales de las categorías subrepresentadas. Por ejemplo, para evitar una representación menor de los hogares pobres sin teléfono, las encuestas telefónicas con frecuencia se complementan con entrevistas de puerta en puerta en los vecindarios pobres. De hecho, para apoyar este esfuerzo, la oficina del censo de Estados Unidos proporciona datos basados en los vecindarios sobre los porcentajes de hogares sin teléfono. Segundo, podemos proceder con el análisis de datos pero estipulando que la población representa de menos o de más a algunos grupos. Por ejemplo, en una encuesta telefónica simplemente anotaríamos que aquellos sin teléfono —los pobres— están subrepresentados en el estudio. Esto, por supuesto, abre la puerta al criticismo. Por último, una muestra no representativa puede ajustarse matemáticamente de manera artificial "ponderando" las categorías de la muestra para acercarlas a sus proporciones poblacionales correctas. La ponderación de la muestra es complicada y se debe hacer con mucho cuidado, y está más allá del alcance de este texto. Basta decir que cada esfuerzo debe hacerse para diseñar un procedimiento muestral que obtenga proporciones correctas de segmentos significativos de la población.

Presentación de datos de pruebas de hipótesis de una muestra única

Es en el área de la representatividad de la muestra donde las pruebas de hipótesis de una muestra única se utilizan con más frecuencia. La tabla 10-2 proviene de un artículo de una revista de investigación científica en donde el autor abordó el punto de la representatividad de la muestra. La muestra consistió de 206 médicos en el condado de Jefferson, Alabama, y el estudio se realizó en un centro académico de ciencias de la salud. La tabla evalúa si la muestra es representativa de los médicos del condado de Jefferson y de los médicos en las áreas metropolitanas de Estados Unidos. El porcentaje reportado de estas dos *poblaciones* son parámetros conocidos tomados de directorios médicos. Estos parámetros conocidos se utilizaron como valores objetivo en pruebas de proporciones de una muestra única. Las probabilidades reportadas bajo "p" son los valores p de cada prueba, donde "NS" indica que no hubo una diferencia significativa entre la población y las proporciones de la muestra al nivel de significación de .05.

TABLA 10-2 | Comparación de categorías de especialidad médica de la muestra con la población de médicos en el condado de Jefferson, Alabama, y con las áreas metropolitanas de Estados Unidos.

	Porcentaje de médicos en				
Categoría de especialidad	**Muestra (%)**	**Condado de Jefferson (%)**	***p***	**Áreas metropolitanas de Estados Unidos (%)**	***p***
Práctica general	7.77	7.95	NS*	10.81	NS
Especialidades médicas	31.07	32.83	NS	28.46	NS
Especialidades quirúrgicas	28.64	35.45	$p < .05$	24.35	NS
Otras especialidades	32.52	23.77	$p < .01$	36.38	NS

*NS = no significativo al nivel de significación de .05, prueba de dos colas.

Fuente: Clair y otros, 1993. Datos de la población de *Physicians Characteristics and Distribution in the U.S.*

La tabla muestra que las cuatro categorías de médicos muestreados fueron representativas de los médicos en áreas metropolitanas de Estados Unidos. Sin embargo, la muestra no fue representativa de los médicos en el condado de Jefferson. Las especialidades quirúrgicas estuvieron subrepresentadas y "otras especialidades" (como medicina nuclear) estuvieron sobrerrepresentadas. En un análisis adicional, se determinó que los médicos del centro académico de ciencias de la salud, muchos de los cuales estuvieron en la categoría "otras especialidades", fueron más probables de responder, tal vez debido a que se sintieron obligados a cooperar con sus colegas del campus que estuvieron realizando el estudio. Para abordar las consecuencias de este sesgo en la muestra, los autores indicaron que sus resultados eran más aplicables a ciudades con centros académicos de ciencias de la salud. En otras palabras, ellos reconocieron que la población a la que sus resultados eran generalizables no consistió de todos los médicos sino de los que reflejaron la composición de su muestra.

Un intervalo de confianza de la media de la población cuando n es pequeña

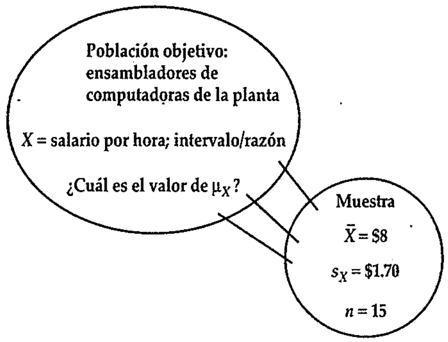
Recuerda que en el capítulo 8 presentamos el cálculo de un intervalo confianza de una media de la población para una muestra grande, donde $n > 121$. Al calcular el término del error, utilizamos puntuaciones críticas de la tabla de la curva normal, como $Z_\alpha = 1.96$ para un intervalo de confianza de 95%. Cuando $n \leq 121$, la puntuación crítica proviene de la tabla de la distribución t. El siguiente problema de ejemplo es el mismo ilustrado en el capítulo 8 para calcular el intervalo de confianza de una media de la población, excepto que el tamaño de la muestra es menor.

Solución para un intervalo de confianza de la media cuando $n \leq 121$ (utilizando puntuaciones t al calcular el término del error)

Problema: estamos realizando un estudio de la estructura salarial de una planta industrial con varios miles de ensambladores de computadoras. Necesitamos obtener una idea aproximada del salario medio por hora de esta población de ensambladores.

Seleccionamos al azar 15 expedientes del personal y registramos los salarios por hora. En esta muestra, encontramos una media de $8.00 y una desviación estándar de $1.70. Calcula el intervalo de confianza de 95% para el salario medio por hora de los ensambladores de la planta.

Paso 1. Pregunta de investigación: dentro de un rango especificado de cantidades en dólares, ¿cuál es el parámetro, μ_X, el salario medio por hora de la población de ensambladores de computadoras?



Paso 2. (error estándar, puntuación *t* crítica y término de error)

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{1.70}{\sqrt{15}} = \$.44$$

$gl = n - 1 = 15 - 1 = 14$. Para confianza de 95%, $t_\alpha = 2.145$

Término de error = $(t_\alpha)(s_{\bar{X}}) = (2.145)(\$.44) = \$.94$

Paso 3. (el LCI y LCS)

$$IC \text{ de } 95\% \text{ de } \mu_X = \bar{X} \pm (2.145)\,(s_{\bar{X}})$$
$$= \text{media muestral} \pm \text{término de error}$$
$$= \$8.00 \pm (2.145)\,(\$0.44) = \$8.00 \pm \$.94$$
$$LCI = \$8.00 - \$0.94 = \$7.06$$
$$LCS = \$8.00 + \$0.94 = \$8.94$$

Paso 4. (interpretación en lenguaje común)

"Estoy 95% seguro de que el salario medio por hora de los ensambladores de computadoras de la planta está entre $7.06 y $8.94".

Paso 5. (interpretación estadística ilustrando la noción de "confianza en el procedimiento")

"Si se realizaran los mismos procedimientos muestrales y estadísticos 100 veces, 95 veces el parámetro poblacional verdadero, μ_X, estará comprendido en los intervalos calculados y 5 veces no lo estará. Por tanto, tengo una confianza de 95% que este intervalo de confianza individual que calculé incluye al parámetro verdadero."

Comparemos los resultados de este cálculo de una muestra pequeña de un intervalo de confianza de una media poblacional con los resultados obtenidos en el capítulo 8.

Del capítulo 8 con $n > 121$	Problema anterior con $n \leq 121$
$\bar{X} = \$8.00$	$\bar{X} = \$8.00$
$s_X = \$1.70$	$s_X = \$1.70$
$n = 129$	$n = 15$
$s_{\bar{X}} = \$.15$	$s_{\bar{X}} = \$.44$
Puntuación crítica: $Z_\alpha = 1.96$	Puntuación crítica = $t_\alpha = 2.145$
Término del error = (1.96)($.15) = $.29	Término del error = (2.145)($.44) = $.94
Intervalo de confianza: $7.71 a $8.29	Intervalo de confianza: $7.06 a $8.94
Precisión (ancho) del intervalo de confianza = $8.29 – $7.71 = $.58	Precisión (ancho) del intervalo de confianza = $8.94 – $7.06 = $1.88

Consistente con la ley de los números grandes (capítulo 7), el error para la muestra menor es mucho mayor que el de la muestra mayor. El tamaño menor de la muestra influye en el cálculo del intervalo de confianza de dos maneras. Primero, la *n* menor resulta en un error estándar mayor de la media, $s_{\bar{X}}$. Segundo, la puntuación crítica para el término del error es mayor (es decir, 2.145 comparada con 1.96). Basta decir que una muestra mayor permite mayor precisión al estimar parámetros de la población.

Insensatez y falacias estadísticas: aspectos del tamaño de la muestra y representatividad de la muestra

El aspecto de la representatividad de la muestra se refiere a si todos los segmentos de una población están representados de manera equitativa en una muestra. Debido a que una prueba de la representatividad requiere un valor objetivo, debemos confiar en parámetros conocidos para formular la hipótesis nula. Se *supone* que si una muestra es representativa con respecto a parámetros conocidos, esa muestra es representativa con respecto a una variedad de opiniones sostenidas por miembros de una población. Por ejemplo, en nuestro ejemplo de los residentes del condado de Delaney supusimos que si el procedimiento de muestreo de la investigadora muestrea de manera correcta el índice de pobreza, es probable que represente correctamente los residentes con opiniones positivas, negativas y neutras sobre el aumento del impuesto. Esas suposiciones no siempre se cumplen. Por ejemplo, si una muestra es un tanto pequeña, quizá no tenga espacio para la variedad de opiniones que existen dentro de la población. Por ejemplo, si el investigador selecciona al azar sólo 10 hogares, aún si representan con precisión las proporciones de cada nivel de ingreso, con tan pocos casos hay una buena posibilidad que no se representen todas las opiniones. Suponga, por ejemplo, que el índice de pobreza de los hogares del condado de Delaney es 22% y que la muestra del investigador es de tamaño 10, con 8 hogares arriba del umbral de pobreza y 2 debajo de éste. Matemáticamente, esto constituye una muestra representativa. Pero ¿podrían sólo 2% de los hogares pobres representar las opiniones de toda la gente pobre? Establecer la representatividad de la muestra para una muestra pequeña es un esfuerzo endeble en el mejor de los casos.

En este punto viene al caso la analogía de probar un platillo muy condimentado, donde el platillo representa una población con una variedad de ingredientes y matices de sabores (u opiniones). Si probamos una muestra grande —un plato lleno de una olla bien mezclada— existe la posibilidad de que obtengamos una probada de cada ingrediente. Sin embargo, si utilizamos la medida de una cuarta parte de una cucharada —que carece de espacio para los frijoles y los trozos de carne— es probable que probemos sólo el caldo.

En una muestra pequeña quizá no haya espacio para la extensión de las opiniones. Imagina que tú has sido acusado de mala conducta y debes comparecer ante un panel de tres jueces. En cualquier día, incluso un panel seleccionado al azar puede componerse de dos o tres jueces muy estrictos o muy indulgentes. La disposición de su caso quizá dependa de la suerte de la selección del panel. Las muestras pequeñas son inherentemente propensas al error en términos de representatividad.

Otra consideración acerca del tamaño de la muestra y de la prueba de la representatividad de la muestra tiene que ver con el hecho que la representatividad de la muestra se establece *fracasando en rechazar* la hipótesis nula (H_0). Rechazar H_0 es más difícil cuando el tamaño de la muestra es pequeño, ya que las muestras pequeñas producen un rango amplio de error en la distribución muestral. Con una muestra muy pequeña el error de muestreo es tan grande que tomaría un efecto de prueba especialmente grande para rechazar H_0. Si fracasamos en rechazar una H_0 falsa y esto ocurriera debido a que simplemente utilizamos una muestra pequeña, hay mucho potencial para cometer un error tipo II. Recuerda que un error tipo II ocurre cuando una prueba estadística falla en rechazar una hipótesis nula falsa (H_0). En este caso, un error tipo II implicaría concluir que la muestra es representativa de la población cuando, de hecho, no lo es. Para evitar este error potencial, los tamaños de las muestras deben ser suficientemente grandes para reducir las posibilidades de cometer un error tipo II. Empleamos el término *poder estadístico* para referirnos a *una probabilidad del estadístico de prueba de no cometer un error tipo II para un nivel de significación dado.* Establecer el poder estadístico es una tarea un tanto complicada. Se aborda en el sitio en la red *The Statistical Imagination* en las extensiones del capítulo 10. Sin embargo, como regla general si tienes una muestra pequeña con un rango de error considerable, digamos, mayor que ±3.0, quizá quieras establecer como punto de corte un efecto de 3% para establecer la representatividad de la muestra. Por ejemplo, supongamos que el tamaño de nuestra muestra del condado de Delaney fue 50. Sustituyendo este tamaño de la muestra en nuestro cuadro de solución anterior resultaría en un error estándar de .06. Tomaría casi el doble de este efecto del tamaño, un efecto de .12, para rechazar H_0. Por tanto, podríamos tener una proporción muestral, P_s, tan baja como .11, que es un efecto de .11 (es decir, .11 − .22 = − .11) y fracasaríamos en rechazar H_0. Concluiríamos que nuestra muestra es representativa. Pero ¿sería razonable creer que podría tener la mitad de la proporción de hogares pobres en su muestra como hay en el país y considerar representativa a esta muestra? Es probable que la muestra pequeña con su rango de error grande haya conducido a un error de tipo II.

¿Qué se puede hacer acerca de las muestras pequeñas? Uno, simplemente elegir un tamaño del efecto, como .03 (3%), como punto de corte para establecer la representatividad. En otras palabras, si el efecto de la prueba, la diferencia entre la proporción muestral y el parámetro poblacional conocido es mayor que .03, asume que la muestra tiene un sesgo. Dos, considera otras formas para detectar sesgos. Por ejemplo, si se dispone de datos, compara características de personas respondientes y no respondientes con algunos parámetros conocidos. Tres, simplemente no pretendas generalizar los datos con una población conocida. Afirma que los datos son exploratorios y advierte a los lectores no poner demasiada fe en los resultados. Hay muchos puntos metodológicos relacionados a la integridad de las muestras. No es muy recomendable realizar pruebas estadísticas a ciegas siguiendo un libro de texto.

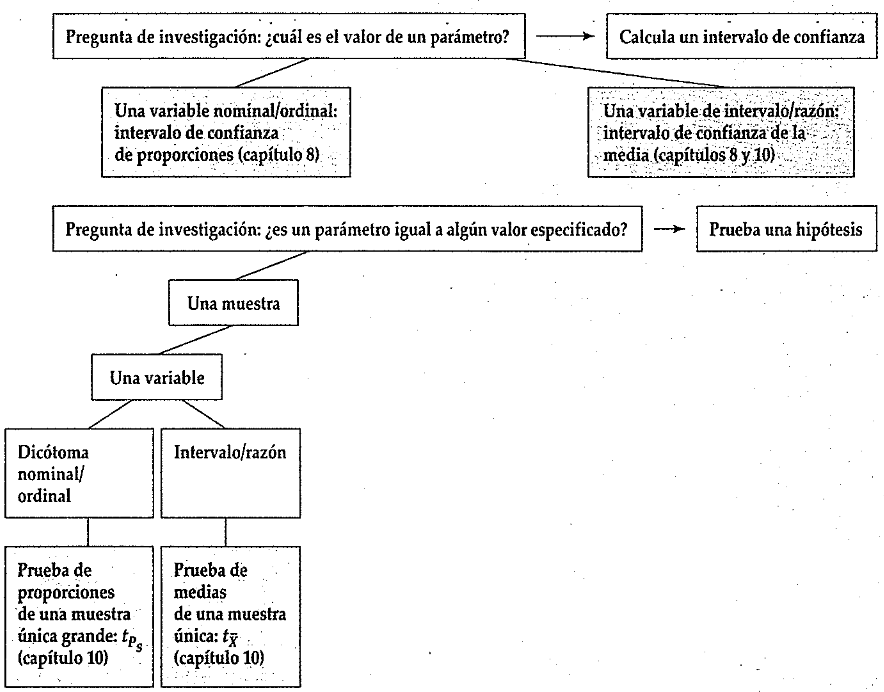
RESUMEN

1. Una prueba de hipótesis de una muestra única se utiliza para responder la pregunta: ¿para una población de interés es un parámetro igual a algún valor objetivo elegido?
2. Los valores objetivo pueden provenir de: *a*) un parámetro poblacional conocido de un grupo de comparación; *b*) parámetros conocidos de un periodo pasado; *c*) un ideal estadístico; *d*) comparar los estadísticos de la muestra de una población muestreada con parámetros poblacionales conocidos para determinar si la muestra es representativa de la población.
3. Una prueba de medias de una muestra única es útil para probar una hipótesis de que la media de X para una población es igual a un valor objetivo. La distribución muestral es la distribución t, que se puede emplear para pruebas de medias de todos los tamaños de muestras pero se debe utilizar cuando $n \leq 121$. Los programas de cómputo se refieren a todas las pruebas de medias de una muestra única, sin importar los tamaños de las muestras, como pruebas t.
4. La distribución t es una distribución aproximadamente normal. La tabla de la distribución t está organizada de manera diferente a la tabla de la curva normal y requiere el cálculo de los grados de libertad.
5. Los grados de libertad son una manera de ajustar las limitaciones en los cálculos estadísticos. Para pruebas de medias, los grados de libertad se basan en el tamaño de la muestra debido a que el cálculo de las medias muestrales con muestras pequeñas se puede distorsionar por puntuaciones extremas.
6. La comprensión de las relaciones entre parámetros hipotéticos, estadísticos muestrales observados, estadísticos de prueba calculados, valores p y niveles alfa mejora la competencia en la prueba de hipótesis. Regla 1: la hipótesis nula se rechaza cuando el efecto de la prueba es lo suficientemente grande que el valor del estadístico de prueba es mayor que la puntuación crítica de prueba. Regla 2: entre mayores sean el efecto de la prueba y el estadístico de la prueba, menor será el valor p. Regla 3: es más fácil rechazar la hipótesis nula con una prueba de una cola que con una de dos colas. Regla 4: entre menor sea el nivel de significación, más difícil será rechazar la hipótesis nula. Regla 5: cuando un resultado muestral observado es en la dirección opuesta a la anticipada por la hipótesis alternativa, de inmediato fracase en rechazar la hipótesis nula.
7. La prueba de proporciones de una muestra única grande es útil para probar una hipótesis que la proporción de una categoría de éxito de una variable nominal/ordinal en una población es igual a un valor objetivo. El tamaño de la muestra debe ser lo suficientemente grande que el menor de P_u y Q_u por n sea mayor que o igual a 5.
8. La prueba de proporciones de una muestra única grande es especialmente útil para probar la representatividad de una muestra. Las agencias gubernamentales, como la oficina del censo de Estados Unidos, tienen muchos parámetros conocidos para variables nominales/ordinales, como edad y género.
9. Para calcular el término del error de un intervalo de confianza cuando $n \leq 121$, se utilizan puntuaciones t en lugar de puntuaciones Z en el término del error. Si todo lo demás es igual, comparada con una muestra grande, una muestra pequeña resulta en un error estándar mayor, un término de error mayor y un intervalo de confianza menos preciso.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 10 del material del texto en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2 incluyen un análisis del poder estadístico y de la importancia de tener una muestra lo suficientemente grande para realizar una prueba de hipótesis de la representatividad de la muestra. Las secuencias de los comandos *SPSS* para procedimientos en este capítulo aparecen en el apéndice D de este texto.

PROCEDIMIENTOS ESTADÍSTICOS ANALIZADOS HASTA ESTE PUNTO



FÓRMULAS PARA EL CAPÍTULO 10

Prueba de medias de una muestra única (prueba t):

Datos: una variable de intervalo/razón X y una muestra y población únicas

Pregunta de investigación: ¿es μ_X (es decir, la media de X en la población) igual a un valor objetivo?

H_0: μ_X = un valor objetivo

Distribución muestral: distribución t con $gl = n - 1$ error estándar estimado utilizando la desviación estándar de la muestra.

Error estándar =

$$s_{\overline{X}} = \frac{s_X}{\sqrt{n}}$$

Efecto de la prueba = $\overline{X} - \mu_X$

Estadístico de prueba [para uso con la tabla de la distribución t aproximadamente normal (tabla estadística C del apéndice B)]:

$$t_{\overline{X}} = \frac{\overline{X} - \mu_X}{s_{\overline{X}}}$$

$$df = n - 1$$

Prueba de proporciones de una muestra única grande:

Datos: una variable nominal/ordinal con $P = p$ [de la categoría de éxito].

Se utiliza cuando: $[(p_{menor})\,(n) \geq 5$. (Si $[(P_{menor})\,(n)] < 5$, vea el capítulo 13.)

Pregunta de investigación: ¿es P_u (es decir, la p [de la categoría de éxito en la población]) igual a un valor objetivo?

H_0: P_u = un valor objetivo.

Distribución muestral: distribución t con $gl = \infty$.

Error estándar =

$$\sigma_{P_s} = \sqrt{\frac{P_u Q_u}{n}}$$

Efecto de la prueba = $P_s - P_u$

Estadístico de prueba [para su uso con la tabla de la distribución t aproximadamente normal (tabla estadística C del apéndice B)]:

$$t_{P_s} = \frac{P_s P_u}{\sigma_{P_s}}$$

PREGUNTAS PARA EL CAPÍTULO 10

1. ¿Cuál es el objetivo de una prueba de hipótesis de una muestra única? ¿Qué tipo de pregunta responde con respecto a una población?
2. Para pruebas de hipótesis de una muestra única, los parámetros objetivo o meta provienen de cuatro fuentes. Menciona cada fuente y proporciona una hipótesis de ejemplo para cada una.
3. Describe la situación (es decir, los criterios de selección) para utilizar una "prueba de medias de una muestra única pequeña".
4. Describe la situación (es decir, los criterios de selección) para utilizar una "prueba de proporciones de una muestra única grande".

5. Relaciona lo siguiente:
   a) σ_{P_s} _____ El error estándar estimado de las medias muestrales
   b) P_u _____ El error estándar de las proporciones muestrales
   c) t_{P_s} _____ La proporción hipotética de la población
   d) P_s _____ El estadístico de prueba (es decir, la distancia en número de errores estándar desde la proporción muestral observada hasta la proporción de la población hipotética)
6. Relaciona lo siguiente:
   a) $\bar{X}$ _____ El error estándar estimado de medias muestrales
   b) $t_{\bar{X}}$ _____ La media de la población hipotética
   c) $s_{\bar{X}}$ _____ La media muestral observada
   d) μ_X _____ El estadístico de prueba (es decir, la distancia en número de errores estándar desde la media muestral observada hasta la media de la población hipotética)
7. Los estadísticos observados y calculados en el paso 4 de los seis pasos de las pruebas de hipótesis nunca deben aparecer en los pasos 1 a 3. ¿Por qué?
8. ¿Cuál es la relación entre el efecto de una prueba de hipótesis y el estadístico de la prueba? En específico, ¿de qué manera mide el estadístico de prueba el efecto de la misma?
9. ¿Cuándo es más probable que se rechace la hipótesis nula, cuando el efecto de la prueba es grande o pequeño? Ilustra tu respuesta trazando una curva de la distribución *t* aproximadamente normal para una prueba de medias de una muestra única.
10. ¿Cuándo es más probable que se rechace la hipótesis nula, cuando el estadístico de la prueba es grande o pequeño? Ilustra tu respuesta trazando una curva de la distribución *t* aproximadamente normal para una prueba de medias de una muestra única.
11. ¿Cuál es la relación entre los tamaños de los valores calculados de los estadísticos de la prueba y sus valores *p*?
12. ¿Cuándo es más fácil rechazar la hipótesis nula, al probarla con una prueba de una cola o de dos colas? Ilustra tu respuesta utilizando una curva de la distribución *t* aproximadamente normal para una prueba de medias de una muestra única.
13. ¿Qué significa la palabra *crítica* en los términos de *región crítica* y *puntuación crítica de prueba*?
14. Proporciona un ejemplo de cómo una muestra no representativa puede conducir a conclusiones erróneas.
15. Para una prueba de hipótesis de la representatividad de la muestra, distingue la población muestreada de la población objetivo. ¿Son iguales o diferentes estas poblaciones cuando la muestra de hecho es representativa de la población objetivo?
16. En general, en una "prueba de medias de una muestra única", cuando la hipótesis nula es verdadera, ¿en qué valor de *X* se centrarán las puntuaciones (es decir, las medias muestrales) en la distribución muestral?
17. Para una prueba de medias de una muestra única, cuando $t_{\bar{X}} > t_\alpha$, ¿el valor *p* es mayor o menor que α?

18. Para una prueba de proporciones de una muestra única grande, cuando $t_{P_s} < t_\alpha$, ¿el valor *p* es mayor que o menor que α? ¿Rechazaremos la hipótesis nula o fracasaremos en rechazarla?
19. Al igual que una curva normal, una distribución *t* es simétrica y su media, mediana y moda son iguales. Sin embargo, ¿por qué decimos que una distribución *t* sólo es *aproximadamente* normal?
20. ¿Qué característica subyacente de la media ocasiona una pérdida de grados de libertad cuando se utiliza la media en una prueba estadística?
21. ¿Qué efecto tiene un aumento en el tamaño de la muestra en el tamaño del error estándar de una distribución muestral?

EJERCICIOS PARA EL CAPÍTULO 10

Conjunto de problemas 10A

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se estipule lo contrario.

10A-1. Para los siguientes valores de parámetros objetivo hipotéticos y estadísticos muestrales observados, calcula el efecto de la prueba. Presenta las fórmulas.

	Valor del parámetro objetivo hipotético (del paso 1 de los seis pasos)	Estadístico muestral observado (del paso 4 de los seis pasos)	Efecto de la prueba
a)	$\mu_X = 32$ años	$\bar{X} = 28.6$ años	
b)	$P_u = .79$	$P_s = .65$	
c)	$\mu_X = 216$ libras	$\bar{X} = 176.4$ libras	
d)	$P_u = .44$	$P_s = .69$	

10A-2. Prueba tu habilidad para utilizar de manera correcta la tabla de la distribución *t* (tabla estadística C del apéndice B). Completa la siguiente la tabla, que representa los resultados de una serie de pruebas *t* de varios tamaños de muestra y niveles de significación. Para cada prueba, estipula: *a*) la "puntuación crítica" (t_α); *b*) una estimación del valor *p* y *c*) si rechazarías o fracasarías en rechazar la hipótesis nula.

	Tamaño de la muestra (*n*)	Nivel de significación (α)	Colas (dirección de la prueba)	Puntuación obtenida ($t_{\bar{X}}$)	Puntuación crítica (t_α)	Valor *p* estimado (*p*)	¿Rechazas o fracasas en rechazar H_0?
a)	25	.05	No direccional	2.068			
b)	17	.05	Una cola	2.550			
c)	32	.001	Dos colas	2.122			
d)	7	.01	Dos colas	–3.219			
e)	14	.05	Direccional	2.398			

10A-3. Supongamos que en una universidad en 1985 una proporción de .47 de los estudiantes con especialidad en sociología eran mujeres. Tú evalúas la composición por género de una muestra aleatoria de 187 estudiantes especializándose en sociología en la misma universidad en la actualidad y determinas que hay 105 mujeres. Prueba una hipótesis para ver si esta proporción ha cambiado desde 1985.

10A-4. Como supervisor de control de calidad de la embotelladora de agua, Mountain Geyser, Inc., quieres probar una hipótesis para determinar si hay una pérdida en el volumen de la botella de agua de 10 onzas durante los procesos de embotellado y entrega. Para el muestreo tú seleccionas al azar 5 botellas de cada una de 7 tiendas detallistas de la entrega anterior. Tus averiguaciones son:

$\bar{X} = 15.6$ onzas $\quad s_X = .7$ onzas $\quad n = 35$ botellas

10A-5. Loureiro y Nayga (2006) examinaron la relación entre la recomendación de un médico, la obesidad y la disminución de peso del paciente. Supongamos que la disminución media de peso de los pacientes en un estudio de dietas recomendadas por médicos es 7.9 libras. Tú decides realizar un proyecto de investigación similar. Prueba la hipótesis de que los pacientes en tu estudio disminuyeron un peso similar cuando se apegaron a los planes de dieta recomendados por los médicos. Tus datos muestrales son:

$n = 27 \quad \bar{X} = 6.7$ libras $\quad s_X = 2.3$ libras

10A-6. En el paso 3 del procedimiento de los seis pasos de una prueba de hipótesis decidimos sobre un nivel de significación (α), que es la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazaremos H_0. En la curva de la distribución muestral, esta es la región crítica con una curva crítica expresada como el número de errores estándar.

a) Para el ejercicio 10A-4, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de onzas?

b) Para el ejercicio 10A-5, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de libras?

10A-7. Utilizando una técnica de muestreo aleatorio, tú realizas una encuesta política de 462 adultos en el área metropolitana de Johnsonville. La siguiente tabla proporciona los parámetros conocidos acerca de la población de Johnsonville de los datos del censo de Estados Unidos así como datos de su muestra.

a) ¿Es representativa esta muestra de la población de Johnsonville con respecto al género?

b) ¿Es representativa esta muestra de la población de Johnsonville con respecto a la raza?

Comparación de la población de Johnsonville con los datos muestrales ($n = 462$)

Característica	Parámetros de Johnsonville de datos del censo de Estados Unidos	Estadísticos muestrales
Género (% mujeres)	53.2	53.0
Raza (% caucásicos)	66.9	65.5

10A-8. Tú quieres calcular una estimación del intervalo de los ingresos medios de los planeadores urbanos en 150 Áreas estadísticas metropolitanas en Sun Belt. Para esto obtienes una muestra aleatoria de 21 planeadores urbanos y determinas un ingreso medio de \$43 571 con una desviación estándar de \$4 792.

a) Siguiendo los cinco pasos del cálculo de un intervalo de confianza, formula el intervalo de confianza de 99% del ingreso medio de los planeadores urbanos.

b) Excepto por el tamaño menor de la muestra, este ejercicio es el mismo que el 8A-3 del capítulo 8. Resuelve ese ejercicio y compara los resultados con los de este ejercicio. Haz un cometario sobre cómo afecta un tamaño pequeño de la muestra la precisión de un intervalo de confianza.

Conjunto de problemas 10B

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de la probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se indique lo contrario.

10B-1. Para los siguientes valores hipotéticos del parámetro objetivo y estadísticos muestrales observados, calcula el efecto de la prueba. Presenta las fórmulas.

	Valor del parámetro objetivo hipotético (del paso 1 de los seis pasos)	Estadístico muestral observado (del paso 4 de los seis pasos)	Efecto de la prueba
a)	$\mu_X = 100$ mazorcas por bushel	$\bar{X} = 113$ mazorcas por bushel	
b)	$P_u = .50$	$P_s = .39$	
c)	$\mu_X = 572$ automóviles robados	$\bar{X} = 591$ automóviles robados	
d)	$P_u = .29$	$P_s = .34$	

10B-2. Prueba tu habilidad para utilizar de manera correcta la tabla de la distribución t (tabla estadística C del apéndice B). Completa la siguiente tabla, que presenta los resultados de una serie de pruebas t de varios tamaños de muestra y niveles de significación. Para cada prueba, estipula: *a*) la "puntuación crítica" (t_α); *b*) una estimación del valor p y *c*) si tú rechazarías o fracasarías en rechazar la hipótesis nula.

	Tamaño de la muestra (n)	Nivel de significación (α)	Colas (dirección de la prueba)	Puntuación obtenida ($t_{\bar{X}}$)	Puntuación crítica (t_α)	Valor p estimado (p)	¿Rechazas o fracasas en rechazar H_0?
a)	23	.05	No direccional	1.720			
b)	9	.01	Direccional	–3.081			
c)	13	.05	Direccional	–1.133			
d)	22	.001	Dos colas	3.141			
e)	11	.001	Dos colas	13.462			

10B-3. Supongamos que la proporción de rusos étnicos en una muestra de 1996 de ciudadanos rusos de la antigua Unión Soviética fue .63. Tú preguntas a una muestra aleatoria de 139 ciudadanos rusos en la actualidad acerca de su identidad étnica y 91 afirman que son rusos étnicos. Prueba una hipótesis para ver si la proporción de rusos étnicos ha cambiado desde 1996.

10B-4. El número de medio de visitas al médico para personas en Estados Unidos mayores de 55 años es 5.2 por año y el número medio de días en deshabilitación (días completos cuando no se pueden realizar funciones normales) es 7.5 días por año (datos ficticios). Prueba la hipótesis que las personas mayores de 55 años en su ciudad tienen la media nacional para visitas al médico. Sus datos muestrales son:

$n = 65$ $\bar{X} = 5.8$ visitas $s_X = 1.8$ visitas

10B-5. Matthews, Jagger y Hancock (2006) examinaron la relación entre el estado socioeconómico (SES) y la esperanza de vida. En un esfuerzo para duplicar estos resultados en una población diferente, tú eliges una muestra de 29 registros de participantes. Dado el nivel del SES de este grupo, su esperanza de vida media deberá ser 75.0 años. Prueba una hipótesis para establecer si esto es cierto. Tus datos de la muestra son:

$n = 29$ $\bar{X} = 71.8$ años $s_X = 9.2$ años

10B-6. En el paso 3 de los seis pasos de una prueba de hipótesis seleccionamos un nivel de significación (α), que es la cantidad del valor *p* debajo de la cual definiremos el resultado muestral como inusual y rechazaremos H_0. En la curva de la distribución muestral, esta es la región crítica con una puntuación crítica expresada como un número de errores estándar.

a) Para el ejercicio 10B-4, ¿cuál es la puntuación crítica expresada en las unidades de puntuación bruta de las visitas al médico?

b) Para el ejercicio 10B-5, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de años?

10B-7. Utilizando una técnica de muestreo aleatorio, tú realizas una encuesta de salida de 485 adultos en el área metropolitana de Commonwealth. La siguiente tabla proporciona parámetros conocidos acerca de la población de Commonwealth de datos del censo de Estados Unidos así como datos de su muestra.

a) ¿Es representativa esta muestra de la población de Commonwealth con respecto al porcentaje viviendo debajo del nivel de pobreza?

b) ¿Es representativa esta muestra de la población de Commonwealth con respecto al género?

Comparación de la población de Commonwealth con los datos muestrales (*n* = 485)

Característica	Parámetros de Commonwealth de datos del censo de Estados Unidos	Estadísticos muestrales
Género (% mujeres)	52.1	54.0
% viviendo debajo del nivel de pobreza	29.0	33.1

10B-8. La Dra. Latisia Latham, una consejera matrimonial, administra la Escala Global de Angustia (EGA), que mide la discordia marital global. Consiste de 43 preguntas de respuesta falso/verdadero con una puntuación total combinada para los dos compañeros (Snyder, Willis y Grady-Fletcher, 1991). Ella te pide proporcionar una estimación aproximada de la puntuación promedio de su clientela. Tú obtienes una muestra aleatoria de 25 parejas y determina una puntuación EGA media de 59 con una desviación estándar de 5.2.

a) Siguiendo los cinco pasos del cálculo de un intervalo de confianza, establece el intervalo de confianza de 95% de la puntuación EGA media de los clientes de la Dra. Latham.

b) Excepto por el tamaño muestral menor, este ejercicio es el mismo que el ejercicio 8B-3 del capítulo 8. Resuelve ese ejercicio y compara los resultados con los de este ejercicio. Haz un comentario sobre cómo afecta un tamaño muestral pequeño la precisión de un intervalo de confianza.

Conjunto de problemas 10C

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se estipule lo contrario.

10C-1. Para los siguientes valores hipotéticos del parámetro objetivo y estadísticos muestrales objetivo, calcula el efecto de la prueba. Presenta las fórmulas.

	Valor hipotético del parámetro objetivo (del paso 1 de los seis pasos)	Estadístico muestral observado (del paso 4 de los seis pasos)	Efecto de la prueba
a)	$\mu_X = 22$ años	$\bar{X} = 21.4$ años	
b)	$P_u = .75$	$P_s = .69$	
c)	$\mu_X = 146$ libras	$\bar{X} = 138.8$ libras	
d)	$P_u = .56$	$P_s = .74$	

10C-2. Prueba tu habilidad para utilizar de manera correcta la tabla de la distribución *t* (tabla estadística C del apéndice B). Completa la siguiente tabla, que representa los resultados de una serie de pruebas *t* de varios tamaños de muestra y niveles de significación. Para cada prueba, estipula: *a*) la "puntuación crítica" (t_α); *b*) una estimación del valor *p* y *c*) si tú rechazarías o fracasarías en rechazar la hipótesis nula.

	Tamaño de la muestra (*n*)	Nivel de significación (α)	Colas (dirección de la prueba)	Puntuación obtenida ($t_{\bar{X}}$)	Puntuación crítica (t_α)	Valor *p* estimado (*p*)	¿Rechazas o fracasas en rechazar H_0?
a)	15	.05	Direccional	2.421			
b)	9	.01	Dos colas	−3.081			
c)	37	.001	Dos colas	1.986			
d)	19	.05	Una cola	2.470			
e)	30	.05	No direccional	2.049			

10C-3. Un investigador determinó que hace dos décadas 53% de la población favorecía alguna forma de control de armas. Prueba la hipótesis que la misma proporción aún la favorece en la actualidad. Tus datos muestrales son: 105 adultos seleccionados al azar de entre los cuales 66 favorecen el control, 32 no lo favorecen y 7 están indecisos o no tienen opinión.

10C-4. Como el supervisor de control de calidad de una compañía de empaque de frituras. Snak Trak, Inc., tú quieres probar una hipótesis para determinar si hay una pérdida en el volumen de los paquetes de frituras de 12 onzas durante los procesos de empaque y entrega. Para la muestra, tú tomas al azar 4 paquetes de cada uno de los seis distribuidores detallistas de la entrega de la semana pasada. Tus averiguaciones son:

$\bar{X} = 11.8$ onzas $\quad s_X = .7$ onzas $\quad n = 24$ paquetes de frituras

10C-5. Kahn y Pearlin (2006) estudiaron el impacto de la tensión financiera durante la vida entre adultos mayores en Estados Unidos. En su población, ellos determinaron un promedio de 5.6 síntomas físicos. Tú emprendes un estudio similar en 29 norteamericanos mayores que siempre han tenido dificultades financieras y, por tanto, se ajustan a una categoría financiera de *alta tensión*. Prueba la hipótesis que los sujetos en su muestra de alta tensión reportan números mayores de síntomas físicos que los sujetos en el otro estudio. Tus datos de la muestra son:

$n = 29 \quad \bar{X} = 6.3$ síntomas $\quad s_X = 1.1$ síntomas

10C-6. En el paso 3 del procedimiento de los seis pasos de una prueba de hipótesis decides sobre un nivel de significación (α), que es la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazaremos H_0. En la curva de la distribución muestral, esta es la región crítica con una puntuación crítica expresada como un número de errores estándar.

a) Para el ejercicio 10C-4, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de onzas?

b) Para el ejercicio 10C-5, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de síntomas?

10C-7. Utilizando una técnica de muestreo aleatorio, tú realizas una encuesta sociológica de 511 adultos en el área metropolitana de Smithville. La siguiente tabla proporciona parámetros conocidos acerca de la población de Smithville de datos del censo de Estados Unidos así como datos de su muestra.

a) ¿Es representativa la muestra de la población de Smithville con respecto al género?

b) ¿Es representativa la muestra de la población de Smithville con respecto al porcentaje de la población con una educación universitaria?

Comparación de la población de Smithville con los datos muestrales ($n = 511$)

Característica	Parámetros de Smithville de datos del censo de Estados Unidos	Estadísticos muestrales
Género (% mujeres)	46.3	48.0
% con educación universitaria	37.0	36.1

10C-8. Tú necesitas calcular una estimación del intervalo de las edades medias de adultos en edad de votar en un distrito congresional, para esto obtienes una muestra aleatoria de 29 adultos en edad de votar y determinas una edad media de 36.3 años con una desviación estándar de 9.7 años.

a) Siguiendo los cinco pasos del cálculo de un intervalo de confianza, establece un intervalo de confianza de 99% de la edad media de adultos en edad de votar.

b) Excepto por el tamaño muestral menor, este ejercicio es el mismo que el ejercicio 8C-3 del capítulo 8. Resuelve ese ejercicio y compara los resultados con los de este ejercicio. Haz un comentario sobre cómo afecta el tamaño de la muestra la precisión de un intervalo de confianza.

Conjunto de problemas 10D

En todas las pruebas de hipótesis, sigue el procedimiento de los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual y las curvas de probabilidad. Por consistencia, redondea los errores estándar a dos lugares decimales. Utiliza $\alpha = .05$ a menos que se estipule lo contrario.

10D-1. Para los siguientes valores hipotéticos del parámetro objetivo y estadísticos muestrales observados, calcula el efecto de la prueba. Presenta las fórmulas.

	Valor hipotético del parámetro objetivo (del paso 1 de los seis pasos)	Estadístico muestral observado (del paso 4 de los seis pasos)	Efecto de la prueba
a)	$\mu_X = 220$ frituras por bolsa	$\bar{X} = 242$ frituras por bolsa	
b)	$P_u = .72$	$P_s = .48$	
c)	$\mu_X = 483$ edificios ocupados	$\bar{X} = 612$ edificios ocupados	
d)	$P_u = .21$	$P_s = .52$	

10D-2. Prueba tu habilidad para utilizar de manera correcta la tabla de la distribución t (tabla estadística C del apéndice B). Completa la siguiente tabla, que presenta los resultados de una serie de pruebas t de varios tamaños de muestra y niveles de significación. Para cada prueba, estipula: *a)* la "puntuación crítica" (t_α); *b)* una estimación del valor p y *c)* si tú rechazarías o fracasarías en rechazar la hipótesis nula.

	Tamaño de la muestra (n)	Nivel de significación (α)	Colas (dirección de la prueba)	Puntuación obtenida ($t_{\bar{X}}$)	Puntuación crítica (t_α)	Valor p estimado (p)	¿Rechazas o fracasas en rechazar H_0?
a)	13	.001	Dos colas	15.651			
b)	27	.001	Dos colas	4.249			
c)	16	.05	Direccional	−1.741			
d)	21	.05	No direccional	2.115			
e)	7	.01	Direccional	−3.994			

10D-3. La proporción de no caucásicos (afroamericanos, asiáticos, hispanos y otras personas no anglosajonas) en la población del estado es .18. Una muestra aleatoria de 95 estudiantes en la universidad más importante del estado revela que 12 son no caucásicos. ¿Es representativa la población estudiantil de la población del estado con respecto a la raza?

10D-4. Supongamos que tú quieres determinar la eficiencia de los sistemas de empaque y distribución de un fabricante importante de papel, por lo que decides probar una hipótesis para determinar si los paquetes de papel copia de 200 piezas salen con menos piezas debido al proceso de empacado. Para conformar tu muestra, tomas al azar 5 paquetes de papel de cada una de las siete cajas de la siguiente entrega que sale de la fábrica. Tus averiguaciones son:

$\bar{X} = 194$ hojas $\quad s_X = 12$ hojas $\quad n = 35$ paquetes

10D-5. Gaugham (2006) examinó la influencia de compañeros en el hábito de tomar bebidas alcohólicas. Los datos de estos tipos de estudios revelan que los adolescentes consumen un promedio de 8.8 bebidas alcohólicas por mes. Tú vives en un área socialmente conservadora y sospechas que allí los adolescentes consumen menos bebidas alcohólicas. Prueba una hipótesis con los datos siguientes para ver si esto es cierto.

$n = 28 \quad \bar{X} = 8.5$ bebidas $\quad s_X = 1.8$ bebidas

10D-6. En el paso 3 del procedimiento de seis pasos de una prueba de hipótesis decidimos sobre el nivel de significación (α), que es la cantidad del valor p debajo de la cual definiremos el resultado muestral como inusual y rechazamos H_0. En la curva de la distribución muestral, esta es la región crítica con una puntuación crítica expresada como un número de errores estándar.

a) Para el ejercicio 10D-4, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de hojas de papel?

b) Para el ejercicio 10D-5, ¿cuál es la puntuación crítica expresada en las unidades de la puntuación bruta de bebidas?

10D-7. Utilizando procedimientos de muestreo aleatorio, tú realizas una encuesta importante demográfica de 512 adultos en la ciudad de Galway. La siguiente tabla proporciona parámetros conocidos acerca de la población de Galway de datos del censo de Estados Unidos así como datos de su muestra.

a) ¿Es representativa la muestra de la población de Galway con respecto al porcentaje de respondientes con hijos?

b) ¿Es representativa la muestra de la población de Galway con respecto al género?

Comparación de la población de Galway con los datos muestrales ($n = 512$)

Característica	Parámetros de Galway de datos del censo de Estados Unidos	Estadísticos muestrales
Género (% mujeres)	54.9	56.6
% de respondientes con hijos	42.0	39.9

10D-8. Tú quieres calcular una estimación del intervalo de las puntuaciones en el Graduate Record Examination (GRE) entre estudiantes graduados de nuevo ingreso en una universidad urbana grande. Tú has reunido una muestra aleatoria de 19 estudiantes y determinado una puntuación GRE media de 1 200 puntos con una desviación estándar de 60 puntos.

a) Siguiendo los cinco pasos del cálculo de un intervalo de confianza, establece el intervalo de confianza al 99% de la puntuación GRE media de los estudiantes graduados de nuevo ingreso.

b) Excepto por el tamaño muestral menor, este ejercicio es el mismo que el ejercicio 8D-3 del capítulo 8. Resuelve ese ejercicio y compara los resultados con los de este ejercicio. Haz un comentario sobre cómo una muestra pequeña afecta la precisión de un intervalo de confianza.

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 10

Si en tu clase se utilizan las aplicaciones computacionales opcionales que acompañan a este texto, abre los ejercicios del capítulo 10 en el sitio en la red *The Statistical Imagination* en www.mhhe.com/ritchey2. Los ejercicios se enfocan en: 1) correr pruebas de hipótesis de una muestra única en *SPSS for Windows*, 2) utilizar pruebas de hipótesis de muestras únicas para probar parámetros objetivo y 3) utilizar pruebas de hipótesis de muestras únicas para examinar la representatividad de muestras. Además, en el apéndice D de este texto se proporciona un repaso breve de las secuencias de comandos *SPSS* para procedimientos estudiados en este capítulo.

CAPÍTULO

11

Relaciones bivariadas: prueba *t* para comparar las medias de dos grupos

RESUMEN DEL CAPÍTULO

Introducción: análisis bivariado

Como vimos en el capítulo 6, la predicción sirve de vínculo entre el análisis estadístico y la teoría de la probabilidad. La capacidad de predecir resultados de la aplicación de técnicas de muestreo, además de otros eventos futuros, constituye una valiosa habilidad. Hacer predicciones constituye una parte importante de la ciencia, mercadotecnia, finanzas, fabricación,

servicios médicos y sociales, juegos de azar y cualquier otro campo en el que se utilice la estadística.

La imaginación estadística va más allá de la mera predicción con el único fin de anticiparse a lo que sucederá: la predicción también permite comprender. Las predicciones científicas dependen de la medición de diversos fenómenos y del empeño de hilar dichas mediciones de forma significativa. Con la ayuda de ideas bien organizadas (teorías), la predicción amplía el entendimiento y viceversa. Por ejemplo, al estudiar la depresión psicológica, si logramos determinar lo que la provoca, podremos identificar a las *poblaciones en riesgo* y tomar medidas preventivas. Por ejemplo, la investigación muestra que las personas que sufren múltiples crisis en sus vidas —diversos *eventos vitales estresantes* en un breve periodo— tienen más riesgo de padecer depresión. El conocimiento de que dichas crisis inducen los síntomas depresivos ha alentado un movimiento para crear clínicas que atienden crisis y grupos de autoayuda. Sencillamente existe una relación entre los eventos vitales estresantes y la depresión.

El análisis bivariado (o de dos variables) implica la búsqueda de relaciones estadísticas entre dos variables. Una **relación estadística** entre dos variables afirma que *las mediciones de una variable tienden a fluctuar de forma coherente con respecto a las mediciones de la otra, lo cual convierte a una de las variables en un buen predictor de la otra.* En la investigación científica, a la variable *predictora* la denominamos *variable independiente* y a la variable *predicha, variable dependiente.* La forma de medir una relación depende de los niveles de medición de las dos variables. Existen tres enfoques comunes para medir relaciones estadísticas:

1. *Prueba de la diferencia de medias:* se comparan las medias de una variable de intervalo o de razón entre las categorías o grupos de una variable nominal u ordinal.
2. *Conteo de las frecuencias de ocurrencias conjuntas* de atributos de dos variables nominales.
3. *Medición de la correlación* entre dos variables de intervalo o de razón.

Estos enfoques se resumen en la tabla 11.1. Aunque este libro se concentra en el enfoque de la diferencia de medias, comencemos analizando brevemente los tres enfoques.

Relación estadística Las mediciones de una variable tienden a fluctuar de forma coherente con respecto a las mediciones de la otra, lo cual convierte a una de las variables en un buen predictor de la otra.

Pruebas de diferencia de medias

Al plantear una hipótesis sobre una relación entre una variable dependiente de intervalo o de razón y una variable independiente dicótoma (de dos grupos) nominal u ordinal, aplicamos el enfoque de la diferencia de medias. La prueba de hipótesis para esta relación se denomina *prueba de la diferencia de medias de dos grupos,* el tema del capítulo. Por ejemplo, los publicistas desean conocer si hay una relación entre la variable nominal *género* y la variable de intervalo/razón *tiempo invertido en ver programas deportivos en televisión.* Plantear una hipótesis sobre esta relación implica afirmar que la media de horas dedicadas a ver la televisión es superior para los hombres que para las mujeres. Si se determina que esto es verdadero, los anuncios de productos masculinos, como las máquinas de afeitar, podrían pasarse durante los programas deportivos.

TABLA 11.1 | Enfoques comunes para medir la relación estadística entre dos variables

Niveles de medida de las dos variables		
Variable independiente	***Variable dependiente***	***Enfoque para medir la relación estadística entre dos variables***
Nominal u ordinal	De intervalo o de razón	Comparar las diferencias de medias de la variable de intervalo/razón entre las categorías de una variable nominal u ordinal (capítulos 11 y 12).
Nominal u ordinal	Nominal u ordinal	Contar las frecuencias de la ocurrencia conjunta de atributos de categoría de dos variables nominales ordinales (capítulo 13)
De intervalo o de razón	De intervalo o de razón	Medir la correlación entre las variables (capítulos 14 y 15).

Para probar si existe una diferencia entre tres o más medias, aplicamos un procedimiento denominado *análisis de varianza* (ANOVA) (véase capítulo 12). Con el ANOVA, las medias se comparan indirectamente determinando cuánta de la varianza se explica por el hecho de pertenecer a un grupo. Por ejemplo, supongamos que tenemos una variable de intervalo/razón denominada *apoyo al concepto de equipo* para el cuidado hospitalario, una escala de medición que consiste en 40 preguntas. Ésta permite discernir si los profesionales médicos de un hospital están convencidos de la necesidad de compartir la responsabilidad del cuidado del paciente. Una teoría que trate sobre las relaciones entre profesionales quizás plantee la pregunta: ¿tratan los médicos a las enfermeras, los farmacéuticos y los terapeutas físicos como compañeros de profesión, como parte de un equipo de colegas, o como subordinados que deben recibir órdenes? La prueba ANOVA compararía las puntuaciones medias que favorecen el concepto de equipo de estos cuatro grupos, con el fin de saber si los médicos apoyan menos el concepto. Con respecto al establecimiento de una relación entre dos variables, el ANOVA se asemeja a la prueba de diferencia de medias con dos grupos, ya que ambas comparan medias (de una variable de intervalo/razón) entre los grupos o categorías de una variable nominal u ordinal.

Ocurrencias conjuntas de atributos

Otra forma de ver la relación entre dos variables consiste en prestar particular atención al análisis de la *ocurrencia conjunta* de los atributos de dos variables nominales. Un atributo es una cualidad o característica de un individuo que se expresa en los nombres de las categorías de una variable nominal u ordinal. En el caso del género, se tienen los atributos hombre y mujer; para la raza, blanco, afroamericano e hispano; y así sucesivamente. Una ocurrencia conjunta de atributos implica dos variables *para cada individuo,* con parejas de atributos de las dos variables. Por ejemplo, Mary Smith es una *mujer blanca;* ella tiene la ocurrencia conjunta de los atributos blanco y mujer. Asimismo, John Jones es un *hombre afroamericano.*

Un ejemplo de pregunta de investigación tiene que ver con la relación entre dos variables nominales/ordinales es la siguiente: ¿es más probable que las mujeres apoyen los servicios de cuidado infantil subsidiados por el gobierno que los hombres? Expresado en términos de una hipótesis alternativa: existe una relación entre el género (hombre frente a mujer) y el apoyo a los servicios de cuidado infantil subsidiados por el gobierno (sí frente a no). Si se acepta esta hipótesis alternativa, será como consecuencia de que las ocurrencias conjuntas de los atributos de mujer —sí y hombre— no ocurren más significativamente de lo esperado. La identificación de este tipo de relaciones es importante para la política pública. Por ejemplo, el conocimiento de la conformación de una comunidad en cuanto a género, permite hacer mejores predicciones relacionadas con la posibilidad de que haya apoyo para los servicios de cuidado infantil subsidiados con los impuestos. Para determinar relaciones entre dos variables nominales aplicamos una prueba chi cuadrada (capítulo 13).

Correlación

Una relación entre dos variables de intervalo/razón indica que las puntuaciones en una variable tienden a cambiar coherentemente, o *se correlacionan,* con las puntuaciones de la otra. Por ejemplo, existe una correlación entre las variables de experiencia frecuente de eventos vitales estresantes (como la muerte del ser amado, un divorcio o la pérdida del empleo) y la de depresión psicológica. Decir que estas dos variables están correlacionadas significa que es posible que los individuos que experimentan muy pocos eventos vitales estresantes tendrán puntuaciones bajas en las medidas de depresión psicológica, mientras que aquellos que experimentan una gran cantidad de tales eventos probablemente tendrán puntuaciones altas en las medidas de depresión. Los estadísticos de correlación se analizan en los capítulos 14 y 15.

En el caso de dos variables de nivel ordinal, aplicamos una prueba de *correlación de rangos ordenados.* Esta permite saber si los sujetos de estudio que ocupan los primeros lugares en una medida también tienden a mantenerlos en la otra. Por ejemplo, al comparar 50 estados, ¿existe una correlación entre el lugar que ocupa un estado en lo que se refiere a nivel de pobreza y el que ocupa en cuanto a tasas de delitos? De esta manera, en el caso de los 50 estados, si uno de ellos se encuentra entre los primeros 10 lugares en lo que se refiere a pobreza, ¿se encuentra también entre los primeros 10 en cuanto a la tasa de delitos? Esta prueba se incluye en el sitio web.

En todas las preguntas de investigación que tienen que ver con la relación entre dos variables, hipotetizamos una relación entre una variable independiente y una variable dependiente (revisa el capítulo 1). La variable dependiente es la que nos interesa principalmente —la variable que intentamos explicar— y la variable independiente es la que se utiliza para predecir puntuaciones de la variable dependiente. Por ejemplo, ¿originan las circunstancias vitales estresantes (variable independiente) depresión psicológica (variable dependiente)? ¿Contribuyen las condiciones de pobreza (variable independiente) a la incidencia de actos delictivos (variable dependiente)? En esta obra sólo se analizan casos de dos variables o pruebas estadísticas bivariadas. Las relaciones entre tres o más variables son temas para textos avanzados de estadística *multivariada.* Concentremos nuestra atención en el enfoque relacionado con la diferencia de medias para el caso de dos grupos.

Prueba de diferencia de medias (prueba t) para dos grupos con muestras independientes

Una prueba de diferencia de medidas para dos grupos simplemente compara las medias de una variable de intervalo/razón para dos grupos o categorías de una variable nominal/ordinal. Por ejemplo, tal vez nos interese saber si la media de las puntuaciones del examen de admisión a la universidad es más alta en el caso de los estudiantes de la universidad de la ciudad que en los estudiantes de la universidad estatal. O quizás deseemos demostrar que no existe diferencia en la media de los niveles de energía de los individuos pertenecientes a un grupo experimental que ingiere una píldora multivitamínica y los que pertenecen al grupo control, a quienes sólo se les administran placebos (píldoras falsas).

Por supuesto, la variable en la cual se calcula la media debe ser una variable de intervalo/razón. Además, en esta prueba de medias de dos grupos esta variable típicamente es la variable dependiente.

Los dos grupos de comparación de la variable independiente pueden provenir de poblaciones separadas, como la universidad de la ciudad y la universidad estatal. Sin embargo, se puede emplear la misma prueba si los dos grupos constituyen categorías de una variable dicotómica nominal/ordinal proveniente de una sola población (como la variable de género, con las categorías hombre y mujer). En dicha situación, comparamos las poblaciones separadas de hombres y mujeres incluidos en la población total.

Para ilustrar una prueba de diferencia de medias para dos grupos, supongamos que hemos entablado una discusión con un amigo sobre si los hombres o las mujeres obtienen mejores calificaciones en cierta universidad, cuyo campus alberga aproximadamente a 9 000 estudiantes. Argumentamos que no existe diferencia en los promedios de calificaciones de estas dos categorías de estudiantes. Nuestro amigo, que ha sido influenciado por mitos tradicionales relativos al hecho de que las mujeres son inferiores a los hombres desde un punto de vista intelectual, apuesta $10 al hecho de que los hombres se desempeñan mejor que las mujeres. Sospechamos que no está al tanto de cuánto han cambiado los roles de género. Por ejemplo, aunque alguna vez el campus de la universidad fue el bastión de la dominación masculina, ahora las mujeres rebasan en número los hombres en las universidades de Estados Unidos. Así que decidimos aceptar su apuesta.

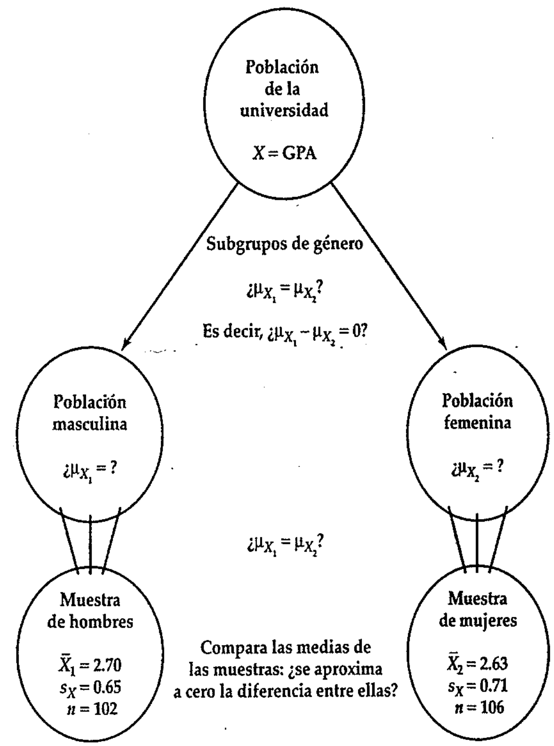
Para resolver el problema de las calificaciones, seleccionamos aleatoriamente los registros de 208 estudiantes, 102 hombres y 106 mujeres. Descubrimos que la media de las calificaciones de los hombres es de 2.70, con una desviación estándar de 0.65; en el caso de las mujeres, la media es de 2.63, con una desviación estándar de 0.71. Nuestro amigo exclama: "¿Lo ves? ¡Te lo dije. Ahora paga!" Nosotros respondemos: "espera un momento. Las diferencias entre muestras no siempre implican diferencias en la población". Entonces intentamos explicarle que es necesario probar la hipótesis de que existe una diferencia significativa entre el promedio de los estudiantes varones y el de las mujeres *en la población total* de dicha universidad. Necesitamos convencer a nuestro amigo de que la diferencia de 0.07 entre el promedio que se encontró en las muestras quizá se deba al error de muestreo.

En esta prueba de hipótesis, el promedio constituye la variable dependiente y el género, la variable independiente, ya que estamos interesados en saber si el género predice el promedio. Por supuesto, nuestro interés se centra en la población estudiantil de la universidad. Con el fin de mantener nuestro enfoque en la población, al llevar a cabo una prueba de medias para dos grupos, resulta conveniente considerar a los subgrupos en el contexto de sus poblaciones separadas. En otras palabras, imaginemos a los hombres y a las mujeres como dos poblaciones distintas en el campus, como lo muestra la figura 11.1.

La pregunta de investigación tiene que ver con las poblaciones de los dos grupos y busca precisar si sus parámetros para el promedio de las calificaciones (es decir, μ_X) son diferentes. No obstante, la hipótesis nula consiste en que las dos μ_X son iguales, ya que es esta hipótesis la que nos proporciona resultados definidos o esperados en el muestreo. Específicamente, si los parámetros de datos poblacionales son iguales, debemos esperar que las medias de las dos muestras tengan una diferencia de cero, con un pequeño error de muestreo. Es decir, si $\mu_{X_1} = \mu_{X_2}$, entonces al restar μ_{X_2} de ambos miembros de esta ecuación, obtenemos $\mu_{X_1} - \mu_{X_2} = 0$.

La figura 11.1 muestra que $\mu_X = ?$ para cada población, lo cual significa que no conocemos la media real de las calificaciones ya sea de hombres o de mujeres. Esto destaca un

FIGURA 11.1
Prueba de igualdad de dos medias poblacionales (μ_X)



punto interesante: a diferencia de la prueba de medias para una sola muestra del capítulo 10, con una prueba de diferencia de medias para dos grupos *no* es necesario contar con un valor objetivo para un parámetro de *X* para toda la población o sus subgrupos. Sin embargo, para todas las hipótesis debemos ser capaces de predecir un parámetro y describir la distribución muestral con respecto a este punto. El parámetro conocido que se requiere para la prueba de diferencia de medias para dos grupos es cero. Sin importar los valores reales de la media de calificaciones de los estudiantes varones o mujeres, si estos valores son iguales, la diferencia es cero. Cero, es decir, *la diferencia de las medias*, y no las medias mismas, constituye el valor objetivo de esta prueba de hipótesis.

La diferencia de las medias de dos grupos implica una prueba *t*. Esta se enfoca en la diferencia calculada entre dos medias muestrales, $\bar{X}_1 - \bar{X}_2$, en la que $\bar{X}_1$ y $\bar{X}_2$ son las medias de los grupos 1 y 2, respectivamente. Al describir la distribución muestral, podemos predecir que la diferencia entre dos medias muestrales cualesquiera dos muestras extraídas al azar es cero, con cierto grado de error de muestreo. La prueba se enfoca en la pregunta de si la diferencia observada entre estas medias muestrales para nuestra observación única de datos, reflejá una diferencia real en las medias de la población de los subgrupos, o si sencillamente se debe al error de muestreo. En otras palabras, determinamos si el efecto de la prueba resulta estadísticamente significativa. Por ejemplo, si encontramos una diferencia significativa entre los promedios de los estudiantes universitarios de los hombres y las mujeres, concluiremos que dicha diferencia constituye un efecto del género.

La prueba t relacionada con la diferencia entre dos medias se aplica cuando se satisfacen los siguientes criterios:

Cuando utilizar la prueba de diferencia de medias (prueba t) para dos grupos independientes (distribución t)

En general: al probar una hipótesis entre una variable independiente nominal/ordinal dicotómica y a una variable dependiente de intervalo/razón:

1. Hay dos variables provenientes de una población y una muestra, una de las variables es de nivel de medición de intervalo/razón y la otra es una variable nominal/ordinal dicotómica, *o* hay dos poblaciones o muestras y una variable de intervalo/razón; las muestras son representativas de sus poblaciones.
2. La variable de intervalo/razón es la variable dependiente.
3. Los dos grupos son independientes entre sí; es decir, no constan de los mismos individuos.
4. En el caso de la variable de intervalo/razón, debe suponerse que las varianzas (o desviaciones estándar) de las *poblaciones* de las cuales provienen los dos grupos son iguales y así deben evidenciarse en las varianzas de las muestras. Si *en las muestras,* la varianza de un grupo es más del doble de la magnitud de la varianza del otro, se requieren ajustes en el cálculo del error estándar de la distribución muestral.

El tercer criterio —la independencia de los grupos— distingue esta prueba estadística de aquella en la que el mismo grupo de individuos se compara en dos variables o en dos tiempos diferentes con respecto a una variable. Posteriormente en este capítulo aprenderemos sobre la prueba t en el caso de una diferencia de medias entre grupos "no independientes".

En textos de estadística, el cuarto criterio recibe el nombre de *suposición de varianzas iguales* (o desviaciones estándar iguales, sin olvidar que la varianza es la desviación estándar al cuadrado). Si las varianzas poblacionales no son iguales, se deben llevar a cabo ajustes en la prueba estadística. Es decir que se emplean fórmulas estadísticas para la prueba t, dependiendo de que las varianzas (o desviaciones estándar) de las *poblaciones* de las que provienen los dos grupos sean iguales entre sí. Comenzaremos ejemplificando la prueba estadística en la que las varianzas son iguales. Posteriormente explicaremos la razón por la que deben realizarse ajustes en la prueba cuando las varianzas de los dos grupos no son iguales.

El error estándar y la distribución muestral para la prueba t de la diferencia entre dos medias

Para entender cómo se forma la distribución muestral para una prueba de diferencia de medias con dos grupos, imaginemos que contamos frijoles, muestreando repetidamente entre dos poblaciones con las mismas medias con respecto a alguna variable X de intervalo/razón. Para cada muestra calcularemos la media de X y restaremos estas dos medias muestrales para obtener la diferencia entre las medias: $\bar{X}_1 - \bar{X}_2$. Si representamos los dos grupos con los

subíndices 1 y 2, los cálculos caerán en la dirección positiva o negativa de acuerdo con el siguiente criterio:

- cuando $\bar{X}_1 > \bar{X}_2$, la diferencia será positiva;
- cuando $\bar{X}_1 < \bar{X}_2$, la diferencia será negativa;
- cuando $\bar{X}_1 = \bar{X}_2$, la diferencia será cero.

Si la hipótesis nula que establece que las dos medias poblacionales son iguales es verdadera, en el muestreo repetido esperamos no caer en el lado alto (positivo) tan a menudo como esperamos no caer en el lado bajo (negativo). De esta manera, la distribución muestral de una cantidad grande de diferencias de medias muestrales es simétrico y se centra alrededor de cero. La forma corresponde aproximadamente a la de una distribución t normal, y se utiliza la tabla de la distribución t (tabla C, apéndice B) para obtener el valor p: la probabilidad del resultado muestral suponiendo que las dos medias poblacionales son iguales.

La distribución muestral se centra en una diferencia de cero entre las dos medias poblacionales (es decir, la diferencia entre los parámetros $\mu_{X_1} - \mu_{X_2}$). El error estándar de la distribución muestral se calcula utilizando las varianzas (es decir, las desviaciones estándar al cuadrado) y los tamaños de las dos muestras. Cuando la varianza de una muestra no es mayor al doble del tamaño de la otra, esto sugiere que las varianzas de las dos poblaciones son iguales y "suponemos la igualdad de las varianzas". (A la igualdad de las varianzas también se le da el nombre de *homogeneidad de las varianzas* u *homoscedasticidad.*)

En este caso, en el que se asume la igualdad de las varianzas, el error estándar de la diferencia de las medias se calcula promediando las dos varianzas. Esto se denomina estimación del error estándar con varianzas agrupadas y tiene la siguiente fórmula.

Cálculo del error estándar de la diferencia entre dos medias (estimación con varianzas agrupadas, utilizada cuando las varianzas de las dos poblaciones parecen iguales)

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{(n_1-1)s_{X_1}^2 + (n_2-1)s_{X_2}^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

con $gl = n_1 + n_2 - 2$

donde

$s_{\bar{X}_1-\bar{X}_2}$ = estimación con varianzas agrupadas del error estándar de la diferencia entre dos medias

n_1 = tamaño de la muestra del grupo 1

n_2 = tamaño de la muestra del grupo 2

$s_{X_1}^2$ = varianza del grupo 1

$s_{X_2}^2$ = varianza del grupo 2

Nota que el *subíndice* del símbolo para esta fórmula del error estándar es $\bar{X}_1 = \bar{X}_2$. Este error estándar es la desviación estándar de la distribución de las *diferencias* entre dos medias muestrales.

El estadístico de la prueba *t* se calcula de la siguiente manera:

Cálculo de una prueba *t* de la diferencia entre dos medias poblacionales

$$t_{\bar{X}_1-\bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$$

en la cual

$t_{\bar{X}_1-\bar{X}_2}$ = número de errores estándares que la diferencia entre dos medias muestrales se desvía de la diferencia hipotética de cero

$\bar{X}_1$ = media muestral del grupo 1

$\bar{X}_2$ = media muestral del grupo 2

$s_{\bar{X}_1-\bar{X}_2}$ = error estándar de las diferencias entre dos medias

Como todas las pruebas estadísticas, este estadístico de prueba está diseñado para contestar preguntas relacionadas con parámetros. La hipótesis nula para esta prueba siempre se referirá al hecho de que las dos medias poblacionales son iguales; es decir,

$$H_0: \mu_{X_1} = \mu_{X_2}$$

Esto es, no hay diferencia entre las dos medias poblacionales.

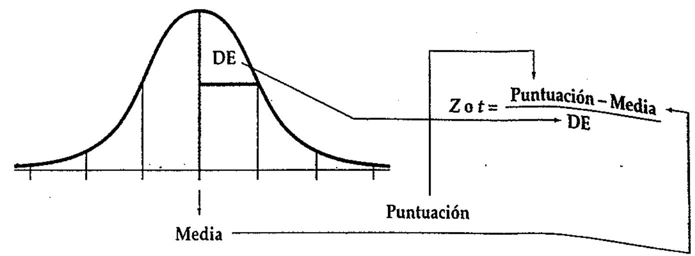
Sin embargo, observa que los parámetros μ_{X_1} y μ_{X_1} no figuran en una fórmula de la prueba *t*. Esto sucede en virtud de que cuando las dos medias poblacionales son iguales, la diferencia entre ellas es cero, lo cual hace que se cancelen en la fórmula. De hecho la fórmula completa para el estadístico de prueba $t_{\bar{X}_1-\bar{X}_2}$ se reduce a la fórmula anterior, de la siguiente manera:

$$t_{\bar{X}_1-\bar{X}_2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{X_1} - \mu_{X_2})}{s_{\bar{X}_1-\bar{X}_2}}$$

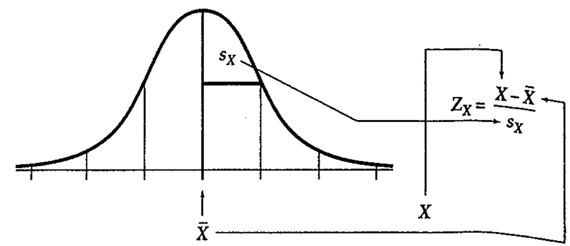
$$= \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{X}_1-\bar{X}_2}} - 0 = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$$

Esto es consistente con todos nuestros estadísticos de prueba. El numerador es un cálculo del efecto de la prueba, la diferencia entre lo que observamos en la muestra y lo que se espera cuando la hipótesis nula es verdadera. En este caso observamos $\bar{X}_1 - \bar{X}_2$ y esperamos cero, ya que $\mu_{X_1} - \mu_{X_2} = 0$ cuando las medias poblacionales son iguales. Al calcularse, el estadístico de prueba simplemente anula el cero por ser redundante.

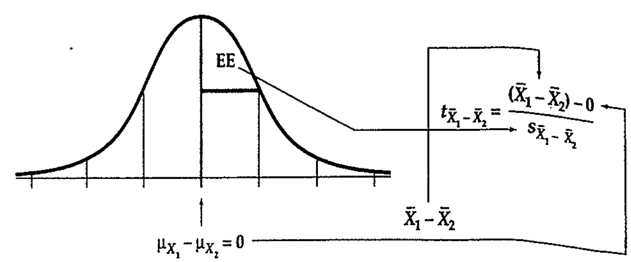
Una mirada a los términos de esta fórmula en lo que se refiere a la curva de distribución revela que se trata de una distribución más que mide la distancia a la que cae el estadístico muestral observado con respecto a un estadístico esperado. Cualquier curva de distribución Z o de distribución *t* tiene una media y una desviación estándar (DE), y una puntuación de intervalo/razón se mide a lo largo de su eje horizontal de la siguiente manera:



Cuando la medida de intervalo/razón es una puntuación bruta *X* entonces se calcula Z_X, como en el capítulo 6.



En el caso de la distribución muestral para la diferencia de medias, la fórmula adquiere forma de la misma manera, pero con la desviación estándar en lugar del error estándar (EE) de la diferencia entre dos medias de la siguiente manera:



Ahora procedamos a realizar los seis pasos de la inferencia estadística para la prueba de la diferencia de medias para dos grupos.

Los seis pasos de la inferencia estadística para la prueba de la diferencia de medias para dos grupos

Breve lista de verificación de los seis pasos de la inferencia estadística

PREPARACIÓN DE LA PRUEBA

Formula la pregunta de investigación. Elabora diagramas conceptuales que describan las especificaciones, incluyendo las poblaciones y muestras bajo estudio, las variables (por ejemplo, $X = \ldots$, $Y = \ldots$) y sus niveles de medición, así como los estadísticos dados o calculados y los parámetros. Establece el procedimiento adecuado de la prueba estadística.

SEIS PASOS

Empleando el símbolo H para presentar la *hipótesis:*

1. Formula H_0 y H_A y estipula la dirección de la prueba.
2. Describe la distribución muestral.
3. Determina el nivel de significancia (α) y especifica el valor crítico de la prueba.
4. Observa los resultados de la muestra cuestión y calcula los efectos de la prueba, el estadístico de prueba y el valor p.
5. Toma la decisión de rechazo.
6. Interpreta y aplica las mejores estimaciones en términos comunes.

Solución para la prueba de la diferencia de medias para dos grupos independientes (prueba *t*)

Pregunta de investigación: la apuesta de $10 con nuestro amigo gira en torno a la pregunta: ¿es superior el promedio de los estudiantes varones de la universidad que el de las estudiantes mujeres? *Especificaciones:*

Procedimiento estadístico: prueba *t* de la diferencia entre dos medias poblacionales; distribución *t;* se asumen varianzas iguales del promedio de calificaciones en las poblaciones de hombres y mujeres. *Especificaciones:* proporciona las especificaciones aquí como se describe en la figura 11.1.

SEIS PASOS

1. H_0: $\mu_{X1(estudiantes\ varones)} = \mu_{X2(estudiantes\ mujeres)}$ (o $\mu_{X_1} - \mu_{X_2} = 0$)

   Es decir, no existe una diferencia entre los promedios de las calificaciones de los estudiantes hombres y mujeres.

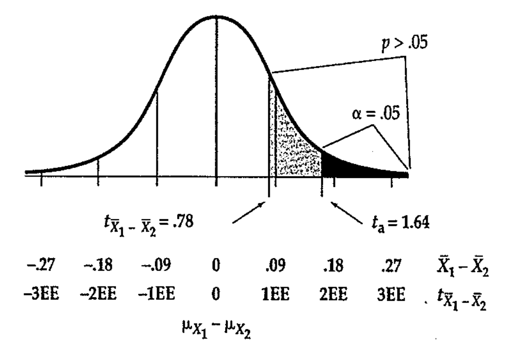
   H_A: $\mu_{X1(estudiantes\ varones)} > \mu_{X2(estudiantes\ mujeres)}$ (o $\mu_{X_1} - \mu_{X_2} > 0$). Una cola

   Es decir, los estudiantes del sexo masculino tienen un promedio de calificación superior a las estudiantes del sexo femenino.

2. *Distribución muestral.* Si H_o es verdadera y las muestras de 102 estudiantes varones y 106 mujeres se extraen repetidamente de sus poblaciones en la universidad, las diferencias entre las medias muestrales $\overline{X}_1 - \overline{X}_2$ se centrarán alrededor de cero como una distribución *t* aproximadamente normal con $gl = n_1 + n_2 - 2 = 206$, y un error calculado como se indica a continuación. (Esta ecuación toma en cuenta el hecho de que las varianzas poblacionales son iguales.)

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{(n_1-1)s_{X_1}^2 + (n_2-1)s_{X_2}^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

$$= \sqrt{\frac{(101).65^2 + (105).71^2}{102+106-2}}\sqrt{\frac{102+106}{(102)(106)}} = .09$$



3. *Nivel de significancia:* $\alpha = 0.05$. Una cola. Situación crítica de la prueba $= t_\alpha = 1.64$. (Marcada sobre la curva.)
4. *Observación:* efecto de la prueba $= \overline{X}_1 = \overline{X}_2 = 0.07$ GPA puntos

   Estadístico de prueba:

$$t_{\overline{X}_1-\overline{X}_2} = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1-\overline{X}_2}} = \frac{2.70 - 2.63}{0.09} = \frac{0.07}{0.09} = 0.78 \text{ EE}$$

   valor *p:* *p* [se extraen medias muestrales ($\overline{X}$) con una diferencia tan inusual o más inusual que 0.07, cuando la diferencia en las medias o nacionales (μ) es cero] > 0.05. (Este valor *p* se sombreó en la curva del paso numérico.)
5. *Decisión de rechazo:* [$|t_{\overline{X}_1-\overline{X}_2}| < |t_\alpha|$ (es decir, 78 < 1.64); de esta manera, $p > \alpha$ (es decir, $p > 0.05$). Se falla en rechazar H_0.
6. *Interpretación:* no existe una diferencia real entre los promedios de los estudiantes varones y mujeres en la universidad. *Mejor estimación:* los promedios son iguales. La diferencia observada de 0.07 puntos en las muestras fue resultado del error de muestreo normal esperado. *Respuesta:* el promedio de los hombres no es más alto que el de las mujeres. ¡Nuestro amigo debe pagar!

He aquí algunos comentarios relacionados con la prueba de diferencia de medias para dos grupos.

- En el paso 1 definimos que la H_0 se refería a "las medias son iguales". Según se indicó, podríamos haberla definido como "la diferencia entre las medias es cero". El cero es lo que centra la curva del paso 2
- En el paso 2, al calcular el error estándar, debemos ejercer cuidado para distinguir la desviación estándar de la varianza. Si en un problema se da la desviación estándar, ésta debe elevarse al cuadrado para obtener la varianza; sin embargo, si en un problema se da la varianza, la operación de elevar al cuadrado no se requiere.
- En el paso 2, los grados de libertad se calculan mediante la operación $gl = n_1 + n_2 - 2$. *Se pierde* un grado de libertad en el cálculo de la media de cada muestra. (Recuerda que en el cálculo de una prueba t con una única muestra, $gl = n - 1$; véase capítulo 10.)
- En el paso tres, observa que revisamos la tabla de la distribución t para obtener el valor crítico t (es decir, t_α), el valor t dónde comienza la región crítica. Dados los tamaños de las muestras y las desviaciones estándar para los dos grupos, t_α constituye el número de errores estándares a partir de cero a los que la diferencia calculada de las medias muestrales se debe encontrar antes de que comencemos a sospechar que las medias de las dos poblaciones no son iguales. En otras palabras, con el muestreo repetido, una diferencia entre medias que se encuentre a 1.64 errores estándares del valor de cero hipotético se presenta menos de 5 veces en 100 muestras extraídas. Un valor t observado y calculado (es decir, $t_{\bar{X}_1 - \bar{X}_2}$) igual o mayor que 1.64 resultaría inusual en poblaciones con medias iguales. En este ejemplo, nuestro valor t observado ($t_{\bar{X}_1 - \bar{X}_2}$) de 0.78 no es tan grande; es menor que 1.64. De esta manera, la diferencia de 0.07 entre las medias de las muestras no es significativamente diferente del valor hipotético de cero. Concluimos que este efecto de la prueba no es inusual y sólo se debe al error de muestreo; por tanto, permitimos que la hipótesis nula se preserve. Recuerda que existe una relación inversa entre el valor de un estadístico de prueba y la probabilidad de rechazar la hipótesis nula. Cuanto mayor sea el valor t calculado, con mayor probabilidad rechazaremos de hipótesis relacionada con el hecho de que las medias son iguales.
- En el paso 5, *fallamos en rechazar* la hipótesis estadística relativa a la igualdad de las varianzas en las poblaciones de los dos grupos. Demostramos que la diferencia observada en las muestras, dados los tamaños muestrales de 102 y 106, es muy pequeña, y se puede esperar que ocurra con frecuencia al calcular la diferencia utilizando datos de muestras.

Cuando las varianzas de las poblaciones (o desviaciones estándares) parecen radicalmente diferentes

Como antes observamos, la prueba t relacionada con las diferencias de medias se vale de las fórmulas anteriores solamente si las varianzas de las poblaciones son iguales. Los ajustes de la fórmula de la prueba t son necesarios en el caso de que estas varianzas sean muy diferentes. Una regla práctica estriba en que podemos asumir que las varianzas *poblacionales* son iguales si la varianza de la *muestra* de un grupo no es superior al doble del tamaño de la del otro grupo. Si se supera este límite, se requiere una fórmula diferente del error estándar, como más adelante se indica. Sin embargo, la aplicación de esta fórmula modificada depende de otros factores, como el tamaño de las muestras, si estas son de tamaños similares, de los tamaños de las desviaciones estándares en relación con sus medidas y si cualquiera de sus distribuciones se encuentra sesgada. Estas complicaciones se pueden evitar utilizando la computadora para llevar a cabo los cálculos. Los programas de computadora pueden llevar a cabo las pruebas t de ambas maneras y ofrecen pautas para elegir el resultado apropiado. Dadas las complejidades implicadas, pocas veces calculamos a mano las pruebas t. De todas formas, para dar realce al enfoque proporcional y lineal, vale la pena discutir el significado de la suposición de que las varianzas son iguales y las consecuencias que surgen cuando no se puede asumir esto.

¿Por qué queremos saber si las varianzas (o desviaciones estándar) de las poblaciones son iguales? Con el error estándar de cualquier prueba estamos estimando el error de muestreo. Como consecuencia, si las varianzas de la población *no* son iguales, los resultados del muestreo repetido se encontrarán más dispersos y el error de muestreo será mayor. El cálculo del error estándar para varianzas diferentes considera este hecho. Si ignoramos este error adicional, es posible que lleguemos a la conclusión incorrecta de que existe una diferencia de medias entre las dos poblaciones cuando, de hecho, una diferencia grande *observada* en las medias de la muestra se debe a una diferencia grande en las varianzas.

Para ilustrar esto, supongamos que tenemos dos poblaciones con puntuaciones medias iguales de coeficiente intelectual (CI). La población corresponde a una preparatoria de clase media alta, cuyos estudiantes tienen un CI medio de 120 puntos con una desviación estándar de 6; de esta manera, la varianza es de $6^2 = 36$. Las puntuaciones del CI tienen una distribución normal y podemos esperar que casi todos los estudiantes caigan a 3 desviaciones estándar de 120; de esta manera, las puntuaciones en bruto oscilarían aproximadamente de 102 a 138. El muestreo repetido de esta población dará como resultado una distribución muestral de medias con un error estándar relativamente pequeño, y la mayoría de las medias muestrales caerán cerca de 121. Supongamos que un investigador llamado Carl toma una muestra de esta población para probar la hipótesis de que la media es igual a 120 puntos de CI. Él obtiene una media muestral de 120.4. Esta media del grupo 1 se encuentra muy cerca de 120. Está dentro del error de muestreo esperado y, de esta manera, el investigador concluye correctamente que la media de la población es de 120.

La segunda muestra corresponde a una preparatoria suburbana con un programa especial, la cual también tiene una media de CI de 120, aunque esta población cuenta con mayor diversidad, las puntuaciones se encuentran más dispersas. Esta tiene una desviación estándar de 15 puntos de CI con un rango de 75 a 165 de esta manera, su varianza es de $15^2 = 225$, una varianza seis veces más grande que la de la otra preparatoria (es decir, la razón de 225 a 36 es de 6.25). Si llevamos a cabo un *muestreo repetido* de esta población, las medias en esta distribución muestral también se centrarán alrededor de 120, pero las observaciones tendrán una dispersión mayor. Dado que las medias muestrales son sensibles a las puntuaciones extremas, a menudo resultarán más allá de 120. El error estándar de esta distribución muestral será grande. Supongamos que otra investigadora, Carolyn, muestrea esta población para probar la hipótesis de que la media del CI de esta escuela es de 120. Obtiene una media muestral de 118 puntos de CI. Ya que la varianza y el error estándar en el grupo 2 de la población son muy grandes, resulta que esta media muestral, aunque se encuentra a dos puntos de 120, no es inusual en el muestreo repetido, dada la amplia dispersión de las puntuaciones; por consiguiente, Carolyn concluye correctamente que su población, así como la de Carl, posee una media de 120 puntos de CI.

Supongamos ahora que Carl se entera del ejercicio de Carolyn y le pregunta sobre sus resultados. Ella sólo menciona que encontró una media de 118. Carl no le pregunta acerca del tamaño de la varianza y asume equivocadamente es igual a la de la población que él utilizó. Compara sus resultados y encuentra una diferencia entre las medias de 2.4 puntos de CI:

$$\bar{X}_1 - \bar{X}_2 = 120.4 - 118 = 2.4 \text{ puntos de CI}$$

Esta diferencia le parece muy grande a Carl, ya que el resultado de Carolyn parece muy alejado de 120 si se utiliza el error estándar pequeño de Carl. Éste concluye que la media poblacional de Carolyn es diferente de 120 y, por tanto, significativamente diferente de su media muestral única de 120.4. Desafortunadamente, Carl esté llegando a una conclusión equivocada. Las dos pruebas realizadas independientemente determinaron de forma correcta que no eran diferentes de 120 y, por tanto, no eran diferentes entre sí. Sin embargo, al comparar las dos medias, Carl malinterpretó que la dispersión grande en la población y la distribución muestral de Carolyn revelaban una diferencia significativa entre las medias. En realidad, la gran diferencia en las varianzas fue la responsable de la dispersión de 2.4 puntos de CI entre las dos medias muestrales. Si Carl llevara a cabo una prueba de diferencia de medias utilizando la estimación del error estándar con varianzas agrupadas y supusiera equivocadamente que las varianzas son iguales, concluiría erróneamente que la media de su muestra es significativamente diferente de la de Carolyn. Incluso cuando dos poblaciones tienen las mismas medias y las varianzas de los grupos difieren, las diferencias grandes *en las medias muestrales observadas* pueden ser consecuencia de una diferencia grande en las varianzas. Esto puede conducir a interpretaciones equivocadas.

Conscientes de este posible escollo, cuando una varianza poblacional resulta mucho más grande que la de la otra, llevamos a cabo ajustes en los cálculos del error estándar, así como en los grados de libertad de la prueba t. El error estándar para varianzas desiguales recibe el nombre de *estimación* del error estándar *con varianzas separadas* de la diferencia entre medias. El símbolo del subíndice en esta fórmula dice *separadas* con el fin de indicar que estamos conscientes de que la suposición de varianzas iguales no se cumple. Excepto por la anotación relativa al hecho de que se utiliza la estimación de varianzas separadas, la fórmula de la prueba t *parecerá* la misma a la que se utiliza en la estimación del error estándar con varias agrupadas.

Cálculo del error estándar de las diferencias entre dos medias (estimación separada, utilizada cuando las varianzas de la población parecen diferentes)

$$s_{\bar{X}_1 - \bar{X}_2\,(separadas)} = \sqrt{\frac{s_{X_1}^2}{n_1 - 1} + \frac{s_{X_2}^2}{n_2 - 1}}$$

con

$$gl_{(separadas)} = \frac{\left(\dfrac{s_{X_1}^2}{n_1 - 1} + \dfrac{s_{X_2}^2}{n_2 - 1}\right)^2}{\left(\dfrac{s_{X_1}^2}{n_1 - 1}\right)^2\left(\dfrac{1}{n_1 + 1}\right) + \left(\dfrac{s_{X_2}^2}{n_2 - 1}\right)^2\left(\dfrac{1}{n_2 + 1}\right)} - 2$$

en la que

$s_{\bar{X}_1 - \bar{X}_2(separadas)}$ = estimación con varianzas separadas del error estándar de la diferencia entre dos medias

n_1 = tamaño de la muestra del grupo 1

n_2 = tamaño de la muestra del grupo 2

$s_{X_1}^2$ = varianza del grupo 1

$s_{X_2}^2$ = varianza del grupo 2

Una determinación verdadera del hecho de que las varianzas poblacionales son iguales requiere la comprobación de una hipótesis separada, inclusive antes de calcular la prueba t de la diferencia entre las medias. Esta prueba adicional, junto con el cálculo complicado de los grados de libertad indicado antes, complica las cosas al grado que resulta más razonable confiar en una computadora. El procedimiento de la prueba t del programa que se incluye en este texto está diseñado para probar la suposición de la igualdad de las varianzas antes de correr la prueba t relacionada con la diferencia de medias. Resulta aconsejable la utilización de una computadora cuando las varianzas parecen diferentes. Sin embargo, entender los principios que se encuentran detrás de la suposición de la igualdad de varianzas es un ejercicio importante en lo que se refiere al enfoque proporcional.

Prueba de la diferencia de medias para dos grupos con muestras no independientes o relacionadas

A veces se requiere comparar los estadísticos de los conjuntos de puntuaciones de los mismos individuos. Por ejemplo, queremos comparar una muestra de individuos que han completado dos escalas de medición que se califican de la misma manera. Supongamos que tenemos escalas para medir la felicidad en el matrimonio y la satisfacción con la vida en general en una muestra de mujeres recién casadas. Las dos escalas van de 1 a 100; las puntuaciones altas indican un alto grado de felicidad matrimonial o de satisfacción con la vida. Tenemos la intención de determinar si existe una diferencia significativa entre las puntuaciones de las dos escalas para cada individuo de la muestra. Tenemos dos "grupos" o conjuntos de puntuaciones, pero no tenemos dos grupos de individuos. Además las puntuaciones de las dos escalas no son independientes entre sí. Es posible que la puntuación que obtenga un individuo en lo que se refiere a felicidad en el matrimonio resulte un buen predictor de la puntuación de dicho individuo en lo que se refiere a la satisfacción con la vida. De esta manera, nos referimos a estos conjuntos de puntuaciones como *no independientes* uno del otro. A esta muestra le damos el nombre de *muestras relacionadas, apareadas o no independientes.*

Un diseño común de muestra relacionada es el diseño experimental de *antes-después* o *test-retest*, en el cual se mide una variable dos veces en el caso de los mismos individuos con algún tipo de intervención entre las pruebas. Por ejemplo, supongamos que una compañía se encuentra interesada en saber si sus centros de trabajo son "ambientes hostiles" para las trabajadoras como consecuencia de insinuaciones de naturaleza sexual, comentarios obscenos y bromas "sucias" por parte de los trabajadores. Para crear un ambiente más cómodo para las empleadas y evitar demandas por acoso sexual, la empresa establece un programa de sensibilización para los trabajadores con el fin de instruir a los varones en lo que atañe al problema. El programa tiene una duración de seis meses y requiere que todos los empleados asistan a sesiones periódicas de discusión, en las que las empleadas relatan experiencias negativas relacionadas con la discriminación de género. El objetivo general de la capacitación consiste en enseñar a hombres y mujeres que el acoso sexual tiene que ver más con el poder y la dominación masculina que con la sexualidad en sí misma. Si el programa da resultados, la capacitación de sensibilización con respecto al género debería influir positivamente en las actitudes de los hombres, sensibilizándolos respecto a los daños que los comentarios de naturaleza sexual causan en sus compañeras.

Para evaluar la eficacia del programa de capacitación seleccionamos una muestra aleatoria de 15 empleados del género masculino. Utilizamos una escala de sensibilidad con respecto al género que consta de 20 preguntas, y se les solicita que contesten cada pregunta en una escala que va desde estar muy en desacuerdo, hasta estar muy de acuerdo. La escala tiene un nivel de medición de intervalo y las puntuaciones varían de cero a 100; las puntuaciones más altas indican una alta sensibilidad. La escala se aplica a los 15 empleados el *día 1* del programa de capacitación, antes de que se celebre cualquier sesión de sensibilización. Estas puntuaciones *antes del tratamiento* representan los datos de la línea base. Después de un mes de sesiones de capacitación, se aplica de nuevo la escala de sensibilización con respecto al género a estos 15 varones. Esta medición de reaplicación representa la puntuación *después del tratamiento*. Enseguida restamos las puntuaciones antes del tratamiento a las que se obtuvieron después del tratamiento para medir la mejoría en la línea base. Si resulta una mejora significativa, esto sugiere que la capacitación de sensibilización tiene un efecto positivo en las actitudes de los hombres.

Ya que sólo observamos a 15 empleados de una población más grande, cualquier diferencia que obtengamos entre las medidas antes y después del tratamiento tal vez sólo sea resultado del error de muestreo. De hecho, si la capacitación no contribuyó en absoluto a mejorar las actitudes, la segunda medición será equivalente a extraer una segunda muestra de la población. La diferencia de las puntuaciones antes y después del tratamiento será cero. Sin embargo, pueden esperarse puntuaciones un tanto diferentes en el muestreo repetido. De esta manera, debemos probar una hipótesis para determinar si cualquier diferencia entre las mediciones se debe sencillamente al error de muestreo.

Podríamos sentirnos tentados a llevar a cabo una prueba *t* para grupos *independientes* con el fin de evaluar la diferencia entre las medias de las puntuaciones de sensibilidad con respecto al género antes y después del tratamiento. No obstante, no contamos con 30 individuos, sólo con 15; por consiguiente, no tenemos 28 grados de libertad. Los individuos de ambos grupos no son independientes entre sí, se trata de los mismos individuos. Las puntuaciones que obtengan en la segunda prueba seguramente dependen de las que obtengan

en la primera. Es decir que si un hombre obtiene una puntuación alta en la primera prueba, es posible que también obtenga una puntuación alta en la segunda. Lo que en realidad nos interesa es saber si cada puntuación ha cambiado y si la diferencia entre las puntuaciones antes y después del tratamiento se orienta en sentido positivo (es decir si los individuos se han vuelto más sensibles, no menos). De esta manera reconoceremos que el tamaño real de la muestra es de 15 y consideraremos a cada individuo como un solo caso con dos puntuaciones de prueba relacionadas.

Para establecer una hipótesis nula —una que nos permita predecir resultados muestrales cuando ésta es verdadera— nos concentramos en la diferencia entre las puntuaciones de sensibilidad antes y después del entrenamiento. Nuestra hipótesis nula consiste en que el entrenamiento de sensibilización con respecto al género *no* tiene efecto. Si la hipótesis nula es verdadera, podemos predecir que no hay cambios en las puntuaciones y esperar que la media de las diferencias entre las puntuaciones antes y después sea cero. En este caso, D representa la diferencia entre las puntuaciones antes y después del tratamiento; nuestra hipótesis nula consiste en que la media de las diferencias sea cero:

$$H_0: \mu_D = 0$$

Es decir que en la población de trabajadores del género masculino no existe diferencia en la media de las puntuaciones de sensibilidad en las mediciones antes y después del tratamiento. La capacitación no tiene ningún efecto.

Observa que esta afirmación es como una prueba de medias con una sola muestra por el hecho de que tiene un valor objetivo de cero. En realidad, se trata de una prueba de medias con una sola muestra, aunque implica dos *grupos* de puntuaciones. Es una prueba relacionada con el hecho de si la *media de las diferencias* entre las puntuaciones relacionadas es significativamente diferente de cero.

La tabla 11.2 muestra la hoja de cálculo de computadora para calcular la media de las diferencias entre las puntuaciones relacionadas. En este ejemplo de 15 casos, la prueba es, en esencia, una prueba de medias para una sola muestra (como se ilustra del capítulo 10).

Los criterios para aplicar esta prueba son los siguientes:

Cuándo aplicar una prueba de diferencia de medias para dos grupos (prueba *t*) con muestras no independientes o relacionadas (distribución *t*, $gl = n - 1$)

En general: se aplica para comprobar la hipótesis de que las puntuaciones de una variable de intervalo/razón difieren dos puntos en el tiempo en el caso de los mismos individuos.

1. Hay una población con una muestra representativa de ella.
2. Existen dos variables de intervalo/razón con el mismo diseño en sus puntuaciones, o una única variable medida dos veces en los mismos individuos pertenecientes a la muestra.
3. Hay un valor objetivo de la variable con el cual podemos comparar la media de las diferencias entre los dos conjuntos de puntuaciones (normalmente este valor objetivo será cero para una prueba de no diferencias entre las dos puntuaciones).

TABLA 11.2 | Hoja de cálculo de computadora a calcular la media y la desviación estándar de las diferencias entre las puntuaciones relacionadas

X = puntuación en la escala de sensibilidad con respecto al género

Especificaciones			Cálculos		
Número del individuo	(A) Puntuación antes del entrenamiento	(B) Puntuación después del entrenamiento	D (diferencia = B − A)	$(D-\bar{D})$	$(D-\bar{D})^2$
1	47	53	6	.73	.53
2	39	38	−1	−6.27	39.31
3	52	54	2	−3.27	10.69
4	48	56	8	2.73	7.45
5	45	49	4	−1.27	1.61
6	42	51	9	3.73	13.91
7	48	54	6	.73	.53
8	50	56	6	.73	.53
9	45	54	9	3.73	13.91
10	44	51	7	1.73	2.99
11	46	44	−2	−7.27	52.85
12	45	54	9	3.73	13.91
13	43	53	10	4.73	22.37
14	47	55	8	2.73	7.45
15	41	39	−2	−7.27	52.85
$n = 15$				$\Sigma(D-\bar{D}) = 0.05$*	
			$\Sigma D = 79$	$\Sigma(D=\bar{D})^2 = 240.89$	

* Diferente de cero como consecuencia del error de redondeo.

El cálculo de la media y de la desviación estándar de las diferencias entre puntuaciones procede de la siguiente manera:

$$\bar{D} = \frac{\Sigma D}{n} = \frac{79}{15} = 5.27 \text{ puntos de la escala de sensibilidad}$$

$$s_D = \sqrt{\frac{\Sigma(D-\bar{D})^2}{n-1}} = \sqrt{\frac{240.89}{14}} = 4.15 \text{ puntos de la escala de sensibilidad}$$

El estadístico de la prueba es el mismo que aquel para una prueba *t* con una muestra, excepto por el hecho de que los símbolos corresponden al cálculo de las diferencias (D). En el siguiente cuadro, observa que los cálculos del error estándar y del estadístico de la prueba *t* se asemejan a las fórmulas aplicadas en el caso de una prueba de medias con una sola muestra (capítulo 10).

Cálculo del error estándar de las diferencias entre puntuaciones relacionadas

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}}$$

donde,

$s_{\bar{D}}$ = error estándar de las diferencias entre puntuaciones relacionadas

s_D = desviación estándar de las diferencias entre puntuaciones relacionadas

n = tamaño de la muestra

El efecto de esta prueba, como el de cualquier prueba de hipótesis, consiste en la diferencia entre la observación de la muestra en cuestión y el parámetro hipotético, en este caso $\bar{D} - \mu_D$. Como la diferencia con respecto al parámetro hipotético es cero, este efecto de prueba se reduce a $\bar{D}$, y esta es la forma en que el efecto aparece en el numerador del estadístico de prueba.

Cálculo de la prueba *t* de la diferencia entre puntuaciones relacionadas

$$t_{\bar{D}} = \frac{\bar{D}}{s_{\bar{D}}}$$

$$gl = n - 1$$

donde,

$t_{\bar{D}}$ = número de errores estándar que una diferencia media muestral entre las puntuaciones relacionadas se desvía de la diferencia media hipotética de cero

$\bar{D}$ = media de las diferencias entre las puntuaciones relacionadas de la muestra

$s_{\bar{D}}$ = error estándar de las diferencias entre puntuaciones relacionadas

gl = grados de libertad

n = tamaño de la muestra = número de puntuaciones relacionadas

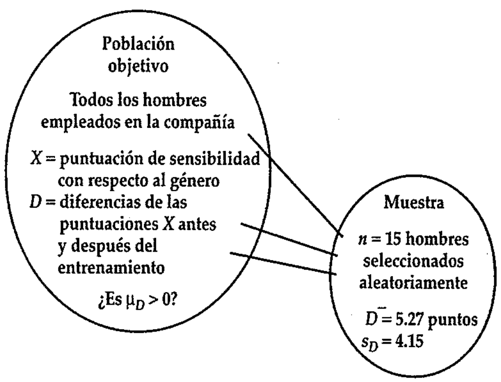
Utilicemos los datos y los cálculos de la tabla 11.2 y sigamos los seis pasos de la inferencia estadística.

Seis pasos de la inferencia estadística para la prueba de la diferencia de medias para dos grupos con muestras no independientes o relacionadas

Solución para la prueba de diferencia de medias de dos grupos (prueba *t*) para muestras no independientes o relacionadas

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿mejoran significativamente las puntuaciones de sensibilidad con respecto al género después de las sesiones de entrenamiento de sensibilización? ¿Resultaron eficaces las sesiones? ¿Debería establecerse el programar en el caso de los empleados varones? *Procedimiento estadístico:* prueba *t* de la diferencia entre puntuaciones relacionadas. *Especificaciones:*



SEIS PASOS

1. H_0: $\mu_{D(\text{todos los empleados varones reciben el entrenamiento})} = 0$ (es decir, la diferencia en las puntuaciones y el entrenamiento no resulta efectivo)

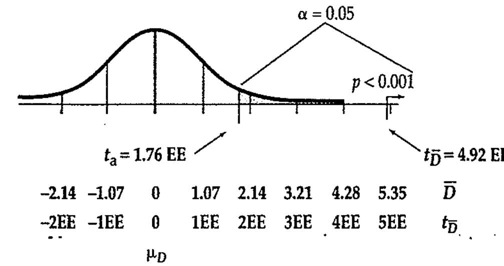
   Esto es, el entrenamiento de sensibilización con respecto al género no resulta eficaz; no hay mejora en las puntuaciones de sensibilidad con respecto al género después del entrenamiento.

   H_A: $\mu_{0(\text{todos los empleados varones que reciben el entrenamiento})} = 0$. De una cola

   Esto es, el entrenamiento de sensibilización con respecto al género resulta eficaz; el entrenamiento de género contribuye a mejorar las puntuaciones de sensibilidad con respecto al género.

2. *Distribución muestral:* si H_0 resulta verdadera y se extraen repetidamente muestras de tamaño 15 de la población de empleados varones que participan en el entrenamiento, las medias muestrales de las diferencias entre las puntuaciones apareadas ($\bar{D}$) se centran en torno a cero como una distribución *t* aproximadamente normal, con $gl = n - 1 = 14$ y un error estándar:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{4.15}{\sqrt{15}} = 1.07 \text{ puntos de la escala de sensibilidad}$$



3. *Nivel de significancia:* $\alpha = 0.05$. De una cola. Puntuación crítica de la prueba = $t_\alpha = 1.76$. (Región crítica marcada en la curva.)
4. *Observación:*

   Efecto de la prueba: $\bar{D} - \mu_D = 5.27 - 0 = 5.27$ de la escala de sensibilidad

   Estadístico de la prueba $= t_{\bar{D}} = \dfrac{\bar{D}}{s_{\bar{D}}} = \dfrac{5.27}{1.07} = 4.92$ EE

   valor *p*: *p* [de extraer una muestra, la diferencia media entre las puntuaciones ($\bar{D}$) tan inusual o más inusual que 5.27 puntos cuando la verdadera media de las diferencias de la población (μ_D)es 0] < 0.001. (Este valor *p* se señala en la curva del paso 2.) .
5. Decisión de rechazo: $|t_{\bar{D}}| > |t_\alpha|$ (es decir, $4.92 > 1.76$); por consiguiente, $p < \alpha$ (es decir que $p < 0.05$). Recháces H_0 y acéptese H_A al nivel de confianza de 95%.
6. *Interpretación:* el entrenamiento de sensibilización con respecto al género parece ser efectivo. *Mejor estimación del efecto:* las sesiones de entrenamiento de sensibilización respecto al género dan como resultado una mejora promedio de 5.27 puntos en las puntuaciones de sensibilidad con respecto al género. *Respuesta:* el programa debe establecerse para todos los empleados varones.

Significancia práctica frente a significancia estadística

El problema relacionado con el entrenamiento de sensibilización con respecto al género descrito anteriormente destaca un punto importante relativo a la comprobación de hipótesis estadística. Esta prueba de hipótesis se enfoca en el hecho de si se debe atribuir una diferencia observada en los datos muestrales al error de muestreo o a un verdadero efecto del entrenamiento en la población. En este ejemplo concluimos —con un 95% de confianza— que la diferencia de 5.27 en la escala de sensibilidad respecto al género entre las mediciones antes y después de tratamiento fue tan grande que es.posible que no sea resultado del muestreo esperado. Ahora bien, ¿resulta significativa esta diferencia de 5.27 puntos en términos prácticos? La escala de sensibilidad respecto al género tenía posible rango en las puntuaciones de 100. En la línea base —antes del entrenamiento de sensibilización—, la puntuación media de la escala fue de 45.47. Los hombres sólo parecían moderadamente sensibles en ese momento.

Después de tratamiento, la puntuación media fue de 50.73 puntos. Los hombres aún parecían moderadamente sensibles. Es más, con un cuestionario de 20 preguntas, un individuo de obtener la diferencia promedio de 5 subiendo el nivel de acuerdo en sólo 5 de las 20 preguntas. ¿Es suficiente esto para conseguir una diferencia en el comportamiento de los hombres en la oficina? Tal vez no. Este incremento de 5 en la sensibilidad quizá no tenga influencia del comportamiento. Es necesario investigar más, quizás analizar los reportes de acoso y los hechos relacionados con la percepción de las mujeres. *Una prueba de hipótesis determina la significancia en términos del error de muestreo probable. Nos dice sencillamente si la diferencia de muestras es tan grande que probablemente exista una diferencia en las poblaciones.* Este hecho no garantiza el hecho de que la diferencia signifique algo en términos prácticos. Cualquier estadístico debe analizarse en función de ideas teóricas y circunstancias prácticas. La significancia práctica, la significancia teórica (es decir, si los resultados respaldan una teoría), así como la significancia estadística constituyen temas independientes.

Los cuatro aspectos de las relaciones estadísticas

El fin último de la investigación científica consiste en expresar declaraciones empíricamente probadas que expliquen un fenómeno para mejorar nuestra comprensión de la relación de éste con otros fenómenos. Estas declaraciones adquieren la forma de teoría, la cual describe las relaciones entre las variables medidas. Una teoría, así como su lista de hipótesis, se prueba llevando a cabo predicciones que afirmen que las mediciones realizadas en una variable se relacionan de alguna forma con las mediciones realizadas en otras. Como lo indicamos antes, estas relaciones pueden adoptar la forma de medias más altas que se relacionan estadísticamente con un grupo o con otro, con una alta frecuencia de ocurrencias conjuntas para dos variables nominales o con una correlación entre dos variables de intervalo o de razón. Una vez que concluimos que las dos variables se encuentran relacionadas, podemos hacer más afirmaciones sobre la relación entre ellas. Un análisis exhaustivo de descubrimientos estadísticos toma en cuenta cuatro *aspectos de una relación* entre variables. Estos cuatro aspectos deben analizarse en el último paso de los seis de la inferencia estadística.

Existencia de una relación

El primer aspecto de la relación estadística es la *existencia*. La existencia de una relación responde a la pregunta: sobre la base del análisis estadístico, ¿podemos concluir que existe una relación entre dos variables en el caso de todos los individuos *de la población*? Por ejemplo, ¿existe una relación entre la preferencia religiosa y el tiempo que se dedica a orar entre los estadounidenses? ¿Existe una relación entre los niveles de pobreza y las tasas de delitos en las ciudades de Estados Unidos? ¿Hay alguna relación entre el tiempo que pasan viendo la televisión y el promedio de calificaciones que obtienen los estudiantes universitarios? La existencia de una relación tiene que ver con la población de sujetos de estudio. Ten en cuenta que los datos de muestras y los estadísticos calculados a partir de éstas sólo proporcionan estimaciones de parámetros poblacionales. Básicamente, al determinar si existe una relación, se establece el hecho de que las conclusiones estadísticas de la muestra no son resultado del error de muestreo sencillamente. En otras palabras, la existencia de una relación se determina rechazando la hipótesis nula. Lo primero que decimos en el paso seis es si existe una relación.

Dirección de la relación

El segundo aspecto de una relación entre dos variables tiene que ver con la dirección, aunque este aspecto no se aplica a todos los análisis bivariados. La dirección de una relación tiene que ver con la pregunta: cuando la variable independiente incrementa, ¿qué ocurre con la variable dependiente, incrementa o disminuye? En este caso utilizamos los términos *positivo* y *negativo*. Una relación positiva es aquella en la que el incremento de una variable se relaciona con el incremento en la otra. Por ejemplo, existe una relación positiva entre el ingreso familiar y el cuidado dental preventivo: *a mayor* ingreso familiar, *más* cuidado dental preventivo reciben sus miembros. Una relación negativa es aquella en la cual el incremento de una variable se relaciona con el decremento en la otra. Por ejemplo, existe una relación negativa entre los niveles de ingreso en las vecindades y las tasas de delito: *a mayor* ingreso, *menor* tasa de delitos. La dirección se especifica en la hipótesis alternativa como una prueba de una cola. La dirección constituye un aspecto muy sencillo en la relación entre dos variables de intervalo o de razón, tema que se analiza con mayor detalle en el capítulo 15.

Fuerza de la relación, poder predictivo y reducción proporcional del error

El tercer aspecto de una relación estadística es la *fuerza*. La fuerza de la relación entre dos variables indica el grado en que se reducen los errores al predecir las puntuaciones de una variable dependiente. Una medición de la fuerza de una relación nos indica el **poder predictivo,** es decir, en qué grado predice la variable independiente los resultados de la variable dependiente. ¿La variable relacionada explica mucho o poco de la variable dependiente? Por ejemplo, ¿en qué grado predice el promedio de preparatoria, el promedio de la universidad? ¿Constituye un fuerte indicador una precisión de 50% de las veces, por ejemplo, o constituye un indicador débil una precisión de apenas 10% de las veces? ¿Predicen mejor el promedio de calificaciones en la universidad otras variables, como el nivel de comprensión de lectura? Prestar atención a la fuerza en la relación es de utilidad con el fin de comparar los efectos negativos de diversas variables independientes sobre una variable dependiente.

Otra forma de analizar el poder predictivo de una variable independiente en las puntuaciones una variable dependiente consiste en preguntarnos: ¿hasta qué punto se reduce del error en las puntuaciones de la variable dependiente por el hecho de conocer las puntuaciones en la variable independiente? Este enfoque se refiere a la reducción proporcional del error (RPE), y resulta de gran utilidad con una variable dependiente de intervalo o de razón, en la cual se ha calculado la media. Recordemos que la diferencia entre la media y cualquier puntuación individual en una distribución de puntuaciones corresponden a la puntuación de la desviación. Por ejemplo, sea X = salario de los gerentes de nivel medio de una empresa. La media es de \$50 000. Si alguien nos reta a adivinar el sueldo de Jacob Smith, nuestra mejor estimación sería la media de \$50 000. Ahora supongamos que averiguamos que éste gana \$60 000. Erramos por \$10 000, que corresponde a la desviación de Jacob con respecto a la media general de la empresa. ¿Cómo podemos explicar esta puntuación de desviación de \$10 000? ¿Podemos encontrar variables que nos permitan reducir el error con el fin de lograr una mejor estimación del salario? Tal vez el salario *alto* de Jacob se relacione con el hecho de que tiene mayor antigüedad en la compañía que los demás gerentes de nivel medio. Probamos la hipótesis de que el tiempo de servicio en la compañía se relaciona con el nivel del salario. Al aplicar la prueba de diferencia de medias, comparamos los salarios medios de *los empleados de mayor antigüedad* con los que han sido contratados recientemente. Supongamos que el salario promedio de los que cuentan con mayor antigüedad es de \$58 000, es decir, \$8 000 más que el promedio de la mayoría de los empleados de la empresa. Ya que Jacob tiene muchos años de antigüedad, ahora podemos llevar a cabo una mejor estimación de su salario —\$58 000—, la media de la empresa de \$50 000 más \$8 000 por el largo historial de servicio. Al estimar su sueldo en \$58 000, nos desviamos de su salario real de \$60 000 por sólo \$2 000. Hemos explicado los \$8 000 de la desviación de Jacob de \$10 000 con respecto a la media de \$50 000. Conocer la relación entre la antigüedad (variable independiente) y el salario (variable dependiente), nos permitió reducir el error de predicción un 80% (es decir, \$8 000/\$10 000).

En el caso de toda una muestra, la reducción proporcional del error depende de la determinación de las relaciones que expliquen la varianza en la muestra, el promedio de la suma de las puntuaciones de desviación al cuadrado (véase capítulo 5). En el capítulo 12 mostraremos la forma de explicar partes de la varianza de la variable dependiente, identificando variables independientes relacionadas. Preguntamos: ¿qué proporción de la varianza se explica por el hecho de conocer dicha relación? El grado al que se explica la varianza constituye una reducción proporcional en el error. Como veremos del capítulo 15, cuando la variable independiente y la variable dependiente son de intervalo o de razón, la reducción proporcional del error se calcula con precisión.

Aplicaciones prácticas de las relaciones

El cuarto aspecto de una relación entre dos variables se refiere a sus aplicaciones prácticas esto implica la descripción de la forma en que el conocimiento de una relación entre dos variables nos ayuda tanto a comprender un fenómeno como a aplicar los resultados a circunstancias prácticas. Al describir la naturaleza de una relación, evitamos el lenguaje estadístico expresando las conclusiones de la investigación en el lenguaje común, especialmente cuando las presentamos al público en general. Trasladamos nuestras conclusiones a la vida revelando su valor para mejorar la sociedad o la vida de los individuos. Respondemos a preguntas como ¿para qué? y, ahora que se ha determinado una relación estadística, ¿cuál es la utilidad de dicho conocimiento?

Un hallazgo científico resulta de particular valor si puede cambiar la vida de la gente. La naturaleza de la relación se enfoca en la forma en que puede aplicarse el conocimiento científico de tal manera que una variable prediga a la otra. En las mejores circunstancias, se demuestra que la relación es *causal.* Es decir que es posible demostrar que alterar una variable independiente da como resultado directo un cambio en la variable dependiente. La naturaleza describe con exactitud qué grado de cambio en la puntuación de la variable independiente provoca qué grado de cambio en la puntuación de la variable dependiente. Por ejemplo, en la relación entre la ingesta de aspirina y la disminución de la inflamación de las articulaciones, ¿qué cantidad de aspirina provoca qué grado de disminución de la inflamación? La causalidad se determina mejor con datos longitudinales —datos recabados a través del tiempo. Este es el caso, ya que, lógicamente, X puede dar origen a Y solamente si ocurre antes de Y. Nuestra prueba de dos grupos no independientes sobre la efectividad del entrenamiento de sensibilidad en actitudes con respecto al género constituye una situación en la que pudiéramos estudiar la causalidad. No obstante, la mayor parte de la investigación en ciencias sociales confía en datos de estudios transeccionales, datos recogidos en un punto del tiempo. Aunque las relaciones causales con frecuencia quedan confirmadas por hallazgos realizados a partir de datos transeccionales, dichas interpretaciones deben realizarse con cuidado.

En circunstancias en las que se determina la existencia de una relación estadística, pero en las que la causalidad no resulta clara, al describir la naturaleza de una relación sencillamente se presenta la esencia de la conclusión proporcionando información empírica específica. Por ejemplo, si determinamos una relación entre el género y la preferencia por películas de clasificación A, ¿qué dice esto en términos comunes? ¿Están los hombres o las mujeres más interesados en dichas películas? (Tú puedes adivinar la respuesta.) Informamos precisamente esos porcentajes muestrales de hombres y mujeres que manifiestan gran interés en películas de clasificación A. De esta manera, ofrecemos la mejor estimación disponible de la forma en que una variable dependiente puede ajustarse para los efectos de una variable independiente. El efecto de la prueba (es decir, la diferencia entre lo que se observa y lo que se plantea estadísticamente como hipótesis) por lo general constituye el componente más significativo al describir la naturaleza de la relación. Con el fin de recordar esta parte importante y práctica de la comprobación de hipótesis, en el sexto paso de la inferencia estadística no sólo formulamos un enunciado general relacionado con la existencia de la relación, sino que también proporcionamos la mejor estimación basada en cálculos de los efectos de la prueba estadística.

Los cuatro aspectos de una relación entre dos variables

Existencia: sobre la base del análisis estadístico de la muestra, ¿es posible concluir que existe una relación entre dos variables para todos los individuos de la población?

Dirección: ¿puede esperarse que la variable dependiente aumente o disminuya cuando la variable independiente aumenta?

Fuerza: ¿hasta qué punto se reducen los errores al predecir las puntuaciones de una variable dependiente cuando una variable independiente se utiliza como predictor?

Naturaleza: en términos prácticos y corrientes, ¿cómo nos permite el conocimiento una relación entre dos variables entender y predecir los resultados de la variable dependiente?

Cuándo aplicar los diversos aspectos de las relaciones

La comprobación de una hipótesis para establecer la existencia de una relación representa el primer paso en cualquier análisis. Si *no* se encuentra una relación entre dos variables, los otros tres aspectos de una relación resultan irrelevantes; es obvio que si no existe dicha relación, entonces la relación no tiene dirección, fuerza ni naturaleza. Además, cuando se determina una relación entre dos variables, puede suceder que la fuerza y la dirección no resulten ni significativas ni útiles. Conforme avancemos por los capítulos que quedan, indicaremos qué aspectos de una relación son de utilidad para cada prueba bivariada.

Aspectos relevantes de las relaciones para las pruebas de diferencia de medias para dos grupos

Con la prueba de diferencia de medias para dos grupos independientes o relacionados, solamente se aplican dos aspectos de una relación: existencia y naturaleza. La *existencia* de una relación entre una variable independiente dicotómica y una variable dependiente de intervalo o de razón se establece utilizando una de las pruebas t descritas en este capítulo. Cuando se rechaza la hipótesis nula relativa al hecho de que no existe diferencia entre las medias, podemos concluir que existe una relación.

Existencia de una relación utilizando una prueba de diferencia de medias para dos grupos

Pruébese la hipótesis nula que indica que $\mu_{X_1} = \mu_{X_1}$ (en la que *X* representa la variable de intervalo o de razón, y 1 y 2 designan grupos, respectivamente). Es decir que no existe diferencia entre las medias de las dos poblaciones. Utilícese el estadístico de prueba:

Grupos independientes	Datos relacionados
$t_{\bar{X}_1-\bar{X}_2} = \dfrac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$	$t_{\bar{D}} = \dfrac{\bar{D}}{s_{\bar{D}}}$

Cuando se rechaza la hipótesis nula, se afirma que existe una relación.

Con la prueba *t* de grupos independientes, la *dirección* de una relación no es relevante. Por ejemplo, si se determina una diferencia entre los ingresos medios de hombres y mujeres en una empresa, no diríamos que un incremento en la masculinidad se relaciona con un aumento del ingreso, ya que una persona es hombre o mujer (dicho esto, debemos señalar que algunos investigadores utilizan la frase "en la dirección de los hombres" para indicar que los hombres perciben un ingreso mayor). Sin embargo, en una prueba de grupos relacionados, quizás afirmemos que observamos un efecto de tratamiento que resultó positivo en la dirección de mejora. Esto es lo que hicimos en el ejemplo anterior relacionado con la sensibilidad respecto al género, empleando una prueba de una cola.

La *fuerza* de la relación para la prueba *t* tampoco se aplica. Es posible calcularla, aunque la simplicidad de la prueba de diferencia de medias para dos grupos permite que nos molestemos en calcular.

Dirección y fuerza de una relación utilizando pruebas de diferencias de medias para dos grupos

No es aplicable.

En el caso de cualquier prueba estadística, la naturaleza de la relación depende de nuestra descripción del efecto de la prueba. Normalmente el efecto de la prueba se encuentra en el numerador del estadístico de la prueba. En la prueba de diferencia de medias para dos grupos, el efecto consiste en la diferencia entre las medias de las muestras; es decir que el efecto de la variable independiente sobre la variable dependiente = $\bar{X}_1 - \bar{X}_2$. Cuando se determina que existe una relación, esta diferencia entre las medias de las muestras se considera como una estimación de la diferencia entre medias poblacionales. Con frecuencia esta cantidad recibe el nombre de efecto de la pertenencia a un grupo. Por ejemplo, supongamos que probamos la hipótesis nula relativa al hecho de que los pesos medios de hombres y mujeres son iguales. Hallamos una diferencia de 25 libras en los pesos medios y rechazamos la hipótesis nula. De esta manera concluimos que el efecto de ser hombre sobre el peso corporal es de 25 libras: en promedio los hombres tienden a pesar 25 libras más que las mujeres.

Aplicaciones prácticas de una relación utilizando pruebas de diferencia de medias para dos grupos

Describe el efecto de la prueba en términos comunes; el efecto de la variable independiente sobre la variable dependiente es el siguiente:

Grupos independientes	Datos relacionados
$\bar{X}_1 - \bar{X}_2$	$\bar{D}$

Para resumir, los aspectos de una relación se reportan por lo común en el caso de la prueba de diferencia de medias para dos grupos: existencia y naturaleza. Recordemos que si la hipótesis nula que se refiere al hecho de que no existe diferencia entre las medias no se rechaza, no existe una relación de las dos variables. Cuando sucede esto, la naturaleza de la relación es irrelevante al permitir que la hipótesis nula se mantenga, declaramos que el efecto de la prueba no es real en las poblaciones y que la diferencia observada entre las medias de las muestras sencillamente se debe al error de muestreo.

Insensatez y falacias estadísticas: fijar la atención en las diferencias de las medias mientras se ignoran las diferencias en las varianzas

Las pruebas de diferencia de medias se utilizan con frecuencia. Incluso en el análisis multivariado, las pruebas bivariadas, como las que se presentaron en este capítulo, representan el primer paso para entender plenamente la naturaleza de los datos. Desafortunadamente el énfasis en la búsqueda de diferencias entre medias puede dar origen a que se pase por alto el interés en las dispersiones de las puntuaciones entre dos grupos. Como vimos en este capítulo, es importante buscar diferencias en la dispersión, ya que las varianzas diferentes en la comparación de grupos provocan un error de muestreo adicional y se requieren ajustes al describir la distribución muestral. No obstante, además de esta necesidad estadística, se puede aprender mucho en relación con la naturaleza de dos grupos cuando se determina que sus varianzas son diferentes.

Supongamos que comparamos los niveles educativos de hombres y mujeres sin hogar y no encontramos diferencias en la media de años de educación. Sin embargo, supongamos que encontramos una diferencia significativa en las varianzas, según las cuales las mujeres exhiben mayor dispersión en sus niveles de educación. Como en el caso de los hombres, muchas de las mujeres tienen entre 6 y 14 años de instrucción escolar, pero entre ellas abundan más las que tienen escasa educación (menos de 6 años), aunque también hay más con educación universitaria (16 años o más). Dicha diferencia en las dispersiones de las distribuciones resulta bastante reveladora. ¿Por qué hay más individuos con poca educación entre las mujeres sin hogar? Esto se debe a que muchas de ellas provienen de familias que dependían de la asistencia familiar del gobierno. En su lucha por la subsistencia económica, abandonaron la escuela muy temprano por trabajos mal pagados o como consecuencia de un embarazo u otras circunstancias. Esto coloca a las mujeres en grave riesgo de perder su hogar. Los hombres que crecen en dichas circunstancias tiene otras oportunidades (como trabajar en la construcción), económicamente más lucrativas, que los alejan de la pérdida de sus hogares. Ahora bien, ¿por qué hay más mujeres que hombres sin hogar y con mayor educación? Porque muchas de estas mujeres son víctimas del abuso marital y este fenómeno no es raro, inclusive entre mujeres con educación universitaria. Puede que haya otras razones que expliquen la diferencia en las varianzas de los niveles educativos de hombres y mujeres sin hogar. Si un investigador centra demasiado su atención en observar e interpretar exclusivamente las diferencias entre las medias, quizás pierda la oportunidad de entender mejor la verdadera naturaleza de la relación entre dos variables. Prestar demasiada atención a las diferencias entre varianzas, así como a la diferencia entre medias, constituye una vía importante para el próximo nivel investigación.

RESUMEN

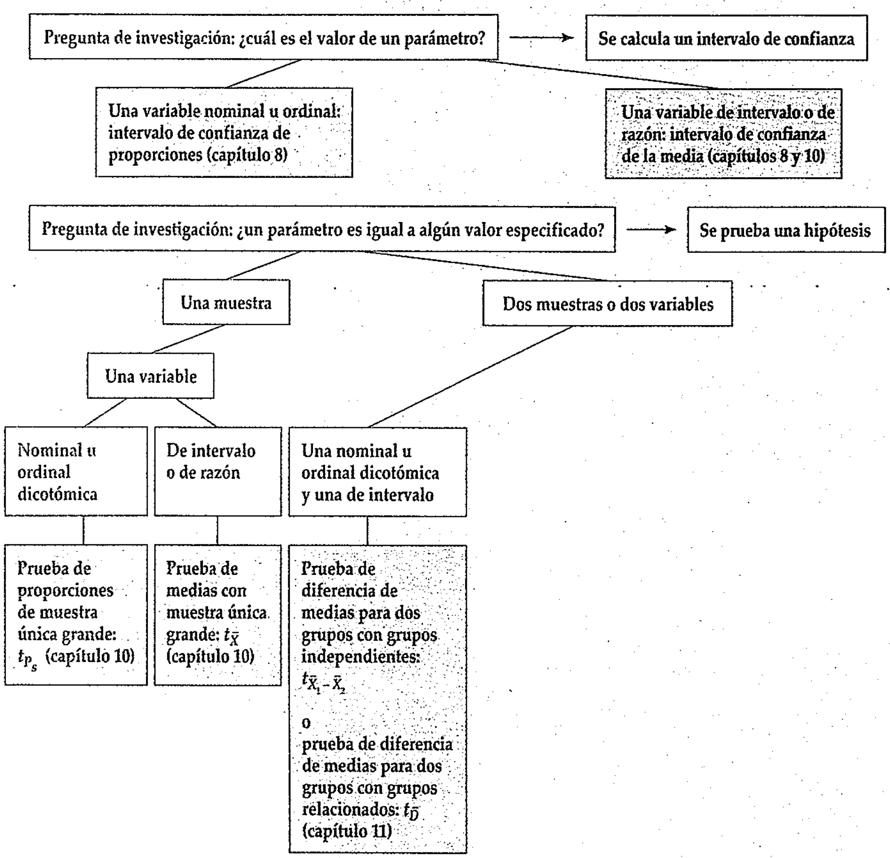
1. El análisis bivariado (dos variables) implica la búsqueda de relaciones estadísticas entre dos variables.
2. Una relación estadística entre dos variables afirma que las mediciones de una variable tienden a fluctuar en congruencia con las mediciones de la otra, lo cual convierte a una de las variables en un buen predictor de la otra. La variable predictora es la variable independiente y la variable predicha es la variable dependiente.
3. Existen tres enfoques comunes para medir relaciones estadísticas: *a)* la prueba de diferencia de medias (se comparan medias de una variable de intervalo o de razón entre las categorías o grupos de una variable nominal u ordinal); *b)* el conteo de frecuencias de ocurrencias conjuntas de atributos de dos variables nominales; *c)* la medición de la correlación entre dos variables de intervalo o de razón.
4. Una prueba de diferencia de medias de dos grupos para muestras independientes (prueba *t*) se aplica para probar la hipótesis que indica que las medias de una variable difieren entre dos poblaciones o entre dos categorías de una variable nominal u ordinal. Los dos grupos son independientes entre sí; es decir que no consisten en los mismos individuos.
5. La distribución muestral para una prueba de diferencia de medias de dos grupos es la aproximadamente normal distribución *t*. El cálculo del error estándar depende de que sea posible suponer que las varianzas poblacionales son iguales. Esto recibe el nombre de *suposición de igualdad de las varianzas.*

6. Cuando una varianza muestral no es mayor que el doble del tamaño de la otra, esto sugiere que las dos varianzas poblacionales son iguales, y asumimos la igualdad de las varianzas. La igualdad de varianzas también recibe el nombre de *homogeneidad de varianzas* u *homoscedasticidad.* Cuando se asumen varianzas iguales, se utiliza una estimación con varianzas agrupadas del error estándar.
7. La heterogeneidad de las varianzas, o heteroscedasticidad, se presenta cuando las varianzas parecen ser diferentes. Cuando esto sucede, se requieren ajustes en el error estándar y los grados de libertad. Esto recibe el nombre de *estimación del error estándar con varianza separadas.*
8. La prueba de diferencia de medias de dos grupos para muestras no independientes o relacionadas se utiliza para probar una diferencia de medias entre dos conjuntos de puntuaciones de los mismos sujetos de investigación, como dos preguntas de un cuestionario o puntuaciones medidas en dos instantes de tiempo. La distribución muestral es la aproximadamente normal distribución *t*.
9. Es importante distinguir entre significancia práctica y estadística. Una prueba de hipótesis determina la significancia estadística en términos de un probable error de muestreo. La significancia práctica tiene que ver con el hecho de que un hallazgo estadísticamente significativo en realidad signifique algo en las aplicaciones de los resultados a la realidad.
10. Existen cuatro aspectos de las relaciones estadísticas: *a) existencia:* sobre la base del análisis estadístico de una muestra, ¿podemos concluir que existe una relación entre dos variables entre todos los individuos en la población?; *b) dirección:* ¿puede esperarse que la variable dependiente aumente o disminuya conforme la variable independiente incrementa? *c) fuerza:* ¿en qué grado se reducen los errores al predecir las puntuaciones de una variable dependiente cuando una variable independiente se utiliza como variable predictora?; *d) naturaleza* (de los resultados): en términos cotidianos y prácticos, ¿cómo nos ayuda el conocimiento de una relación entre dos variables a entender y predecir resultados de la variable dependiente?
11. Solamente dos aspectos de una relación se aplican a una prueba de diferencia de medias de dos grupos, la existencia y la naturaleza.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 11 del material del texto disponible en del sitio web *The Statistical Imagination* en www.mhhe.com/rithcey2 incluyen un análisis adicional sobre la razón por la que se requieren modificaciones en una prueba *t* cuando las varianzas de los dos grupos son diferentes.

PROCEDIMIENTOS ESTADÍSTICOS CUBIERTOS HASTA AQUÍ



FÓRMULAS PARA EL CAPÍTULO 11

Prueba de diferencia de medias para dos grupos (prueba t) con grupos independientes:

Especificación: una variable dependiente X de intervalo/razón comparada para dos grupos que constan de diferentes individuos (es decir, grupos independientes) que se obtienen de: 1) una variable dicotómica nominal u ordinal de una muestra y una población; 2) dos poblaciones y muestras.

Pregunta de investigación: ¿son diferentes las medias de X en las poblaciones de los dos grupos?

$$H_0: \mu_{X_1} = \mu_{X_2} \text{ (es decir, } \mu_{X_1} - \mu_{X_2} = 0)$$

Distribución muestral: distribución t; error estándar estimado de una de dos formas dependiendo de que las varianzas de las dos poblaciones parezcan iguales.

Error estándar: si las varianzas de las dos poblaciones parecen iguales, utiliza la estimación del error estándar con varianzas agrupadas:

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{(n_1-1)s_{X_1}^2 + (n_2-1)s_{X_2}^2}{n_1+n_2-2}}\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

con

$$gl = n_1 + n_2 - 2$$

Si las varianzas de las dos poblaciones parecen diferentes (es decir que la varianza de un grupo es el doble de la del otro), utiliza la estimación del error estándar con varianza separada:

$$s_{\bar{X}_1-\bar{X}_2 (separada)} = \sqrt{\frac{s_{X_1}^2}{n_1-1} + \frac{s_{X_2}^2}{n_2-1}}$$

$$gl_{(separada)} = \frac{\left(\frac{s_{X_1}^2}{n_1-1} + \frac{s_{X_2}^2}{n_2-1}\right)^2}{\left(\frac{s_{X_1}^2}{n_1-1}\right)^2\left(\frac{1}{n_1+1}\right) + \left(\frac{s_{X_2}^2}{n_2-1}\right)^2\left(\frac{1}{n_2+1}\right)} - 2$$

Efecto de la prueba: $\bar{X}_1 - \bar{X}_2$

Estadístico de la prueba [para utilizarse con la tabla de la distribución t aproximadamente normal (tabla estadística C, apéndice B)] determinar si existe una relación:

$$t_{\bar{X}_1-\bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1-\bar{X}_2}}$$

Determinación de los aspectos de una relación:

Dirección: *normalmente no es aplicable*

Fuerza: no es aplicable

Naturaleza: especifica la diferencia entre medias de grupos:

$$\bar{X}_1 - \bar{X}_2$$

Prueba de la diferencia de medias para dos grupos (prueba t) con muestras relacionadas:

Especificaciones: 1) dos variables de intervalo/razón con el mismo tipo de puntuación medidas en los mismos individuos o 2) una sola variable de intervalo o de razón medida dos veces en los mismos individuos de la muestra (es decir, grupos no independientes).

Pregunta de investigación: ¿son diferentes las medias de X en el caso de las dos variables o ambas mediciones?

H_0: $\mu_D = 0$

Distribución muestral: distribución t de la distribución de diferencias de medias ($\bar{D}$, con $gl = n - 1$)

Error estándar:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}}$$

Efecto de la prueba: $\bar{D}$ (es decir, $\bar{D} - 0 = \bar{D}$)

Estadístico de la prueba [para utilizarse con la tabla de la distribución t aproximadamente normal (tabla estadística C, apéndice B)]: determina si existe una relación:

$$t_{\bar{D}} = \frac{\bar{D}}{s_{\bar{D}}}$$

$$gl = n - 1$$

Determinación de los aspectos de una relación:

Dirección: normalmente no es aplicable

Fuerza: no es aplicable

Naturaleza: informe la diferencia media $\bar{D}$

PREGUNTAS PARA EL CAPÍTULO 11

1. Estudia la tabla 11.1 lo suficiente como para que puedas reproducirla.
2. Describir la situación en la que utilizamos una *prueba de diferencia de medias para dos grupos* con grupos independientes.
3. Con una prueba de diferencia de medias para dos grupos debemos suponer la igualdad de las varianzas. ¿Por qué debe considerarse este hecho?
4. Explica la diferencia entre las pruebas de diferencia de medias para dos grupos independientes y relacionados.
5. ¿Cómo se formula la hipótesis nula para una prueba de diferencia de medias con dos grupos? ¿Por qué debe formularse de esta forma?
6. Los efectos estadísticos de la prueba implican los estadísticos de la muestra y los parámetros de la población. El efecto de la prueba se encuentra en el numerador del estadístico de la prueba. En el caso de las pruebas de diferencia de medias para dos grupos, ¿por qué *no* aparecen los parámetros en las fórmulas de los estadísticos de la prueba?
7. Con una prueba de diferencia de medias para dos grupos con grupos independientes, en la cual se suponen varianzas poblacionales iguales, ¿por qué se calculan de la siguiente manera los grados de libertad?

$$gl = n_1 + n_2 - 2$$

8. La existencia de una relación entre dos variables determina lo siguiente: sobre la base del análisis estadístico de una __________, ¿podemos concluir que existe una relación entre dos variables entre todos los individuos de la __________?
9. La existencia de una relación se determina comprobando la hipótesis, que establece si los hallazgos estadísticos de la muestra son consecuencia del error de __________.
10. En lo que se refiere a la dirección de una relación, una relación positiva es aquella en la cual el incremento en una de las variables se relaciona con un __________ en la otra. Una relación negativa es aquella en la cual el incremento en una de las variables se relaciona con un __________ en la otra.
11. La __________ de una relación entre dos variables establece en qué grado se reducen los errores cuando se predicen las puntuaciones de una variable dependiente. Esta medición nos da una indicación del __________, es decir, qué tan adecuadamente la variable independiente predice los resultados de una variable dependiente.
12. La __________ de la relación implica describir la forma en que el conocimiento de una relación entre dos variables nos ayuda a comprender un fenómeno y a aplicar los resultados en circunstancias prácticas.
13. Al describir la __________ de una relación, evitamos el lenguaje estadístico presentando los hallazgos de la investigación en lenguaje común, particularmente cuando presentamos los resultados al público en general.
14. Una prueba estadística de __________ resulta de utilidad en el diseño experimental *antes-después* o *test-reset,* en el que una variable se mide dos veces en el caso de los mismos individuos con algún tipo de intervención en las pruebas.
15. Indica la diferencia entre la significancia estadística y la significancia práctica. Menciona ejemplos.
16. En el caso de la prueba de diferencia de medias con dos grupos, ¿qué aspectos de una relación son relevantes? ¿Cómo se determinan estos aspectos?
17. Relaciona lo siguiente:

    a) $s_{\bar{X}_1 - \bar{X}_2}$ ______ Número de errores estándar que la diferencia de las medias de las muestras independientes se desvía de la diferencia de la media hipotética de cero

    b) $t_{\bar{D}}$ ______ Error estándar de las diferencias entre puntuaciones apareadas

    c) $t_{\bar{X}_1 - \bar{X}_2}$ ______ Número de errores estándar que la diferencia de las medias de las muestras entre puntuaciones apareadas se desvía de la diferencia media hipotética de cero

    d) $s_{\bar{D}}$ ______ Estimación del error estándar de la diferencia entre dos medias con varianzas agrupadas

EJERCICIOS PARA EL CAPÍTULO 11

Conjunto de problemas 11A

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, diagrama conceptual, curvas de probabilidad y aspectos adecuados de una relación. Para lograr mayor consistencia, redondea los errores estándares a dos decimales. Utiliza $\alpha = .05$ a menos que se indique otra cosa.

11A.1 En un restaurante se lleva a cabo una encuesta entre dos grupos de mujeres elegidas al azar: las que trabajan en casa y las que trabajan en otra parte. En la encuesta se les pregunta cuántas veces prepararon comida en casa durante las últimas dos semanas. De acuerdo con las respuestas que siguen (ficticias), determina si las mujeres que trabajan fuera de casa preparan comida con menor frecuencia que las que trabajan en casa. Asume la igualdad de las varianzas poblacionales.

Situación de trabajo	Número de comidas preparadas en casa durante las últimas dos semanas
En casa	9
En casa	10
En casa	11
En casa	8
En casa	9
En casa	12
En casa	14
En casa	10
En casa	12
En casa	13
En casa	9
En casa	10
En casa	12
En casa	14
En casa	10
Fuera de casa	8
Fuera de casa	10
Fuera de casa	8
Fuera de casa	7
Fuera de casa	9
Fuera de casa	9
Fuera de casa	12
Fuera de casa	8
Fuera de casa	10
Fuera de casa	10
Fuera de casa	7
Fuera de casa	12
Fuera de casa	6
Fuera de casa	10
Fuera de casa	9
Fuera de casa	8
Fuera de casa	10
Fuera de casa	9

11A.2 Supongamos que se seleccionan dos muestras aleatorias de 40 empresas para comparar el ingreso medio por hora de los trabajadores sindicalizados y los no sindicalizados. Las 40 empresas sin sindicato ofrecían un salario medio de \$10.80 con una varianza de \$2.50, mientras que las 40 compañías con sindicato ofrecían un salario medio de \$ 11.90, con una varianza de \$ 2.50. ¿Constituyen estos datos una evidencia suficiente que sugieran que a un trabajador se le paga mejor en una compañía sindicalizada? Asume la igualdad de las varianzas poblacionales.

11A.3 En el caso de una encuesta aleatoria de 641 adultos, determina si existe alguna diferencia de género en el grado de apoyo relacionado con el control de armas. El apoyo se mide en una escala sobre actitudes con respecto al control de armas (X), que posee un nivel de medición de intervalo (datos ficticios). Una puntuación alta indica una actitud más favorable con respecto al control de armas. Asume la igualdad de las varianzas poblacionales.

Hombres	Mujeres
$\bar{X} = 6.2$	$\bar{X} = 6.5$
$s_X = 1.3$	$s_X = 1.4$
$n = 324$	$n = 317$

11A.4 Una comparación de la expectativa de vida en muestras aleatorias de 40 países en vías de desarrollo y 31 países industrializados revela los siguientes datos (ficticios). Las desviaciones estándar son significativamente diferentes. ¿Existe una diferencia significativa en la expectativa de vida media entre los países con estos dos niveles de desarrollo económico? X = expectativa de vida de un residente al nacer.

Países en vías de desarrollo	Países industrializados
$\bar{X} = 66.1$ años	$\bar{X} = 76.7$ años
$s_X = 28.9$ años	$s_X = 4.2$ años
$n = 40$	$n = 31$

11A.5 Las lesiones del cojín rotador del brazo y hombro son comunes en los atletas. Un instituto de medicina deportiva se encuentra probando la eficacia de una nueva terapia para mejorar el grado de movimiento del brazo. La nueva terapia se administra a un grupo de tratamiento experimental, y los tratamientos tradicionales, a un grupo control. Los sujetos, son asignados aleatoriamente a cada grupo, y el grado de movimiento se mide con un instrumento graduado en centímetros. El régimen del tratamiento experimental incluye sesiones de tres veces por semana por seis semanas con mediciones en el tiempo 1 (es decir, al inicio del tratamiento) y en el tiempo 2 (es decir, al concluir el tratamiento). Los resultados (ficticios) aparecen en la siguiente tabla. Prueba las hipótesis para responder primero las preguntas *a* y *c*, y posteriormente la pregunta *d*. Ten cuidado de utilizar las pruebas para datos relacionados cuando sea conveniente. Utiliza $\alpha = .01$ y asume que las varianzas poblacionales son iguales. Y = grado del movimiento del brazo.

a) Si los miembros de los dos grupos se asignaran en realidad de forma aleatoria, entonces no habría una diferencia significativa en el grado del movimiento del brazo para el tiempo uno. ¿Fue este el caso?

b) ¿Hubo alguna diferencia significativa entre los grupos experimental y de control en el tiempo 2?

c) ¿Hubo alguna diferencia significativa en el grado de movimiento del brazo en los tiempos 1 y 2 en el caso de los sujetos del experimento?

d) Tomando en cuenta los resultados de los incisos *a*), *b*) y *c*), ¿parece que hay una mejora con el nuevo tratamiento en comparación con el tratamiento tradicional?

	Grupo control ($n = 35$)	Grupo experimental del tratamiento ($n = 35$)
Tiempo 1	$\bar{Y}$– = 16.9 cm	$\bar{Y}$ = 17.1 cm
	s_Y = 3.6 cm	s_Y = 3.7 cm
Tiempo 2	$\bar{Y}$ = 32.1 cm	$\bar{Y}$ = 35.5 cm
	s_Y = 3.4 cm	s_Y = 3.4 cm
Diferencia media por individuo	$\bar{D}$ = 14.7 cm	$\bar{D}$ = 18.2 cm
	s_D = 3.1 cm	s_D = 3.0 cm

Conjunto de problemas 11B

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, diagrama conceptual, curvas de probabilidad y aspectos adecuados de una relación. Para lograr mayor consistencia, redondea los errores estándares a dos decimales. Utiliza $\alpha = 0.05$, a menos que se indique otra cosa.

11B-1 En el próximo reclutamiento de jugadores de la National Football League (NFL), el equipo que terminó en último lugar la temporada pasada busca la manera de cubrir las posiciones en las que se requiere experiencia con jugadores que resulten productivos en términos de puntuación. El equipo toma muestras de los registros de jugadores de fútbol colegial y tabulan la cantidad de anotaciones logradas el año pasado por posición (datos ficticios). ¿Qué posición es más productiva, la de corredor, la del receptor abierto? Asume la igualdad de las varianzas poblacionales.

Posición	Anotaciones
Corredor	10
Corredor	15
Corredor	12
Corredor	13
Corredor	16
Corredor	11
Corredor	13
Corredor	13
Corredor	14
Corredor	12
Corredor	12
Corredor	15
Corredor	11
Corredor	13
Receptor abierto	8
Receptor abierto	12
Receptor abierto	10
Receptor abierto	10
Receptor abierto	9
Receptor abierto	13
Receptor abierto	11
Receptor abierto	8
Receptor abierto	8
Receptor abierto	10
Receptor abierto	12
Receptor abierto	15
Receptor abierto	10
Receptor abierto	12
Receptor abierto	9

11B-2 Shallowstone Pictures es la compañía líder de dibujos animados en Estados Unidos. Los ejecutivos de la compañía desean saber a qué población deben dirigirse las películas, a la gente joven o la gente mayor. Realizan encuestas en poblaciones de estudiantes universitarios y personas retiradas. La muestra de 61 estudiantes universitarios arrojó un promedio de 23.45 películas por año con una varianza de 6.86 películas, y la muestra de 61 personas retiradas arrojó un promedio de 21.79 películas por año con una varianza de 6.86 películas. ¿Indican estos datos que los estudiantes universitarios ven más películas que las personas retiradas? Asume la igualdad de las varianzas poblacionales.

11B-3 Tú deseas investigar si la depresión es más alta entre los estudiantes universitarios de primer año que entre los estudiantes de segundo año. Tu teoría consiste en que los estudiantes de primer año deben adaptarse a nuevas circunstancias y que el estrés da origen a más casos de depresión. La depresión constituye una medida de nivel de intervalo, la Escala de Depresión del Centro de Estudios Epidemiológicos,[1] varía de 0 a 60. Las puntuaciones altas (en puntos de la escala CESD) indican niveles más altos de depresión. ¿Apoyan la teoría los estadísticos que aparecen enseguida? Asume la igualdad de las varianzas poblacionales.

Estudiantes de primer año	Estudiantes de segundo año
$\bar{X} = 9.42$	$\bar{X} = 9.13$
$s_X = 2.18$	$s_X = 2.29$
$n = 169$	$n = 174$

11B-4 Las investigaciones demuestran que incluso en el caso de las parejas que perciben dos ingresos, las mujeres realizan más trabajo en casa de manera significativa que los hombres (véase Hersch y Stratton, 2000). Tú deseas reproducir este estudio y tomas muestras de poblaciones de hombres y mujeres casados, todos los cuales laboran 40 horas semanales fuera de casa. El trabajo en casa (X) se mide en horas y las desviaciones estándares *son* significativamente diferentes entre sí. ¿Apoyan los datos que aparecen a continuación las conclusiones anteriores?

Hombres	Mujeres
$\bar{X} = 23.24$	$\bar{X} = 29.15$
$s_X = 11.05$	$s_X = 4.12$
$n = 57$	$n = 52$

11B-5 Un psicólogo trata pacientes con altos niveles de ansiedad y se interesa por saber si la lectura de un libro de autoayuda reduce el nivel de ansiedad más que la terapia convencional que él proporciona a sus pacientes. Toma muestras de su población de pacientes y asigna aleatoriamente la lectura del libro a individuos del grupo experimental, y a los individuos del grupo de control no les asigna la lectura. Mide la ansiedad en el tiempo 1 (antes) y en el tiempo 2 (después) en el caso del grupo experimental que ha leído el libro. En la siguiente tabla aparecen los resultados ficticios obtenidos. Realiza una prueba de las hipótesis para responder las preguntas *a*) a *c*), y en seguida responde la pregunta *d*). Ten cuidado de utilizar una prueba para muestras relacionadas cuando sea apropiado, utiliza $\alpha = 0.01$ y asume que las varianzas poblacionales son iguales. Y = nivel de ansiedad.

a) Si los miembros de los dos grupos se asignaron en realidad de forma aleatoria, entonces no debería haber diferencia significativa entre sus niveles de ansiedad en el tiempo 1. ¿Fue este el caso?

b) ¿Hubo una diferencia significativa en los niveles medios de ansiedad entre los grupos experimental y de control en el tiempo 2?

c) ¿Hubo una diferencia significativa en los niveles medios de ansiedad en los tiempos 1 y 2 en el caso de los individuos del grupo experimental?

[1] Center for Epidemiological Studies Depression Scale (CESD). (N. del T.)

d) Tomando en cuenta los resultados de los incisos *a* a *c*, ¿resulta mejor la lectura del libro que el tratamiento convencional?

	Grupo control (*n* = 40)	Grupo experimental del tratamiento (*n* = 40)
Tiempo 1	$\bar{Y} = 26.61$ puntos	$\bar{Y} = 25.95$ puntos
	$s_Y = 4.33$ puntos	$s_Y = 4.87$ puntos
Tiempo 2	$\bar{Y} = 19.12$ puntos	$\bar{Y} = 15.57$ puntos
	$s_Y = 3.76$ puntos	$s_Y = 2.64$ puntos
Diferencia media por individuo	$\bar{D} = 7.49$ puntos	$\bar{D} = 10.38$ puntos
	$s_D = 2.81$ puntos	$s_D = 3.63$ puntos

Conjunto de problemas 11C

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, diagrama conceptual, curvas de probabilidad y aspectos adecuados de una relación. Para lograr mayor consistencia, redondea los errores estándar a dos decimales. Utiliza $\alpha = .05$, a menos que se indique otra cosa.

11C-1 Dos movimientos sociales que surgieron en la década de los ochenta para aumentar el conocimiento del público relacionado con los peligros de beber antes de manejar fueron Madres Contra Conductores Ebrios (MCCE) y Retiro de Conductores Intoxicados (RCI) (McCarthy y Wolfson, 1996). Supón que los datos que aparecen a continuación provienen de una encuesta aplicada a una muestra de presidentes de diferentes periodos de las dos organizaciones. X = número de apariciones públicas el año pasado. ¿Existe una diferencia significativa en las medias de X? Asume la igualdad de las varianzas poblacionales.

Organización del presidente	X
MCCE	41
MCCE	29
RCI	10
MCCE	33
RCI	24
RCI	26
MCCE	45
MCCE	39
MCCE	33
RCI	26
MCCE	28
RCI	23

MCCE	45
RCI	10
MCCE	26
RCI	19
MCCE	37
RCI	15
MCCE	32
MCCE	32
RCI	20
RCI	14
MCCE	36
MCCE	38
RCI	24

11C-2 Se tomaron muestras aleatorias de 100 adultos de dos grupos étnicos en una ciudad grande y se les preguntó a éstos acerca de la cantidad de años que asistieron a las escuelas públicas. Analiza si existe una diferencia significativa entre las dos medias poblacionales. Asume la igualdad de las varianzas poblacionales.

	Grupo étnico 1	Grupo étnico 2
Media	7.4 años	8.2 años
Desviación estándar	2.1 años	2.4 años
n	100	100

11C-3 Orbuch y Eyster (1997) solicitaron a esposas de parejas negras y de parejas blancas que evaluaran la participación de sus esposos en tareas tradicionalmente femeninas, como preparar los alimentos, lavar platos, limpiar la casa, lavar ropa y cuidar niños. Se aplicó una escala de seis puntos en la que una puntuación alta indica una mayor participación por parte del esposo. Los estudios anteriores demostraron que los maridos negros participan más. ¿Se observa esto en los datos de Orbuch y Eyster? Asume la igualdad de las varianzas poblacionales. X = puntuación de participación del esposo.

Negros	Blancos
$\bar{X} = 1.54$	$\bar{X} = 1.47$
$s_X = 0.36$	$s_X = 0.37$
$n = 199$	$n = 174$

11C-4 Una empresa de electrónica exitosa reciente ha estado comercializando dos modelos de reproductores de discos compactos (CD), uno con un cambiador de discos rotatorio y el otro con un cambiador fijo. Las muestras aleatorias de recientes compradores devolvieron las encuestas con una escala de satisfacción de múltiples de activos, con un nivel de medición de intervalo. Las desviaciones estándar son significativamente diferentes entre sí. ¿Existe una diferencia significativa en la satisfacción media del comprador en el caso de los dos modelos? Y = puntuación en la escala de satisfacción del producto.

Compradores del modelo rotatorio	Compradores del modelo fijo
$\bar{Y} = 31.1$	$\bar{Y} = 28.1$
$s_Y = 2.4$	$s_Y = 4.2$
$n = 149$	$n = 167$

11C-3 En el momento de iniciar un programa de pérdida de peso, se solicitó a los participantes que identificaran en una lista de alimentos aquellos que contenían un alto nivel de grasas saturadas (es decir, alimentos AGS). Terminado el programa de educación nutricional, se mostró a los mismos participantes la lista de alimentos de nuevo y se les encargó la misma tarea. ¿Resultó eficaz del programa educativo para incrementar el conocimiento de los participantes con relación a los alimentos AGS? Asume que las varianzas poblacionales son iguales. X = número de alimentos AGS identificados correctamente en la lista (datos ficticios).

	Número de alimentos AGS identificados	
Individuo	**Antes del programa**	**Después del programa**
1	7	11
2	4	9
3	8	14
4	7	12
5	5	11
6	2	7
7	6	15
8	5	12
9	7	10
10	8	13
11	5	11
12	4	10
13	7	13
14	6	10
15	8	12
16	4	8
17	7	14
18	6	11
19	6	10
20	8	13

Conjunto de problemas 11D

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, diagrama conceptual, curvas de probabilidad y aspectos adecuados de una relación. Para lograr mayor consistencia, redondee los errores estándares a dos decimales. Utiliza $\alpha = 0.05$, a menos que se indique otra cosa.

11D-1 La ABC Cab Company compite con XYZ Cabs en la ciudad de Nueva York. ABC calcula que la mejor forma de ganar más dinero consiste en recoger más pasaje. Lleva a cabo una encuesta con la población de conductores de XYZ Cabs y con sus propios conductores de taxi para determinar cuántos pasajeros recogen en un día. Utiliza los siguientes datos (ficticios) para determinar si existe una diferencia significativa en la cantidad media de pasajeros recogidos entre las dos compañías de taxis. X = cantidad de pasajeros. Asume la igualdad de las varianzas poblacionales.

Compañía de taxis	X
XYZ	19
XYZ	15
ABC	25
XYZ	22
ABC	26
XYZ	23
ABC	28
ABC	27
ABC	31
XYZ	15
ABC	26
XYZ	16
XYZ	23
ABC	21
XYZ	18
XYZ	26
ABC	30
XYZ	24
ABC	22
XYZ	25
ABC	31
ABC	27
XYZ	19
XYZ	21
XYZ	23
ABC	27
XYZ	16
ABC	24
ABC	26
XYZ	21
ABC	24

11D-2 La BodyMax Fitness Company mantiene un ardid publicitario mediante el que acusa a la compañía Heavy Lift de fabricar mancuernas más ligeras de lo que supone que deben pesar. La BodyMax Fitness Company pesa mancuernas de 25 libras de Heavy Lift y una de las que ella misma fábrica. La mancuerna de Heavy Lift pesa 23.6 libras y la mancuerna de BodyMax pesa exactamente 25 libras. Como investigador independiente, tú deseas cerciorarte de la veracidad de la acusación. Tomas muestras de mancuernas fabricadas por ambas compañías reguladas con pesos de 25 libras y, de hecho, el peso medio de las mancuernas de BodyMax es ligeramente superior al de las mancuernas de HeavyLift. De acuerdo con los siguientes datos, determina si las mancuernas de HeavyLift son en realidad más ligeras que las mancuernas de BodyMax. En otras palabras, ¿existe una diferencia significativa entre los pesos de las mancuernas?

	Mancuernas de BodyMax (grupo 1)	Mancuernas de HeavyLift (grupo 2)
Media	25.00 libras	24.70 libras
Desviación estándar	1.27 libras	1.13 libras
n	35 pesos	35 pesos

11D-3 DuBois y Steverthorn (2005) reportaron que contar con un mentor estaba asociado con el incremento de la autoestima en los adolescentes. Tú deseas repetir su estudio, así que entrevistas una población de adolescentes entre los 12 y 17 años y obtienes los siguientes estadísticos. La autoestima (X) se mide utilizando una escala de autoestima. ¿Poseen los adolescentes con mentores una autoestima más alta que los que no los tienen? Asume la igualdad de las varianzas poblacionales.

Con mentor	Sin mentor
$\bar{X} = 13.35$	$\bar{X} = 10.02$
$s_X = 1.39$	$s_X = 1.45$
$n = 104$	$n = 110$

11D-4 Tú deseas determinar si las mujeres en Estados Unidos ganan menos que los hombres por realizar la misma cantidad y tipo de trabajo. Entrevistas a capturistas de 100 empresas y obtienes los siguientes datos. Determina si existe una diferencia estadísticamente significativa en el ingreso medio entre hombres y mujeres. Las desviaciones estándar son significativamente diferentes. Y = pago por hora en dólares.

Hombres	Mujeres
$\bar{X} = 12.50$	$\bar{X} = 10.75$
$s_X = 1.25$	$s_X = 2.50$
$n = 50$	$n = 50$

11D-5 Las investigaciones han mostrado que el movimiento, como correr, caminar o levantar pesas incrementa los niveles de proteína MGF en los músculos (Deschens, 2004). Para replicar este estudio, tú tomas muestras de la población de adultos entre 35 y 45 años y mides los niveles de proteína en el jefe en los músculos de las piernas. Enseguida se somete a los individuos a un régimen de peso-entrenamiento de 10 semanas concentrado en las extremidades inferiores. Al final de las 10 semanas, tú mides de nuevo los niveles de proteína MGF. ¿Hubo un incremento de proteína MGF entre los individuos? X = unidades de proteína MGF.

	Niveles de proteína MGF	
Individuo	**Antes del programa**	**Después del programa**
1	11	17
2	9	16
3	15	20
4	13	16
5	11	21
6	9	16
7	10	15
8	14	17
9	11	19
10	10	16
11	15	21
12	14	19
13	9	16
14	10	16
15	13	21
16	12	15
17	14	18
18	11	17
19	11	19
20	8	16
21	16	19
22	11	20
23	14	18
24	9	14
25	11	16

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 11

Si en tu clase se utilizan las aplicaciones opcionales en computadora que acompañan el texto, abre los ejercicios del capítulo 11 localizados en el sitio web de *The Statistical Imagination*, www.mhhe.com/ritchey2. Los ejercicios relacionados con la prueba de diferencia de medias para dos grupos en el *SPSS* para Windows se enfocan en la elección de las secuencias de comandos de prueba apropiadas y de interpretar los datos correctamente resulta de particular interés distinguir los resultados de las pruebas con varianzas iguales y diferentes. Además, el apéndice D del texto contiene una vista rápida de las secuencias del comando *SPSS* para los procedimientos cubiertos en este capítulo.

CAPÍTULO

12

Análisis de varianza: diferencias entre las medias de tres o más grupos

RESUMEN DEL CAPÍTULO

Introducción

Justo antes del fin del milenio, en Estados Unidos dio inicio un movimiento nacional para reducir la asistencia del gobierno a madres solteras pobres e impulsarlas a buscar empleo. En la mayor parte de los estados, la asistencia familiar nunca ha sido suficiente para alejar a las familias de la pobreza; ya que se trata de una medida temporal que la mayoría de los beneficiarios utilizan durante un breve periodo. Aún al calor de la retórica política, algunos políticos y ciudadanos piensan que vivir "con asistencia social" significa unas largas vacaciones pagadas.

Los investigadores y trabajadores sociales que conocen los programas estatales de asistencia se mantienen escépticos en lo que refiere a dicho estereotipo público. De hecho, una mirada de cerca a la vida cotidiana de los beneficiarios de la asistencia familiar revela un estilo de vida caracterizado por una lucha por la supervivencia económica. Por ejemplo en una muestra de 214 madres que vivían de la asistencia social en cuatro ciudades, Edin y Lein (1997) midieron "el gasto mensual en diversión" (GMD) para conocer el grado de "esplendidez" con la que viven dichas madres. Hallaron una medida de GMD de apenas \$22. Este promedio bajo de gastos desafía el estereotipo de que los beneficiarios de la asistencia social están tomando unas largas vacaciones.

No obstante, el objetivo principal del estudio de Edin y Lein consistía en observar si la medida del GMD difería de ciudad en ciudad. El costo de vida varió en las ciudades, lo cual sugirió que los gastos en diversión también variarían. En términos teóricos, la pregunta de investigación era la siguiente: ¿existe una relación entre la ciudad en la que reside una madre que vive de la asistencia social y su GMD? La ciudad de residencia constituye una variable, X, con un nivel de medición nominal. El GMT representa la segunda variable, Y, que es de nivel de razón. También es la variable dependiente y el interés se centra en su media.

El diseño estadístico para comparar tres o más medias grupales es el análisis de varianza (ANOVA) de un factor. El ANOVA constituye una extensión de la prueba de diferencia de medias para dos grupos (prueba t), que es estudio del capítulo 11, pero con el ANOVA partimos de una perspectiva un tanto diferente. Para ilustrar la forma en la que funciona el ANOVA, utilicemos datos similares a los de Edin y Lein. Para simplificar analizaremos una muestra de 15 madres de tres ciudades. Resulta claro que nuestro interés se centra en saber si la media del GMD (μ_Y) difiere en lo que se refiere a *las poblaciones* de madres que viven de la asistencia social en las tres ciudades. Es decir, ¿difieren sus parámetros, μ_Y, en el caso de que Y = GMD? Esta pregunta de investigación se ilustra gráficamente en la figura 12-1.

Los datos aparecen en la tabla 12-1, en la que Y = GMD y X es la ciudad de residencia. En la esquina inferior derecha de la tabla 12-1 encontramos que la media del GMD de las 15 madres es de \$22. Esta media total de la muestra recibe el nombre de *gran media*. Por encima de la media total se encuentran las medias de las tres ciudades: Boston (\$28), Chicago (\$24) y Charleston (\$14); éstas se denominan *medias grupales*.

Para probar las diferencias entre las medias de las tres ciudades, podríamos aplicar la prueba t (capítulo 11); pero el proceso requeriría tres pruebas, una para cada uno de los siguientes pares de ciudades: Boston-Chicago, Boston-Charleston y Chicago-Charleston. Comparar las medias de, por ejemplo, seis ciudades resultaría aún más complicado y se requerirían 15 pruebas t. La naturaleza engorrosa de calcular conjuntos grandes de pruebas t desafió a los estadistas obligándonos a crear el ANOVA, una extensión de la prueba t.

Como en el caso de la prueba t, en el ANOVA la hipótesis estadística se enuncia de la siguiente manera: no existen diferencias entre las medias grupales. En el caso de nuestras madres que viven de la asistencia social, enunciamos la hipótesis estadística de la siguiente manera:

$$H_0: \mu_{Y(Boston)} = \mu_{Y(Chicago)} = \mu_{Y(Charleston)}$$

donde Y = GMD. Este enunciado reúne los requisitos de una hipótesis nula. Cuando es verdadera, la diferencia entre las medias será cero, con un error de muestreo predecible.

Cálculo de los efectos principales

Aunque el enunciado de la hipótesis nula es en esencia el mismo que con la prueba t, con el ANOVA adoptamos un enfoque ligeramente diferente. En lugar de comparar cada media grupal con las demás, el ANOVA compara cada media grupal con la media total, lo cual tiene sentido. Si las medias poblacionales son iguales para los tres grupos, la media de todos los casos combinados será la misma. De esta manera, la hipótesis nula puede enunciarse nuevamente de la siguiente manera:

$$H_0: \mu_{Y(Boston)} = \mu_{Y(Chicago)} = \mu_{Y(Charleston)} = \mu_{Y(total)}$$

FIGURA 12-1

¿Son los gastos mensuales en diversión los mismos en las poblaciones de madres que viven de la asistencia pública en las tres ciudades?

Poblaciones: madres que viven de la asistencia social en tres ciudades

Y = gastos mensuales en diversión (GMD)

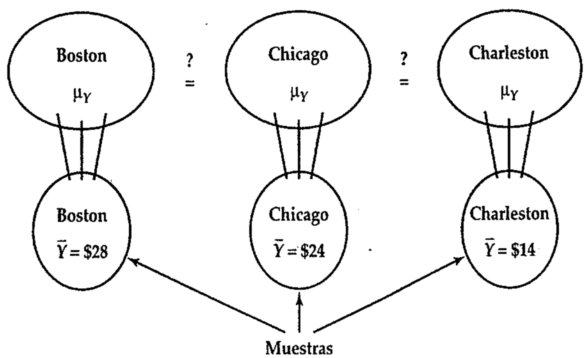


TABLA 12-1 | Hoja de cálculo relacionada con los gastos mensuales en diversión (GMD) de 15 madres que viven de la asistencia social en tres ciudades

Especificaciones			Cálculos		
Caso	Ciudad (X)	GMD (Y)	$(Y_{(cada\ caso)} - \bar{Y}_{(total)})$ puntuación de desviación	$(Y_{(cada\ caso)} - \bar{Y}_{(total)})^2$	Medias grupales
1	Boston	\$33	\$ 11	\$121	
2	Boston	30	8	64	$\bar{Y}_{(Bos.)} = \frac{\Sigma Y}{n} = \frac{140}{5} = \$28$
3	Boston	28	6	36	
4	Boston	26	4	16	
5	Boston	23	1	1	
6	Chicago	26	4	16	$\bar{Y}_{(Chic.)} = \frac{\Sigma Y}{n} = \frac{120}{5} = \$24$
7	Chicago	19	-3	9	
8	Chicago	24	2	4	
9	Chicago	22	0	0	
10	Chicago	29	7	49	
11	Charleston	14	-8	64	$\bar{Y}_{(Char.)} = \frac{\Sigma Y}{n} = \frac{70}{5} = \$14$
12	Charleston	19	-3	9	
13	Charleston	16	-6	36	
14	Charleston	12	-10	100	
15	Charleston	9	-13	169	
	$\Sigma Y_{(total)} = \$330$		$\Sigma(Y_{(cada\ caso)} - \bar{Y}_{(total)})^2 = \$694$		$\bar{Y}_{(total)} = \frac{\Sigma Y_{(total)}}{n} = \frac{330}{15} = \$22$
$n = 15$			$\Sigma(Y_{(cada\ caso)} - \bar{Y}_{(total)}) = \$0$		

Por otra parte, si las tres medias son iguales a la media total, la diferencia entre cualquier media grupal y la media total será cero. En el ANOVA, estas *diferencias entre cada media grupal y la media total* son los efectos de la prueba, que en el caso del ANOVA reciben el nombre de **efectos principales**.

Cálculo del efecto principal de una media grupal

Efecto principal de una media grupal $= \bar{Y}_{(grupo)} - \bar{Y}_{(total)}$

= diferencia entre una media grupal y la media de todas las naciones en la muestra

donde

Y = una variable de nivel de intervalo/razón

$\bar{Y}_{(grupo)}$ = media de Y para un grupo (es decir, una categoría de la variable de nivel nominal)

$\bar{Y}_{(total)}$ = media de todas las puntuaciones en la muestra

Recordemos que el efecto de prueba de una variable es la diferencia entre lo que se observa en la muestra (paso cuatro de los seis pasos de la inferencia estadística) y lo que se hipotetiza estadísticamente (paso uno de los seis pasos). Cuando la hipótesis nula resulta verdadera, no hay ninguna diferencia entre las medias de las tres ciudades y la media total. La media del GMD de cada ciudad sería de \$22, y *todos los efectos principales serían cero.* De esta manera, la hipótesis nula puede verse de otra forma:

$$H_0: \bar{Y}_{(cualquier\ grupo)} - \bar{Y}_{(total)} = 0$$

No obstante, podremos ver que en la tabla 12-1, las medias muestrales de las tres ciudades no son iguales. De hecho, en el caso de las poblaciones de madres que viven de la asistencia social en las tres ciudades, la hipótesis alternativa se refiere al hecho de que la media del GMD no es la misma (es decir, los efectos principales no son cero). En el caso de nuestras madres que viven de la asistencia social, los efectos principales son los siguientes:

$$\text{Efecto principal en el GMD por vivir en Boston} = \bar{Y}_{(Boston)} - \bar{Y}_{(total)} = \$28 - \$22 = \$6$$

$$\text{Efecto principal en el GMD por vivir en Chicago} = \bar{Y}_{(Chicago)} - \bar{Y}_{(total)} = \$24 - \$22 = \$2$$

$$\text{Efecto principal en el GMD por vivir en Charleston} = \bar{Y}_{(Charleston)} - \bar{Y}_{(total)} = \$14 - \$22 = -\$8$$

Enfocarse en la media total y comparar la media de cada ciudad con ésta, es una manera indirecta de comprobar la diferencia entre cualquier número de medias grupales. Al utilizar el ANOVA, determinamos si los efectos principales son significativamente diferentes de cero. Esta prueba gira en torno al hecho de que los efectos principales observados son tan grandes que resulta poco probable que se deban al error de muestreo.

Por consiguiente, en términos matemáticos, con el ANOVA la hipótesis nula puede enunciarse de varias formas que comunican el mismo significado: no existe diferencia entre las medias grupales; todas las medias son iguales; todas las medias son iguales a la media total y todos los efectos principales son cero. Si rechazamos la hipótesis nula y aceptamos la hipótesis estadística, afirmamos que alguno o todos los efectos principales son significativamente diferentes de cero. Esto indica, a su vez, que por lo menos dos de las medias poblacionales difieren entre sí. Además, al aceptar la hipótesis alternativa afirmamos que existe una relación entre la variable independiente de nivel nominal/ordinal y la variable dependiente de nivel de intervalo/razón.

Modelo lineal general: prueba de la significancia estadística de los efectos principales

Como resultado, Edin y Lein hallaron que, de hecho, existía una relación entre la ciudad de residencia y el GMD. Los efectos de la ciudad de residencia en el GMD fueron significativamente diferentes de cero. Concluyeron que en promedio *las poblaciones* de madres que vivían de la asistencia social en Boston tenían gastos en diversión $6 superiores a la media total de $22; en Chicago, los gastos eran $2 mayores y en Charleston eran $8 menores.

¿Cómo se emplean dichos efectos para llegar a esta conclusión? La respuesta a esta pregunta consiste en demostrar que las puntuaciones individuales del GMD en la muestra se determinan en parte por el efecto principal de la ciudad de residencia de un individuo. Esto se logra enfocándose en las desviaciones individuales con respecto a la media total y observando qué cantidad de desviación se debe al efecto de la ciudad de residencia. Recuerda que, de acuerdo con el capítulo 5, una puntuación de desviación es la diferencia entre la puntuación de un individuo y la gran media.

Cálculo de una puntuación de desviación

Puntuación de desviación para un caso $= Y_{(cada\ caso)} - \overline{Y}_{(total)}$

= diferencia entre la puntuación de un caso y la media de todas las puntuaciones de la muestra

donde

Y = una variable de nivel de intervalo o de razón

$Y_{(cada\ caso)}$ = puntuación Y de un caso individual

$\overline{Y}_{(total)}$ = media de todas las puntuaciones de la muestra

Analicemos la puntuación de desviación del caso 1 de la tabla 12-1 de la señora Jones. Si Y = gasto mensual en diversión (GMD), su GMD o puntuación Y es de $33. Por tanto, su puntuación de desviación es

$$\text{Puntuación de desviación de la señora Jones} = Y_{(Sra.\ Jones)} - \overline{Y}_{(total)} = \$33 - \$22 = \$11$$

¿Por qué gasta la señora Jones $11 más que el promedio en diversión? Si el GMD se relaciona con la ciudad de residencia, podemos argüir que $6 de esta desviación de $11 se *explica* por las diferencias en la media del GMD entre grupos de ciudades de residencia. El grupo de residencia de la señora Jones es Boston. La diferencia entre la media del GMD en Boston y la media total es de $6. Éste corresponde al efecto principal para Boston, el *costo extra* del GMD por vivir en Boston. En el caso de la señora Jones, $6 representa la parte de sus $11 de desviación, que se explica por las diferencias entre las medias grupales. Ésta es la *desviación entre grupos* de la señora Jones; es decir, la misma para todas las madres del grupo de Boston, la cual es igual al efecto principal para Boston:

$$\text{Desviación entre grupos de la señora Jones} = \text{efecto principal en el GMD por vivir en Boston}$$
$$= \overline{Y}_{(Boston)} - \overline{Y}_{(total)} = \$28 - \$22 = \$6$$

Aunque vivir en Boston explica $6 de la puntuación de desviación de $11 de la señora Jones, esto deja un adicional de $5 que no se explica por la ciudad de residencia. Ésta es la diferencia entre la puntuación Y de la señora Jones y la media para Boston. Esta cantidad recibe el nombre de *desviación dentro del grupo*, ya que, incluso, dentro de su grupo de Boston, la señora Jones gasta $5 más que el promedio. La desviación dentro del grupo indica la razón por la que un individuo no tiene la media dentro de su grupo. Representa la diferencia entre la puntuación Y de un individuo y la media del grupo de ese individuo. Mientras que la desviación *entre* grupos es la misma para todos los casos en un grupo, la desviación *dentro* del grupo varía entre los miembros de este:

Cálculo de una puntuación de desviación dentro del grupo de un caso

Desviación dentro del grupo $= Y_{(caso\ de\ un\ grupo)} - \overline{Y}_{(grupo)}$

= diferencia entre la puntuación de un caso y la media de todas las puntuaciones en su grupo

donde

Y = una variable ordinal de nivel de intervalo o de razón

$Y_{(caso\ de\ un\ grupo)}$ = puntuación Y de un caso individual en un grupo

$\overline{Y}_{(grupo)}$ = media de todas las puntuaciones en el grupo de dicho caso

Por consiguiente,

$$\text{Desviación dentro del grupo}_{(Sra.\ Jones)} = Y_{(Sra.\ Jones)} - \overline{Y}_{(Boston)} = \$33 - \$28 = \$5$$

La desviación dentro del grupo también se denomina *desviación no explicada.* Aunque podemos explicar que $6 de los $11 de la señora Jones por encima del promedio del GMD se deben a que reside en Boston, no sabemos la razón por la que gasta $5 adicionales por encima del promedio de Boston. Podría ser una consecuencia de los costos de transporte, por tener niños que comen demasiado o por alguna otra razón. Por el momento se miden estas otras variables y, por tanto, la desviación dentro del grupo de la señora Jones queda sin explicación.

A fin de cuentas, no podemos explicar cada dólar de los $33 que la señora Jones gastó en diversión. Sin embargo, es posible *reportarla* si nos concentramos en su puntuación de desviación de \$11 y separando las partes explicadas y las no explicadas por el hecho de que reside en Boston:

$$Y_{(Señora\ Jones)} = \$33 = \overline{Y}_{(total)} + \text{su puntuación de desviación}$$

$$= \overline{Y}_{(total)} + [Y_{(Señora\ Jones)} - \overline{Y}_{(total)}]$$

$$= \overline{Y}_{(total)} + [\overline{Y}_{(Boston)} - \overline{Y}_{(total)}] + [Y_{(Señora\ Jones)} - \overline{Y}_{(Boston)}]$$

$$= \$22 + \$6 + \$5$$

\$6: Desviación explicada entre grupos (la parte de la puntuación de desviación explicada por la residencia en Boston)

\$5: Desviación no explicada dentro del grupo (la parte de la puntuación de desviación explicada por otras variables no medidas)

De esta manera podemos explicar las puntuaciones Y (es decir, GMD) de todos los casos de una muestra. Por ejemplo, para el caso 15, una madre de Charleston que recibe asistencia social:

$$Y_{(caso\ 15)} = \$9 = \overline{Y}_{(total)} + [\overline{Y}_{(Charleston)} - \overline{Y}_{(total)}] + [Y_{(caso\ 15)} - \overline{Y}_{(Charleston)}]$$

$$= \$22 + (-\$8) + (-\$5)$$

(–\$8): Desviación explicada (efecto de recidir en Charleston)

(–\$5): Desviación no explicada (efecto de dos variables)

Este enfoque para el análisis, que se centra en las desviaciones con respecto a la media de una variable dependiente Y y que explica sus puntuaciones de desviación $Y - \overline{Y}$, recibe el nombre de *modelo lineal general* o, más adecuadamente, *modelo general de efectos aditivos*. Éste sencillamente indica que la mejor predicción de cualquier variable dependiente Y es su media más los efectos de una variable independiente X. En el caso del ANOVA o cualquier otro procedimiento estadístico, el modelo general de efectos aditivos se representa matemáticamente de la siguiente manera:

Modelo lineal general

$$Y_{cada\ caso} = \overline{Y}_{total} + \text{efecto explicado de } X + \text{error no explicado}$$

donde

Y = variable dependiente de intervalo o de razón

X = variable independiente relacionada con Y

$Y_{(cada\ caso)}$ = puntuación de Y en la muestra

$\overline{Y}_{(total)}$ = medias de todas las puntuaciones Y en la muestra

El modelo general de efectos aditivos separa (o *descompone*) cada puntuación Y en tres partes:

1. La *cantidad de la puntuación Y explicada por la media total* (es decir, la media de toda las puntuaciones Y).
2. La *cantidad de su puntuación de desviación explicada por X* (es decir, el efecto principal para la categoría X).
3. La *cantidad de su puntuación de desviación no explicada por X* (es decir, error).

La tabla 12-2 muestra la descomposición de cada uno de los 15 casos de nuestra muestra de madres que viven de la asistencia social, en función de estas tres partes. Esta clase de tabla recibe el nombre de *tabla de descomposición*. En el caso de cada ciudad de residencia, la columna A constituye la parte de una puntuación Y explicada por la media total. La columna B representa el efecto principal del grupo: la parte de la puntuación de desviación explicada por la ciudad de residencia (X). La columna C lista la parte de la puntuación Y de cada individuo que no se explica por la ciudad de residencia: el error no explicado.

Determinación de la significancia estadística de los efectos principales utilizando el ANOVA

¿Cómo determinamos el hecho de que la diferencia de \$6 entre la media del GMD de los residentes de Boston y la gran media no se debe simplemente al error de muestreo? ¿Es posible

TABLA 12-2 | Descomposición de los efectos de la ciudad de residencia (X) en el gasto mensual total en diversión [GMD (Y)] de madres que viven de la asistencia social, n = 15

Todas las columnas (A): media total = $\overline{Y}_{(total)} = \frac{\Sigma Y}{n} = \frac{330}{15} = \$22$

Boston			Chicago			Charleston		
(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
$Y = \overline{Y}_{(total)}$ +	Efecto principal de X	+ error	$Y = \overline{Y}_{(total)}$ +	Efecto principal de X	+ error	$Y = \overline{Y}_{(total)}$ +	Efecto principal de X	+ error
\$33 = 22 +	6	+ 5	\$26 = 22 +	2	+ 2	\$14 = 22 +	(–8)	+ 0
\$30 = 22 +	6	+ 2	\$19 = 22 +	2	+ (–5)	\$19 = 22 +	(–8)	+ 5
\$28 = 22 +	6	+ 0	\$24 = 22 +	2	+ 0	\$16 = 22 +	(–8)	+ 2
\$26 = 22 +	6	+ (–2)	\$22 = 22 +	2	+ (–2)	\$12 = 22 +	(–8)	+ (–2)
\$23 = 22 +	6	+ (–5)	\$29 = 22 +	2	+ 5	\$ 9 = 22 +	(–8)	+ (–5)
$\overline{Y}_{(Bos.)} = \frac{\Sigma Y}{n} = \frac{140}{5} = \$28$			$\overline{Y}_{(Chic.)} = \frac{\Sigma Y}{n} = \frac{120}{5} = \$24$			$\overline{Y}_{(Char.)} = \frac{\Sigma Y}{n} = \frac{70}{5} = \$14$		
Columna (B):			Columna (B):			Columna (B):		
Desviación entre grupos = efecto principal para Boston			Desviación entre grupos = efecto principal para Chicago			Desviación entre grupos = efecto principal para Charleston		
$\overline{Y}_{(Bos.)} - \overline{Y}_{(total)} = \$28 - \$22 = \$6$			$\overline{Y}_{(Chic.)} - \overline{Y}_{(total)} = \$24 - \$22 = \$2$			$\overline{Y}_{(Char.)} - \overline{Y}_{(total)} = \$14 - \$22 = -\$8$		
Columna (C):			Columna (C):			Columna (C):		
Desviación dentro del grupo = error $= Y - \overline{Y}_{(Bos.)}$			Desviación dentro del grupo = error $= Y - \overline{Y}_{(Chic.)}$			Desviación dentro del grupo = error $= Y - \overline{Y}_{(Char.)}$		

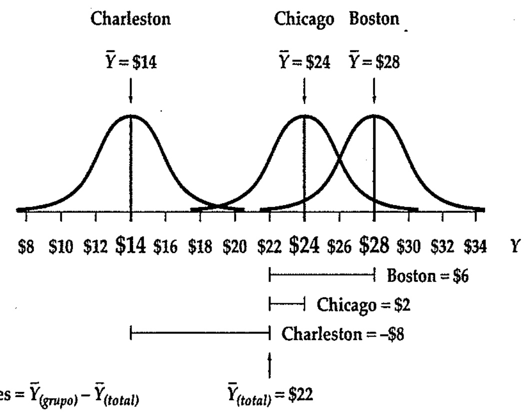
que en el muestreo repetido la media de otra muestra de Boston resulte diferente? En otras palabras, ¿cómo determinamos que los efectos principales y, por tanto, las diferencias entre las medias grupales son estadísticamente significativas? ¿Existe realmente una relación entre la ciudad de residencia y el GMD en las poblaciones de estudio, o una segunda muestra daría como resultado efectos principales radicalmente distintos?

Para probar si los efectos de la ciudad de residencia son reales en las poblaciones, debemos incluir una medida sumaria —un estadístico— que explique la variación en todos los casos. Si en realidad existen diferencias entre las medias grupales, como regla general las madres de Boston que viven de la asistencia social deberían tener un GMD de alrededor de $6 por encima de la media total. Asimismo, los efectos de Chicago y Charleston deben reflejarse en el patrón de puntuaciones del GMD para los individuos de dichas ciudades. Con el

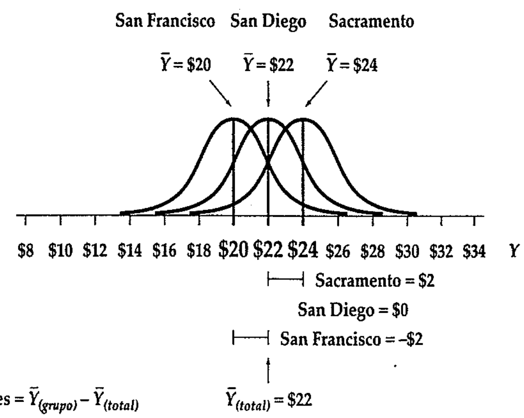
FIGURA 12-2

Comparación de la dispersión de las distribuciones de las puntuaciones cuando los efectos principales de los grupos son grandes (*A*) y (*B*)

(A) Cuando las medias son significativamente diferentes: los efectos principales son relativamente grandes; las puntuaciones se agrupan en torno a sus respectivas medias grupales.



(B) Cuando las medias no son significativamente diferentes: los efectos principales son relativamente pequeños; las puntuaciones se agrupan en torno a la gran media.



ANOVA, estamos afirmando que la *dispersión de las puntuaciones* —Boston en el extremo superior y Charleston en el extremo inferior— es consecuencia de los efectos de residencia en dichas ciudades. El patrón general de la dispersión de las puntuaciones debería mostrar una agrupación de los casos sobre la base de la ciudad de residencia. Este hecho se describe en la figura 12-2(*A*), en la que las puntuaciones del GMD de los residentes de Charleston se agrupan $8 debajo de la media total de $22, y los residentes de Chicago y Boston se agrupan $2 y $6 por encima de la media total, respectivamente.

Cuando en efecto se agrupan los casos de cada ciudad, como en el caso de la figura 12-2(*A*), esto indica que las desviaciones de las puntuaciones individuales en un grupo (es decir, ciudad) son la consecuencia principal de pertenecer a un grupo. Los casos de un grupo varían alrededor de su media grupal.

Cuando la agrupación no se presenta con un patrón relacionado con la pertenencia a un grupo, como lo muestra la figura 12-2(*B*), entonces las puntuaciones sencillamente varían en torno a la media total. Por ejemplo, supongamos que estamos llevando a cabo una investigación en Sacramento, San Francisco y San Diego, California, ciudades con economías y programas de asistencia familiar semejantes (datos ficticios). Las desviaciones de la gran media no reciben ninguna influencia aleatoria —que se presenta en cualquier dirección— por la pertenencia de grupo. Este es el patrón que se presenta cuando la hipótesis nula que se refiere al hecho de que no hay diferencias entre estas medias es verdadera. El ANOVA es la prueba estadística que determina si la agrupación de los casos se asemeja más a la figura 12-2(*A*), en la que las puntuaciones de los grupos se acumulan en torno de medias diferentes, o si se parece más a la figura 12-2(*B*), donde las puntuaciones, prescindiendo de la ciudad de residencia, tienden a agruparse en torno a la media total.

Como su nombre lo indica, el ANOVA se enfoca en las varianzas de las puntuaciones. Recordemos nuevamente (véase capítulo 5) que una puntuación de desviación para un caso individual es la diferencia entre su puntuación y la media total. Sin embargo, para obtener una medida sumaria para la muestra completa, debemos elevar al cuadrado las puntuaciones de desviación. Éste es el caso, ya que la suma de las desviaciones (sin elevar al cuadrado) siempre es cero, como el caso de la tabla 12-1. La suma de las puntuaciones de desviación elevadas al cuadrado constituyen la *variación* o *suma de cuadrados*. Finalmente, recordemos que la *varianza* es la variación dividida entre el tamaño de la muestra para generar un promedio de la variación. Las puntuaciones de desviación, la variación y la varianza miden la forma en que se dispersan las puntuaciones con respecto a la media. El ANOVA se enfoca en la variación y, posteriormente, en la varianza de la muestra como un todo. En seguida determina si esas medidas de dispersión se explican por las diferencias en el GMD promedio entre las tres ciudades o sencillamente son el resultado del error de muestreo aleatorio.

De la misma forma como descompusimos la puntuación de desviación de la señora Jones, el ANOVA resume la descomposición en el caso de toda la muestra. Comparemos la descomposición individual con la descomposición sumaria:

Explicación del caso individual como un efecto de pertenencia a un grupo:

$$\text{Puntuación de desviación de la señora Jones} = Y_{(Sra.\ Jones)} - \bar{Y}_{(total)}$$
$$= \underbrace{\bar{Y}_{(Boston)} - \bar{Y}_{(total)}}_{\substack{\uparrow \\ \text{Desviación} \\ \text{explicada}}} + \underbrace{(Y_{(Sra.\ Jones)} - \bar{Y}_{(Boston)})}_{\substack{\uparrow \\ \text{Desviación} \\ \text{no explicada}}}$$

Explicación de las desviaciones de todos los casos juntos:

Variación total = suma total de cuadrados

$$= \Sigma(Y_{(cada\ caso)} - \bar{Y}_{(total)})^2$$

$$= \underbrace{\Sigma(\bar{Y}_{(grupo)} - \bar{Y}_{(total)})^2}_{\text{Variación explicada}} + \underbrace{\Sigma(Y_{(cada\ caso)} - \bar{Y}_{(grupo)})^2}_{\text{Variación no explicada}}$$

Nota que en estas mediciones sumarias calculamos tres tipos de sumas de cuadrados: la suma total de cuadrados (SC_T), la suma de cuadrados entre grupos o *explicada* (SC_E) y la suma de cuadrados dentro de los grupos o *no explicada* (SC_D).

Tipos de variación o sumas de cuadrados

SC_T= variación total = suma total de cuadrados

$$= \Sigma(Y_{(\text{cada caso})} - \bar{Y}_{(\text{total})})^2$$

SC_E= suma de cuadrados entre grupos

$$= \Sigma(\bar{Y}_{(\text{grupo})} - \bar{Y}_{(\text{total})})^2$$

SC_D= suma de cuadrados dentro de los grupos

$$= \Sigma(Y_{(\text{cada caso de un grupo})} - \bar{Y}_{(\text{grupo})})^2$$

variación explicada (variación por los efectos de grupo

= variación no explicada o error (es decir, la variación no explicada por la pertenencia a un grupo, sino por otras variables no medidas)

Los cálculos de la variación explicada y de la variación no explicada se llevan a cabo con los datos de las tablas 12-1 y 12-2. La variación total, o *suma total de cuadrados* (SC_T) es el mismo cálculo que realizamos en el capítulo cinco cuando determinamos la desviación estándar. En el caso de los datos de la tabla 12-1, vemos que la SC_T es de \$694. Para ilustrarlo, analicemos cómo se acumuló dicho total:

$$SC_T = \Sigma(Y_{(cada\ caso)} - \bar{Y}_{(total)})^2$$

$$= \underbrace{11^2 + 8^2 + 6^2 + 4^2 + 1^2}_{\text{Boston}} + \underbrace{4^2 + (-3)^2 + 2^2 + 0^2 + 7^2}_{\text{Chicago}} + \underbrace{(-8^2) + (-3^2) + (-6^2) + (-10^2) + (-13^2)}_{\text{Charleston}} = 694$$

Observa que los residentes de Boston tienden a desviarse en la dirección positiva y que los residentes de Charleston, en la dirección negativa.

La parte correspondiente a la SC_T explicada por los efectos del grupo (es decir, la ciudad de residencia) se denomina *suma de cuadrados explicada.* Ésta se calcula como la suma de los efectos del grupo elevados al cuadrado. La suma de cuadrados explicada también se denomina *suma de cuadrados entre grupos* (SC_E), ya que se debe a diferencias entre las medias grupales.

Ya que cada individuo dentro de un grupo (es decir, la ciudad) tienen la misma puntuación de efecto, la SC_E se calcula de manera directa. Estos representan los efectos de la prueba listados en la columna (B) para cada ciudad de la tabla 12-2. De esta manera, en el caso de todas las madres de Boston que viven de la asistencia social, el efecto al cuadrado es \$$6^2$; para las de Chicago, \$$2^2$ y para las de Charleston, –\$$8^2$. Así, en el caso de la muestra completa, el cálculo de la suma de cuadrados explicada es:

$$SC_E = \Sigma(\bar{Y}_{(grupo)} - \bar{Y}_{(total)})^2$$

$$= \underbrace{6^2 + 6^2 + 6^2 + 6^2 + 6^2}_{\text{Boston}} + \underbrace{2^2 + 2^2 + 2^2 + 2^2 + 2^2}_{\text{Chicago}} + \underbrace{(-8)^2 + (-8^2) + (-8^2) + (-8^2) + (-8^2)}_{\text{Charleston}} = 520$$

Estos cálculos se pueden abreviar, ya que cada caso en un grupo tiene el mismo efecto.

Cálculo de la suma de cuadrados entre grupos (o explicada) (SC_E)

$$SC_E = \Sigma(\bar{Y}_{(grupo)} - \bar{Y}_{(total)})^2 = \Sigma[(n_{(grupo)})(\text{efecto del grupo}^2)]$$

donde

SC_E = suma de cuadrados entre grupos (o explicada)

$\bar{Y}_{(grupo)}$ = media de Y para un grupo o categoría de X

$\bar{Y}_{(total)}$ = media de Y para todas las puntuaciones en la muestra

$n_{(grupo)}$ = número de casos en un grupo o categoría de X

Efecto del grupo = $(\bar{Y}_{(grupo)} - \bar{Y}_{(total)})$

De esta manera, en el caso de la muestra de madres que viven de la asistencia social:

$$SC_E = \Sigma(\bar{Y}_{(grupo)} - \bar{Y}_{(total)})^2 = \Sigma[(n_{(grupo)})\ (\text{efecto del grupo}^2)]$$

$$= (5)\ (6^2) + (5)\ (2^2) + (5)\ (-8^2) = 180 + 20 + 320 = 520$$

En la columna C de la tabla 12-2, listamos la parte de la puntuación Y de cada individuo que no se explica por la ciudad de residencia. Ésta es la desviación dentro del grupo, la diferencia entre la puntuación Y de un individuo y la media del grupo de dicho individuo. Estas desviaciones dentro del grupo también se elevan al cuadrado y se suman para obtener la *suma de cuadrados dentro de los grupos*, o SC_D. En el caso de las puntuaciones del GMD de las madres que viven de la asistencia social en la tabla 12-2,

$$SC_D = \Sigma(Y_{(cada\ caso\ del\ grupo)} - \bar{Y}_{(grupo)})^2 = \overbrace{5^2 + 2^2 + 0^2 + (-2^2) + (-5^2)}^{\text{Boston}}$$
$$+ \underbrace{2^2 + (-5^2) + 0^2 + (-2^2) + 5^2}_{\text{Chicago}} + \underbrace{0^2 + 5^2 + 2^2 + (-2^2) + (-5^2)}_{\text{Charleston}} = 174$$

Como podemos ver, el cálculo de la SC_D resulta un tanto engorroso. No obstante, existe una forma menos complicada de calcular esta suma. Nota que la suma de cuadrados entre grupos y la suma de cuadrados dentro de un grupo es igual a la suma total de cuadrados:

$$SC_T = SC_E + SC_D$$

Es decir, 694 = 520 + 174. Por tanto, una vez que hemos calculado la SC_T y la SC_E podemos calcular rápidamente la SC_D.

Cálculo de la suma de cuadrados dentro de los grupos (SC_D) (o no explicada)

$$SC_D = SC_T - SC_E$$

donde

SC_D = suma de cuadrados dentro de los grupos (o no explicada) de Y

SC_T = suma total de cuadrados (o variación) de Y

SC_E = suma de cuadrados entre grupos (o explicada) de Y

De esta manera, en el caso de la muestra de madres que viven de la asistencia social,

$$SC_D = SC_T - SC_E = 694 - 520 = 174$$

¿Qué nos dicen los tamaños relativos de estas sumas? Si las medias grupales difieren, los efectos de la ciudad de residencia y, por tanto, la SC_E serán grandes. ¿Cuán grande es una SC_E estadísticamente significativa? Al comprobar una hipótesis nula relativa a la igualdad de medias grupales, no basta con observar solamente el tamaño de la SC_E, ya que ese tamaño depende en gran medida de la cantidad total de casos en una muestra (n). Es decir, prescindiendo del hecho de que las medias grupales difieran, cuantos más casos se incluyan en los cálculos, mayores serán los tres tipos de sumas de cuadrados. De forma similar, la cantidad de grupos (K) influye en los cálculos de las sumas de cuadrados; es decir que cuanto más grupos se incluyan en la hipótesis, más efectos de prueba deberán calcularse, elevarse al cuadrado y sumarse. De esta manera debemos tomar en cuenta el tamaño de la muestra y la

cantidad de grupos; por tanto, las varianzas se calculan dividiendo estas sumas de cuadrados entre sus respectivos grados de libertad. En el ANOVA las varianzas que resultan reciben el nombre de *varianzas de los cuadrados medios, medias cuadráticas* o simplemente *cuadrados medios*. Los grados de libertad entre grupos (gl_E) son $K - 1$; los grados de libertad dentro de grupo (gl_D) son $n - K$. Los gl_E son $K - 1$, ya que una vez que se calculan las medias y los efectos en el caso de todos, excepto un grupo, la última media grupal y su efecto quedan fijos. (Nota que en la tabla 12-2, los efectos suman cero; si se conocen dos efectos, el otro queda determinado matemáticamente.) Los gl_D reflejan el hecho de que cuando todos, excepto un caso, se conocen dentro de un grupo, el último caso queda determinado matemáticamente. De esta manera, 1 grado de libertad se pierde en el caso de cada grupo. Para la suma de cuadrados entre grupos, la varianza del cuadrado medio es la siguiente:

Cálculo de la varianza del cuadrado medio entre grupos (es decir, la varianza explicada)

$$CM_E = \frac{SC_E}{gl_E} = \frac{SC_E}{K-1}$$

donde

CM_E = cuadrado medio entre grupos o *varianza explicada*

SC_E = suma de cuadrados entre grupos (es decir, variación explicada)

gl_E = grados de libertad entre grupos

K = número de grupos comparados

En el caso de la muestra de madres que viven de la asistencia social,

$$CM_E = \frac{SC_E}{gl_E} = \frac{SC_E}{K-1} = \frac{520}{2} = 260$$

En el caso de la suma de cuadrados dentro del grupo, la varianza de cuadrados medios dentro de grupos es la siguiente:

Cálculo de la varianza del cuadrado medio dentro de los grupos (es decir, varianza no explicada)

$$CM_D = \frac{SC_D}{gl_D} = \frac{SC_D}{n-K}$$

donde

CM_D = varianza del cuadrado medio dentro de los grupos, o *varianza no explicada*

SC_D = suma de cuadrados dentro de los grupos (es decir, variación no explicada)

gl_D = grados de libertad dentro del grupo

n = tamaño total de la muestra

K = número de grupos que se comparan

En el caso de la muestra de madres que viven de existencia social,

$$CM_D = \frac{SC_D}{gl_D} = \frac{SC_D}{n - K} = \frac{174}{12} = 14.50$$

***Estadístico de prueba de la razón* F**

En el ANOVA, la forma para calcular la probabilidad de los resultados de la muestra implica determinar la razón de la varianza del cuadrado medio explicada entre la varianza del cuadrado medio no explicada. Esto se denomina **estadístico de la razón *F***, cuya fórmula es la siguiente:

Cálculo del estadístico de la razón *F*

$$F = \frac{CM_E}{CM_D}$$

donde

F = estadístico de la razón F

CM_E = varianza del cuadrado medio entre grupos (o varianza explicada)

CM_D = varianza del cuadrado medio dentro de los grupos (o varianza no explicada)

En el caso de la muestra de mujeres que viven de la asistencia social,

$$F = \frac{CM_E}{CM_D} = \frac{260}{14.50} = 17.93$$

Una razón F calculada siempre será positiva, ya que al elevar el cuadrado el numerador y el denominador, se eliminan los signos negativos. Con el fin de organizar estos cálculos, normalmente la razón F se presenta en la *tabla de fuentes de variación* que distingue entre sumas de cuadrados entre grupos, dentro de los grupos y suma total. La tabla 12-3 es la tabla de fuentes de variación correspondiente al GMD de las madres que viven de la asistencia social.

TABLA 12-3 | Tabla de fuentes de variación para el análisis de varianza con los datos de la tabla 12-2

Fuente de variación	SC	gl	Varianza de los cuadrados medios: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	520	$K - 1 = 2$	260	17.93
Dentro de grupos (SC_D)	174	$n - K = 12$	14.50	
Total (SC_T)	694	$n - 1 = 14$	49.57	

Precisamente como las pruebas t miden la significancia de los efectos de la prueba, la razón F evalúa si los efectos principales que se observan en las medias grupales de la muestra son significativamente diferentes de cero, los efectos hipotéticos. Cuando los efectos principales son grandes, la SC_E y el CM_E son grandes. Ya que el CM_E se localiza en el numerador, cuando ésta es grande, el estadístico de la razón F también será grande. Cuanto mayor sea la razón F, mayor será la probabilidad de que la hipótesis nula se rechace. Como lo demostraremos cuando llevemos a cabo en la ANOVA con los seis pasos de la inferencia estadística, una razón F de 17.93 es bastante grande. Concluiremos que existe una diferencia significativa en el GMD de por lo menos dos ciudades.

Cómo resulta la razón* F *cuando las medias grupales no son significativamente diferentes

Antes de completar la prueba de hipótesis para la muestra de madres que viven de la asistencia social, intentemos obtener una perspectiva más adecuada sobre el ANOVA y la razón F. Analicemos el caso en el que las medias grupales no son significativamente diferentes. En otras palabras ¿cuál es el valor de F cuando no se rechaza la hipótesis nula relativa a la igualdad de las medias poblacionales?

TABLA 12-4 | Descomposición de los efectos de la ciudad de residencia (X) sobre el total de los gastos mensuales en diversión [GMD (Y)] para el ejemplo hipotético de la diferencia de grupos no significativa (madres que viven de la asistencia social, $n = 15$)

Todas las columnas (A): media total $= \bar{Y}_{(total)} = \frac{\Sigma Y}{n} = \frac{330}{15} = \$22$

Sacramento			San Diego			San Francisco		
(A)	(B) Efecto principal	(C)	(A)	(B) Efecto principal	(C)	(A)	(B) Efecto principal	(C)
$Y = \bar{Y}_{(total)}$ +	de X	+ error	$Y = \bar{Y}_{(total)}$ +	de X	+ error	$Y = \bar{Y}_{(total)}$ +	de X	+ error
$\$22 = 22$ +	2	+ (−2)	$\$27 = 22$ +	0	+ 5	$\$25 = 22$ +	(−2)	+ 5
$\$26 = 22$ +	2	+ 2	$\$24 = 22$ +	0	+ 2	$\$22 = 22$ +	(−2)	+ 2
$\$24 = 22$ +	2	+ 0	$\$22 = 22$ +	0	+ 0	$\$20 = 22$ +	(−2)	+ 0
$\$19 = 22$ +	2	+ (−5)	$\$20 + 22$ +	0	+ (−2)	$\$18 = 22$ +	(−2)	+ (−2)
$\$29 = 22$ +	2	+ 5	$\$17 = 22$ +	0	+ (−5)	$\$15 = 22$ +	(−2)	+ (−5)
$\bar{Y}_{(Sacra.)} = \frac{\Sigma Y}{n} = \frac{120}{5} = \$24$			$\bar{Y}_{(Diego)} = \frac{\Sigma Y}{n} = \frac{110}{5} = \$22$			$\bar{Y}_{(Fran.)} = \frac{\Sigma Y}{n} = \frac{100}{5} = \$20$		
Columna (B):			Columna (B):			Columna (B):		
Desviación entre grupos = efecto principal para Sacramento			Desviación entre grupos = efecto principal para San Diego			Desviación entre grupos = efecto principal para San Francisco		
$\bar{Y}_{(Sacra.)} - \bar{Y}_{(total)} = \$24 - \$22 = \$2$			$\bar{Y}_{(Diego)} - \bar{Y}_{(total)} = \$22 - \$22 = \$0$			$\bar{Y}_{(Fran.)} - \bar{Y}_{(total)} = \$20 - \$22 = -\$2$		
Columna (C):			Columna (C):			Columna (C):		
Desviación dentro del grupo = error $= Y - \bar{Y}_{(Sacra)}$			Desviación dentro del grupo = error $= Y - \bar{Y}_{(Diego)}$			Desviación dentro del grupo = error $= Y - \bar{Y}_{(Fran.)}$		

TABLA 12-5 | Tabla de fuentes de variación para el análisis de varianza con los datos de la tabla 12-4

Fuente de variación	SC	gl	Varianza de los cuadrados medios: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	40	$K - 1 = 2$	20	1.38
Dentro de grupos (SC_D)	174	$n - K = 12$	14.50	
Total (SC_T)	214	$n - 1 = 14$	15.29	

La tabla 12-4 presenta dicho escenario para los datos ficticios de las tres ciudades de California. Como ilustración, la gran media del GMD aún es de $22, pero las medias grupales no son significativamente diferentes de $22. Observa que en las distribuciones del GMD de los tres grupos, todas las puntuaciones se encuentran dentro de un rango similar, cerca de 20 y cerca de 30. Las medias se encuentran muy cerca, lo cual sugiere que las puntuaciones quizás provengan de la misma población (es decir, una muestra nacional de madres que viven de la asistencia social, cuyo GMD es de $22). De acuerdo con estos nuevos datos, los efectos principales de la ciudad de residencia [columnas (B)] son pequeños en comparación con los efectos dentro del grupo [columna (C)]. La tabla 12-5 representa la tabla de fuentes de variación de los datos de la tabla 12-4. Con estos efectos principales pequeños, la razón F es de apenas 1.39 en comparación con una razón F de 17.93 según los datos originales de la tabla 12-2.

La tabla de las fuentes de variación muestra que cuando las medias no son significativamente diferentes, la suma de cuadrados entre grupos SC_E es relativamente pequeña. A su vez, esto da como resultado una pequeña puntuación de la razón F. Además, nota que, ya que las 15 puntuaciones se acumulan en torno a la media total de $22, la suma total de cuadrados es relativamente pequeña.

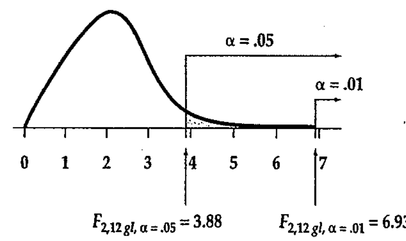
Volvamos a la figura 12-2. La parte (B) proporciona una ilustración gráfica de los datos de la tabla. Los efectos principales pequeños en la muestra sugieren que, de hecho, las medias poblacionales de los grupos son las mismas e iguales a la media total de $22. No hay una acumulación definida de las puntuaciones. Las pequeñas diferencias en el GMD medio de las ciudades son consecuencia del error de muestreo.

La razón F como distribución muestral

Como se observó anteriormente, la razón F se utiliza para determinar la significancia estadística con el ANOVA. La razón F constituye una distribución muestral, la cual puede describirse por medio de una curva como en la figura 12-3. Con el muestreo repetido, la razón F puede calcularse en el caso de todos los resultados posibles del muestreo *cuando la hipótesis nula es verdadera.* Estos resultados se representan por medio del área bajo la curva, que, como en el caso de las curvas de la distribución normal y de la distribución t, suman un total de 1.00, o 100%. Observa que la curva de la distribución F se encuentra sesgada y que todas las puntuaciones son positivas. Esto se debe a la operación de elevar al cuadrado en la ecuación de la razón F.

Con el ANOVA, la hipótesis nula consiste en que las medias poblacionales de los grupos son iguales. Cuando, de hecho, la hipótesis nula es verdadera y se muestrea repetidamente, las medias grupales de las muestras diferirán poco y los efectos principales calculados serán pequeños. Estos efectos principales pequeños —que se deben al error de muestreo aleato-

FIGURA 12-3
Valores críticos de la distribución F para 2 y 12 grados de libertad a los niveles de significancia de 0.05 y 0.01



rio— dan origen a un resultado pequeño de la razón F. En la curva de la distribución F, esta alta frecuencia de puntuaciones F pequeñas se hace evidente en la acumulación de los resultados muestrales en el extremo inferior de la curva (lo cual da origen al sesgo positivo que se encuentra a la derecha en la figura 12-3).

En el caso de las muestras relacionadas con las madres que viven de la asistencia social, la hipótesis nula consiste en que las madres de Charleston, Chicago y Boston tienen el mismo GMD promedio. Si esto resulta verdadero, el muestreo repetido proporcionará una descripción de los tamaños de las medias muestrales y de las diferencias de las medias que se presentan, por ejemplo, el 95% de los casos. Con el muestreo repetido, a veces Boston aparecerá en la parte superior de las ilustraciones. Otras veces, Charleston o Chicago tendrán medias muestrales un poco mayores. Si la hipótesis nula relativa al hecho de que no hay diferencias en las medias resulta verdadera, la mayor parte de las veces las medias de los tres grupos se encontrarán cerca de la media total. Cuando se calculan repetidamente efectos principales, sumas de cuadrados y razones F, los resultados mostrarán que la distribución de las razones F tiene un límite inferior de cero y que carece de límite superior.

Ahora bien, hemos señalado que el tamaño de la razón F tiene la influencia de los grados de libertad: el tamaño de la muestra y la cantidad de grupos. De esta manera, una distribución muestral de razones F adquiere su forma de acuerdo con los grados de libertad. Las tablas estadísticas D y E, del apéndice B, contienen los valores de la razón F con diversos grados de libertad para los niveles de significancia de 0.05 y 0.01, respectivamente. La tabla incluye los valores críticos de la razón F. En el margen superior se encuentran los grados de libertad para el CM_E, el numerador de la razón F. A la izquierda se localizan los grados de libertad para el CM_D, el denominador de la razón F. Los valores de la razón F en la tabla son como las puntuaciones t; son valores críticos de F para los niveles de significancia de 0.05 y 0.01. Por ejemplo, en los datos de las tablas 12-2 y 12-4, tenemos $(K - 1) = 2$ grados de libertad entre grupos y $(n - K) = 12$ grados de libertad dentro de los grupos. A partir de la tabla estadística D, se deduce que el valor crítico de la razón F para 2 y 12 grados de libertad al nivel de 0.05 es

$$\text{Valor crítico para } F_{2,\,12\,gl,\,\alpha = .05} = 3.88$$

Esto significa que cuando la hipótesis nula referente a medias grupales iguales es verdadera, con el muestreo repetido la razón F será igual o excederá 3.88 sólo el 5% de las veces. En la tabla estadística E, del apéndice B, el valor crítico correspondiente a la razón F con un nivel de significancia de 0.01 es de 6.93. La figura 12-3 identifica las regiones críticas e ilustra la forma de la distribución de la razón F para 2 y 12 grados de libertad.

En el caso de nuestros datos originales de la tabla 12-2, obtuvimos una razón F de 17.93. Esto es mayor que el valor crítico de la razón F al nivel de 0.01, un valor de 6.93. Por consiguiente, para el paso 4 de los seis pasos la inferencia estadística, el valor p calculado sería

$p < .01$. Y si estamos realizando la prueba con un nivel de significancia de 0.05, se rechaza la hipótesis nula puesto que

$$|F_{observada}| > |F_{\alpha}| \text{ (es decir, } 17.93 > 3.88)$$

De esta manera, $p < \alpha$, (es decir, $p < .05$).

Sin embargo, en el caso de nuestro ejemplo hipotético relacionado con el hecho de que no hay diferencias significativas (tablas 12-4 y 12-5), no rechazaríamos la hipótesis nula porque

$$|F_{observada}| < |F_{\alpha}| \text{ (es decir, } 1.39 < 3.88)$$

De esta manera, $p > \alpha$, (es decir, $p > .05$).

Estos datos no significativos ilustran el hecho de que en el muestreo repetido no es poco usual obtener efectos principales de \$2, \$0 y –\$2, cuando, de hecho, no hay diferencias entre las tres medias poblacionales.

Aspectos relevantes de una relación para el ANOVA

Existencia de la relación

La existencia de la relación para el ANOVA se determina utilizando la razón *F* para probar la hipótesis nula relativa a medias grupales iguales como se acaba de describir. (Los seis pasos completos de la inferencia estadística se presentarán más adelante.) La hipótesis alternativa indica que las medias grupales no son iguales. Si se rechaza la hipótesis nula y se acepta la hipótesis alternativa, existe una relación entre la ciudad de residencia y el GMD.

Existencia de una relación utilizando el ANOVA

Se prueba la hipótesis nula:

$$H_0: \bar{Y}_{(grupo\ 1)} = \bar{Y}_{(grupo\ 2)} = \bar{Y}_{(grupo\ 3)} = \ldots = \bar{Y}_{(total)}$$

Por tanto, los efectos principales = 0. Es decir que no existe relación entre *X* y *Y*.
Utilícese el estadístico de prueba de la razón *F*:

$$F = \frac{CM_E}{CM_D}$$

Dirección de la relación

Toma en cuenta que abordamos la dirección, fuerza y naturaleza de una relación entre variables sólo cuando existe dicha relación. Como la variable independiente en el ANOVA típicamente es de nivel de medición nominal, la dirección no tiene significado. Por ejemplo, no tiene sentido afirmar que la señora Jones se encuentra más en la dirección de Boston que en la de Chicago o Charleston. Ella es de Boston o no lo es. Esta falta de significado de la dirección también se encuentra implícita por el hecho de que el estadístico de la razón *F* siempre es de una cola, aunque no direccional, ya que sus cálculos implican elevar al cuadrado los signos negativos. Por tanto, la dirección de la prueba de ANOVA no es aplicable.

Dirección de una relación utilizando el ANOVA No es aplicable

Fuerza de la relación

Con el ANOVA, si se encuentra una relación en una muestra, sencillamente indica que los efectos principales son significativamente diferentes de cero en la población. La fuerza de la relación se encuentra en la pregunta independiente: ¿cuán grandes son estos efectos principales en términos prácticos? ¿Difieren un poco o mucho las medias grupales? Por ejemplo, si la ciudad de residencia se relaciona con el GMD, ¿mejora este hecho poco o mucho nuestra comprensión y predicción del GMD?

Una fuerte relación es aquella en la que el conocimiento de los efectos principales del grupo permite realizar predicciones precisas de la variable dependiente. Por ejemplo, supongamos que todas las madres que viven de la asistencia social en Boston gastan exactamente \$28 en GMD, las de Chicago gastan exactamente \$24 y las de Charleston gastan exactamente \$14. Las varianzas o dispersiones de las puntuaciones en torno a las medias grupales (es decir, las varianzas dentro de los grupos) serían cero. Conocer la ciudad de residencia sería un predictor perfecto del GMD y no habría error en las predicciones del GMD particular. En un caso tan poco probable, si se nos proporciona la ciudad de residencia de una madre, podemos predecir perfectamente su GMD: los efectos principales explicarán la varianza total del GMD. Una relación fuerte es aquella en la que una elevada proporción de la varianza total en la variable dependiente de intervalo/razón es explicada por la variable de grupo.

Existen diversas medidas de la fuerza de la relación para el ANOVA, pero con frecuencia ninguna se informa, ya que cada una de ellas se debe utilizar con cautela. Cualquier medida se puede encontrar sesgada como consecuencia del tamaño de la muestra y otros aspectos relacionados con el error de muestreo. Sin embargo, la media conservadora es la *razón de correlación*, ε^2 (que se pronuncia "épsilon al cuadrado") (Blalock, 1979, pp. 373-374). Toma en cuenta que estamos utilizando datos muestrales y que ε^2 resulta conservador por el hecho de que es poco probable que se sobreestime la fuerza de la relación en la población. La razón de correlación constituye una medida de la reducción proporcional del error (RPE). Ésta proporciona una idea del grado de exactitud con el que puede predecirse la variable dependiente (en este caso el GMD) utilizando el conocimiento de la variable independiente (en este caso, la ciudad de residencia). Su fórmula es la siguiente:

Cálculo de la razón de correlación ε^2 para medir la fuerza de una relación utilizando el ANOVA

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T}$$

donde

ε^2 = razón de correlación de la fuerza de la relación
CM_T = varianza total del cuadrado medio total
CM_D = varianza del cuadrado medio dentro de los grupos (o varianza no explicada)

Observa que esta fórmula no mide la proporción de la varianza explicada (es decir, CM_E) directamente, ya que el CM_E depende en gran medida del número de grupos utilizado en el cálculo. Más bien, considera una proporción de 1 (o 100%) como la varianza total por explicar y resta de ella la proporción no explicada (es decir, CM_D/CM_T). Además, con esta fórmula, la razón de correlación ε^2 posee límites definidos; ésta varía de cero a 1.00 y, por consiguiente, siempre es positiva. Cuando los efectos principales de la ciudad de residencia explican por completo la varianza del GMD, los efectos dentro de los grupos serán 0, lo cual deja al CM_D en 0, lo cual da como resultado una ε^2 de 1:

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T} = 1 - \frac{0}{CM_T} = 1 - 0 = 1$$

En el otro extremo, cuando los efectos del grupo son 0, el caso poco probable en el que todas las puntuaciones de los individuos sean iguales a la media total de $22 de GMD, el CM_E será igual a cero. Esto dará como resultado que el CM_D sea igual al CM_T. Por consiguiente,

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T} = 1 - 1 = 0$$

En situaciones reales, rara vez observamos relaciones perfectas entre variables. Una ε^2 caerá entre 0 y 1.00, y cuanto mayor sea, más fuerte es la relación. Por ejemplo, si utilizamos los datos de la tabla 12-3, la tabla de fuentes variación para la muestra de madres que viven de la asistencia social,

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T} = 1 - \frac{14.50}{49.57} = 1 - .2935 = .7065$$

En forma de porcentaje, esta cifra es (.7065)(100) = 70.65 por ciento. Podemos concluir que 70.65% de la varianza del GMD se explica por la ciudad de residencia. Desafortunadamente, la interpretación de ε^2 debe hacerse con cautela. *De hecho, no debe calcularse ni reportarse a menos que todos los grupos tengan aproximadamente el mismo número de casos y las varianzas dentro de cada grupo sean aproximadamente iguales.*

Fuerza de una relación utilizando el ANOVA

Se calcula el porcentaje de la varianza de Y explicada por el conocimiento de X utilizando la razón de correlación:

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T}$$

Aplicaciones prácticas de la relación

Si se determina que existe una relación entre las variables, podemos describir sus aplicaciones prácticas. Nos concentramos en lo esencial —las madres que viven de la asistencia social, sus ciudades de residencia y su GMD— y proporcionamos las mejores estimaciones de la variable dependiente. Sencillamente, ¿cómo permite el conocimiento de la relación entre la ciudad de residencia y el GMD obtener mejores predicciones del GMD de las madres que viven de la asistencia social?

TABLA 12-6 | Diagrama de árbol de las diferencias entre medias para las comparaciones de la prueba de rango

Medias grupales	*Diferencias*	
$\bar{Y}_{(Charleston)} = \$14$		
	$10	$14
$\bar{Y}_{(Chicago)} = \$24$		
	$4	
$\bar{Y}_{(Boston)} = \$28$		

Primero hacemos las mejores estimaciones en un nivel grupal reportando la media total, las medias grupales y los efectos principales. Segundo, proporcionamos ejemplos de las mejores estimaciones para los individuos. Finalmente, especificamos qué medias grupales son significativamente diferentes de otras. Esto requiere el cálculo adicional de lo que se denomina una *prueba de rango.*

Pruebas de rango Las pruebas de rango constituyen un paso adicional necesario con el ANOVA porque con el ANOVA el rechazo de la hipótesis nula sólo indica que *por lo menos dos* de las medias grupales son significativamente diferentes entre sí. En particular, cuando se tiene un número grande de grupos, las pruebas de rango proporcionan una forma más rápida que en el caso de una serie de pruebas t para identificar qué medias grupales son significativamente diferentes de otras.

Una prueba de rango determina el grado en que una diferencia entre medias (es decir, un rango de diferencias) es estadísticamente significativa. Si los tamaños de las muestras y las desviaciones estándar de las medias grupales no son muy diferentes, podemos comenzar suponiendo que la media grupal más pequeña y la más grande son significativamente diferentes, en el ejemplo del GMD, aquellas entre Charleston y Boston. Pero, quizás otras diferencias también sean significativas, como las que encontramos entre Charleston y Chicago. Una prueba de rango nos dice el grado al que deben estar alejadas dos medias grupales antes de suponer que son diferentes en las poblaciones. Las pruebas de rango también reciben el nombre de *pruebas de comparación múltiple.*

Existen diversas pruebas de rango. Una prueba conservadora consiste en la fórmula de la diferencia altamente significativa de Tukey, la prueba DHS (Tukey, 1953). La fórmula de DHS es conservadora en virtud de que es poco probable que por error nos diga que existe una diferencia cuando en realidad no es así.

Ya que las pruebas de rango comparan cada grupo con los demás, comenzamos por elaborar un diagrama de árbol que ordene las medias de menor a mayor e indique las diferencias entre las medias grupales (tabla 12-6). En la tabla 12-6 se observa que la diferencia entre Charleston y Chicago es de $10; la diferencia entre Charleston y Boston es de $14 y la diferencia entre Chicago y Boston es de $4. La DHS nos indica con exactitud la magnitud de la diferencia entre dos medias muestrales para suponer que realmente existe una diferencia entre las medias de las dos poblaciones. Por ejemplo, ¿es estadísticamente significativa una diferencia de $4?

Cálculo de la prueba de rango de la diferencia altamente significativa (DHS) de Tukey

$$DHS = q\sqrt{\frac{CM_D}{n_{(por\ grupo)}}}$$

donde

DHS = valor de la prueba de rango de la diferencia altamente significativa: ¿cuán grande debe ser la diferencia entre dos medias muestrales para suponer que la diferencia entre las medias poblacionales en realidad existe?

q = valor crítico de la prueba de rango en la tabla estadística F en el apéndice B, para un nivel específico de significancia y de grados de libertad

CM_D = varianza del cuadrado medio dentro de los grupos

$n_{(por\ grupo)}$ = tamaño de la muestra de los grupos (suponiendo que las n de las muestras son iguales); si la cantidad de individuos no es igual en todos los grupos, se deben llevar a cabo cálculos complicados para obtener un promedio del tamaño de los grupos

Grados de libertad (para utilizar con la tabla estadística F en el apéndice B)

gl para el $CM_D = n_{(total)} - K$ (listada a la izquierda de la tabla estadística F)

donde

$n_{(total)}$ = tamaño total de la muestra

K = número de grupos

gl para $n_{(por\ grupo)} = K$ (listada en el margen superior de la tabla estadística F como *número de grupos*)

donde

K = número de grupos

La tabla estadística F en el apéndice B proporciona los valores de q: los valores críticos para los niveles de significancia al .05 y .01. Para obtener q en el problema en cuestión, elijamos el nivel de significancia al .05. Hemos calculado los grados de libertad para el CM_D cuando llevamos a cabo la prueba de la razón F, y, en este caso, hay 12. Existen 3 grados de

libertad para $n_{(por\ grupo)}$. Al observar la tabla, el valor q para 12 y 3 grados de libertad es de 3.77. Por consiguiente, el cálculo de la DHS es

$$DHS = q\sqrt{\frac{CM_D}{n_{(por\ grupo)}}} = 3.77\sqrt{\frac{14.50}{5}} = 6.42$$

Por consiguiente, una diferencia de por lo menos $6.42 entre cualquier par de medias resulta estadísticamente significativa. En la tabla 12-6 vemos que el GMD promedio es significativamente diferente entre Charleston y Chicago (es decir, $10 > $ 6.42) y entre Charleston y Boston (es decir, $14 > $6.42). El GMD de Chicago no es significativamente diferente del de Boston (es decir $4 ≈ $6.42). Con el cálculo de la prueba de rango, tenemos todo lo que se requiere para abordar las aplicaciones prácticas de la relación entre la ciudad de residencia y el GMD. Ahora apliquemos los seis pasos de la inferencia estadística.

Aplicaciones prácticas de una relación utilizando el ANOVA

1. Proporciona las mejores estimaciones de la media total, las medias grupales y los efectos principales.
2. Proporciona ejemplos de las mejores estimaciones de Y para casos individuales en la población:

$$\begin{aligned} Y'_{(cada\ caso)} &= \bar{Y}_{(total)} + \text{efecto (explicado) de } X \\ &= \bar{Y}_{(total)} + \text{efecto principal del grupo } X \end{aligned}$$

3. Utilice pruebas de rango para determinar qué medias grupales son significativamente diferentes entre sí.

Los seis pasos de la inferencia estadística para el ANOVA de un factor

Ahora que hemos adquirido un sentido de la lógica del ANOVA de un factor, aplicamos los seis pasos de la inferencia estadística. Primero repasemos los criterios para seleccionar el ANOVA.

Cuándo utilizar el análisis de varianza de un factor (ANOVA) para probar diferencias de medias entre tres o más grupos (distribución *F*)

En general: Se prueba una hipótesis entre una variable independiente nominal/ordinal con tres o más categorías y una variable dependiente de intervalo/razón.

1. Número de variables, muestras y poblaciones: *a*) una población con una sola variable dependiente de intervalo/razón; se comparan medias para tres o más grupos de una variable independiente nominal/ordinal. La muestra de cada grupo debe ser representativa de su subpoblación. O, *b*) una sola variable dependiente

de intervalo/razón cuya media se compara entre tres o más poblaciones utilizando muestras representativas.

2. Tamaño de la muestra. Por lo general no implica ningún requisito. No obstante, la variable dependiente de intervalo/razón no debería estar demasiado sesgada dentro de cualquier muestra grupal. Además, las pruebas de rango no son confiables a menos que los tamaños muestrales de los grupos sean aproximadamente iguales. Estas restricciones son menos importantes cuando los tamaños muestrales de los grupos son grandes.
3. Las varianzas (y desviaciones estándar) de los grupos son iguales. Ésta es la misma limitación para la prueba t (véase el material relacionado con la igualdad de las varianzas en el capítulo 11).

El requisito de igualdad de las varianzas es necesario en cualquier prueba de diferencia de medias, incluyendo la prueba t y la prueba de la razón F. Una gran dispersión de las puntuaciones en un grupo puede conducirnos a creer equivocadamente que existe una diferencia entre las medias cuando, en realidad, no es el caso. Recuerda que en el capítulo 11 se indicó que cuando existe una diferencia grande entre las varianzas de los grupos (y, por consiguiente, entre las desviaciones estándar), los grados de libertad deben ajustarse. Con el ANOVA ésta resulta una operación difícil que conviene realizar con la computadora. Sin embargo, afortunadamente con el ANOVA y la prueba de la razón F, si las muestras son grandes (cada grupo > 30), es menos probable que las varianzas diferentes influyan en los resultados de la prueba.

Breve lista de verificación de los seis pasos de la inferencia estadística

PREPARACIÓN DE LA PRUEBA

Formula la pregunta de investigación. Elabora diagramas conceptuales que describan las especificaciones, incluyendo las poblaciones y muestras de estudio, las variables (por ejemplo, $X = ...$, $Y = ...$) y sus niveles de medición, así como los estadísticos y parámetros calculados o dados. Indica el procedimiento de la prueba estadística adecuado.

SEIS PASOS

Se utiliza la letra H para representar la *hipótesis*:

1. Indica la H_0 y la H_A y estipula la dirección de la prueba.
2. Describe la distribución muestral.
3. Determina el nivel de significancia (α) y especifica el valor crítico de la prueba.
4. Observa los resultados de la muestra en cuestión y calcula los efectos de la prueba, el estadístico de la prueba y el valor p.
5. Toma la decisión de rechazo.
6. Interpreta y aplica los resultados y proporciona las mejores estimaciones en términos comunes.

Solución para el análisis de varianza de un factor (ANOVA; distribución F)

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿existe una relación entre la ciudad de residencia y el gasto mensual en diversión (GMD)? Es decir, ¿existen diferencias significativas en el GMD promedio entre las madres que viven de la asistencia social en Boston, Chicago y Charleston? *Especificaciones*: variables: Y = GMD, la variable dependiente, nivel de razón; X = ciudad de residencia, una variable nominal con tres categorías. *Población*: madres que viven de la asistencia social en tres ciudades. *Muestra:* $n = 15$ madres que viven de la asistencia social en Boston ($n = 5$), Chicago ($n = 5$) y Charleston ($n = 5$). *Procedimiento estadístico*: ANOVA de un factor, prueba F de diferencia entre tres o más medias muestrales; se asumen varianzas iguales del GMD en las subpoblaciones de las ciudades. *Observación*: Datos en la tabla 12-1. La pregunta de investigación se describe en la figura 12-1.

SEIS PASOS

1. H_0: $\mu_{Y(Boston)} = \mu_{Y(Chicago)} = \mu_{Y(Charleston)} = \mu_{Y(total)}$

   Por consiguiente, los efectos principales = 0.
   Es decir, no existe relación entre la ciudad de residencia y el GMD.

   H_A: $\mu_{Y(Boston)} \neq \mu_{Y(Chicago)} \neq \mu_{Y(Charleston)} = \mu_{Y(total)}$

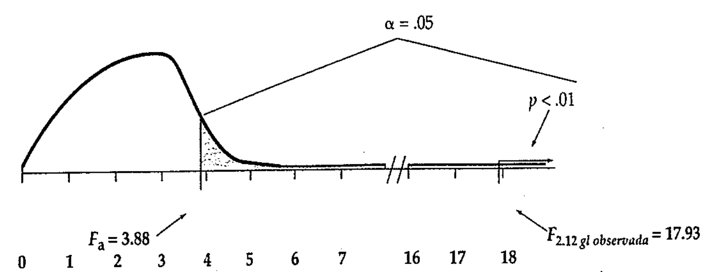
   Por consiguiente, los efectos principales $\neq$ 0.
   Es decir, existe una relación entre la ciudad de residencia y el GMD.

2. *Distribución muestral*: si la H_0 es verdadera y se extraen repetidamente muestras de tamaño 5 de las poblaciones de madres que viven de la asistencia social a las tres ciudades, la distribución adquiere la forma de la distribución F con

$$gl_B = K - 1 = 3 - 1 = 2$$

y

$$gl_D = n - K = 15 - 3 = 12$$



3. *Nivel de significancia*: $\alpha = .05$ (no direccional). $F_\alpha = 3.88$ crítica (para 2 y 12 *gl*, de la tabla estadística D, apéndice B).
4. *Observaciones*: de la hoja de cálculo en la tabla 12-1.

Efectos de la prueba: primero calcula las medias y la variación total:

$$\text{Media total} = \overline{Y}_{(total)} = \$22$$

Medias grupales:

$$\overline{Y}_{(Boston)} = \$28 \qquad \overline{Y}_{(Chicago)} = \$24 \qquad \overline{Y}_{(Charleston)} = \$14$$

$$\text{Variación total: } SC_T = \Sigma(Y_{(cada\ caso)} - \overline{Y}_{(total)})^2 = 694$$

Segundo, calcula los efectos principales:

$$\text{Efecto principal para Boston: } \overline{Y}_{(Boston)} - \overline{Y}_{(total)} = \$28 - \$22 = \$6$$

$$\text{Efecto principal para Chicago: } \overline{Y}_{(Chicago)} - \overline{Y}_{(total)} = \$24 - \$22 = \$2$$

$$\text{Efecto principal para Charleston: } \overline{Y}_{(Charleston)} - \overline{Y}_{(total)} = \$14 - \$22 = -\$8$$

Tercero, calcula la suma de cuadrados entre los grupos y dentro de los grupos:

$$\begin{aligned}\text{Suma de cuadrados entre grupos} = SC_E &= \Sigma(\overline{Y}_{(grupo)} - \overline{Y}_{(total)})^2 \\ &= \Sigma[(n_{(grupo)})\text{ (efecto del grupo}^2)] \\ &= (5)(6)^2 + (5)(2^2) + (5)(-8^2) \\ &= 180 + 20 + 320 = 520\end{aligned}$$

$$\text{Suma de cuadrados dentro de los grupos} = SC_D = SC_T - SC_E = 694 - 520 = 174$$

Cuarto, calcula las varianzas de los cuadrados medios (utilizando los grados de libertad del paso 2):

$$\text{Varianza del cuadrado medio entre grupos} = CM_E = \frac{SC_E}{K-1} = \frac{520}{2} = 260$$

$$\text{Varianza del cuadrado medio dentro de los grupos} = CM_D = \frac{SC_D}{n-K} = \frac{174}{12} = 14.5$$

Quinto, calcula el estadístico de la prueba:

$$\text{Estadístico de la prueba} = F = \frac{CM_E}{CM_D} = \frac{260}{14.50} = 17.93$$

Sexto, realiza un resumen en una tabla de fuentes de varización:

Fuente de variación	***SC***	***gl***	**Varianza de los cuadrados medios: *CM = SC/gl***	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	520	$K-1=2$	260	17.93
Dentro de grupos (SC_D)	174	$n-K=12$	14.50	
Total (SC_T)	694	$n-1=14$	49.57	

Séptimo, calcula el valor *p* (utilizando las tablas estadísticas D y E del apéndice B): *valor p* [efectos principales tan inusuales o más inusuales que los que se observan cuando, de hecho, no hay diferencias entre las medias grupales] <.01. (Este valor *p* se indica en la curva en el paso 2.)

5. Decisión de rechazo: $|F_{observada}| > |F_\alpha|$ (es decir, $17.93 > 3.88$)

Por consiguiente, $p < \alpha$ (es decir, $p < .05$). Rechaza H_0 y acepta H_A en el nivel de confianza de 95%.

6. *Interpretación*: aspectos de la relación y mejores estimaciones.

Existencia: existe una relación entre la ciudad de residencia y el GMD;
Razón F = 17. 93; $p < .01$.
Dirección: no aplicable.
Fuerza:

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T} = 1 - \frac{14.50}{49.57} = 1 - .2935 = .7065$$

(.7065)(100) = 70.65 por ciento.

Por consiguiente, el 70.65% de la varianza del GMD es explicado por el conocimiento de la ciudad de residencia.

Aplicaciones prácticas:

a) Medias y efectos principales: media total = \$22; medias grupales: Boston = \$28, Chicago = \$24 y Charleston = \$14. Efectos principales: Boston = \$6, Chicago = \$2, Charleston = –\$8

b) La mejor estimación del GMD de las madres que viven de la asistencia social es

$$\begin{aligned}Y'_{(cada\ caso)} &= \overline{Y}_{(total)} + \text{efecto de } X \text{ (explicado)} \\ &= \$22 + \text{efecto principal del grupo } X\end{aligned}$$

(donde Y' es una estimación calculada de Y).

Por ejemplo, la mejor estimación del GMD para la señora Jones de Boston es \$22 + efecto de Boston = \$22 + \$6 = \$28.

c) *Prueba de rango:* el GMD medio de las madres que viven de la asistencia social en Charleston es significativamente diferente al de las madres en Boston y Chicago (sobre la base de la DHS de Tukey y la tabla 12-6 anterior).

Respuesta a la pregunta de investigación: existe una relación entre la ciudad de residencia y los gastos mensuales en diversión entre las madres que viven de la asistencia social. Las madres en Boston y Chicago tienen un gasto mensual promedio mayor en diversión que las de Charleston.

TABLA 12-7 | Ingreso mensual de asistencia social, distribución del gasto mensual y déficit presupuestal de las madres que viven de la asistencia social entre ciudades

Gastos/ingreso promedio mensual	Boston $n = 75$	Chicago $n = 75$	Charleston $n = 75$	Significancia
Ingreso mensual de asistencia social	$696	$599	$493	†
Gasto total medio	$927	$1 003	$891	†
Vivienda	239	289	224	†
Alimentos	217	288	249	†
Otros gastos esenciales	372	365	372	*
En diversión	28	24	14	†
Otros gastos no esenciales	70	37	31	†
Déficit presupuestal mensual medio	$231	$404	$398	†

Nota: Gastos en dólares para 1991. Las categorías quizás no sumen el total como consecuencia del error de redondeo.

* $p < .05$

† $p < .01$

Pruebas F no direccionales de diferencias entre medias.

Fuente: Modificada de Edin y Lein (1977). Reimpreso con autorización de la American Sociological Association.

Presentación tabular de resultados

En un estudio real como el de Edin y Lein (1997), los resultados del ANOVA pueden presentarse al público o a audiencias profesionales en un formato como el que aparece en la tabla 12-7. Este formato destaca una comparación de grupos, mientras que la tabla incluye diversas variables dependientes. Los grupos se enlistan a lo largo del margen superior y las variables en la columna izquierda. Aunque algunas de las cifras corresponden a las de Edin y Lein, la tabla 12-7 incluye datos ficticios.

Aplicaciones multivariadas del modelo lineal general

Aunque los cálculos rebasan el alcance de este libro, vale la pena destacar la utilidad del modelo lineal general en lo que se refiere a la reducción proporcional del error cuando se utilizan diversas variables independientes para explicar una variable dependiente de intervalo/razón. Cada variable independiente adicional incrementa la precisión de las mejores estimaciones de la variable dependiente y reduce el error predictivo. Por ejemplo, supongamos que tenemos datos adicionales relativos a la señora Jones. Descubrimos que ella no cuenta con transporte personal y que sus hijos comen mucho, dos variables que se sabe afectan el GMD familiar. Tiene un GMD especialmente alto. Con los datos adicionales, explicamos esta puntuación alta de Y de la siguiente manera:

$$Y_{(Sra.\ Jones)} = \overline{Y}_{(total)} + [Y_{Sra.\ Jones} - \overline{Y}_{(total)}]$$
$$= \overline{Y}_{(total)} + \text{efecto de } X_1 + \text{efecto de } X_2 + \text{efecto de } X_3 + \text{error}$$

Esto es,

$$\$33 = \$22 + \$6 + \$2 + \$1 + \$2$$
$$= \$22 + \$9 + \$2$$

donde Y = GMD; X_1 = ciudad de residencia, efecto principal de Boston = \$6; X_2 = efecto principal de no tener transporte personal = \$2 y X_3 = efecto principal de tener niños que comen mucho = \$1. Dichos efectos combinados explican \$9 de su desviación de \$11 por encima de la media total. Solamente \$2 del GMD de la señora Jones de \$33 quedan sin explicación. De manera ideal, si se identifican todas las variables que influyen en el GMD, es posible explicar y predecir perfectamente el GMD de cada madre que vive de la asistencia social, lo cual reduce el error predictivo a cero.

Cuando consideramos más de una variable predictora, la fuerza de una relación se torna particularmente importante. Al comparar la fuerza de diversas relaciones, podemos establecer qué variables predictoras resultan mejores. Hemos encontrado que la ciudad de residencia mejora las predicciones del GMD; sin embargo, puede haber otras variables que caracterizan la vida de las madres que viven de la asistencia social que resultan más adecuadas para mejorar las predicciones. Por ejemplo, el GMD puede estar relacionado con el ingreso, la cantidad de ayuda que se recibe del estado, la cobertura del seguro de salud y otros beneficios. Como ya lo señalamos, el cálculo de la fuerza de la relación para la ciudad de residencia al predecir el GMD no resulta de valor en sí mismo. La razón de correlación ε^2 resulta más útil cuando se calcula y compara para diversas variables dependientes.

El análisis multivariado también toma en cuenta las relaciones entre las variables independientes. Por ejemplo, supongamos que no solamente la señora Jones, sino casi todas las madres de Boston, carecen de transporte personal. Este efecto ya se combinó con el efecto de Boston de \$6. El análisis multivariado permite separar los efectos de la prueba de diversas variables independientes, incluso cuando algunas se relacionan con otras.

Semejanzas entre la prueba t y la prueba de la razón F

Tanto la prueba t como la prueba de la razón F comparan medias grupales. De hecho, incluso una comparación de medias de dos grupos puede probarse con el estadístico de la razón F, en lugar de hacerlo con una prueba t. Como lo hemos destacado en este capítulo, la prueba de la razón F pondera la varianza explicada con respecto a la varianza no explicada:

$$F = \frac{\text{varianza explicada}}{\text{varianza no explicada}}$$

Cuando la hipótesis nula es verdadera (es decir, cuando no existe una diferencia significativa entre las medias), la varianza no explicada en el denominador es, en esencia, un error estándar. Describe la varianza del error de muestreo que puede esperarse cuando la variable independiente no tiene efecto. Asimismo, en el caso de la prueba t, el denominador del estadístico de la prueba es un error estándar, que describe la variación del error cuando la hipótesis nula es verdadera. Además, la variación en las diferencias entre las medias se calcula en el numerador de ambas pruebas, aunque de forma un tanto diferente. Estas similitudes matemáticas son tales que una prueba de razón F para una comparación de dos grupos es igual al resultado de la prueba t elevado al cuadrado y, por lo tanto, t es igual a la raíz cuadrada de la razón F. De esta manera,

$$F = t^2 \qquad t = \sqrt{F}$$

Esta consistencia entre las medidas también revela que la distribución de la razón F es otra distribución aproximadamente normal, lo que refleja la tendencia natural de los resultados muestrales de tomar una forma predecible.

Insensatez y falacias estadísticas: individualización de los hallazgos grupales

Comparar las medias de diversos grupos resulta muy informativo. Sin embargo debe tenerse precaución en recordar que los resultados constituyen generalizaciones estadísticas: enunciados que se aplican a poblaciones y subgrupos, no a individuos (véase el capítulo 2). Nuestro análisis se concentra en los promedios entre grupos. Se debe evitar considerar los resultados de forma estereotipada, suponiendo, por ejemplo, que cada madre que vive de la asistencia social en Boston gasta más que las madres de las demás ciudades. De hecho, en las puntuaciones individuales de la tabla 12-1 se observa que una de las madres de Charleston tiene el mismo GMD que una de Chicago. Aunque las medias son significativamente diferentes, existe una superposición en las distribuciones de las puntuaciones de las tres ciudades. Es más adecuado aplicar el análisis estadístico a grupos que a individuos.

La imaginación estadística hace hincapié en la percepción del fenómeno general: observar el bosque (grupo), así como los árboles (individuos). Analizar una hipótesis única, sola y fuera del contexto de una teoría más general, representa solamente un paso en la dirección del entendimiento completo. El análisis multivariado, la siguiente fase del estudio posterior a este curso, resulta esencial para obtener una comprensión completa. Cuando se consideran y controlan otras variables, la perspectiva de los individuos cambia hacia los sistemas sociales más amplios. Por ejemplo, los resultados de nuestro ANOVA del GMD nos dicen mucho respecto de las ciudades, así como de los residentes que viven de la asistencia social cuando se relaciona con otros resultados. Por ejemplo, observemos que en la tabla 12-7, además de los gastos, se incluyen los ingresos mensuales y déficits presupuestales (el déficit entre gastos e ingresos). Los gastos totales son bajos en el caso de Charleston, aunque también lo es el ingreso total de asistencia social. Podríamos plantear la hipótesis de que el GMD bajo de las madres de Charleston constituye una combinación del bajo costo de la diversión y un menor ingreso para gastar en ésta. La cantidad que gasta una madre en diversión puede tener poco que ver con su actitud con respecto a la diversión o con su moderación al administrar el escaso presupuesto familiar. En cambio, las diferencias que encontramos entre las tres ciudades quizás tengan que ver con las características de las ciudades, no con las de las madres.

Otro aspecto relacionado con la interpretación correcta consiste en evitar la *falacia ecológica*. Este error consiste en llegar a conclusiones relacionadas con el comportamiento individual sobre la base del análisis de grupos, como familias, comunidades, estados y naciones. La ecología es el estudio del ambiente, la realidad externa a los seres humanos. El análisis ecológico conduce a conclusiones relacionadas con estas grandes entidades, no necesariamente con los individuos que las conforman. Hace más de cien años, Emile Durkheim, uno de los primeros sociólogos con orientación científica, realizó un excelente estudio ecológico relacionado con las tasas de suicidio (Durkheim, 1951 [1897]). Él encontró que los condados predominantemente católicos de Europa tenían tasas de suicidio más bajas que los condados predominantemente protestantes. Una primera mirada a estos datos sugiere lo siguiente: 1) la tasa de suicidios entre los católicos es más baja; 2) las prohibiciones religiosas relacionadas con el suicidio en la teología católica desalientan la comisión del acto. Sin embargo, estas conclusiones se concentran en la elección individual y pueden resultar engañosas o equívocas. De hecho, podría suceder que la tasa de suicidios entre los católicos sea superior. Tal vez los católicos en los condados predominantemente protestantes cometen suicidio para escapar de la opresión. Las bajas tasas de suicidios en los condados católicos tal vez sólo indiquen la ausencia de opresión.

Para evitar la falacia ecológica, Durkheim interpretó sus hallazgos con cautela y se concentró en los sistemas sociales de los condados, no en las características individuales de las víctimas de suicidio. Por medio de un control cuidadoso de muchas variables, el estudio de Durkheim evita la falacia ecológica. Durkheim ideó un argumento convincente relacionado con el hecho de que las tasas de suicidio tienen una gran influencia, no tanto del punto de vista de los individuos, sino de los sistemas sociales en los que éstos residen. Argüía que la religión católica, con el énfasis que da a los intereses de la comunidad sobre los de los individuos, alentaba la solidaridad social y proporcionaba una red de asistencia social que ayudaba a los individuos en tiempos difíciles. Su tesis principal era que el suicidio no constituye un acto individual, sino un fenómeno social.

La imaginación estadística requiere que las conclusiones siempre consideren la perspectiva más amplia. Deben evitarse las conclusiones estereotipadas. Las generalizaciones estadísticas deben interpretarse por lo que son: promedios, enunciados sumarios sobre un grupo o una categoría de personas. El ANOVA, con su énfasis en la varianza entre grupos y dentro de ellos, revela claramente que un promedio constituye solamente un aspecto del trabajo estadístico básico. Los individuos varían con respecto a los promedios, e ignorar esta variación implica una falta de imaginación estadística.

RESUMEN

1. El análisis de varianza (ANOVA) se aplica para comparar tres o más medias grupales. En lugar de comparar cada media grupal con las demás (como harías en una prueba t), el ANOVA compara cada media grupal con la media total, que es la media para todos los casos en la muestra.
2. En el ANOVA, las diferencias entre cada media grupal y la media total constituyen los efectos de la prueba, y reciben el nombre de *efectos principales*. Cuando los efectos principales son cero, no existen diferencias entre las medias. En la prueba de hipótesis, la hipótesis nula establece que las medias son iguales, lo cual quiere decir que los efectos principales son iguales a cero.
3. El ANOVA hipotetiza acerca de las diferencias entre medias, pero sus cálculos se basan en la explicación de la varianza con respecto a la media total. Recordemos que la diferencia entre una puntuación individual y la media total constituye una puntuación de desviación. El ANOVA determina si resulta factible decir que parte de una puntuación de la desviación individual puede explicarse por la pertenencia del individuo a una categoría (grupo) de la variable independiente. Por consiguiente, el foco en el ANOVA consiste en la explicación de las puntuaciones de desviación. Recuerda que las puntuaciones de desviación se elevan al cuadrado y se promedian para obtener la varianza. De ahí el nombre de *análisis de varianza*.
4. El modelo lineal general establece que la mejor predicción de cualquier variable dependiente, Y, es su media total más un ajuste para los efectos de una

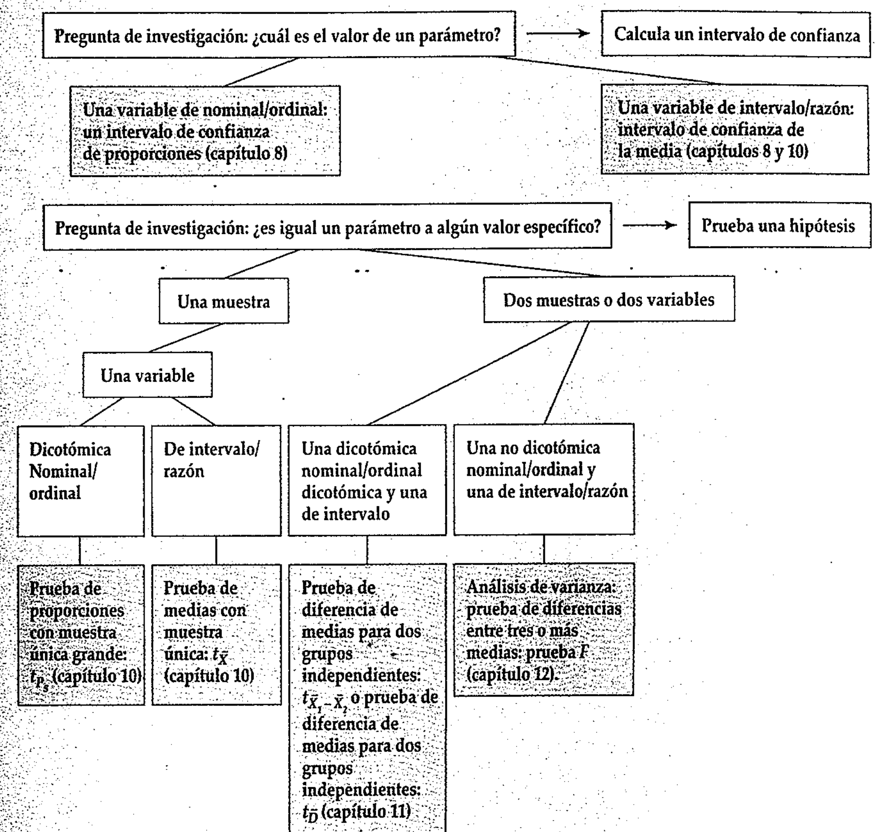
variable independiente, X. Una puntuación individual se descompone de la siguiente manera: Y = la media total más el efecto explicado de una categoría de X, más el error no explicado. Estas partes individuales de una puntuación se elevan al cuadrado, se suman y se promedian entre los grados de libertad para obtener las varianzas para todos los casos. La razón de la varianza explicada y la no explicada comprende la razón F, que es el estadístico de prueba para el ANOVA. Los cálculos para el ANOVA se resumen en una *tabla de fuentes de variación.*

5. El valor p se determina utilizando curvas de distribución F, apéndice B, tablas D y E.
6. Aspectos relevantes de la relación para el ANOVA: *a)* la existencia se determina utilizando la razón F para probar la hipótesis nula relacionada con la igualdad de medias grupales. Cuando se rechaza H_0, existe una relación. *b)* La dirección no es aplicable (porque la variable dependiente es nominal). *c)* La fuerza se establece con la razón de correlación, ε^2 (epsilon al cuadrado). Una relación fuerte es aquella en la que una alta proporción de la varianza total en la variable dependiente de intervalo o de razón se explica por la variable de agrupación. *d)* Las aplicaciones prácticas se describen mediante: 1) las mejores estimaciones a nivel de los grupos reportando la media total, las medias grupales y los efectos principales; 2) proporcionando ejemplos de las mejores estimaciones para los individuos; 3) utilizando pruebas de rango para especificar qué medias grupales son significativamente diferentes de otras.
7. Con el ANOVA, el rechazo de la hipótesis nula simplemente indica que por lo menos dos de las medias grupales son significativamente diferentes. Las pruebas de rango determinan específicamente qué pares de medias difieren entre sí. Las pruebas de rango establecen qué tanta diferencia entre medias resulta estadísticamente significativa. Entre las diversas pruebas de rango disponibles, una de naturaleza conservadora es la Diferencia Altamente Significativa de Tukey (HSD).
8. La falacia ecológica es el error que consiste en obtener conclusiones de un comportamiento individual sobre la base de un análisis de grupos. Porque un individuo pertenece a un grupo con determinada media, esto no significa que todo individuo se ajuste a esta puntuación. Los estadísticos muestrales se aplican al grupo, y aplicarlos a un individuo es estereotipar.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 12 del material del texto disponibles en el sitio web *The Statistical Imagination*, www.mhhe.com/ritchey2 incluyen una breve introducción al ANOVA multivariado, que recibe el nombre de ANOVA de N factores.

PROCEDIMIENTOS ESTADÍSTICOS CUBIERTOS HASTA AQUÍ



FÓRMULAS PARA EL CAPÍTULO 12

El análisis de varianza de un factor (ANOVA), prueba F de la diferencia de medias entre tres o más grupos:

Especificaciones: la comparación de una variable dependiente Y de intervalo/razón entre tres o más grupos obtenidos de: 1) las categorías de una variable X nominal/ordinal de una muestra y población; 2) tres o más poblaciones y muestras.

Pregunta de investigación: ¿son diferentes las medias de Y en las poblaciones de los grupos?

$$H_0: \mu_{Y_1} = \mu_{Y_2} = \mu_{Y_3} = \ldots = \mu_{Y(total)}$$

Distribución muestral: distribución F con $gl_E = K - 1$ y $gl_D = n - K$
Efectos de la prueba: calcula los efectos principales:

$$\text{Efecto principal de la media grupal} = \overline{Y}_{(grupo)} - \overline{Y}_{(total)}$$

Estadístico de prueba: distribución F [para utilizar con la tabla F (tablas estadísticas D y E, apéndice B); determinar si existe una relación]:
Cálculos para generar una tabla de fuentes de las variaciones con el estadístico de la razón F:
Cálculo de las sumas de cuadrados:

$$SC_T = \Sigma[Y_{(cada\ caso)} - \overline{Y}_{total}]^2$$

$$SC_E = \Sigma(\overline{Y}_{(grupo)} - \overline{Y}_{(total)})^2 = \Sigma[(n_{(grupo)})\ (\text{efecto de grupo}^2)]$$

$$SC_D = SC_T - SC_E$$

Cálculos de las varianzas de los cuadrados medios:

$$CM_E = \frac{SC_E}{gl_E} = \frac{SC_E}{K-1}$$

$$MSV_D = \frac{SS_D}{gl_D} = \frac{SS_D}{n-K}$$

El estadístico de la razón F mismo:

$$F = \frac{CM_E}{CM_D}$$

Tabla de fuentes de variación:

Fuente de variación	SC	gl	Varianza de los cuadrados medios: CM = SC/gl	$F = \frac{CM_E}{CM_D}$
Entre grupos	SC_E	$K - 1$	CM_E	$F = \frac{CM_E}{CM_D}$
Dentro de grupos	SC_D	$n - K$	CM_D	
Total	SC_T	$n - 1$	CM_T	

Determinación de los aspectos de una relación:
Existencia: prueba de la razón F
Dirección: por lo general no aplicable
Fuerza: razón de correlación ε^2:

$$\varepsilon^2 = 1 - \frac{CM_D}{CM_T}$$

Aplicaciones prácticas:

1. Reporta las medias grupales y los efectos principales
2. Mejor estimación de Y:

$$Y'_{(cada\ caso)} = \overline{Y}_{(total)} + \text{el efecto (explicado) de } X.$$

3. Pruebas de rangos DHS de Tukey para determinar qué medias de grupos son diferentes entre sí:

$$DHS = q\sqrt{\frac{CM_D}{n_{(por\ grupo)}}}$$

PREGUNTAS PARA EL CAPÍTULO 12

1. Tanto la prueba t (capítulo 11) como el análisis de varianza (ANOVA) permiten establecer diferencias entre medias. Explica la diferencia de enfoque entre las dos pruebas estadísticas.
2. Especifica la fórmula del modelo lineal general. Explica qué significa cada parte del modelo.
3. Describe por lo menos tres formas de enunciar la hipótesis nula para una prueba de ANOVA.
4. ¿Cuáles son los efectos principales en la prueba de ANOVA? ¿Para qué aspecto de una relación son importantes?
5. En términos comunes, explica qué son las variaciones dentro de los grupos y entre los grupos.
6. Supongamos que se prueba la hipótesis relacionada con el hecho de que las medias de una variable difieren en lo que se refiere a poblaciones y después se rechaza la hipótesis nula. ¿Garantiza esto que cada una de las cuatro medias sea significativamente diferente de las otras tres? Explica.
7. ¿Qué es la falacia ecológica? Cita un ejemplo.

EJERCICIOS PARA EL CAPÍTULO 12

Conjunto de problemas 12A

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Por consistencia, redondea los cálculos a dos decimales. Utiliza $\alpha = 0.05$, a menos que se estipule otra cosa.

12A-1. Con los siguientes datos, calcula los efectos principales de la facultad de pertenencia sobre el promedio (GPA) de los alumnos. Muestra símbolos y fórmulas.

Facultad	Promedio GPA
Artes y ciencias	2.74
Ingeniería	2.54
Negocios	2.62
General	2.63

12A-2. El modelo lineal general ordena las puntuaciones de una variable dependiente de intervalo/razón (Y) en partes explicadas y partes no explicadas por una variable independiente (X). Con los siguientes estadísticos, aplica este modelo para explicar las puntuaciones Y que aparecen en la lista. Y = meses en prisión; X = tipo de delito.

$\overline{Y}_{(total)} = 55$ meses $\overline{Y}_{(asalto\ a\ mano\ armada)} = 71$ meses

$\overline{Y}_{(robo\ simple)} = 27$ meses $\overline{Y}_{(homicidio)} = 133$ meses

Número de identificación del preso	X	Y
1	Robo simple	22
2	Homicidio	156
3	Asalto a mano armada	79
4	Homicidio	131
5	Asalto a mano armada	67
6	Robo simple	37

12A-3. Con frecuencia, la ideología política se mide en una escala conservador-liberal. Supón que tú cuentas con una escala de ideología con un nivel de intervalo, con puntuaciones que van de cero (extremadamente conservador) a 15 (extremadamente liberal). De acuerdo con los siguientes datos de una encuesta, prevé la hipótesis relacionada con el hecho de que varias categorías raciales o étnicas difieren en lo que se refiere a ideología política. Asume la igualdad de las varianzas poblacionales.

Categoría racial o étnica	Puntuación de ideología conservador-liberal
Angloamericano	7
Angloamericano	8
Angloamericano	7
Angloamericano	5
Angloamericano	4
Angloamericano	8
Angloamericano	4
Afroamericano	10
Afroamericano	10
Afroamericano	7
Afroamericano	8
Afroamericano	9
Afroamericano	8
Afroamericano	10
Hispano	8
Hispano	7
Hispano	10
Hispano	9
Hispano	8
Hispano	12
Hispano	11

12A-4. En Estados Unidos, aproximadamente una de cada cuatro personas tiene obesidad, sobrepeso serio que pone a una persona en riesgo de sufrir efectos físicos adversos en su salud, como diabetes y enfermedades del corazón. La obesidad también tiene efectos psicológicos adversos, como hacer sentir mal a las víctimas por la impresión que sus cuerpos provocan en otros (Friedman y Bronwell, 1995). Supongamos que se comparan tres grupos de personas con diferente peso en una escala de insatisfacción corporal, instrumento de sondeo con un nivel de intervalo/razón, con puntuaciones que van de 0 a 30. Tomando en cuenta la altura, género y complexión de los individuos, se les clasifica como normales, casi obesos (20% a 30% por encima del peso normal) y obesos (más del 30% por encima del peso normal). ¿Afecta la obesidad la satisfacción con respecto a la apariencia corporal? Asume la igualdad de las varianzas poblacionales.

Grupo de peso	Escala de puntuación de insatisfacción corporal
Rango normal	11
Casi obeso	15
Casi obeso	13
Obeso	16
Rango normal	9
Casi obeso	14
Obeso	19
Obeso	17
Rango normal	13
Casi obeso	16
Obeso	15
Rango normal	12
Casi obeso	11
Obeso	15
Rango normal	10

12A-5. En la mayoría de las prisiones existe una variedad de tratamientos y programas de rehabilitación, como la asesoría relacionada con el abuso de sustancias, la asesoría psicológica y espiritual, así como programas académicos y vocacionales. Una cuestión interesante radica en saber si los oficiales de correcciones de distintas razas se oponen a dichos programas orientados a los reclusos y, por esa razón, adoptan una actitud más punitiva hacia el cumplimiento de una condena (Jackson y Ammen, 1996). De acuerdo con los siguientes estadísticos, prueba la hipótesis de que existen diferencias entre las puntuaciones obtenidas con una escala de actitud punitiva entre los oficiales blancos, afroamericanos e hispanos de las correccionales. Asume la igualdad de varianzas poblacionales.

Raza	Media	Desviación estándar	n
Blanco	27.90	3.09	30
Afroamericano	21.77	3.39	30
Hispano	25.58	3.03	30
Total	25.08	4.03	90

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	567.12	K − 1 = 2	283.56	28.19
Dentro de los grupos (SC_D)	875.60	n − K = 87	10.06	
Total (SC_T)	1442.72	n − 1 = 89	16.21	

La DHS de Tukey al nivel de significancia de .05 = 1.95 puntos de la escala de actitud punitiva.

12A-6. Haijar y Kotchen (2003) estudiaron lecturas de la presión sanguínea sistólica en Estados Unidos y descubrieron que éstas eran significativamente más altas en el caso de los adultos de ciertas regiones geográficas. Supongamos que tú deseas replicar sus conclusiones y haces un sondeo entre adultos de tres regiones. De acuerdo con los siguientes estadísticos (ficticios), prueba la hipótesis de que existen diferencias en las unidades de presión sanguínea sistólica entre las diferentes regiones: sur, occidente y noreste. Asume la igualdad de las varianzas poblacionales.

Región	Media	Desviación estándar	n
Sur	126.04 unidades	.65 unidades	20
Occidente	123.09 unidades	.77 unidades	20
Noreste	125.22 unidades	.61 unidades	20
Total	124.79 unidades	1.42 unidades	60

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	92.76	$K - 1 = 2$	46.38	98.68
Dentro de los grupos (SC_D)	26.54	$n - K = 57$	.47	
Total (SC_T)	119.30	$n - 1 = 59$	2.02	

DHS de Tukey al nivel de significancia de .05 = 0.438.

Conjunto de problemas 12 B

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Por consistencia, redondea los cálculos a dos decimales. Utiliza $\alpha = 0.05$, a menos que se estipule otra cosa.

12B-1. Arthur y Graziano (1996) analizaron los predictores de personalidad implicados en accidentes automovilísticos. Una medida de nivel de intervalo fue la concientización: entendida como el sentido de obligación para respetar las normas sociales en todos aspectos de la vida. De acuerdo con los siguientes datos modificados, calcula los efectos principales del involucramiento de los accidentes automovilísticos en la concientización. Muestra símbolos y fórmulas.

Involucramiento en accidentes automovilísticos	Concientización media
Choque sin culpabilidad	122.70
Choque y culpabilidad	109.41
Sin choque	134.63
General	123.11

12B-2. El modelo lineal general ordena las puntuaciones de una variable dependiente de intervalo/razón (Y) en partes explicadas y partes no explicadas por una variable independiente (X). Con los siguientes estadísticos, aplica el modelo para explicar las puntuaciones Y que aparecen en la lista. Y = millas de viajero frecuente acumuladas por los empleados de ACME, Inc.; X = Clasificación del trabajo del empleado.

$$\bar{Y}_{(total)} = 16489 \text{ millas}$$

$$\bar{Y}_{(vicepresidentes)} = 9737 \text{ millas}$$

$$\bar{Y}_{(representantes\ de\ ventas)} = 26391 \text{ millas}$$

$$\bar{Y}_{(ingenieros)} = 13655 \text{ millas}$$

Caso	X	Y
John Callahan	Vicepresidente de finanzas	3248
Michael Windom	Vicepresidente de manufactura	11522
Antonio Williams	Ingeniero de materiales	21467
Arlene Slater	Ingeniero químico	2487
Kathy Schaefer	Representante de ventas de la costa este	24829
Charles Brown	Representante de ventas del suroeste	35663

12B-3. Al investigar sobre los peligros de la cafeína, un investigador agrega dos tipos de cafeína (la que se encuentra en el café y la que se encuentra en el chocolate) al suministro de agua de grupos de ratas criadas en laboratorio. Por lo general, esta especie sobrevive cerca de 13 meses. El suministro de agua del grupo control de ratas no fue alterado con cafeína. ¿Afecta la cafeína el tiempo de vida de las ratas? Prueba la hipótesis con los siguientes datos. Asume la igualdad de las varianzas poblacionales.

Grupo de tratamiento	Días que vivió la rata
Cafeína de café	398
Cafeína de café	372
Cafeína de café	413
Cafeína de café	419
Cafeína de café	408
Cafeína de café	393
Cafeína de café	387
Cafeína de café	414
Cafeína de chocolate	401
Cafeína de chocolate	389
Cafeína de chocolate	413
Cafeína de chocolate	396
Cafeína de chocolate	406
Cafeína de chocolate	378
Cafeína de chocolate	382
Cafeína de chocolate	417
Control (sin cafeína)	412
Control (sin cafeína)	386
Control (sin cafeína)	394
Control (sin cafeína)	409
Control (sin cafeína)	415
Control (sin cafeína)	401
Control (sin cafeína)	384
Control (sin cafeína)	398

12B-4. Al igual que Guth y cols. (1995), pretendemos analizar si las ideas religiosas influyen en los puntos de vista de una persona en relación con el ambiente. Comparamos clérigos de tres denominaciones —evangélico, protestante y católico— con la suposición de que los líderes religiosos de una denominación en particular tienen creencias religiosas similares. Nuestra variable dependiente constituye una escala de nivel de intervalo/razón, la cual mide las actitudes positivas con respecto al ambientalismo —apoyo a los esfuerzos gubernamentales para controlar la contaminación—. (Una puntuación alta indica mucho apoyo.) ¿Existe alguna relación entre las creencias religiosas y el ambientalismo? Asume la igualdad de las varianzas poblacionales.

Clérigo	Puntuación en la escala ambientalista
Ministro protestante	26
Sacerdote católico	30
Ministro protestante	24
Ministro evangélico	14
Sacerdote católico	25
Ministro evangélico	12
Ministro protestante	31
Sacerdote católico	34
Ministro protestante	22
Ministro evangélico	23
Ministro protestante	28
Sacerdote católico	28
Sacerdote católico	24
Ministro evangélico	17
Ministro protestante	32
Sacerdote católico	25
Ministro evangélico	22
Ministro evangélico	19

12B-5. La Agencia de Protección Ambiental monitorea el riesgo tóxico de los condados por medio del registro de la cantidad de veces que las industrias liberan sustancias químicas al aire o a los ríos (Rogge, 1996). Supongamos que tenemos los siguientes datos relacionados con la cantidad de descargas tóxicas durante el año pasado en 60 condados con ingresos promedios bajos, moderados y altos. ¿Se relaciona el nivel de ingreso de un condado con el riesgo tóxico que experimenta la población? Asume la igualdad de las varianzas poblacionales.

Nivel de ingreso del condado	Media de la cantidad de descargas tóxicas en 1996	Desviación estándar	*n*
Bajo	252.65	19.68	20
Moderado	159.10	17.87	20
Alto	129.95	27.49	20
Total	180.57	57.07	60

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	164377.43	$K-1=2$	82188.72	168.63
Dentro de los grupos (SC_D)	27781.30	$n-K=57$	487.39	
Total (SC_T)	192158.73	$n-1=59$		

DHS de Tukey con un nivel de significancia al .05 = 17.67 descargas tóxicas.

12B-6. Pinquart y Sorensen (2005) reportan que los cuidadores afroamericanos manifiestan niveles de tensión inferiores a los de sus homólogos hispanos ¿confirman sus hallazgos los datos ficticios que aparecen en seguida? En otras palabras, prueba la hipótesis de que existen diferencias étnicas en la tensión que se reporta (medida en una *Escala de Tensión*).

Origen étnico	Puntuación media en la escala de tensión	Desviación estándar	n
Afroamericano	22.35	1.97	17
Blanco	26.06	1.64	17
Hispano	28.65	1.69	17
Total	25.69	3.13	51

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	340.28	$K-1=2$	170.14	54.18
Dentro de los grupos (SC_D)	150.71	$n-K=48$	3.14	
Total (SC_T)	490.98	$n-1=50$	9.82	

DHS de Tukey con un nivel de significancia al .05 = 1.15.

Conjunto de problemas 12C

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Por consistencia, redondea los cálculos a dos decimales. Utiliza $\alpha = 0.05$, a menos que se estipule otra cosa.

12C-1. De acuerdo con los siguientes datos, calcula los efectos principales del nivel de ingresos por la cantidad de visitas al médico cada año (datos ficticios). Muestra símbolos y fórmulas.

Nivel de ingresos	Cantidad media de visitas al médico por año
Ingreso alto	5.12
Ingreso moderado	4.75
Ingreso bajo	1.87
General	3.91

12C-2. El modelo lineal general ordena las puntuaciones de una variable dependiente de intervalo/razón (Y) en partes explicadas y partes no explicadas por una variable independiente (X). Con los siguientes estadísticos, aplica el modelo para explicar las puntuaciones Y que aparecen en la lista. Y = cantidad de horas invertidas en internet a la semana; X = categoría de edad.

$\bar{Y}_{(total)} = 18.56$ horas $\quad \bar{Y}_{(adolescentes)} = 26.67$ horas

$\bar{Y}_{(mediana\ edad)} = 18.67$ horas $\quad \bar{Y}_{(mayores)} = 10.00$ horas

Número de individuo	X	Y
1	Adolescente	27
2	Mayor	10
3	Mayor	12
4	Mediana edad	16
5	Adolescente	24
6	Mediana edad	21
7	Mayor	8
8	Mediana edad	19
9	Adolescente	30

12C-3. Tú te encuentras estudiando la relación entre la ocupación y el nivel de depresión medida en la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D). Con los siguientes datos ficticios, prueba la hipótesis de que la depresión varía entre las diferentes ocupaciones. Asume la igualdad de las varianzas poblacionales.

Ocupación	Puntuación CES-D
Cajero de banco	6
Cajero de banco	9
Cajero de banco	11
Cajero de banco	6
Cajero de banco	4
Cajero de banco	5
Cajero de banco	3

Paramédico	14
Paramédico	18
Paramédico	15
Paramédico	13
Paramédico	19
Paramédico	14
Paramédico	13
Profesor universitario	8
Profesor universitario	13
Profesor universitario	9
Profesor universitario	9
Profesor universitario	12
Profesor universitario	7
Profesor universitario	14

12C-4. Los investigadores han encontrado que los vecindarios menos prósperos cuentan con menos lugares para adquirir alimentos saludables (Lewis, Sloane, Nascimento, Diamant y cols., 2005). Supongamos que tú deseas replicar el estudio en tu comunidad. Realizas una encuesta en diversos vecindarios clasificados de acuerdo con tres niveles de ingresos (X). Registras el número de lugares en cada vecindad donde se pueden comprar alimentos saludables (Y). ¿Influye el nivel de ingresos del vecindario en el número de opciones de alimentos saludables? Asume la igualdad de las varianzas poblacionales.

X Nivel de ingresos del vecindario	Y Cantidad de lugares que venden alimentos saludables
Ingreso bajo	7
Ingreso alto	14
Ingreso bajo	4
Ingreso medio	9
Ingreso medio	10
Ingreso alto	12
Ingreso bajo	6
Ingreso alto	15
Ingreso medio	13
Ingreso bajo	8
Ingreso alto	10
Ingreso bajo	10
Ingreso medio	11
Ingreso alto	13
Ingreso bajo	5
Ingreso alto	10
Ingreso medio	8
Ingreso medio	10

12C-5. Los investigadores han descubierto que los afroamericanos tienden a tener niveles más altos de involucramiento religioso que los blancos (Hunt y Hunt 2001). Supongamos que tú deseas replicar este estudio e incluyes hispanos. Mides el involucramiento religioso con base en el número de veces que una persona asiste a la iglesia cada mes. Utilizando los datos ficticios que se incluyen a continuación, prueba la hipótesis de que existen diferencias raciales en el involucramiento religioso entre blancos, hispanos y afroamericanos. Asume la igualdad de las varianzas poblacionales.

Raza	Media	Desviación estándar	n
Blanco	4.08	2.08	24
Hispano	5.21	2.47	24
Afroamericano	6.88	2.87	24
Total	5.39	2.50	72

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	94.70	$K - 1 = 2$	47.35	7.59
Dentro de los grupos (SC_D)	430.42	$n - K = 69$	6.24	
Total (SC_T)	525.11	$n - 1 = 71$	7.40	

DHS de Tukey con un nivel de significancia al .05 = 1.44 veces por mes.

12C-6. Una organización local de Rugby afirmó recientemente que los jugadores de rugby se encuentran en mejores condiciones físicas que otros atletas profesionales. Tú deseas probar esta afirmación, así que encuestas atletas profesionales que juegan rugby, baloncesto y fútbol. Mides la destreza atlética con una serie de pruebas físicas y sumas las puntuaciones para obtener un índice de destreza atlética general. Prueba la hipótesis de que los jugadores de rugby poseen mayor destreza atlética general. Asume la igualdad de las varianzas en lo que se refiere a destreza atlética entre las tres subpoblaciones.

Deporte	Índice de destreza atlética general	Desviación estándar	n
Rugby	66.72	3.91	25
Baloncesto	65.92	3.20	25
Fútbol	65.72	3.52	25
Total	66.12	3.55	75

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	14.00	$K - 1 = 2$	7.00	.55
Dentro de los grupos (SC_D)	909.92	$n - K = 72$	12.64	
Total (SC_T)	923.92	$n - 1 = 74$	12.49	

HSD de Tukey con un nivel de significancia al .05 = 2.42 puntos en el índice de destreza atlética general.

Conjunto de problemas 12 D

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Por consistencia, redondea los cálculos a dos decimales. Utiliza $\alpha = 0.05$, a menos que se estipule otra cosa.

12D-1. La clase de automóvil que una persona posee probablemente tenga que ver con la cantidad de gasolina que gasta cada mes, ya que algunos automóviles recorren más millas que otros. De acuerdo con los siguientes datos ficticios, calcula los efectos principales de la clase de automóvil que se posee en relación con el gasto mensual de gasolina. Muestra símbolos y fórmulas.

Clase de automóvil que se posee	Costo mensual de gasolina (en dólares)
SUV	188.21
Económico	107.87
Mediano	131.26
General	142.45

12D-2. El modelo lineal general ordena las puntuaciones de una variable dependiente de intervalo/razón (Y) en partes explicadas y partes no explicadas por una variable independiente (X). Con los siguientes estadísticos, aplica el modelo para explicar las puntuaciones Y que aparecen en la lista. Y = ingreso anual; X = nivel de educación.

$$\overline{Y}_{(total)} = 49\,257 \text{ dólares}$$
$$\overline{Y}_{(universidad)} = 47\,352 \text{ dólares}$$
$$\overline{Y}_{(preparatoria)} = 25\,167 \text{ dólares}$$
$$\overline{Y}_{(posgraduado)} = 78\,492 \text{ dólares}$$

Caso	X	Y
George Youngblood	Graduado de universidad	46 725
Amber Moore	Graduado de secundaria	27 316
Geoff Thomas	Posgraduado	67 256
Chris Kyden	Graduado de preparatoria	35 951
Nicole Owens	Posgraduado	81 247
Elise Palmer	Grados de universidad	57 124

12D-3. Se ha reportado que las vitaminas antioxidantes refuerzan los niveles de inmunidad en el cuerpo humano. Tú deseas probar la hipótesis y administras dos tipos de suplementos antioxidantes: antioxidantes fabricados (como las píldoras) o antioxidantes naturales (como la fruta). Para esta comparación, incluyes un grupo control al cual no se le administran antioxidantes. Prueba la hipótesis con los siguientes datos (ficticios). Asume la igualdad de las varianzas poblacionales.

Grupo de tratamiento	Nivel de inmunidad
Fabricado	21
Fabricado	23
Fabricado	19
Fabricado	17
Fabricado	26
Fabricado	28
Fabricado	16
Fabricado	23
Natural	27
Natural	19
Natural	18
Natural	28
Natural	21
Natural	23
Natural	24
Natural	20
Control (sin antioxidantes)	15
Control (sin antioxidantes)	17
Control (sin antioxidantes)	19
Control (sin antioxidantes)	14
Control (sin antioxidantes)	15
Control (sin antioxidantes)	13
Control (sin antioxidantes)	20
Control (sin antioxidantes)	15

12D-4. En un esfuerzo por atraer adeptos, un grupo local de iglesias sin denominación reportó que sus miembros se encuentran más satisfechos que los miembros de las iglesias locales protestante y católica. Tú deseas probar esta afirmación, así que muestreas las poblaciones de las iglesias sin denominación, protestante y católica. De acuerdo con los siguientes datos (ficticios), prueba la hipótesis de que existe una diferencia en lo que se refiere a la satisfacción entre los miembros de las iglesias protestante, católica y sin denominación. Asume la igualdad de las varianzas poblacionales.

Afiliación religiosa	Puntuación de satisfacción
Protestante	20
Católico	27
Sin denominación	22
Católico	16
Protestante	15
Católico	28
Sin denominación	24
Sin denominación	23
Protestante	14
Católico	13
Sin denominación	27
Protestante	23
Sin denominación	21
Católico	15
Sin denominación	15
Católico	19
Protestante	16
Protestante	22

12D-5. La hipótesis de contacto establece que el incremento en la interacción entre grupos da como resultado actitudes más positivas entre ellos. Por ejemplo, los estudios demuestran que los adolescentes que tienen mayor contacto con personas de mayor edad tienden a manifestar actitudes más positivas con respecto a ellas (Meshel y McGlynn 2004). Supongamos que tú deseas verificar esta teoría. Tú encuestas adolescentes de categorías variables de contacto. Mides sus actitudes hacia las personas de mayor edad con una Escala de Percepciones Positivas hacia el Envejecimiento (EPPE), en la que una puntuación alta indica una actitud más positiva. Prueba la hipótesis de que el contacto torna más positivas las actitudes de los adolescentes con respecto a las personas de mayor edad.

Contacto con personas de mayor edad	Escala de Percepciones Positivas hacia el Envejecimiento	Desviación estándar	n
Bajo	10.00	2.42	15
Moderado	17.60	2.47	15
Alto	22.27	3.01	15
Total	16.62	2.65	45

Fuente de variación	SS	df	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	1150.04	$K - 1 = 2$	575.02	82.02
Dentro de los grupos (SC_D)	294.53	$n - K = 42$	7.01	
Total (SC_T)	1444.58	$n - 1 = 44$	32.83	

HSD de Tukey con un nivel de significancia al .05 = 1.96.

12D-6. Una cantidad considerable de investigaciones han demostrado las disparidades raciales en lo que se refiere a las sentencias judiciales. Un estudio muestra que es más probable que a los afroamericanos y a los hispanos se les clasifique como delincuentes habituales y pasen más tiempo en prisión (Crawford, 2000; Crawford, Chiricos y Kleck, 1998). Supongamos que tú deseas estudiar el efecto de la raza en el tiempo de sentencia. Encuestas a varios delincuentes que han cometido el mismo crimen y registras sus sentencias en años. De acuerdo con los siguientes estadísticos (ficticios), prueba la hipótesis de que hay una mayor probabilidad de que los afroamericanos y los hispanos cumplan sentencias más largas que los blancos. Asume la igualdad de las varianzas poblacionales.

Raza	Media	Desviación estándard	n
Blanco	4.50	1.07	30
Hispano	5.63	.96	30
Afroamericano	6.13	1.20	30
Total	5.42	1.27	90

Fuente de variación	SC	gl	Varianza del cuadrado medio: $CM = SC/gl$	$F = \frac{CM_E}{CM_D}$
Entre grupos (SC_E)	42.02	$K - 1 = 2$	21.01	17.96
Dentro de los grupos (SC_D)	101.93	$n - K = 87$	1.17	
Total (SC_T)	143.96	$n - 1 = 89$	1.62	

DHS de Tukey con un nivel de significancia al .05 = 0.56 años.

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 12

Si en tu clase se utilizan las aplicaciones opcionales en computadora que acompañan el texto, abre los ejercicios del capítulo 12 localizados en el sitio web de *The Statistical Imagination*, www.mhhe.com/ritchey2 para aprender más sobre el ANOVA de un factor en *SPSS* para Windows. Los ejercicios hacen énfasis en la interpretación apropiada de la salida para el ANOVA y en las formas de abordar aspectos de la relación. Además, el apéndice D del texto contiene una vista rápida de las secuencias del comando *SPSS* para los procedimientos cubiertos en este capítulo.

CAPÍTULO

13 Variables nominales: las distribuciones chi cuadrada y binomial

RESUMEN DEL CAPÍTULO

Introducción: enfoque proporcional relacionado con el estatus social

En las investigaciones de las ciencias sociales y del comportamiento, las variables con niveles de medición nominal y ordinal se utilizan con frecuencia. En particular, mucha de la investigación se refiere a la forma en la que el estatus social de los individuos de grupos y sociedades influye en sus oportunidades en la vida. Las posiciones sociales definen un *estatus*, palabra griega que significa "rango o posición dentro de un grupo". El estatus social indica la cantidad relativa de privilegios y autoridad que se confieren a un individuo que ocupa una posición social dentro de un grupo, comunidad o sociedad. A una persona que ocupa una posición con un estatus alto, como la de presidente de una corporación, se le otorgan más recompensas y privilegios de las que serán individuos de posiciones inferiores, como los programadores de computadoras.

Todo individuo ocupa una variedad de posiciones sociales, como la de estudiante, esposa, padre, tesorero del club y empleado de oficina. Algunas posiciones sociales se encuentran abiertas de tal manera que se les puede elegir o ganar, aunque la competencia para ellos resulte difícil. Éstas se denominan posiciones ganadas, e incluyen la de esposo, doctor y vicepresidente de un banco. Otras posiciones sociales se confieren —asignan o *estampan* sobre los individuos en el momento de nacer—. Por ejemplo, ser hombre y blanco tiene sus privilegios en la sociedad norteamericana; sin embargo (a menos que se cambie sexo o de color) estas posiciones son fijas y no se eligen. Las posiciones asignadas se pueden identificar con facilidad como consecuencia de indicadores físicos, como el color de la piel y los rasgos biológicos distintivos de sexo o herencia étnica. La investigación en biología no es clara en lo que se refiere a los efectos de las posiciones sociales asignadas en la conducta. Por ejemplo, no resulta claro que si las mujeres son en realidad más emotivas que los hombres. No obstante, resulta importante entender los efectos de las posiciones sociales asignadas ya que éstas tienen consecuencias en los individuos. Si los hombres o las mujeres son más emotivos es un asunto aparte si la mayoría cree el mito de que las mujeres lo son. Las posiciones sociales asignadas son indicadores sociales que influyen en la forma en que, por ejemplo, se trata a las mujeres y a los miembros de las minorías. El estatus social limita las oportunidades de quienes ocupan posiciones minoritarias, como se hace evidente a la luz del racismo y el sexismo. Desde un ángulo positivo, ya que las posiciones asignadas con frecuencia resultan visibles, éstas son de utilidad para colocar productos y enfocar las intervenciones sociales. Por ejemplo, los clubes de golf se anuncian en las revistas para caballeros, y las campañas contra el cáncer de mama se anuncian en revistas dirigidas a las mujeres.

Como consecuencia de la importancia del estatus, las investigaciones sociales con frecuencia comienzan determinando los efectos de las variables de estatus sobre las variables dependientes de un estudio. Por ejemplo, en los estudios sobre la conducta criminal, las primeras hipótesis en probarse quizás tengan que ver con relaciones demográficas (la demografía es el estudio de la población de una sociedad). ¿Son los hombres o las mujeres los que cometen actos criminales con mayor probabilidad, o son los jóvenes o viejos; quienes viven en áreas rurales o en áreas urbanas, etc.? Una vez que se calculan los estadísticos, resulta vital que se les interprete estadísticamente —como porcentajes de un grupo de observaciones—, en lugar de individualmente. Por ejemplo, las estadísticas del Departamento de Justicia de Estados Unidos revelan de forma coherente que a los hombres se les arresta por crímenes violentos casi 90% de las veces; sin embargo, esto no significa que todo hombre constituya un criminal en potencia o que las mujeres no cometan crímenes. Una interpretación adecuada consiste en que hay mayor probabilidad de que los hombres cometan un crimen violento. Interpretar los datos de otra manera sería estereotipar —aplicar una generalización estadística a un individuo—. Los estereotipos son comunes en todas las sociedades y conducen al prejuicio y a la injusticia. Por ejemplo, si un oficial de policía aprehendió a un hombre y a una mujer que supuestamente asaltaron un banco, ¿podría ser que la mujer fuera la autora intelectual y líder del ilícito? Tiende a suponerse que es el hombre quien lo hizo.

Las investigaciones relacionadas con el estatus tienen el potencial de reforzar estereotipos y, por tanto, debe procederse con cuidado para evitar este escollo. Cuando se presentan las estadísticas como caracterizaciones de individuos, en lugar de generalizaciones probabilísticas, abundan las malas interpretaciones de los datos. Por ejemplo, en un perfil demográfico de personas sin hogar en Midcity, Estados Unidos, los datos pueden revelar que 70% de las personas sin hogar en la comunidad son hombres; 60% son afroamericanos y 50% son adictos a las drogas. Sería incorrecto reportar que la típica persona sin hogar en Midcity es un hombre afroamericano adicto a las drogas o al alcohol. De hecho, puede ser que haya más probabilidad de que las mujeres sin hogar resulten adictas a las drogas, ya que las mujeres que se encuentran desamparadas, aunque no sean adictas a una sustancia, tienen más posibilidades de que las acojan sus parientes. Los investigadores científicos deben ejercer cuidado para no reforzar estereotipos. El rasgo fundamental de la imaginación estadística consiste en considerar las observaciones aisladas en relación con un todo más amplio.

El capítulo se enfoca principalmente en las variables nominales, muchas de las cuales constituyen indicadores del estatus. Comenzamos con la prueba chi cuadrada de una relación entre dos variables nominales. Interpretaremos cuidadosamente conclusiones utilizando el lenguaje de las proporciones y porcentajes del grupo entero. Las conclusiones se referirán a la categoría (es decir, el grupo), no al individuo. Además, los hallazgos representarán con precisión las relaciones complejas entre las variables. Estas precauciones se deben al hecho de que en la prensa popular se habla mucho de relaciones entre indicadores de estatus, como género, raza, identidad étnica y otros fenómenos que resultan consecuencias, como las tasas de crimen. El error de presentar estos resultados de forma proporcional se suma a las desventajas de las minorías, como la afroamericana, la árabe-americana y las mujeres. En otras palabras, los reportes populares a menudo carecen de imaginación estadística. Este capítulo también abarca la prueba de la distribución binomial, que constituye una prueba de proporciones para una muestra pequeña.

Tablas cruzadas: comparación de frecuencias de dos variables nominales u ordinales

Un recurso común para presentar datos consiste en la **tabla cruzada o tabla de contingencia**, la cual compara *dos variables nominales/ordinales a la vez*. Dichas tablas *cruzadas* son esenciales para probar una hipótesis sobre la relación entre estas variables. Por ejemplo, ¿existe una relación entre el género (X) y la asistencia a la iglesia (Y) entre estudiantes universitarios? Los resultados (ficticios) del estudio en el recinto universitario aparecen en la tabla 13-1.

La tabla 13-2 se vale de lenguaje convencional para describir las partes de una tabla cruzada, en este caso los datos de la tabla 13-1. En la parte superior de la tabla se encuentran las *casillas*. El número que se encuentra en una casilla representa la frecuencia de una ocurrencia conjunta (o simplemente *frecuencia conjunta*) de las categorías de las dos variables. Una ocurrencia conjunta en la combinación de categorías *para una sola variable*. Por ejemplo, digamos que Charles asiste a la iglesia; entonces éste se cuenta como parte de la frecuencia conjunta de la casilla *hombres que asisten*, en la que hay 66 individuos. La frecuencia conjunta de hombres que no asisten es de 134; en el caso de las mujeres que asisten es de 94 y en el de las mujeres que no asisten, de 146. Las sumas en los márgenes de la tabla reciben el nombre de totales marginales. El total de la columna en el caso de los hombres nos indica que hay 200 hombres en la muestra; asimismo, hay 240 mujeres. Los totales de renglón de asistencia y de no asistencia son de 160 y 230, respectivamente. El gran total es el tamaño de la muestra (n), que aparece en la esquina inferior derecha. Observa que los totales de columna y la de renglón suman el gran total.

> **Frecuencia conjunta (u ocurrencia conjunta de categorías)** En el caso de un solo individuo, la ocurrencia conjunta de las categorías de dos variables nominales/ordinales, como hombre que asiste a la iglesia o mujer que no asiste a la iglesia.

TABLA 13-1 | Asistencia a la iglesia el mes pasado por género entre estudiantes universitarios, $n = 440$ (datos ficticios)

Especificaciones			
	Género (X)		
Asistencia a la iglesia (Y)	**Hombres**	**Mujeres**	**Total**
Asistieron	66	94	160
No asistieron	134	146	280
Total	200	240	440

TABLA 13-2 | Lenguaje de la tabla cruzada: asistencia a la iglesia el mes pasado por género entre estudiantes universitarios, $n = 440$

	Género (X)			
Asistencia a la iglesia (Y)	**Hombres**	**Mujeres**	**Totales**	
Asistieron	66	94	160	
	Casilla (contienen frecuencias conjuntas)			Totales renglón (marginales)
No asistió	134	146	280	
Total	200	240	440	
	Totales de columna (marginales)			Gran total (n)

Si existe una relación entre el género y la asistencia a la iglesia, uno u otro género tienen una frecuencia de asistencia a la iglesia mucho más alta. Sin embargo, los números en bruto de la tabla pueden resultar confusos, ya que hay más mujeres que hombres en la muestra. Debemos preguntarnos, ¿cuántos? Por consiguiente, utilizamos los totales marginales para calcular *porcentajes* de hombres y mujeres que asisten a la iglesia:

$$p\text{ [de hombres que asistieron a la iglesia el mes pasado]} = \frac{\#\text{ hombres que asisten}}{\#\text{ total de hombres}} + \frac{66}{200}$$
$$= 0.3300$$

$$\%\text{ [de hombres que asistieron a la iglesia el mes pasado]} = (p)(100) = (0.3300)(100)$$
$$= 33.00\%$$

Asimismo, el porcentaje de mujeres que asistió a la iglesia el mes pasado es de 94/240 (100) = 39.17%. Después de calcular estos porcentajes, vemos claramente que un porcentaje de mujeres más alto asistió a la iglesia el mes pasado. A propósito, estos porcentajes reciben el nombre de *porcentaje de columna*, ya que se basan en los totales marginales de columna. Es decir que el **porcentaje de columna** *constituye una frecuencia de casilla en forma de porcentaje del total marginal de la columna.* Asimismo, un **porcentaje de renglón** *constituye una frecuencia de casilla en forma de porcentaje del total marginal de renglón.* Por ejemplo, el porcentaje de renglón de asistentes a la iglesia que son hombres es de 66/160 (100) = 41.25%. Estas simples tablas cruzadas proporcionan importantes estadísticas descriptivas y permiten determinar si existe una relación entre dos variables nominales u ordinales.

Cálculo de porcentajes de renglones y columnas de casillas de un tabla cruzada

$$\% \text{ de columna [de ocurrencia conjunta]} = \frac{\# \text{ en una casilla}}{\# \text{ total de columna}}(100)$$

$$\% \text{ de renglón [de ocurrencia conjunta]} = \frac{\# \text{ en una casilla}}{\# \text{ total la renglón}}(100)$$

Prueba chi cuadrada: enfoque en las frecuencias de ocurrencias conjuntas

La existencia de una relación entre dos variables se establece con una prueba de hipótesis denominada prueba chi cuadrada. Supongamos que queremos analizar la relación entre la raza y el partido político de preferencia con los datos ficticios en la tabla cruzada 13-3. Repasemos elementos en la tabla. En la tabla cruzada, colocamos las categorías de la variable independiente (raza) en las columnas y las de la variable dependiente (partido político de preferencia) en las columnas. Denominamos a ésta una tabla de 2 × 3 (que se lee "dos por tres") para indicar el número de categorías de cada variable. El número de renglones multiplicado por el número de columnas representa el número de casillas de frecuencia conjunta, de las cuales hay seis en una tabla de 2 × 3. Las seis posibles parejas de categorías de las variables de raza y preferencia de partido político son: blanco-demócrata, afroamericano-demócrata, blanco-republicano, afroamericano-republicano, blanco-independiente/otro/ninguno y afroamericano-independiente/otro/ninguno. El tamaño de la muestra total, o *gran total*, es de 400, que aparece en la esquina inferior derecha. Los totales marginales a la derecha indican los totales de renglón, con 150 casos en cada uno de los partidos demócrata y republicano y 100 en la categoría independiente/otro/ninguno, que da un gran total de 400 entrevistados. Los totales marginales en la parte inferior indican los totales de columna, de los que 300 individuos de la muestra son blancos y 100 afroamericanos, que, de nuevo, suman un gran total de 400. Si nos enfocamos en las frecuencias conjuntas, hallamos que 96 individuos de la muestra son blancos y demócratas; 123 son blancos y republicanos; 54 son afroamericanos y demócratas, etc. Finalmente, los números entre paréntesis a la derecha de cada frecuencia conjunta representan la frecuencia esperada, que en breve explicaremos.

TABLA 13-3 | Preferencia de partido político por raza (las frecuencias esperadas están en paréntesis)

Partido político (Y)	Raza (X) → Blanco	Afroamericano	Totales de renglón
Demócrata	96 (112.5)	54 (37.5)	150
Republicano	123 (112.5)	27 (37.5)	150
Independiente/otro/ninguno	81 (75.0)	19 (25.0)	100
Totales de columna	300 (300)	100 (100)	400

Si existe una relación entre la raza y la preferencia por algún partido político, esperamos encontrar que un porcentaje más alto de cierta raza prefiere un partido. Podríamos hipotetizar que un porcentaje más alto de afroamericanos se identifica como demócrata. Si esto es verdad, en una muestra aleatoria de adultos norteamericanos esperamos encontrar frecuencias particularmente altas de la ocurrencia conjunta afroamericano-demócrata.

Como en el caso de cualquier hipótesis, la hipótesis puede enunciarse de tal manera que nos permita conocer qué resultados muestrales esperar cuando la hipótesis es verdadera. Con la prueba chi cuadrada, enunciamos nuestra hipótesis nula de *no relación* entre las dos variables. Como veremos en un momento, cuando éste es el caso, el estadístico chi cuadrada dará un valor de cero dentro del error de muestreo. Por consiguiente, establecemos el paso número 1 de la prueba de la siguiente manera:

$$H_0: \chi^2 = 0$$

Es decir, que no existe una relación entre la raza y la preferencia por algún partido político.

$$H_A: \chi^2 > 0$$

Es decir, que existe una relación entre la raza y la preferencia por algún partido político. De una cola, pero no direccional.

Como en el caso de cualquier hipótesis estadística, este enunciado nos permite hacer predicciones de los resultados del muestreo repetido. En este caso, si suponemos que no existe una relación, podemos utilizar las frecuencias marginales para predecir las *frecuencias esperadas* de cada casilla. En seguida comparamos estas frecuencias esperadas con nuestras *frecuencias observadas*, las frecuencias conjuntas reales encontradas en los datos de la muestra e incluidas en la tabla cruzada. Si las frecuencias observadas son aproximadamente iguales a las esperadas, con un pequeño error de muestreo, mantenemos la hipótesis de *no relación* y concluimos que la raza no tiene nada que ver con la preferencia por algún partido político. Sin embargo, si existe una gran diferencia entre las frecuencias observadas y las esperadas, comenzamos a sospechar que hay una relación entre las variables. La hipótesis de la chi cuadrada nos dice que la suma de las diferencias entre las frecuencias de las casillas observadas y las esperadas es tan grande que no se debe sencillamente al resultado de un error de muestreo. Como más tarde veremos con detalle, la prueba es no direccional. Sin embargo, cualquier prueba chi cuadrada es de una cola, ya que este estadístico se basa en números elevados al cuadrado, que siempre son positivos.

Cálculo de las frecuencias esperadas

¿Cómo se calculan las frecuencias esperadas? En la tabla 13-3 podemos ver que tres cuartos (300 de 400) de los individuos de la muestra son blancos. Podemos predecir que *si la raza no tiene nada que ver con la preferencia por un partido político, es decir, si no hay relación entre la raza y la preferencia por algún partido político*, podemos esperar que tres cuartos de los demócratas, tres cuartos de los republicanos y tres cuartos de los independientes/otro/ninguno sean blancos. En otras palabras, los blancos serán representados entre los partidos políticos en proporción a sus números en la población general. Asimismo, dado que un cuarto de la muestra está conformada por afroamericanos, podemos esperar que un cuarto de cada una de las categorías de los partidos políticos sea afroamericana. Las frecuencias esperadas nos indican cuántos casos caen en cada casilla si cada casilla es proporcional a las frecuencias marginales —el caso en el que las dos variables no se encuentran relacionadas y, por consiguiente, las casillas se llenan de manera aleatoria.

La frecuencia esperada, E, para cada casilla, se calcula precisamente con la siguiente fórmula:

Cálculo de frecuencias esperadas de una tabla cruzada

$$E_{casilla} = \frac{\text{(total marginal de columna para la casilla)}\ \text{(total marginal de renglón para la casilla)}}{\text{gran total}}$$

donde

$E_{casilla}$ = frecuencia esperada de una casilla

Total marginal de columna para la casilla = casos totales para la categoría de columna de la casilla

Total marginal de renglón para la casilla = total de casos a la categoría de renglón en la celda

Gran total = tamaño de la muestra total

Por ejemplo, la frecuencia esperada de demócratas blancos es

$$E_{demócratas\ blancos} = \frac{\text{(número total de blancos) (número total de demócratas)}}{\text{gran total}}$$

$$= \frac{(300)\,(150)}{(400)} = 112.5 \text{ casos}$$

Diferencias entre las frecuencias observadas y esperadas

En todas las pruebas de hipótesis, la diferencia entre lo que se observa en los datos de la muestra real y lo que se hipotetiza en el caso de la hipótesis nula constituye el *efecto* de la prueba. El cálculo del estadístico chi cuadrada se basa en la medición de las diferencias entre las frecuencias observadas en las frecuencias esperadas; las frecuencias esperadas son las frecuencias conjuntas que se presentan en el muestreo repetido cuando *no* hay relación entre las dos variables. En el caso de muchas fórmulas estadísticas, estas diferencias se encuentran en el numerador del estadístico de prueba. De la misma manera, el denominador de la mayoría de los estadísticos de prueba constituye una medida del error de muestreo normal cuando la hipótesis nula es verdadera. Por consiguiente, las frecuencias esperadas se incluyen en el denominador.

Cálculo del estadístico de prueba chi cuadrada

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

donde

χ^2 = una medida de la probabilidad de las diferencias entre las frecuencias observadas esperadas

O = frecuencia observada de una casilla

E = frecuencia esperada de una casilla

La tabla 13-4 contiene una hoja de cálculo de computadora para calcular el estadístico chi cuadrada. Los datos provienen de la tabla 13-3.

Una mirada de cerca a esta fórmula nos da un sentido de la proporción de la forma en la que estos valores se relacionan con los resultados estadísticos reales. Cuando los efectos (es decir, las diferencias entre las frecuencias observadas y esperadas) son grandes, esto

TABLA 13-4 | Hoja de cálculo de computadora para calcular el estadístico chi cuadrada utilizando los datos de la tabla 13-3

Especificaciones		Cálculos			
Casilla (X, Y)	**O**	**E**	**(O − E)**	**[(O − E)²/E]**	**(O − E)²**
Demócrata blanco	96	112.5	−16.5	272.25	2.42
Demócrata afroamericano	54	37.5	16.5	272.25	7.26
Republicano blanco	123	112.5	10.5	110.25	.98
Republicano afroamericano	27	37.5	−10.5	110.25	2.94
Independiente blanco	81	75.0	6.0	36.0	.48
Independiente afroamericano	19	25.0	−6.0	36.0	1.44
Totales	400	400.0	0.0		χ^2 = 15.52

sugiere que los casos no se distribuyen aleatoriamente entre las casillas y la hipótesis nula debería rechazarse. Diferencias más grandes aparecen cuando algunas de las casillas (a menudo las casillas en las esquinas diagonales opuestas) *se cargan* como consecuencia de que existe una relación entre las dos variables. Analicemos una de las diferencias de casilla con mayor detenimiento. La *frecuencia observada* real de demócratas blancos es de 96 casos. La diferencia entre esta frecuencia observada de 96 y la frecuencia esperada de 112.5 es de −16.5 casos (es decir, aproximadamente 17 menos de lo esperado, dados los números totales de demócratas blancos). Parece que los blancos se encuentran subrepresentados entre los demócratas en proporción con sus números en la población general. Como veremos, se presentan efectos semejantes para raza respecto a la preferencia de partido político en otras casillas.

Sin embargo, es necesario considerar la posibilidad de que las diferencias para esta casilla y otras sean resultado del error de muestreo aleatorio. Aun cuando no hay una relación entre la raza y la preferencia por un partido político, en el muestreo repetido obtendremos una variedad de distribuciones de casilla. Una primera muestra podría abarcar un poco menos demócratas blancos de lo esperado; pero la siguiente muestra podría abarcar un poco más. Dichas fluctuaciones de una muestra a la siguiente serían el resultado del error de muestreo normal esperado. Nuestra prueba de hipótesis de la relación entre la raza en la preferencia por un partido político depende de la muestra única que extrajimos. La prueba chi cuadrada responde a la pregunta: ¿cuán inusual es observar en el muestreo repetido brechas entre las frecuencias observadas y en las esperadas cuando la raza no tiene nada que ver con la preferencia por un partido político? En otras palabras, ¿resultan significativos los aspectos de raza en la preferencia por un partido político? Una distribución aleatoria de frecuencias —en proporción con los totales marginales— es lo que se presenta en el muestreo cuando no existe una relación entre las dos variables. La hipótesis alternativa probablemente se acepta cuando los efectos de la prueba son grandes (es decir, cuando existen diferencias grandes entre las frecuencias observadas y las esperadas).

El estadístico chi cuadrada y la tabla de la distribución chi cuadrada (tabla estadística G, apéndice B) nos permite calcular la probabilidad exacta (valor p) de las diferencias entre las frecuencias observadas y las esperadas cuando la hipótesis nula de *no relación* es verdadera. Al comparar esta probabilidad a nuestro nivel alfa, o rechazamos la hipótesis nula o no la rechazamos. Si no la rechazamos, concluimos que no hay razón para creer que raza y la preferencia partido político se encuentran relacionados; que las diferencias entre las frecuencias observadas y esperadas son el resultado del error de muestreo. Sin embargo, si rechazamos la hipótesis nula, concluimos que existe una relación entre la raza y la preferencia por un partido político. También describiríamos las aplicaciones prácticas de esta relación, señalando efectos fuertes —las casillas en las que existen grandes discrepancias entre las frecuencias observadas y las esperadas—. Por ejemplo, señalaríamos que un porcentaje desproporcionadamente bajo de blancos se identifican como demócratas, en contraste con un porcentaje desproporcionadamente alto de afroamericanos.

Los grados de libertad para la prueba chi cuadrada

Para leer la tabla de la distribución chi cuadrada (tabla estadística G, apéndice B), debemos calcular el estadístico de prueba chi cuadrada y los grados adecuados de libertad. En el caso de la prueba chi cuadrada, los grados de libertad se determinan por el número de columnas y filas que hay en la tabla cruzada.

Cálculo de los grados de libertad para la prueba chi cuadrada

$$gl = (r - 1)(c - 1)$$

donde

gl = grados de libertad para la prueba chi cuadrada
r = número de renglones o filas en la tabla cruzada
c = número de columnas en la tabla cruzada

En el caso de una tabla de 2 × 3, como la tabla 13-3 referente a la raza y preferencia política,

$$gl = (r - 1)(c - 1) = (3 - 1)(2 - 1) = 2$$

Recordemos que los grados de libertad representan el número de oportunidades en el muestreo para compensar las limitaciones en un estadístico o en una prueba estadística. Por ejemplo, en el capítulo 10, observamos en el caso de las pruebas de medias, que las puntuaciones extremas tienden a apartar a la media del verdadero parámetro. En el caso de la prueba chi cuadrada, como lo indica la fórmula para gl, el número de casillas en una tabla cruzada influye en el tamaño del estadístico calculado. Una tabla cruzada se puede distorsionar si, en este muestreo particular, tiene un exceso o falta de frecuencias en comparación con su verdadera ocurrencia en la población. Por ejemplo, supongamos que en la población hay muchos demócratas blancos, que en esta muestra sencillamente faltan. Así como una puntuación extrema *distorsiona* una media, una o más casillas cargadas aleatoriamente en exceso o por defecto en una tabla cruzada puede distorsionar el cálculo del estadístico de la prueba chi cuadrada. El cálculo de los grados de libertad compensa esta tendencia. Un análisis cuidadoso de la fórmula para gl para el estadístico chi cuadrada revela que cuántas más categorías tengan las variables en una tabla cruzada, más altos serán los grados de libertad. Por el contrario, las tablas con menos casillas, o las tablas de 2 × 2 y 2 × 3, cuentan con menos grados de libertad. Con estas pequeñas tablas, si por azar la muestra incluye una casilla sobrecargada o poco cargada, existen pocas oportunidades (es decir, otras casillas) para equilibrar este desafortunado evento muestral. El muestreo repetido nos dice que con el paso del tiempo, cuando se extrae una gran cantidad de muestras, en ocasiones se muestrea en exceso una casilla y en ocasiones se hace de manera insuficiente. Con unas cuantas casillas, el muestreo excesivo de una sola casilla puede distorsionar los resultados. De hecho, si conocemos la frecuencia esperada de una casilla en la tabla cruzada de 2 × 2, es posible utilizar los totales marginales para calcular las otras tres. En otras palabras, una casilla tiene libertad para variar independientemente de la forma en que se calcula el estadístico; de esta manera, hay 1 grado de libertad.

Requerimiento de la frecuencia mínima por casilla Sin embargo, con la chi cuadrada, un incremento de los grados de libertad no siempre constituye una ventaja. Este estadístico de prueba tiene una limitación que restringe la cantidad de categorías utilizadas. La frecuencia esperada de cada casilla en una tabla cruzada debe ser por lo menos de 5 o, de lo contrario, se deben llevar a cabo ajustes en el cálculo del estadístico chi cuadrada. Las casillas de las frecuencias esperadas pequeñas generan un error de muestreo adicional, el cual puede llevar a una decisión de rechazo incorrecta y a una conclusión equivocada. Cuando los tamaños de la casilla son pequeños, denominamos a este hecho *una reducción de casilla.* La experiencia del autor sugiere que para evitar la reducción de una casilla, el tamaño global de la muestra debe ser igual al número de casillas multiplicado por 12. De esta manera, en una tabla simple de 2 × 2, se requieren aproximadamente 48 casos; en una tabla de 2 × 3, 72 casos, etc. Aún con un tamaño grande en la muestra global, la reducción de la casilla puede representar un problema en algunas variables. Por ejemplo, en el caso de la variable de preferencia religiosa, pudiera ser que tuviéramos una muestra de 100 individuos, aunque sólo 6 de ellos sean musulmanes. Si tratamos de descomponer esta variable según el género, por ejemplo, no existe ninguna posibilidad de que las frecuencias conjuntas de hombre-musulmán y mujer-musulmana tengan el mínimo requerido de 5 casos.

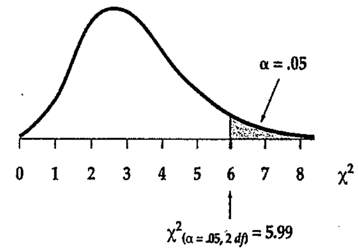
Cuando no hay suficientes casos en una casilla, pueden llevarse a cabo diversas correcciones. Primero, es posible reducir la cantidad de casillas eliminando las categorías poco densas. Por ejemplo, en el caso de la preferencia religiosa, la categoría musulmán podría eliminarse o incluirse en una categoría con la denominación *otra*, con otras religiones con bajas frecuencias, como la hindú. Por supuesto, esto requiere que los resultados se reporten con las modificaciones. Una segunda alternativa al tratar con la reducción de una casilla consiste en absorber la categoría de baja frecuencia incluyéndola en otra categoría que posea un significado teórico. Por ejemplo, en el caso de la variable raza, la categoría *otra* podría combinarse con *afroamericano* para crear una nueva categoría denominada *no blanco.* Esto resulta adecuado y la variable raza es, en esencia, representativa (una sustituta adecuada) del estatus mayoría-minoría. Una tercera alternativa consiste en llevar a cabo una corrección en el cálculo de la chi cuadrada, restando 0.5 al calcular cada una de las diferencias entre las frecuencias observadas y las esperadas. Esto se denomina *corrección de Yates por continuidad*; ésta realiza un ajuste como resultado de la inconsistencia de los tamaños de las casillas. Una cuarta alternativa consiste en utilizar otra prueba estadística denominada *prueba exacta de Fisher* (Blalock, 1979, 292). Ésta resulta de utilidad en situaciones en las que se presentan muchas frecuencias menores de 5, pero ninguna de las cuales puede eliminarse o combinarse de manera justificada.

La corrección de Yates y la prueba exacta de Fisher complican los cálculos al grado de que el tiempo que se emplea en dominarlos resultaría más útil para aprender a calcular los estadísticos de una tabla cruzada en computadora. Los programas de computadora para tablas cruzadas generalmente proporcionan un conteo de las casillas con frecuencias pequeñas y presentan una chi cuadrada con la corrección de Yates y la prueba exacta de Fisher, junto con el estadístico chi cuadrada, con el fin de que se les pueda sustituir cuando sea adecuado. Una comparación de estos estadísticos en los resultados de la computadora revelará que, cuando el tamaño de la muestra es muy grande, unas cuantas casillas con baja frecuencia tienen poco impacto en el estadístico chi cuadrada que resulta. En otras palabras, en el caso de las muestras grandes, el valor corregido de la chi cuadrada será aproximadamente el mismo del valor no corregido, y ambos estadísticos tendrán los mismos valores *p*.

Distribución muestral de la chi cuadrada y sus regiones críticas

La fórmula de la chi cuadrada la dedujo a principios del siglo XX un especialista en estadística llamado Karl Pearson, a quien también se le conoce por los estadísticos de los capítulos 14 y 15. Es probable que haya deducido esta fórmula matemáticamente; pero imaginemos cómo pudo haber creado la fórmula empíricamente —por muestreo repetido, como, de hecho, lo hicieron los contadores de frijoles de la antigüedad. Toma una caja con cantidades iguales de frijoles surtidos —una población de frijoles—. Considera a los frijoles blancos como demócratas blancos; a los frijoles rojos, como afroamericanos demócratas; a los frijoles pintos, como blancos independientes, y así sucesivamente. Además toma una muestra aleatoria de 400 frijoles, clasifícalos y cuenta los frijoles de cada clase con el fin de obtener las frecuencias observadas de la tabla cruzada. Después de calcular las frecuencias esperadas de cada casilla, resta las frecuencias observadas y las esperadas, eleva al cuadrado los resultados y divide entre la frecuencia esperada. En seguida suma los tres cocientes para obtener el estadístico de la χ^2. Repite esto, por decir algo, 10000 veces y representa los cálculos en forma de curva de probabilidad o histograma suavizado. La forma de esta distribución de probabilidades aparece en la figura 13-1. Observa que ésta es de una cola. Siempre será así, ya que al elevar al cuadrado las diferencias, se elimina cualquier signo negativo. No obstante, la prueba no es direccional, como veremos más adelante.

FIGURA 13-1
Forma de la curva de distribución de probabilidad chi cuadrada con la región crítica para $\alpha = .05$ y 2 grados de libertad



Como lo indicamos antes, la forma particular de la distribución de la chi cuadrada para un problema determinado depende del número de grados de libertad, los cuales quedan determinados por la cantidad de casillas que hay en la tabla cruzada. La figura 13-1 muestra el valor crítico de la chi cuadrada con un nivel de significancia de .05 con 2 grados de libertad. Este valor crítico implica que si no hay relación entre las dos variables, con el muestreo repetido el estadístico chi cuadrada resultará tan alto o mayor que 5.99, solamente 5 por ciento de las veces. El otro 95% de las veces, la chi cuadrada resultará inferior a 5.99; la mayoría de los resultados caerán ligeramente por encima de cero; esto es justamente lo que se espera cuando las frecuencias observadas y las esperadas son las mismas con cierto error de muestreo.

La prueba chi cuadrada se emplea en las siguientes circunstancias:

Cuándo utilizar la prueba chi cuadrada referente a la relación entre dos variables nominales

En general: al comprobar una hipótesis de relación entre dos variables nominales.

1. Hay una población con una muestra representativa de ésta.
2. Se tienen dos variables, las dos con un nivel de medición nominal/ordinal.
3. La frecuencia esperada de cada casilla en la tabla cruzada es de por lo menos 5.

Los seis pasos de la inferencia estadística para la prueba chi cuadrada

Utilicemos los datos de la tabla 13-1 y su hoja de cálculo (tabla 13-2) para contestar la pregunta de investigación sobre si existe una relación entre la raza y la preferencia por algún partido político. La hipótesis nula se refiere al hecho de que no existe una relación. Ya que ambas variables son nominales y las frecuencias esperadas son superiores a 5, se puede utilizar la prueba chi cuadrada.

Breve lista de verificación de los seis pasos de la inferencia estadística

PREPARACIÓN DE PRUEBA

Formula la pregunta de investigación. Elabora diagramas conceptuales e indica las especificaciones, incluyendo las poblaciones y muestras bajo estudio, las variables (por ejemplo, $X = \ldots, Y = \ldots$) y sus niveles de medición, así como los estadísticos y parámetros dados o calculados. Señala el procedimiento adecuado para la prueba estadística.

SEIS PASOS

Representemos la *hipótesis* con la letra H.

1. Enuncia la H_0 y la H_A, y estipula la dirección de la prueba.
2. Describe la distribución muestral.
3. Especifica el nivel de significancia (α) y el valor crítico de la prueba.
4. Observa los resultados reales de la muestra y calcula los efectos de la prueba, el estadístico de la prueba y el valor p.
5. Toma la decisión de rechazo.
6. Interpreta, aplica los resultados y proporciona las mejores estimaciones en términos familiares.

Solución para la prueba chi cuadrada para una relación entre dos variables nominales/ordinales

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿están determinadas razas más a favor de ciertos partidos políticos? Es decir, ¿existe alguna relación entre la raza y la preferencia por un partido político? Especificaciones: *Variables:* X = raza, nivel nominal; Y = preferencia por un partido político, nivel nominal. *Población:* población de adultos estadounidenses. *Muestra:* $n = 400$ adultos. *Procedimiento estadístico:* prueba chi cuadrada de la relación entre dos variables nominales. *Observación:* frecuencias observadas y frecuencias esperadas calculadas ($E_{casilla}$), donde:

$$E_{casilla} = \frac{\text{(total marginal de columna a la casilla) (total marginal de renglón para la casilla)}}{\text{gran total}}$$

(*Nota:* en algunos problemas relacionados con chi cuadrada, renunciamos a los diagramas de las poblaciones y, en su lugar, nos concentramos en la tabla cruzada.)

Partido político (Y) / Raza (X) →	Blanco	Afroamericano	Totales de renglón
Demócrata	96 (112.5)	54 (37.5)	150
Republicano	123 (112.5)	27 (37.5)	150
Independiente/otro/ninguno	81 (75.0)	19 (25.0)	100
Totales de columna	300 (300)	100 (100)	400

SEIS PASOS

1. H_0: $\chi^2 = 0$

   Es decir, no existe relación entre la raza y la preferencia por algún partido político.

   H_A: $\chi^2 > 0$

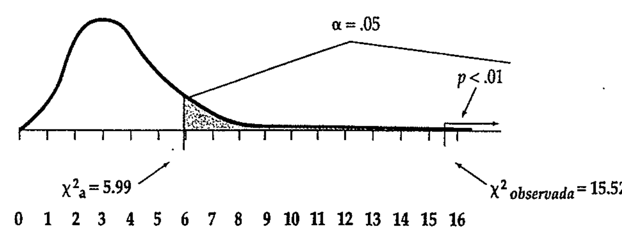
   Es decir que existe una relación entre la raza y la preferencia por algún partido político. Prueba de una cola no direccional.

2. *Distribución muestral:*

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$df = (r-1)(c-1) = (3-1)(2-1) = 2$$

   Si la H_0 es verdadera y se toman repetidamente muestras de tamaño 400 de la población y se colocan en una tabla cruzada de raza y preferencia política, los cálculos del estadístico chi cuadrada generarán la siguiente distribución chi cuadrada:



3. *Nivel de significancia:* $\alpha = .05$ (no direccional). De una cola $gl = 2$. Valor crítico de la prueba $= \chi^2_\alpha = 5.99$ (En la tabla estadística G, apéndice B.) (Área sombreada sobre la curva.)

4. *Observaciones:*

Efecto de la prueba y estadísticos de la prueba:

Especificaciones	Cálculos				
Casilla (X, Y)	O	E	(O – E)	$(O-E)^2$	$[(O-E)^2/E]$
Blanco demócrata	96	112.5	–16.5	272.25	2.42
Afroamericano demócrata	54	37.5	16.5	272.25	7.26
Blanco republicano	123	112.5	10.5	110.25	.98
Afroamericano republicano	27	37.5	–10.5	110.25	2.94
Blanco independiente	81	75.0	6.0	36.0	0.48
Afroamericano independiente	19	25.0	–6.0	36.0	1.44
Totales	400	400.0	0.0		χ^2 = 15.52

Valor *p*: *p* [se extrae una muestra con diferencias entre las frecuencias observadas y las esperadas tan inusuales o más inusuales que las observadas cuando, de hecho, no existe ninguna relación entre las variables] <.001 (área señalada en la curva en el paso 2).

5. *Decisión de rechazo:* $|\chi^2_{observada}| > |\chi^2_{\alpha}|$ (es decir, 15.52 > 5.99); $p < \alpha$; .001 < .05. Rechácese H_0 y acéptese H_A un nivel de confianza de 95%.
6. *Interpretación:* Aspectos de la relación y mejores estimaciones. *Existencia:* hay una relación entre la raza y la preferencia por algún partido político: $\chi^2 = 15.52, p < .001$.
   Aplicaciones prácticas: una cantidad desproporcionadamente alta de afroamericanos prefieren el partido demócrata. *Mejores estimaciones:* cincuenta y cuatro por ciento de los afroamericanos son demócratas en comparación con el 32% de los blancos. Sólo 27% de los afroamericanos son republicanos en comparación con el 41% de blancos.

Aspectos relevantes de una relación para la prueba chi cuadrada

Observa que en el último paso de la prueba de hipótesis solamente se reducen sin problemas dos de los cuatro aspectos de una relación: existencia y aplicaciones prácticas. La existencia de una relación se establece con la fórmula de la prueba chi cuadrada, comprobando la hipótesis de no relación entre las dos variables nominales. Si se determina una relación, se describe su naturaleza reportando los *efectos* (es decir, las diferencias entre las casillas de las frecuencias observadas y esperadas) de las casillas que destacan de la tabla cruzada. En términos familiares, se reportan las mejores estimaciones por medio del cálculo de los porcentajes de columna de las casillas seleccionadas. Recuerda que, de acuerdo con el capítulo 2, un porcentaje de columna *es la frecuencia de una casilla representada en forma de porcentaje con respecto al total marginal de columna:*

$$\% \text{ de columna [de frecuencia conjunta]} = \frac{\#\text{ en una casilla}}{\#\text{ total en la columna}} \times 100$$

Para calcular el porcentaje de afroamericanos demócratas:

$$\% \text{ [de afroamericanos demócratas]} = \frac{\#\text{ de afroamericanos demócratas}}{\#\text{ total de afroamericanos}} \times 100$$

$$= \frac{54}{100} \times 100 = 54\%$$

Se llevan a cabo cálculos semejantes relacionados con los porcentajes de columna en el caso de blanco-demócrata, afroamericano-republicano y blanco-republicano. Reportamos los porcentajes de columna que sugieren aspectos relevantes, en este caso, el hecho de que uno de los partidos políticos es desproporcionadamente afroamericano y el otro, desproporcionadamente blanco.

En el caso de las variables nominales, la dirección de una prueba no tiene significado. Por ejemplo, con la variable de género, carece de sentido el hecho de si los hombres son más positivos o negativos; o si es hombre o mujer. (La chi cuadrada puede utilizarse con variables ordinales y, en este caso, la dirección a veces tiene sentido.)

Existen medidas de la fuerza de una relación para el estadístico de la prueba chi cuadrada; no obstante, pocas veces se reportan, ya que dichos reportes están plagados de errores potenciales. Cada medida requiere un uso adecuado y se aplica solamente en circunstancias muy específicas (véase Lee y Maykovich, 1995, 129).

Aspectos de una relación cuando se utiliza el estadístico chi cuadrada

Existencia: prueba la hipótesis nula que indica que $\chi^2 = 0$; es decir que no existe una relación entre *X* y *Y*.

Utiliza el estadístico de la prueba chi cuadrada:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$gl = (r-1)(c-1)$$

Dirección: no es aplicable.

Fuerza: normalmente no se informa.

Aplicaciones prácticas: reporta los porcentajes de columna para las casillas seleccionadas que transmiten con claridad la esencia de las conclusiones.

Utilización de la chi cuadrada como prueba de diferencia de proporciones

Cuando se presentan las frecuencias conjuntas de dos variables nominales en una tabla cruzada, resulta muy fácil calcular sus porcentajes. Supongamos que contamos con los datos de la tabla 13-5, que corresponden a un estudio relacionado con pacientes de edad avanzada que van por primera vez al médico en una clínica. A cada paciente lo acompaña un cuidador, por lo general un miembro de la familia, particularmente la esposa o una hija. Algunos de los pacientes junto con sus cuidadores van a una clínica de atención primaria, la cual proporciona

TABLA 13-5 | Tipo de clínica a la que se canaliza al cuidador con un grupo de apoyo (frecuencias esperadas entre paréntesis)

Canalizado a un grupo de apoyo	Tipo de clínica →	Atención primaria	Geriátrica	Totales de renglón
Sí		7 (15.83)	23 (14.17)	30
No		50 (41.17)	28 (36.83)	78
Totales de columna		57 (57.00)	51 (51.00)	108

servicios a la gente de cualquier edad. Los demás asisten a una clínica geriátrica especializada en pacientes de mayor edad. Como consecuencia de que los cuidadores con frecuencia padecen estrés, la clínica geriátrica exige que se les incluya en el proceso de tratamiento y que se les proporcione apoyo, como el hecho de canalizarlos a grupos de apoyo para cuidadores.

Con el fin de evaluar esta suposición, entrevistamos muestras aleatorias de cuidadores en ambas clínicas. Los datos de la tabla 13-5 comparan la cantidad de cuidadores en cada clínica canalizados a grupos de apoyo. De esta manera, una variable nominal se refiere al tipo de clínica, y la otra la referencia a un grupo de apoyo. Podemos analizar la pregunta de investigación relacionada con el hecho de que existe una relación entre el tipo de clínica y la probabilidad de canalización a un grupo de apoyo. Si los parientes de pacientes se canalizan a una clínica geriátrica en una tasa significativamente mayor, la suposición de esa clínica está justificada.

En la tabla 13-5 enfoquemos nuestra atención en el porcentaje de cuidadores de la clínica geriátrica canalizados a grupos de apoyo, calculando el porcentaje de columna:

$$\text{\% [de cuidadores de clínica geriátrica canalizados a un grupo de apoyo]} = \frac{\text{\# en la casilla geriátrica sí}}{\text{\# total en clínica geriátrica}} \times 100$$

$$= \frac{23}{51}(100) = 45.10\%$$

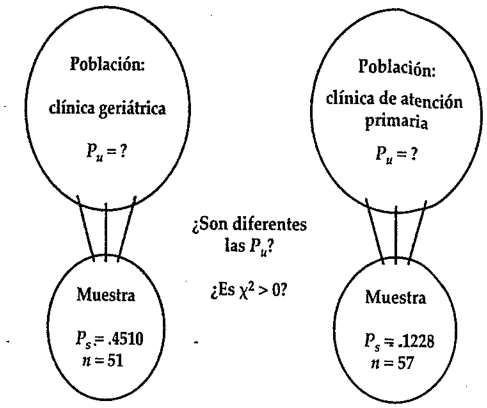
Asimismo, el porcentaje de la columna de cuidadores de clínica de atención primaria canalizados a grupos de apoyo es de 12.28% [es decir (7/57) (100)]. Aparentemente existe una gran diferencia entre estos dos porcentajes; sin embargo, debemos recordar que se trata de datos muestrales. Cualquier diferencia entre los dos porcentajes quizás se deba al error de muestreo; así que debemos probar una hipótesis para determinar si estos porcentajes son significativamente diferentes para las poblaciones de cuidadores de las dos clínicas. Desde un punto de vista conceptual, este problema de investigación se puede concebir de acuerdo con la figura 13-2. Aunque las fracciones se multiplican por 100 para obtener los porcentajes de columna, por coherencia denominamos a esta acción una *prueba de proporciones* y utilizamos el símbolo P_u en la figura 13-2.

Utilizando la prueba chi cuadrada, podemos analizar la pregunta de investigación relativa a saber si existe una diferencia significativa en el porcentaje de cuidadores canalizados a grupos de apoyo de acuerdo con el tipo de clínica.

Para ser prácticos, nos regimos por la solución del recuadro anterior y resolvemos este problema. La hipótesis nula (H_0) se refiere al hecho de que no existe una diferencia entre las clínicas en lo que se refiere a la canalización a grupos de apoyo, lo cual quiere decir que la chi cuadrada es igual a cero. La hipótesis alternativa consiste en que la chi cuadrada es

FIGURA 13-2

Comprobación de la diferencia de proporciones o porcentajes entre grupos



mayor que cero. Sin embargo, si $P = p$ [canalizado a grupos de apoyo], la H_0 también puede enunciarse de la siguiente manera:

$$H_0: P_{u(\text{Cuidadores en clínicas geriátricas})} = P_{u(\text{Cuidadores en clínicas de atención primaria})}$$

En el paso 4 observamos un valor de la chi cuadrada de 14.43 con 1 grado de libertad. El valor p es menor que .001; por consiguiente, en el paso 5 rechazamos la H_0 al nivel de significancia de .05. El paso 6 de la inferencia estadística se lee:

6. *Interpretación:* aspectos de la relación y mejores estimaciones.

   Existencia: hay una relación significativa entre el tipo de clínica y el porcentaje de cuidadores canalizados a un grupo de apoyo; $\chi^2 = 14.43$; $p < .001$.

   Aplicaciones prácticas: los cuidadores en las clínicas geriátricas tienen mayor probabilidad de que se les canalice a grupos de apoyo. *Mejores estimaciones:* 45.10% de los cuidadores que acompañan a los pacientes a la clínica geriátrica son canalizados a grupos de apoyo, en comparación a sólo 12.28% de los cuidadores que acompañan a los pacientes a la clínica de atención primaria.

   Respuesta: los cuidadores que acompañan a los pacientes a la clínica geriátrica tienen mayor probabilidad de ser canalizados a un grupo de apoyo. Las suposiciones de la clínica están justificadas.

En las primeras etapas del análisis de datos, después de que cada variable se examina en lo que se refiere a peculiaridades como sesgos, la siguiente etapa consiste en aplicar pruebas bivariadas. Cuando se comparan dos o más grupos, la prueba t para la diferencia de medias con dos grupos (capítulo 11), la prueba del análisis de varianza (ANOVA, capítulo 12) y la prueba chi cuadrada proporcionan una manera breve de presentar resultados de forma tabular, como se muestra en la siguiente sección.

Presentación tabular de los datos

Como se ha señalado en varios capítulos, el objetivo principal de la investigación normalmente consiste en explicar la variación de una variable resultado —una sola variable dependiente—. Sin embargo, a veces el foco se centra en una sola variable nominal/ordinal independiente —dos o más grupos— y los grupos se comparan en el caso de diversas varia-

TABLA 13-6 | Comparación de los cuidadores de clínicas de atención primaria y clínicas geriátricas en relación con servicios recibidos, variables demográficas y resultados de estrés ($n = 108$)

Variables		Cuidadores que acompañan pacientes a: Clínica de atención primaria ($n = 57$)	Clínica geriátrica ($n = 51$)	Significancia
Servicios recibidos				
Porcentaje canalizado a grupos de apoyo		12.28%	45.10%	‡
Porcentaje al que se le informó de servicios de asesoría		5.26	43.14	‡
Porcentaje al que se le informó de agencias de asesoría que brindan apoyo a cuidadores		14.04	45.10	‡
Porcentaje canalizado a asesoría legal		1.75	23.53	†
Porcentaje aconsejado sobre seguridad en el hogar		10.53	52.94	‡
Datos demográficos				
Edad del cuidador	Media	55.1 años	61.2 años	*
	Desviación estándar	12.5	12.1	
Años de educación	Media	10.6 años	11.1 años	NS
	Desviación estándar	3.1	3.2	
Resultados de estrés				
Tensión en el cuidador	Media	17.2	21.2	‡
	Desviación estándar	4.7	6.5	‡
Depresión psicológica	Media	17.0	18.7	NS
	Desviación estándar	11.2	11.5	

NS = no significativo.
* $p < .05$.
† $p < 0.01$.
‡ $p < 0.001$.
Fuente: Clair, Rithey y Allman, 1993.

bles dependientes. Por ejemplo, en la ilustración de la prueba chi cuadrada relacionada con las diferencias de las proporciones, comparamos dos grupos de cuidadores. La tabla 13-6 incluye datos adicionales para comparar estos dos grupos. La significancia estadística de las diferencias entre porcentajes se determinó con pruebas chi cuadrada. Esta tabla también compara dos grupos respecto a variables de razón/intervalo (es decir, diferencias de medias) utilizando pruebas t de muestra independiente (capítulo 11).

Una tabla similar a la tabla 13-6 abarca una gran cantidad de información en una página. Primero, muestra que la clínica geriátrica se preocupa más de los cuidadores de los pacientes. Indica que casi la mitad de los cuidadores son canalizados a grupos de apoyo o servicios de asesoría y agencias comunitarias, y a casi la mitad se le da consejo sobre seguridad en el hogar, un tema importante entre las personas de edad. En la clínica de atención primaria, a reducidos porcentajes se les advierte sobre estos asuntos. Segundo, los datos en la parte inferior de la tabla —basados en pruebas t— son importantes para conducir análisis multivariados adicionales. Los cuidadores de los pacientes de la clínica geriátrica tienen una edad media más alta de aproximadamente seis años. Aunque la tasa de depresión psicológica no es estadísticamente diferente entre los grupos, los cuidadores de clínicas geriátricas experimentan mayor tensión en el papel de cuidadores. Estas conclusiones permiten explicar la razón por la cual la mayor parte de los cuidadores de clínicas geriátricas se canalizan a grupos de asesoría y apoyo. Tal vez sus parientes se encuentran más enfermos, lo cual genera mayor tensión en los cuidadores. Y quizás la falta de servicios para los cuidadores en las clínicas de atención primaria es resultado del hecho de que especialmente los pacientes enfermos son canalizados a las clínicas geriátricas. Además, se debe llevar a cabo un análisis adicional con controles relacionados con la severidad de la condición del paciente, la edad del cuidador y otras variables. Una de las principales funciones del análisis bivariado —como el que aparece en la tabla 13-6— consiste en guiar el proceso de investigación a la siguiente etapa.

Un par de comentarios técnicos resultan pertinentes en lo que se refiere a la forma en que se diseñó la tabla 13-6. Observa que en la clínica de atención primaria, sólo aproximadamente 2% de los cuidadores fueron canalizados a asesoría legal y sólo 5% fue notificado sobre otros servicios de asesoría. Estos pequeños porcentajes requirieron una corrección de Yates en el cálculo del estadístico chi cuadrada. Finalmente, observa en el caso de la variable carga del cuidador, que no solamente las medias grupales sino también las desviaciones estándares grupales resultaron significativamente diferentes. Hubo una dispersión en la distribución significativamente más amplia en las puntuaciones de atención de los cuidadores en las clínicas geriátricas. La prueba t relativa a las diferencias en las medias para la variable de tensión requirió la estimación de la varianza separada del error estándar como consecuencia de la desigualdad de las varianzas (o desviaciones estándar). Algunas revistas científicas requieren notas de pie de página en las tablas para especificar estos puntos estadísticos finos, mientras que otras alegan que sus lectores supondrán que se llevaron a cabo dichas correcciones. En el caso de una audiencia, no existen procedimientos estadísticos que discutan la cuestión. Los datos se aceptan sobre la base de la autoridad del orador. Además, las representaciones gráficas, en particular las gráficas de barras, resultan un recurso más adecuado para comunicar un sentido de proporción a un auditorio.

Prueba de proporciones con una muestra única pequeña: distribución binomial

Como la palabra *proporción* implica, una *prueba de proporciones con muestra única pequeña* se utiliza como variable de nivel nominal, en la que $P = p$ [de la categoría de éxito] de la variable y $Q = p$ [de las categorías de fracaso]. Esta prueba se aplica cuando el menor entre P_u y Q_u, multiplicado por n, resulta menor que 5 (es decir, $[(p_{menor}\,(n)] < 5)]$. De acuerdo con el capítulo 10, recordemos que cuando $[(p_{menor})\,(n)] \geq 5$, se aplica la prueba de proporciones con muestra única grande.

El estadístico de prueba para la prueba de proporciones con muestra única pequeña recibe el nombre de *prueba de la distribución binomial.* El terminó *binomial* significa "dos números", y esta prueba se aplica con variables dicotómicas muy comunes —las que tienen sólo dos categorías—. Ejemplos de variables dicotómicas son el género (masculino/femenino), la clasificación de raza en mayoritaria/minoritaria (blanco/no blanco), el lanzamiento de la moneda (cara/cruz), la actitud con respecto a un asunto (a favor/en contra), la existencia de un atributo (presente/ausente), el resultado de un experimento (con efecto/sin efecto), la efectividad de un tratamiento (eficaz/no eficaz) y los resultados de grupos en riesgo, como las víctimas de un ataque cardiaco (sobrevive/muere). También debe observarse que en el caso de variables de cualquier nivel de medición, con frecuencia *dicotomizamos* una variable dividiendo las puntuaciones en dos grupos. Por ejemplo, en un estudio relacionado con jubilados podría resultar de utilidad comparar a los que se jubilan antes y después de los 65 años (la edad a la que los beneficios del Seguro Social se vuelven asequibles).

La distribución binomial resulta de particular utilidad en experimentos de laboratorio con pocos individuos. Por ejemplo, supongamos que las investigaciones anteriores revelan que se espera que una mitad (una proporción de 0.5000) de los pacientes a los que se les aplica la quimioterapia experimentan náuseas. Se tiene una muestra de ocho pacientes y se espera que cuatro de ellos experimenten náuseas. Se les administra un nuevo fármaco que reduce el malestar. Si éste resulta eficaz, una cantidad de pacientes *significativamente menor* de cuatro debería experimentar náuseas. Supongamos que con el nuevo fármaco sólo uno de ocho (12.50%) experimenta náuseas, lo cual representa tres casos menos de náuseas que los que se esperan sin la aplicación del fármaco, lo que sugiere que el fármaco es eficaz. Ahora bien, ¿en realidad esto significa que el fármaco resulta eficaz? ¿Podría haber surgido este resultado aparentemente favorable del error de muestreo aleatorio? En otras palabras, en un segundo grupo de ocho pacientes a los que se les aplica el nuevo fármaco, siete podrían experimentar náuseas. En un tercer grupo, alguna otra cantidad aleatoria de pacientes puede experimentar náuseas. Una prueba de hipótesis resolvería la pregunta sobre cuán inusual es el resultado en el muestreo repetido. La distribución binomial proporciona una forma rápida de determinar este valor p y de establecer si la reducción observada de la cantidad de pacientes con náuseas es estadísticamente significativa. Esta prueba de hipótesis conduciría a una de las siguientes conclusiones:

1. el fármaco resulta eficaz y reduce las náuseas en una proporción adicional de tres de ocho casos;
2. el fármaco no resulta eficaz. La reducción de casos de náuseas se debe al error de muestreo.

Cuándo aplicar una prueba de proporciones con muestra única pequeña (distribución binomial)

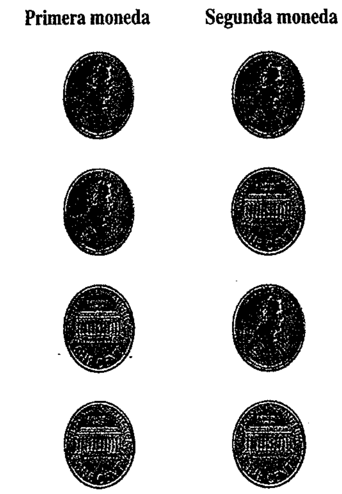
En general: una variable nominal u ordinal dicotómica y una muestra.

1. Sólo hay una variable nominal dicotómica (es decir, quien tiene *solamente dos* categorías, donde $P = p$ [de la categoría de éxito] y $Q = q$ [de la categoría de fracaso]).
2. Existe una sola muestra representativa de una población.
3. Del tamaño de la muestra es tal que $[(p_{menor}\,(n)] < 5)$, donde p_{menor} = el menor de P_u y Q_u.
4. Hay un valor meta de la variable con respecto al cual podemos comparar la proporción de la muestra.

Ecuación de la distribución binomial

Para ilustrar la forma en que se predicen los resultados muestrales en el caso de variables dicotómicas, consideremos el ejemplo que consiste en lanzar dos monedas. En el capítulo 6, cuando estudiamos las probabilidades, indicamos que en el lanzamiento repetido, los resultados de dos monedas lanzadas son predecibles, como lo muestra la figura 13-3.

FIGURA 13-3
Posibles resultados del lanzamiento de dos monedas



Los primeros matemáticos cuentan frijoles o, en este caso, lanzan monedas, con seguridad se valieron de lanzamientos de monedas para descubrir la previsibilidad de eventos muestrales. Quizás lanzaron dos monedas una y otra vez y anotaban la frecuencia de los resultados, cara o cruz. Posteriormente hicieron lo mismo con tres monedas, cuatro, etc. Después de observar un patrón o *ley natural*, estos matemáticos procedieron a crear una fórmula para calcular los resultados con rapidez. En el caso de resultados dicotómicos, como cara o cruz, sí y no, presente y ausente, entre otros, la ecuación binomial describe con rapidez todos los resultados posibles y la probabilidad de cada uno. Esta ecuación se puede aplicar como una distribución muestral.

Una distribución binomial es una ampliación de la siguiente ecuación binomial general:

Ecuación de la distribución binomial

$$(P + Q)^n$$

donde

$P = p$ [de la categoría de éxito]
$Q = p$ [de la categoría de fracaso]
n = tamaño de la muestra

Recuerda que $P + Q = 1$; por tanto, $Q = 1 - P$ y $P = 1 - Q$. Desarrollar la ecuación consiste en elevarla a la potencia representada por n. Por ejemplo, si n es 2, entonces $(P + Q)$ se eleva a la segunda potencia (o sencillamente al cuadrado):

$$(P + Q)^2 = (P + Q)(P + Q) = (P)(P) + (P)(Q) + (Q)(P) + (Q)(Q)$$
$$= P^2 + 2PQ + Q^2$$

Para el lanzamiento de dos monedas, definamos una cara como un éxito y una cruz como un fracaso: $P = p$ [de cara] y $Q = q$ [de cruz]. El tamaño de la muestra es $n = 2$, ya que solamente se lanzan dos monedas. La ecuación desarrollada proporciona todos los posibles resultados y la probabilidad de cada uno, delineando todas las combinaciones posibles de caras y cruces correspondientes a la figura 13-3:

$$(P + Q) = P + 2PQ + Q^2$$

$$= \text{caras, caras} + (2)\ \text{caras, cruces} + \text{cruces, cruces}$$

Los coeficientes de la ecuación (1, 2 y 1) representan la frecuencia con la que ocurre cada combinación en una gran cantidad de pruebas o muestras. (Observa que no se representó el cociente 1, ya que se vuelve redundante e innecesario.) Como p [de cara] = 0.5 y p [de cruz] = 0.5, ambas P y Q = 0.5. De esta manera, la distribución muestral del lanzamiento de dos monedas es la siguiente:

$$\begin{aligned}(P + Q)^2 &= P^2 + 2PQ + Q^2\\ &= (.5)^2 + 2(.5)(.5) + (.5)^2\\ &= .25 + .5 + .25\\ &= \frac{1}{4} + \frac{2}{4} + \frac{1}{4}\end{aligned}$$

En otras palabras, como resulta evidente a partir de la figura 13-3, cuando se lanzan dos monedas, el resultado *dos caras* se presenta una de cuatro veces; la combinación *cara-cruz* ocurre en dos de cuatro veces, o la mitad de las veces; y *dos cruces* ocurre una de cuatro veces.

Fórmula breve para desarrollar la ecuación binomial

El desarrollo de la ecuación binomial resulta una tarea difícil. Por ejemplo, para describir la distribución muestral en el caso del lanzamiento de cuatro monedas, nos enfrentamos a lo siguiente:

$$(P + Q)^4 = (P + Q)(P + Q)(P + Q)(P + Q), \text{etcétera}$$

Afortunadamente, los matemáticos valoran la parsimonia, la *preferencia por la economía y la simplicidad por encima de la complejidad.* Esta actitud se ejemplifica en el enunciado: "La solución más simple es la mejor", y los matemáticos dedican mucho tiempo a simplificar cálculos. Blaise Pascal (1623-1662) ideó un método breve para desarrollar la ecuación binomial. Primero notó que, una vez desarrollada, cualquier ecuación binomial cuenta con $n + 1$ términos, cada uno de los cuales representa una combinación de los resultados P y Q. Por ejemplo, en el lanzamiento de tres monedas, se tienen cuatro términos o combinaciones: 1) tres caras; 2) una cara y dos cruces; 3) una cara y dos caras; 4) tres cruces. Segundo, Pascal descubrió que los coeficientes de la ecuación —los números que indican la frecuencia con la que ocurre una combinación— se rigen por un patrón predecible. Dicho descubrimiento recibe el nombre de *triángulo de Pascal* (tabla 13-7). En el caso de las variables dicotómicas, el triángulo de Pascal permite calcular con rapidez las probabilidades de todos los posibles resultados. Al sumar los coeficientes conforme se desciende, el triángulo proporciona los coeficientes de los resultados para cualquier tamaño muestral n. Además, en el caso muy común en el que P y Q son iguales, en otras palabras, cuando $P = 0.5$ y $Q = 0.5$, la suma de los coeficientes proporciona el número total de combinaciones de resultados y, por consiguiente, el denominador para calcular las probabilidades.

TABLA 13-7 | Triángulo de Pascal de coeficientes de ecuaciones binomiales para tamaños de muestra de 1 a 10

n	Coeficientes	Suma de coeficientes (= 2 a la *n*-ésima potencia)*
	1	
1	1 1	2
2	1 2 1	4
3	1 3 3 1	8
4	1 4 6 4 1	16
5	1 5 10 10 5 1	32
6	1 6 15 20 15 6 1	64
7	1 7 21 35 35 21 7 1	128
8	1 8 28 56 70 56 28 8 1	256
9	1 9 36 84 126 126 84 36 9 1	512
10	1 10 45 120 210 252 210 120 45 10 1	1 024
Etc.		

*Utiliza esta suma de coeficientes como un denominador de las probabilidades sólo cuando $P = Q = 0.5$.

De nuevo, ilustremos el caso de dos monedas empleando los coeficientes para $n = 2$ de la tabla 13-7:

$$(P + Q)^2 = P^2 + 2PQ + Q^2$$

De acuerdo con el triángulo de Pascal: 1 (cara, cara) + 2 (cara, cruz) + 1 (cruz, cruz). Como P y Q son iguales en el caso de las monedas, la probabilidad de un resultado se calcula rápidamente dividiendo el coeficiente para cualquier combinación entre la suma de los coeficientes de la ecuación. De esta manera,

$$p\ [\text{de cara y cara}] = P^2 = \frac{1}{4}$$

Con unos cuantos pasos podemos aplicar el triángulo de Pascal para calcular rápidamente las probabilidades de los resultados. Ilustremos el caso del lanzamiento de cinco monedas, $(P + Q)^5$, donde $P = p$ [de cara], $Q = p$ [de cruz] y $n = 5$ (para cinco monedas).

1. Puesto que $n + 1 = 6$, escribe el término "PQ +" seis veces para obtener la siguiente ecuación parcialmente completa:

$$(P + Q)^5 = PQ + PQ + PQ + PQ + PQ + PQ$$

2. Especifica el exponente de cada término, el cual representará los resultados posibles (posibles combinaciones de caras y cruces). A lo más, se pueden obtener cinco caras, lo cual no da cabida a cruces. De esta manera, los exponentes de P inician con 5 (es decir, el valor de n) y después se reduce a cero conforme avanzamos en la ecuación. Los exponentes de Q inician con cero e incrementan hasta n conforme avanzamos. Por tanto,

debemos iniciar con P, asignando los exponentes 5, en seguida 4 y así sucesivamente, de izquierda a derecha. Asimismo, inserta cero para la primera Q y así sucesivamente, para obtener la siguiente ecuación parcialmente completa:

$$(P + Q)^5 = P^5Q^0 + P^4Q^1 + P^3Q^2 + P^2Q^3 + P^1Q^4 + P^0Q^5$$

Ya que cualquier número elevado a la potencia cero = 1, esta ecuación se puede simplificar omitiendo dichos términos para obtener la siguiente ecuación parcialmente completa:

$$(P + Q)^5 = P^5 + P^4Q^1 + P^3Q^2 + P^2Q^3 + P^1Q^4 + Q^5$$

Esto proporciona todas las combinaciones: P^5 representa la combinación de cinco caras y ninguna cruz; P^4Q^1 representa cuatro caras y una cruz y así sucesivamente, hasta Q^5, que representa ninguna cara y cinco cruces.

3. Utiliza el triángulo de Pascal (tabla 13-7) para obtener los coeficientes para cada término. Una vez que se insertan estos coeficientes, tenemos la ecuación binomial desarrollada para $n = 5$:

$$(P + Q)^5 = P^5 + 5P^4Q^1 + 10P^3Q^2 + 10P^2Q^3 + 5P^1Q^4 + Q^5$$

4. Finalmente, ya que P y Q son iguales, de la columna derecha del triángulo de Pascal (tabla 13-7) obtén la suma de los coeficientes de una muestra de $n = 5$ y divide el resultado entre cada coeficiente para obtener las probabilidades de cada resultado:

$$(P + Q)^5 = P^5 + 5P^4Q^1 + 10P^3Q^2 + 10P^2Q^3 + 5P^1Q^4 + Q^5$$
$$= \frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} + \frac{1}{32}$$

Por tanto, la probabilidad de lanzar cinco monedas y obtener cinco caras, es de 1 en 32; cuatro caras y una cruz, 5 en 32; tres caras y dos cruces, 10 en 32, etcétera.

Con un poco de experiencia, el triángulo de Pascal resulta muy eficaz, en particular cuando P y Q son iguales. Para ilustrar, las siguientes respuestas se pueden obtener rápidamente de la tabla 13-7 sin necesidad de plantear las ecuaciones:

1. p [de lanzar siete veces una moneda y obtener solamente caras] $= \frac{1}{128} = .0078$

2. p [de lanzar nueve monedas y obtener siete caras y dos cruces] $= \frac{36}{512} = .0703$

En los casos en los cuales P y Q *no* son iguales, el triángulo de Pascal también se puede aplicar para obtener coeficientes. No obstante, las sumas a la derecha de la tabla no se aplican y los términos de la ecuación deben multiplicarse. Por ejemplo, supongamos que en una habitación llena de personas, el 70% son hombres ($P = 0.7$) y el 30% son mujeres ($Q = 0.3$). Si seleccionamos de manera aleatoria tres personas de la habitación, ¿qué posibles combinaciones de hombre y mujer aparecerán y cuáles serán sus probabilidades? Consultamos el triángulo de Pascal para construir la ecuación y después insertamos los valores 0.7 y 0.3 para obtener las probabilidades de cada combinación:

$$(P + Q)^3 = P^3 + 3P^2Q^1 + 3P^1Q^2 + Q^3$$
$$= (.7^3) + 3(.7^2)(.3) + 3(.7)(.3^2) + (.3^3)$$
$$= .343 + .441 + .189 + .027$$

Observa que las cuatro probabilidades suman 1.0. Nota también que las dos combinaciones predominantemente masculinas a la izquierda de la ecuación se presentan con mucha mayor frecuencia que las dos combinaciones predominantemente femeninas. Esto tiene sentido en virtud de que la proporción de hombres en la habitación es mucho mayor que la de las mujeres. De hecho, el triángulo de Pascal es un recurso de gran utilidad.

Los seis pasos de la inferencia estadística para una prueba de proporciones de una muestra única pequeña: prueba de la distribución binomial

Ahora utilicemos la ecuación binomial como una distribución muestral para pruebas de hipótesis. Supongamos que en la ficticia Universidad de Overlord, el tema de la violación que se presenta en una cita se ha convertido en una cuestión polémica. En respuesta a la presión de los estudiantes y padres, el director de la universidad anuncia que se convocará a un grupo de enfoque para discutir el tema de la violación en las citas y lo que puede hacerse para evitarla. Algunos estudiantes sospechan que para reducir al mínimo la publicidad negativa relacionada con la universidad, el director recurre a un grupo de enfoque para afirmar que la violación durante las citas no representa un problema. Esperan ansiosamente saber quién formará parte del grupo de enfoque. Los estudiantes anticipan que el director lo saturará de hombres, quienes supuestamente son menos solidarios con respecto al problema que las mujeres.

La proporción de hombres y mujeres en el campus es igual. El director revela su selección, la cual se compone de un grupo de ocho estudiantes, seis de los cuales son hombres. Los estudiantes lo acusan de saturar el grupo de hombres. El director ha declarado públicamente que sencillamente llevó a cabo una selección aleatoria de estudiantes por medio de un programa de computadora y de los expedientes de los estudiantes. ¿Miente el director de la universidad?

¿Cómo se puede utilizar en este caso la distribución binomial? Ya que la proporción de hombres y mujeres en el campus es la misma, cada sexo tiene la misma probabilidad de ser elegido, así como una moneda tiene la misma probabilidad de que sea cara o cruz. Así, una verdadera selección aleatoria de estudiantes es como lanzar ocho monedas con, por así decirlo, hombres como caras y mujeres como cruces. Ahora bien, esto no significa que una selección aleatoria de ocho personas siempre dará como resultado cuatro hombres y cuatro mujeres. Al lanzar ocho monedas, hay nueve posibles resultados, que van de nueve caras a nueve cruces. Esperaríamos obtener combinaciones equilibradas, como cuatro caras y cuatro cruces, 5 y 3 o 3 y 5, ya que tales combinaciones se presentan con mayor frecuencia en lanzamientos repetidos. Al utilizarse con el triángulo de Pascal (tabla 13-7), la distribución binomial ofrece una descripción de todas las combinaciones posibles de hombre y mujer en un grupo de ocho, así como la probabilidad de cada combinación *si la selección es, de hecho, aleatoria.*

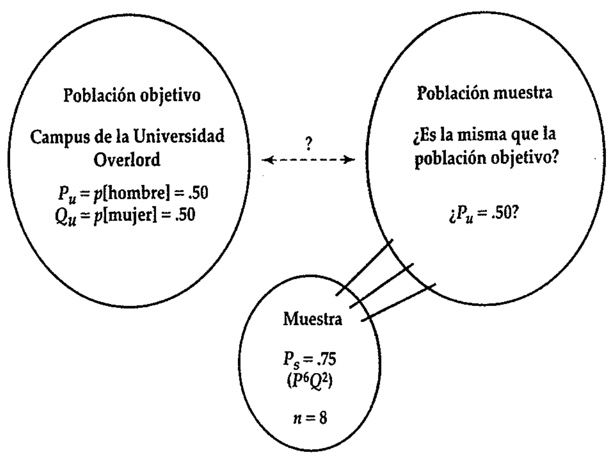
Como con cualquier otra prueba de hipótesis, debe formularse una hipótesis de manera que ésta prediga los resultados muestrales. La distribución binomial proporciona resultados predecibles para una selección aleatoria, en la cual los individuos del sexo masculino y del sexo femenino tienen la misma probabilidad de ser elegidos, y las "monedas" no se encuentran cargadas a favor de alguno de los sexos. Por consiguiente, podemos formular la hipótesis nula de la siguiente manera: el director no miente; el grupo de enfoque se eligió de manera aleatoria. Esta prueba de hipótesis *no* tiene sesgo alguno. Si hallamos que el resultado del director es tan común como el obtenido con una selección aleatoria o *ciega*, nos apegaremos a esta hipótesis nula de no sesgo. La hipótesis alternativa consiste en que el director saturó intencionalmente el grupo de enfoque con hombres y mintió cuando declaró que llevó a cabo una selección aleatoria. El efecto de la prueba es la diferencia entre el número observado en la categoría, P, y el número esperado, suponiendo que no hay ningún sesgo (es decir, suponiendo que P_u es verdadera). Ésta es una prueba de una cola, ya que declara que los hombres tenían una oportunidad *mayor* de ser elegidos. Nuestra decisión de rechazo depende del hecho de que una combinación tan inusual o más inusual que incluye seis hombres y dos mujeres sea poco probable que se presente por casualidad. ¿Se debe el efecto de tener dos hombres más de lo predicho por un resultado de 50-50 al error de muestreo aleatorio o a un sesgo de parte del director de la universidad? En otras palabras, en el lanzamiento repetido de ocho monedas, ¿es inusual obtener seis caras o más? ¿Ocurre esto con mucha frecuencia, como por ejemplo, más del .05 de las ocasiones, o es éste un evento poco probable? Enmarquemos el problema con los seis pasos de la inferencia estadística.

Solución para una prueba de proporciones de una única muestra pequeña (distribución binomial)

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿dice la verdad el director de la Universidad Overlord cuando afirma que su grupo de enfoque fue elegido aleatoriamente (es decir, los hombres y las mujeres tuvieron la misma oportunidad de ser elegidos)? *Procedimiento estadístico:* prueba de proporciones con muestra única pequeña, distribución binomial; valor objetivo: $P_u = 0.50$ (la proporción real conocida de estudiantes varones en la población de estudiantes de la universidad Overlord). *Observación:* $n = 8$. Se seleccionaron seis hombres y dos mujeres para el grupo de enfoque (que corresponde al término P^6Q^2 de la fórmula binomial). Expresado en forma de proporción,

$$P_s = \frac{\#\text{ de hombres}}{n} = \frac{6}{8} = .75$$



SEIS PASOS

1. H_0: $P_{u(\text{población muestreada})} = .50$ (proporción conocida de hombres (P_u) en la población de la Universidad Overlord.
   Es decir que el director no mintió; el grupo de enfoque se eligió aleatoriamente.
   H_A: $P_{u(\text{población muestreada})} > .50$ (la población conocida de hombres (P_u) en la población objetivo de la universidad). De una cola.
   Es decir que el director miente. Los hombres se encuentran sobrerrepresentados, lo cual sugiere que el grupo de enfoque estaba saturado de hombres.
2. *Distribución muestral:* si la H_0 es verdadera y se extraen repetidamente muestras del tamaño 8 de la población estudiantil de la Universidad Overlord, las combinaciones de hombres y mujeres se ajustarán a la distribución binomial con $P = .5$, $Q = .5$, $n = 8$ y con los coeficientes del triángulo de Pascal (tabla 13-7):
$$(P+Q)^8 = P^8 + 8P^7Q^1 + 28P^6Q^2 + 56P^5Q^3 + 70P^4Q^4 + 56P^3Q^5 + 28P^2Q^6 + 8P^1Q^7 + Q^8$$
3. *Nivel de significancia:* $\alpha = .05$. De una cola.
4. *Observación:* $n = 8$
   Efecto de la prueba: seis hombres y dos mujeres; dos hombres más de lo esperado.
   Estadístico de la prueba:
$$= (P+Q)^8 = P^8 + 8P^7Q^1 + 28P^6Q^2 + 56P^5Q^3 + 70P^4Q^4 + 56P^3Q^5 + 28P^2Q^6 + 8P^1Q^7 + Q^8$$
$$= \underbrace{\frac{1}{256} + \frac{8}{256} + \frac{28}{256}}_{37/256\,=\,.1445} + \frac{56}{256} + \frac{70}{256} + \frac{56}{256} + \frac{28}{256} + \frac{8}{256} + \frac{1}{256}$$
   Valor *p:* p [de observar un resultado muestral tan o más inusual que seis hombres y dos mujeres, cuando la proporción real de la población P_u es .50] = 0.1445.
5. *Decisión de rechazo:* $p > \alpha$, $0.1445 > 0.05$. No se rechace la H_0.
6. *Interpretación*
   Existencia: parece que el director no miente; el grupo de enfoque pudo haberse elegido aleatoriamente. *Mejores estimaciones:* la aparente sobrerrepresentación de hombres en el grupo de enfoque podría ser fácilmente consecuencia del error de muestreo aleatorio. Una combinación tan extrema o más extrema que seis hombres y dos mujeres se presenta más de 14% de las veces en dichas elecciones aleatorias. *Respuesta:* no hay razones para creer que el director miente.

Algunas cuestiones relacionadas con esta prueba de hipótesis:

- En el paso 2 observa la eficacia de la ecuación binomial. A diferencia de la prueba Z, la prueba t y la prueba de la razón F, esta distribución muestral no se describe por medio de una curva de frecuencias, lo cual significa que no se requiere una tabla para una curva de frecuencias (como la tabla de la distribución t). Esta sola ecuación concisa reúne los requisitos para describir la distribución muestral. Describe todas las combinaciones posibles de hombres y mujeres en un grupo de ocho, así como la probabilidad de que se presente cada combinación en el muestreo repetido.
- En el paso 4, el término P^6Q^2 de la ecuación binomial representa la combinación de seis hombres y dos mujeres, nuestro resultado muestral. Cuando se presenta un muestreo "honesto", sería más inusual extraer siete hombres y una mujer u ocho hombres y ninguna mujer. De esta manera, el valor p para el resultado muestral incluye las probabilidades de P^7Q^1 y P^8. Se añadieron las probabilidades de todos los términos en la *cola* de la ecuación —los que son tan inusuales o más inusuales que P^6Q^2—.
- Si en el paso cuatro *hubiéramos llevado a cabo una prueba de dos colas*, también habríamos incluido las probabilidades de P^2Q^6, P^1Q^7 y Q^8, ya que habría resultado tan inusual elegir aleatoriamente estas combinaciones sesgadas de mujeres como seleccionar combinaciones sesgadas de hombres. El valor p se había calculado sumando las probabilidades de ambas *colas* de la ecuación de la siguiente manera:

$$(P+Q)^8 = P^8 + 8P^7Q^1 + 28P^6Q^2 + 56P^5Q^3 + 70P^4Q^4 + 56P^3Q^5 + 28P^2Q^6 + 8P^1Q^7 + Q^8$$

$$= \underbrace{\frac{1}{256} + \frac{8}{256} + \frac{28}{256}}_{37/256\,=\,.1445} + \frac{56}{256} + \frac{70}{256} + \frac{56}{256} + \underbrace{\frac{28}{256} + \frac{8}{256} + \frac{1}{256}}_{37/256\,=\,.1445}$$

En otras palabras, puesto que la ecuación es simétrica cuando P y Q son iguales (es decir, las combinaciones se encuentran equilibradas en ambos miembros), el valor p se duplica y resulta $(2)(0.1445) = 0.2890$.

- En los pasos 5 y 6 de este ejemplo no rechazamos la hipótesis nula. En situaciones en las que se rechaza la hipótesis nula, proporciona las mejores estimaciones de P_u o Q_u reportando la categoría que se encuentra sobrerrepresentada; por ejemplo, "los hombres se encuentran sobrerrepresentados".

Insensatez y falacias estadísticas: bajo poder estadístico cuando el tamaño de la muestra es pequeño

En nuestra ilustración de la ecuación binomial encontramos que, aunque su grupo de enfoque parecía cargarse en favor de los hombres, *no se tenía ninguna razón para creer* que el director de la Universidad Overlord mentía cuando declaró que el grupo había sido seleccionado al azar. Aunque le otorgamos el beneficio de la duda, todavía es posible que saturara de hombres el grupo de enfoque.

Para empezar, al comprobar la hipótesis nula de que el director *no* mentía, utilizamos el nivel de significancia de .05. Esto mantuvo la posibilidad de un error de tipo I (es decir, rechazar la hipótesis nula cuando, de hecho, es verdadera) a un nivel ligeramente bajo. En otras palabras sólo aceptamos un 5% de probabilidad para afirmar que mentía si, de hecho, no lo hacía. Los estudiantes que lo acusan de sesgar la muestra quizás preferían un nivel de significancia mayor (por decir algo, .20), ya que ellos ven el asunto desde un punto de vista político, no científico. Quizás arguyan que hicimos que resultara prácticamente imposible rechazar la hipótesis nula, ya que un valor p menor que .05 sólo podría obtenerse si el director hubiera elegido a siete hombres y una mujer, o a ocho hombres:

$$(P+Q)^8 = P^8 + 8P^7Q^1 + 28P^6Q^2 + 56P^5Q^3 + 70P^4Q^4 + 56P^3Q^5 + 28P^2Q^6 + 8P^1Q^7 + Q^8$$

$$= \underbrace{\frac{1}{256} + \frac{8}{256}}_{9/256\,=\,0.0352} + \frac{28}{256} + \frac{56}{256} + \frac{70}{256} + \frac{56}{256} + \frac{28}{256} + \frac{8}{256} + \frac{1}{256}$$

$9/256 = 0.0352$ (que es inferior a $\alpha = 0.05$, de una cola)

Estos estudiantes protestarían con el argumento de que el director habría tenido que seleccionar, no uno, ni dos, sino tres hombres para que se le tildara de mentiroso con semejante nivel alfa bajo.

Lo que en realidad expresan estos estudiantes es su sospecha de que ocurrió un error tipo II (véase capítulo 9). Recuerda que un error tipo II se presenta cuando en una prueba estadística no se rechaza una hipótesis nula falsa. Recuerda, asimismo, que β es la probabilidad de cometer un error tipo II y que éste se encuentra inversamente relacionado con α: cuando α se incrementa, β disminuye, y viceversa. En este caso, un error tipo II habría ocurrido si el director en realidad hubiera mentido, pero concluimos lo contrario.

Si elevamos el nivel de significancia (α), podríamos reducir β —la posibilidad de cometer un error tipo II—. Por ejemplo, si hubiéramos utilizado un $\alpha = .20$, habríamos rechazado la honestidad del director cuando eligió a seis hombres y dos mujeres, ya que el valor p que teníamos para la prueba era de 0.1445, que es inferior a 0.20. Si hubiéramos hecho esto, sin embargo, el director nos podría haber acusado de prejuicio a favor de los estudiantes. Lo que ilustra este embrollo es el hecho de que una muestra pequeña constituye un recurso sumamente débil para determinar si los hechos que se observan son consecuencia del error de muestreo. Ya que la ecuación binomial se utiliza en pequeñas muestras, sus resultados son poco sustanciales. El lenguaje estadístico afirma que un estadístico tan limitado tiene un bajo poder estadístico. El **poder estadístico** implica *la probabilidad de que un estadístico de prueba no incluya un error tipo II para un nivel de significancia determinado.* En términos matemáticos,

$$\text{Poder de una prueba} = 1 - \beta = 1 - p \text{ [de un error tipo II]}$$

De acuerdo con el capítulo 10, calcular el poder de una prueba estadística es muy complejo y rebasa el alcance de este libro. Sin embargo, con la ecuación binomial es fácil ver que los estadísticos de prueba de muestras pequeñas son limitados.

Un caso más sencillo puede ayudarnos a recordar esta limitación. Supongamos que el director de la universidad sólo hubiera elegido a tres estudiantes para el grupo de enfoque. Con un nivel de significancia de .05, ¿sería posible rechazar la hipótesis nula de que mintió? La respuesta es no. Inclusive en el peor de los casos en el que los tres estudiantes elegidos resultaran ser hombres, no habría forma de rechazar la hipótesis nula. De acuerdo con el triángulo de Pascal, con tres personas elegidas al azar, la probabilidad de seleccionar exclusivamente hombres es de 1 en 8, o .1250. Es decir, que en el peor de los casos, un valor p de .1250 es mayor que 0.05. No habría forma de rechazar la hipótesis nula utilizando una muestra tan pequeña de bajo poder y la ecuación binomial. Incluso un resultado que incluya solamente hombres no sería inusual en una honrada selección aleatoria de tres estudiantes.

Existe una segunda razón para mantenerse escéptico con respecto a la conclusión a la que llegamos relacionada con el hecho de que el director no mentía. Como se señaló en el capítulo 1, el trabajo estadístico ocurre en un contexto político y cultural más amplio. Se encuentra sujeto a distorsión por parte de usuarios deshonestos. Nuestra prueba de hipótesis sencillamente determinó que dicho director se encontraba en condiciones para garantizar que su selección fue aleatoria. De hecho, él podría haber colocado a seis hombres de manera intencional en el grupo de enfoque. Quizás sabía que ése era el número máximo de hombres que podrían estar en el grupo sin rechazar una hipótesis nula con un nivel de significancia de .05. Las conclusiones estadísticas basadas en pequeñas muestras ni siquiera son completamente seguras, incluso cuando las conclusiones de la prueba estadística apoyan determinada postura. Ésta es la razón por la cual la ética representa una parte importante del trabajo estadístico en la ciencia. Un especialista en estadística honesto no apoyaría la postura del director sin destacar que es difícil obtener una conclusión sólida con una muestra pequeña.

Suponiendo que, de hecho, el director de la Universidad Overlord fuera honesto, éste podría haber impedido la protesta estudiantil utilizando cuotas de género al formar el grupo de enfoque. Es decir, hubiera podido elegir de forma aleatoria exactamente a cuatro hombres y cuatro mujeres —las proporciones adecuadas de representación del género en la población del campus—. Esto le hubiera ahorrado muchos dolores de cabeza.

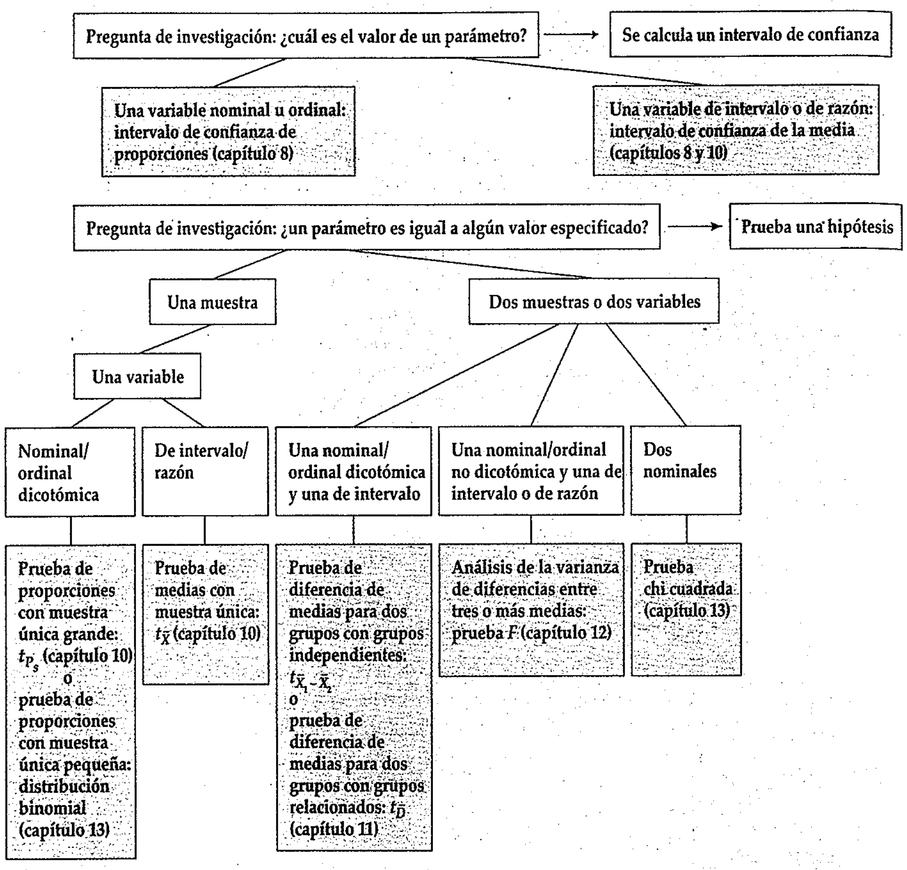
RESUMEN

1. En la investigación de las ciencias sociales, muchas variables de estatus, como al género y la raza, tienen un nivel de medición nominal. La prueba chi cuadrada, que se basa en los cálculos realizados a partir de una tabla cruzada, permite analizar la relación entre dos variables nominales.
2. Los cálculos para la prueba chi cuadrada se llevan a cabo utilizando una tabulación cruzada, la cual contiene las frecuencias de las ocurrencias conjuntas de las categorías de dos variables nominales u ordinales. Las categorías de la variable independiente (X) se colocan en las columnas de la tabla cruzada y las de la variable dependiente (Y) en los renglones.
3. En el caso de un solo individuo, una ocurrencia conjunta de categorías involucra pares de categorías de las dos variables nominales/ordinales, como blanco-demócrata o afroamericano-republicano.
4. Las frecuencias esperadas se calculan y se comparan con las frecuencias observadas reales en las casillas de una tabla cruzada. Las frecuencias esperadas son las frecuencias que se presentarían si no existiera una relación entre las dos variables nominales. Cuando las frecuencias observadas y las esperadas son las mismas, dado un error de muestreo por exceso o defecto, no existe una relación y la ecuación de la chi cuadrada resulta igual a cero. Por tanto, la hipótesis nula consiste en que la chi cuadrada es igual a cero.
5. La distribución muestral es la distribución de la chi cuadrada, en la que los grados de libertad se determinan por medio del número de columnas y filas o renglones de la tabla cruzada: $gl = (r - 1)(c - 1)$.
6. El valor p se obtiene comparando el valor calculado de la chi cuadrada con los valores críticos del estadístico chi cuadrada en la tabla estadística G del apéndice B.
7. Aspectos relevantes de una relación para la prueba chi cuadrada: *a)* Existencia: prueba de H_0 relacionada con el hecho de que $\chi^2 = 0$; es decir que no existe ninguna relación entre X y Y. Si la H_0 se rechaza, existe una relación. *b)* La dirección no es aplicable, porque las variables son de nivel nominal. *c)* La fuerza normalmente no se reporta, ya que las medidas que existen resultan inadecuadas y plagadas de posibles errores. *d)* Las aplicaciones prácticas se describen reportando las diferencias entre las frecuencias observadas y esperadas en el caso de una pareja de casillas destacadas y calculando los porcentajes de columna para las casillas seleccionadas.
8. La prueba chi cuadrada comúnmente se utiliza como una prueba de diferencia de proporciones.
9. La prueba de la distribución binomial es una prueba de proporciones con una muestra única pequeña, la cual se utiliza cuando el tamaño de la muestra es tal que $[(p_{menor}(n)] < 5$, donde p_{menor} = el menor de P_u y Q_u. Contrasta esta prueba de hipótesis con la prueba de proporciones para una muestra única pequeña del capítulo 10.
10. En el caso de la prueba de hipótesis de la distribución binomial, la hipótesis nula se enuncia de la siguiente manera: H_0: P_u = un valor objetivo. El desarrollo de la ecuación y binomial $(P + Q)^n$, proporciona la distribución muestral para eventos dicotómicos. El valor p se calcula directamente con esta ecuación. El triángulo de Pascal proporciona un método corto para desarrollar la ecuación binomial. Permite un cálculo rápido de las probabilidades de todos los resultados posibles, cuando P y Q son iguales a .5. Ésta toma el lugar de una tabla estadística.
11. El poder estadístico es la probabilidad de que un estadístico de prueba no incluya un error tipo II para determinado nivel de significancia. Las muestras pequeñas tienen un poder estadístico bajo. La prueba de la distribución binomial constituye un buen recurso para ilustrar el bajo poder estadístico.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones del capítulo 13 del material del texto disponibles en el sitio web *The Statistical Imagination*, www.mhhe.com/ritchey2 incluyen un amplio material sobre el estadístico de prueba *gama*, que se utiliza para variables de categorías con nivel de medición ordinal.

PROCEDIMIENTOS ESTADÍSTICOS CUBIERTOS HASTA AQUÍ



FÓRMULAS PARA EL CAPÍTULO 13

Prueba chi cuadrada de la relación entre dos variables nominales:

Especificaciones: dos variables nominales/ordinales X y Y o una variable nominal/ordinal para dos grupos. Una tabla cruzada de variables revela por lo menos cinco casos por casilla.

Pregunta de investigación: ¿existe una relación entre las dos variables nominales/ordinales? ¿O hay alguna diferencia entre los dos grupos en las proporciones de una categoría de una variable nominal/ordinal?

H_0: $\chi^2 = 0$

Distribución muestral: la distribución chi cuadrada con

$$gl = (r - 1)(c - 1)$$

Efectos de la prueba: las diferencias entre las frecuencias observadas y las esperadas, $O - E$, mejor ilustradas en los porcentajes de columna.

Estadístico de la prueba: chi cuadrada [con la ayuda de la tabla de la chi cuadrada (tabla estadística G, apéndice B); determina si existe alguna relación]:

Cálculos para obtener el estadístico de la prueba chi cuadrada:

1. Organiza los datos en una tabla cruzada y calcula las frecuencias esperadas de las casillas:

$$E_{casilla} = \frac{(\text{total marginal de columna para la casilla})\,(\text{total marginal de renglón para la casilla})}{\text{gran total}}$$

2. Inserta las frecuencias observadas y esperadas de las casillas en la siguiente hoja de cálculo:

Especificaciones		Cálculos			
Casillas (X, Y)	O	E	$(O - E)$	$(O - E)^2$	$[(O - E)^2/E]$
X-Y casilla	...	...	...	...	...
X-Y casilla	...	...	...	...	...
X-Y casilla	...	...	...	...	...
X-Y casilla	...	...	...	...	...
Totales	n	n	0.0		$\chi^2 = ...$

3. Calcula el estadístico chi cuadrada (de la esquina inferior derecha de hoja de cálculo):

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Determinación de los aspectos de una relación:

Dirección: normalmente no aplicable

Fuerza: no aplicable

Aplicaciones prácticas: calcula los porcentajes de columna de las casillas de interés:

$$\%\text{ de columna [de frecuencia conjunta]} = \frac{\#\text{ en una casilla}}{\#\text{ total en la columna}} \times 100$$

Prueba de proporciones con muestra única pequeña (distribución binomial):

Especificación: una variable dicotómica nominal, donde $P = p$ [de la categoría de éxito] y $Q = q$ [de la categoría de fracaso]; una muestra única representativa de una población; el tamaño de la muestra es tal que $[(p_{menor}(n)] < 5$, donde p_{menor} = el menor de P_u y Q_u y hay un valor objetivo de la variable con el que es posible comparar la proporción de la muestra.

Pregunta de investigación: ¿Es P_u, la proporción de casos en la categoría de éxito, significativamente diferente de un valor objetivo?

H_0: $P_{(población\ muestreada)}$ = valor objetivo conocido

Distribución muestral: distribución binomial.

Efecto de la prueba: diferencia entre el número observado en la categoría, P, y el número esperado, suponiendo que P_u es verdadera.

Estadístico de la prueba: ecuación de la distribución binomial $(P + Q)^n$.

Cálculos para obtener un valor p utilizando la ecuación binomial:

1. Determina si la ecuación binomial es adecuada comprobando que

$$[(p_{menor})\ (n)] < 5$$

(Si $[(p_{menor})\ (n)] \geq 5$, entonces utiliza la prueba de proporciones con muestra única grande del capítulo 10.)

2. Desarrolla la fórmula $(P + Q)^n$ por medio de:
   - *a*) Enlista el término "PQ" $n + 1$ veces;
   - *b*) Especifica los exponentes de cada término (para representar los posibles resultados);
   - *c*) Utiliza el triángulo de Pascal (tabla 13-7) para obtener los coeficientes para término e inserta estos coeficientes para completar la ecuación binomial desarrollada;
   - *d*) Calcula el valor $p = p$ [de combinaciones tan inusuales o más inusuales], agregando las probabilidades en las colas de la ecuación.

PREGUNTAS PARA EL CAPÍTULO 13

1. Un atributo es una cualidad o característica de un individuo, que contiene los nombres de las categorías de una variable nominal/ordinal. Hasta donde lo permitan tus conocimientos, identifica los atributos de las siguientes variables:
   - *a*) nivel de grado de la universidad;
   - *b*) color del cabello;
   - *c*) preferencia religiosa;
   - *d*) especialidad médica.
2. En el caso de un individuo, el apareamiento de atributos de dos variables recibe el nombre de ocurrencia conjunta. ¿A qué variables representa la ocurrencia conjunta mujer-cirujano?
3. Elige dos variables nominales; inventa datos ficticios y elabora una tabla cruzada para ilustrar las frecuencias conjuntas. Rotula las partes de la tabla (casillas, totales, etc.) para mostrar que sabes leer dichas tablas.
4. ¿De qué nivel de medición resultan adecuadas las variables para la utilización de la prueba chi cuadrada?
5. ¿Qué indica una prueba chi cuadrada cuando las frecuencias observadas y las esperadas son iguales?
6. ¿De qué nivel de medición resultan adecuadas las variables para utilizar la prueba de proporciones con muestra única pequeña, y cómo se denomina dicha prueba?

7. ¿Qué significa el término *binomial*? ¿Debe una variable necesariamente ser dicotómica para que se utilice la distribución binomial? Explica tu respuesta.
8. ¿Qué es el poder estadístico? Menciona un ejemplo de conclusión incorrecta basada en una muestra con poder estadístico bajo.

EJERCICIOS PARA EL CAPÍTULO 13

Conjunto de problemas 13A

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, la tabla cruzada y los aspectos adecuados de una relación. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

13A-1. La siguiente tabla cruzada proviene de una muestra aleatoria de médicos. La variable en los renglones indica si el médico es miembro de una especialidad con alto riesgo de que se le demande por negligencia (por ejemplo, obstetricia o neurocirugía). La variable en las columnas, lugar de práctica, indica si se trata de un condado grande (es decir, un condado no aledaño al área metropolitana), una comunidad urbana (es decir, un área metropolitana no afiliada a una escuela médica), un centro académico de ciencias de la salud (es decir, un centro médico asociado a una escuela médica). Calcule los porcentajes de columna para todas las casillas de frecuencias conjuntas.

Riesgo de demanda ↓ / Lugar de práctica →	Condado rural	Comunidad urbana	Centro académico de ciencias de la salud	Totales
Especialidad de alto riesgo	100	462	77	
Especialidad de bajo riesgo	42	279	23	
Totales				

13A-2. De 43 profesores y 55 estudiantes encuestados, 25 profesores y 18 estudiantes se inclinaron por el sistema cuatrimestral en lugar del semestral. Coloca estos datos en una tabla cruzada. Prueba la hipótesis de que existe una relación entre el estatus organizacional y el tipo de sistema académico preferido.

13A-3. Utiliza los siguientes datos de la Encuesta Social General para probar la hipótesis de que quienes se casaron jóvenes tienen mayor probabilidad de divorciarse.

	Casado siendo menor de 20 años	Casado a los 20 años o más
Divorciado	27	52
Nunca divorciado	59	244

13A-4. Organiza los datos de la siguiente hoja de cálculo en forma de tabla cruzada. Prueba la hipótesis de que hay una diferencia significativa entre los residentes del centro de la ciudad y los de los suburbios en términos de apoyo a una política de incremento en los impuestos con el fin de ampliar el departamento de policía.

Lugar de residencia	Apoyo al incremento de impuestos
Suburbano	No
Centro de la ciudad	Sí
Suburbano	No
Suburbano	Sí
Centro de la ciudad	No
Centro de la ciudad	Sí
Suburbano	No
Centro de la ciudad	Sí
Centro de la ciudad	Sí
Suburbano	Sí
Centro de la ciudad	Sí
Centro de la ciudad	No
Suburbano	No
Centro de la ciudad	Sí
Centro de la ciudad	Sí
Suburbano	Sí
Centro de la ciudad	Sí
Suburbano	No
Suburbano	No
Centro de la ciudad	No
Centro de la ciudad	Sí
Suburbano	Sí
Centro de la ciudad	Sí
Centro de la ciudad	No
Suburbano	No
Centro de la ciudad	Sí
Suburbano	No
Centro de la ciudad	Sí
Centro de la ciudad	Sí
Suburbano	Sí
Centro de la ciudad	Sí
Suburbano	Sí
Centro de la ciudad	Sí
Suburbano	No
Centro de la ciudad	No

13A-5. Utiliza el triángulo de Pascal para reportar rápidamente las siguientes probabilidades de lanzamientos de monedas. En este ejercicio no es necesario escribir las ecuaciones.

a) p [de lanzar nueve monedas y obtener sólo caras] =

b) p [de lanzar cuatro monedas y obtener una cara y tres cruces] =

c) p [de lanzar seis monedas y obtener tres caras y tres cruces] =

13A-6. Micro-Medication Corporation se encuentra probando un nuevo fármaco contra el cáncer en ratones genéticamente infectados. Investigaciones anteriores en sus laboratorios demuestran que 50% de estos animales sobrevive seis meses sin ningún tratamiento. Se administra la droga a ocho ratones. Seis de ellos sobreviven seis meses. Prueba la hipótesis de que el tratamiento administrado ofrece una tasa superior de supervivencia que la que se observa con la ausencia del mismo fármaco. Utiliza $\alpha = .001$.

13A-7. Una clínica de fertilidad afirma que cuenta con un procedimiento para elegir el género de un breve. De nueve parejas seleccionadas al azar que querían tener niñas, siete obtuvieron el resultado deseado. Prueba la hipótesis para determinar si el procedimiento en cuestión supera lo esperado por el azar.

13A-8. Supongamos que piensas utilizar la distribución binomial para probar una hipótesis con un nivel de significancia de .01 con una prueba de dos colas. ¿Cuál es el tamaño mínimo de la muestra que se requieren para rechazar posiblemente la hipótesis nula? Muestra los cálculos.

Conjunto de problemas 13B

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, la tabla cruzada, las curvas de probabilidad y los aspectos adecuados de una relación. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

13B-1. La siguiente tabla cruzada proviene de una muestra aleatoria de profesores de ciencias sociales. La variable de renglón, ambiente de enseñanza, indica si un profesor enseña en una universidad pública o privada. La variable de columna, ubicación del trabajo, indica la región de Estados Unidos en la que los profesores laboran. Calcula los porcentajes de columna de todas las casillas de frecuencias conjuntas.

Ambiente de trabajo	Lugar de trabajo →	Noreste	Oeste/ suroeste	Sur	Totales
Universidad pública		192	155	188	
Universidad privada		173	231	55	
Totales					

13B-2. De 64 médicos y 89 pacientes encuestados, 36 médicos y 52 pacientes se pronunciaron a favor de un seguro médico de cobertura universal en el actual sistema médico diversificado. Ordena estos datos en una tabla cruzada. Prueba la hipótesis de que existe una relación entre el estatus médico-paciente y el tipo de sistema de seguro médico preferido.

13B-3. Utiliza los siguientes datos para probar la hipótesis de que los estudiantes de más edad tienen una mayor probabilidad de obtener mejores resultados en las clases de sociología de primer año en una universidad pública grande.

	Matriculados y de menos de 25 años	Matriculados y de 25 años o mayores
Calificación final C o inferior	33	63
Calificación final A o B	72	296

13B-4. Organiza los datos de la siguiente hoja de cálculo en una tabla cruzada. Prueba la hipótesis de que existe una diferencia significativa entre demócratas y republicanos en términos del apoyo al mejoramiento de la protección del ambiente marítimo que alberga especies en peligro.

Partido político	Apoya la protección
Demócrata	No
Republicano	No
Demócrata	Sí
Demócrata	No
Demócrata	Sí
Republicano	Sí
Demócrata	Sí
Republicano	No
Republicano	Sí
Demócrata	Sí
Demócrata	No
Republicano	No
Republicano	Sí
Republicano	No
Demócrata	Sí
Demócrata	Sí
Republicano	Sí
Demócrata	Sí
Republicano	No
Demócrata	Sí
Demócrata	No
Republicano	Sí
Republicano	No
Demócrata	Sí
Demócrata	Sí
Demócrata	Sí
Republicano	Sí
Demócrata	Sí
Demócrata	No
Demócrata	Sí
Republicano	No
Demócrata	Sí
Republicano	No

13B-5. Utiliza el triángulo de Pascal para presentar rápidamente las siguientes probabilidades relativas al lanzamiento de monedas. En este ejercicio no es necesario escribir las ecuaciones.

a) p [de lanzar ocho monedas y obtener sólo cruces] =

b) p [de lanzar 5 monedas y obtener dos caras y tres cruces] =

c) p [de lanzar cuatro monedas y obtener dos caras y dos cruces] =

13B-6. Medic-8 Corporation se encuentra probando un nuevo fármaco cardiovascular en ratones genéticamente infectados. Investigaciones anteriores en el laboratorio muestran que 50% de estos ratones sobreviven seis meses sin tratamiento. A 10 ratones se les administra el fármaco. Siete sobreviven seis meses. Lleva a cabo la prueba de hipótesis de que el tratamiento con el fármaco contribuye a elevar la tasa de supervivencia en el caso de los que no reciben el fármaco. Utiliza $\alpha = .001$.

13B-7. Una clínica de fertilidad afirma que cuenta con un procedimiento para seleccionar el género de un niño. De ocho parejas elegidas aleatoriamente que deseaban tener niñas, cinco obtuvieron el resultado deseado. Prueba la hipótesis para determinar si el procedimiento supera lo esperado por el azar.

13B-8. Supongamos que piensas utilizar la distribución binomial para probar una hipótesis con un nivel de significancia de 0.01 en una prueba de dos colas. ¿Cuál es el tamaño mínimo de la muestra que se requiere para rechazar la hipótesis nula? Muestra los cálculos.

Conjunto de problemas 13C

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, la tabla cruzada, las curvas de probabilidad y los aspectos adecuados de una relación. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

13C-1. La siguiente tabla cruzada proviene de una muestra aleatoria de empleados del gobierno de una ciudad. La variable de las columnas representa la clasificación del sector laboral y la variable de los renglones representa *la obtención de un grado universitario.* Calcula los porcentajes de columna de todas las casillas de frecuencias conjuntas.

Grado universitario	Sector de trabajo →	Servicio	Personal de apoyo	Gerencial/ administrativo	Totales
Sí		6	57	29	
No		41	29	5	
Totales					

13C-2. En una prueba local de degustación, se probó la preferencia de las personas entre el helado de vainilla y el yogurt de vainilla. De 61 de los hombres encuestados, 41 prefirieron el helado. De 75 de las mujeres encuestadas, 62 prefirieron el helado. Elabora una tabla cruzada. ¿Existe una relación entre el género y la elección del postre?

13C-3. El consejo de arte estudiantil de un campus propone la exhibición del trabajo de un artista controversial, conocido por sus pinturas con contenido violento y sexual de iconos religiosos. El consejo encuesta a estudiantes y ex alumnos para evaluar la oposición a la exhibición. De acuerdo con los siguientes datos ficticios, determina si existe una diferencia significativa entre las proporciones de estudiantes y ex alumnos que se oponen a la exhibición.

	Estudiantes	Ex alumnos
Apoyan la exhibición	172	278
Se oponen a la exhibición	60	170

13C-4. Organiza los datos de la siguiente hoja de cálculo en forma de tabla cruzada. Prueba la hipótesis relacionada de que existe una diferencia significativa en el grado de apoyo para privatizar escuelas públicas entre quienes votaron por los republicanos y quienes lo hicieron por los demócratas en las pasadas elecciones (datos ficticios).

Cómo votó en las elecciones presidenciales pasadas	¿Apoya la privatización de las escuelas?
Republicano	Sí
Republicano	No
Demócrata	No
Demócrata	No
Republicano	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Republicano	Sí
Demócrata	Sí
Republicano	Sí
Republicano	No
Demócrata	No
Republicano	Sí
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	No
Demócrata	Sí
Republicano	No
Republicano	Sí
Demócrata	Sí
Republicano	Sí
Republicano	No
Demócrata	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	Sí
Demócrata	No
Republicano	Sí
Demócrata	Sí
Republicano	Sí
Demócrata	Sí
Demócrata	No
Republicano	Sí

13C-5. Como padres de una generación caracterizada por tener muchos hijos, Phillip y Kitty tuvieron siete hijos —todos hombres—. Suponiendo que en un solo nacimiento la probabilidad de tener hombre o mujer es la misma, ¿cuál es la probabilidad de que en una familia de siete hijos todos sean hombres? Aplica el triángulo de Pascal. En este ejercicio no es necesario escribir las ecuaciones.

13C-6. Micro-Medication Corporation se encuentra probando un nuevo fármaco contra el cáncer en ratones genéticamente infectados. Investigaciones anteriores en sus laboratorios demuestran que 50% de estos animales sobrevive seis meses sin ningún tratamiento. Se administra la droga a ocho ratones. Seis de ellos sobreviven seis meses. Prueba la hipótesis de que el tratamiento administrado ofrece una tasa superior de supervivencia que la que se observa con la ausencia del mismo fármaco. Utiliza $\alpha = 0.05$, ya que se sabe el fármaco es seguro.

13C-7. Un grupo de vigilancia de consumidores afirma que la mayor parte de las instituciones de beneficencia gasta más de la mitad de sus donativos en actividades para recaudar fondos. Tú analizas la forma de divulgación pública de diez instituciones de beneficencia seleccionadas al azar y descubres que seis se ajustan a la crítica del grupo de vigilancia. Prueba una hipótesis para determinar si la conclusión del grupo de vigilancia es correcta.

13C-8. Supongamos que piensas utilizar la distribución binomial para probar una hipótesis con un nivel de significancia de 0.05 en una prueba de una cola. ¿Cuál es el tamaño mínimo de la muestra que se requiere para que se rechace la hipótesis nula? Muestra los cálculos.

Conjunto de problemas 13D

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, la tabla cruzada, las curvas de probabilidad y los aspectos adecuados de una relación. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

13D-1. La siguiente tabla cruzada proviene de una muestra aleatoria de adultos de una ciudad grande del sureste. La variable de columna se refiere al *nivel educativo*, y la variable de renglón, al *género*. Calcula los porcentajes de columna de todas las casillas de frecuencias conjuntas.

Género	Máximo grado educativo →	Diploma de bachillerato	Grado de técnico universitario	Grado universitario	Totales
Mujer		14	58	24	
Hombre		17	68	19	
Totales					

13D-2. En una degustación entre el refresco normal y el refresco de dieta, 52 de 88 hombres encuestados prefirieron el refresco normal. De 83 mujeres encuestadas, 62 prefirieron el refresco normal. Construye una tabla cruzada. ¿Existe una relación entre el género y la preferencia por algún refresco?

13D-3. Tú entrevistas a estudiantes y profesores de una universidad para evaluar el apoyo para modificar el formato de interacción en clase del estilo de conferencia al estilo de discusión. De acuerdo con los siguientes datos, determina si existe una diferencia significativa en las proporciones de estudiantes y profesores que apoyan el cambio de formato de clase.

	Profesores	Estudiantes
Apoya el estilo de discusión	90	602
Apoya el estilo de conferencia	55	210

13D-4. Organiza la siguiente hoja de cálculo con datos en forma de tabla cruzada. Prueba la hipótesis de que existe una diferencia significativa en el apoyo a la privatización del Sistema de Seguridad Social entre republicanos y demócratas registrados (datos ficticios).

Partido político	Apoya la privatización del Sistema de Seguridad Social
Demócrata	Sí
Demócrata	No
Republicano	No
Republicano	Sí
Republicano	Sí
Demócrata	Sí
Demócrata	No
Republicano	Sí
Republicano	No
Republicano	Sí
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	No
Demócrata	Sí
Republicano	No
Republicano	Sí
Republicano	No
Demócrata	No
Republicano	Sí
Republicano	Sí
Demócrata	Sí
Demócrata	No
Republicano	Sí
Republicano	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	No
Demócrata	No
Republicano	No
Republicano	Sí
Demócrata	No
Republicano	Sí
Demócrata	Sí
Republicano	Sí

13D-5. Como padres de una generación caracterizada por tener muchos hijos, Sam y Edna tuvieron siete hijos —todos hombres—. Suponiendo que en un solo nacimiento la probabilidad de tener hombre o mujer es la misma, ¿cuál es la probabilidad de que en una familia de seis hijos todos sean niñas? Aplica el triángulo de Pascal. En este ejercicio no es necesario escribir las ecuaciones.

13A-6. Una compañía farmacéutica se encuentra probando un nuevo fármaco contra el cáncer en ratones genéticamente infectados. Investigaciones anteriores en sus laboratorios demuestran que 50% de estos animales sobrevive seis meses sin ningún tratamiento. Se administra la droga a siete ratones. Cinco de ellos sobreviven seis meses. Prueba la hipótesis de que el tratamiento administrado ofrece una tasa superior de supervivencia que la que se observa con la ausencia del mismo fármaco. Utiliza $\alpha = .05$, ya que se sabe que el fármaco es seguro.

13D-7. Alguien te comenta que muchas iglesias gastan más de la mitad de sus donativos en salarios de empleados. Tú analizas las formas de divulgación pública de ocho iglesias seleccionadas aleatoriamente y descubres que cinco se ajustan a la crítica del grupo observación. Prueba una hipótesis para determinar si la conclusión del grupo de observación es correcta.

13D-8. Supongamos que piensas utilizar la distribución binomial para probar una hipótesis con un nivel de significancia de .05 en una prueba de una cola. ¿Cuál es el tamaño mínimo de la muestra que se requiere para que posiblemente se rechace la hipótesis nula? Muestra los cálculos.

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 13

Si en tu clase se utilizan las aplicaciones opcionales en computadora que acompañan el texto, abre los ejercicios del capítulo 13 localizados en el sitio web de *The Statistical Imagination*, www.mhhe.com/ritchey2. Los ejercicios se enfocan en la presentación de tablas cruzadas y procedimientos binomiales en *SPSS* para Windows e interpretar adecuadamente la salida. Además, el apéndice D del texto contiene una vista rápida de las secuencias de los comandos del *SPSS* para los procedimientos cubiertos en este capítulo.

CAPÍTULO 14

Correlación y regresión bivariadas

Parte 1: Conceptos y cálculos

RESUMEN DEL CAPÍTULO

Introducción: superación de las mejores estimaciones de una variable dependiente

A los especialistas en estadística a menudo se les critica por hacer uso de la jerga y concentrarse en la teoría más que en los resultados prácticos. Un buen investigador, ya sea que pruebe o no una teoría, da mucho valor a los aspectos prácticos del análisis —realizando predicciones sobre la manera en que las cosas funcionan a nuestro alrededor—. Al probar una hipótesis que tiene que ver con relaciones entre variables, la descripción de las aplicaciones prácticas constituye la etapa final del análisis, en la que respondemos las siguientes preguntas: ¿Ahora qué? ¿Qué significa todo esta jerigonza sobre hipótesis estadísticas, niveles de significancia y confianza, media de valores p en el mundo real? Nuestras respuestas a estas preguntas deberían expresarse en lenguaje práctico y proporcionar los detalles esenciales sobre la forma en la que los hallazgos se pueden aplicar a situaciones de la vida diaria. Concentrar nuestros esfuerzos en la práctica nos permite contribuir al bienestar de la sociedad (por ejemplo, al mejorar la entrega de servicios) o de sus individuos (por ejemplo, al dar consejos sobre inversiones en la bolsa).

La simple idea de realizar las mejores estimaciones resulta valiosa en la aplicación de conclusiones al trabajo estadístico. Y particularmente valioso es el uso de las variables predictoras (es decir, independientes) para superar las mejores estimaciones de una variable dependiente de particular interés. Las mejores estimaciones se orientan hacia el futuro y se basan en probabilidades que proporcionan confianza en lo que realizamos. En este texto se ha dado énfasis al hecho de que un ingrediente fundamental de la imaginación estadística consiste en el análisis de causas y consecuencias con el fin de predecir eventos futuros. La realización de una mejor estimación enfoca nuestro análisis en el aspecto práctico. Las mejores estimaciones permiten generar resultados relevantes.

Repasemos nuestra experiencia relacionada con la formulación de las mejores estimaciones. Por ejemplo, supongamos que deseamos estimar el peso medio de una población de alumnos de séptimo grado. Al manejar esta única variable, la mejor estimación de su parámetro es un intervalo de confianza: una estimación puntual, como el peso medio de una muestra más menos determinado rango del error predecible. Si determinamos una media muestral de, por ejemplo 90 libras, un intervalo de confianza del 95% permitiría concluir que el peso medio oscila entre 85 y 95 libras. Ahora bien, ¿qué pasaría si quisiéramos llevar a cabo una estimación del peso de un individuo *no visto* de esta población? En este caso contamos con una estimación muy burda —la media muestral del peso de 90 libras—, una estimación puntual. Efectuamos esta estimación no muy convencidos, ya que sabemos que no todos los alumnos de séptimo grado pesan exactamente 90 libras. *Sin más información relacionada con el individuo*, esto es lo mejor que se puede hacer.

Sin embargo, a partir de nuestra experiencia con dos variables, sabemos que la mejor estimación de una variable dependiente mejora si identificamos las variables independientes —variables predictoras— y ampliamos el análisis al nivel bivariado. Por ejemplo, supongamos que Doug es un muchacho de séptimo grado, y sabemos que los muchachos tienden a pesar en promedio 10 libras más que los alumnos de séptimo grado en general—. Como aprendimos en el capítulo 11, dado este efecto del género masculino, es posible mejorar la estimación del peso de Doug agregando 10 libras a la media de 90 libras para obtener una mejor estimación de 100 libras —el peso medio de los muchachos de séptimo grado—. Si introducimos una variable predictora o independiente, X = genero, mejoramos las estimaciones de la variable dependiente, Y = peso. Se aprovecha el conocimiento sobre X para ajustar la presión de las mejores estimaciones de Y.

Una correlación entre dos variables de intervalo/razón

En el caso de una variable independiente nominal, como el género, se ajustan las estimaciones de Y exclusivamente para dos *puntuaciones —las categorías X* de hombre y mujer—. Sin embargo, al examinar la relación entre dos variables de intervalo/razón debemos tratar con una gran cantidad de puntuaciones. Para una muestra de adultos con una variable independiente X, como la estatura, hay 25 *categorías de estatura de los miembros* —puntuaciones que oscilan entre aproximadamente 56 pulgadas (4 pies, 8 pulgadas de estatura) a aproximadamente 80 pulgadas (6 pies, 8 pulgadas de estatura)—. Podríamos aprobar las diferencias de medias entre 25 categorías de estatura; no obstante, esto resultaría engorroso y requeriría una muestra muy grande.

Afortunadamente, no es necesario pensar en cada pulgada como una categoría de estatura separada o una categoría X. Más bien, se considera cada pulgada de estatura como una puntuación X, un incremento en su unidad de medida de intervalo/razón. Si intentamos estimar el peso (Y), es posible mejorar nuestras estimaciones *por cada pulgada adicional de estatura* (X), de una estatura baja de 56 pulgadas a una estatura alta de 80 pulgadas. Ya que sabemos que las personas con mayor estatura tienden a pesar más, se busca una forma de determinar cuánto peso debe agregarse *por cada pulgada adicional de estatura.* La imaginación estadística nos lleva a preguntarnos: ¿cuál es el efecto del incremento de una pulgada de estatura sobre el peso? Sobre la base de dicho conocimiento, efectuamos rápidamente pequeños ajustes de una pulgada a la estatura para obtener las mejores estimaciones del peso.

¿Cómo se realizan estos pequeños ajustes? Las variables de intervalo/razón, con sus unidades físicas de medida, ofrecen muchas ventajas matemáticas y geométricas. Mientras que en el caso de las pruebas estadísticas anteriores no era posible determinar los cuatro aspectos de una relación, en el caso de dos variables de intervalo/razón es posible hacerlo utilizando precisamente las cualidades matemáticas de una recta para calcular Y con cantidades determinadas de X. Dicha recta se utiliza geométricamente cuando se correlacionan dos variables. Una **correlación** constituye *un cambio sistemático en las puntuaciones de dos variables de intervalo/razón.*

> **Correlación** Cambio sistemático en las puntuaciones de dos variables de intervalo/razón.

Dos variables de intervalo/razón se correlacionan cuando las mediciones de una variable cambian simultáneamente con las medidas de la otra. Por ejemplo, existe una correlación entre la estatura (X) y el peso (Y). Las mediciones de una gran cantidad de pulgadas (estatura) tienden a coincidir con las mediciones de una gran cantidad de libras (pesadez). De esta manera, al calcular el peso de una persona alta, se agregan algunas libras; si dicha persona es baja, se restan algunas libras. Las correlaciones entre dos variables de intervalo/razón son muy frecuentes y, por consiguiente, la idea de la correlación resulta atractiva desde un punto de vista intuitivo. Por ejemplo, en una población de adultos, existe una correlación entre la edad y el número de veces que se visita al médico al año, es decir, los adultos de mayor edad tienden a ir más al doctor que los adultos de menos edad. Entre los estudiantes universitarios, el tiempo que invierten en la lectura se correlaciona con el promedio de sus calificaciones; cuanto más tiempo pasan leyendo, obtendrán mejores calificaciones. En términos un tanto diferentes, podemos afirmar que el nivel educativo se encuentra *correlacionado* con el ingreso; a mayor nivel educativo, mayor ingreso. Las dos variables se encuentran correlacionadas si sus medidas cambian simultáneamente, de forma coherente, caso por caso. Se dice que las mediciones se encuentran ordenadas juntas o coordinadas —o, sencillamente, correlacionadas—.

Identificación de una relación lineal

En el caso de dos variables de intervalo/razón, el procedimiento para superar las mejores estimaciones de una variable dependiente (Y), tomando en cuenta su relación con una variable independiente (X) se denomina *correlación lineal simple* o *bivariada y análisis de regresión.* (Es *simple* porque se trata de dos variables. En el caso de tres o más variables recibe el nombre de *análisis de correlación* y *regresión múltiple*, lo cual rebasa el alcance este libro.) La idea central detrás de la correlación lineal y de la regresión simples consiste en aplicar la fórmula a una recta para obtener una mejor estimación de Y (por ejemplo, el peso en libras), dado cualquier valor de X (por ejemplo, cualquier estatura en pulgadas). El símbolo Y' (Y prima) se emplea para referirse a este valor aproximado de Y.

Idea central detrás de la correlación lineal bivariada y del análisis de regresión simple

Aplicación de las fórmulas a una recta con el fin de superar las mejores estimaciones de una variable dependiente (Y) de intervalo/razón para todos los valores de una variable independiente de intervalo/razón (X).

La fórmula de una recta para calcular Y es la siguiente:

$$Y' = a + bX$$

(El lector quizás haya visto símbolos diferentes en libros de geometría, en los que la fórmula de la recta con frecuencia se presenta como $Y' = mx + b$.) Antes de describir el significado de cada uno de los símbolos de la fórmula, es necesario controlar las circunstancias en las que esta fórmula se aplica con precisión. En primer lugar, la fórmula sólo se aplica cuando las dos variables son de niveles de medición de intervalo/razón. En segundo lugar, la fórmula se aplica adecuadamente sólo cuando existe una relación lineal entre X y Y en un diagrama de dispersión, el cual constituye una representación gráfica de los datos. Un **diagrama de dispersión** consiste en una *rejilla en dos dimensiones* (como las líneas del papel milimétrico) *de las coordenadas de dos variables de intervalo/razón*, X y Y. Una **coordenada** es *un punto en un diagrama de dispersión en el que se grafican los valores de X y Y de un caso.* Los estadísticos que acompañan el diagrama de dispersión sólo se aplican en casos en los que las coordenadas se rigen por un patrón lineal —aquel en el que las coordenadas del diagrama de dispersión caen dentro del patrón de forma de cigarro que se aproxima a la forma de una recta—. Las consecuencias de no identificar adecuadamente una relación lineal se analizan en la sección de este capítulo *Insensatez y falacias estadísticas.*

Representación gráfica de la relación entre dos variables de intervalo/razón

Diagrama de dispersión: rejilla bidimensional de las coordenadas de dos variables de intervalo/razón, X, Y.

Coordenada: punto en un diagrama de dispersión en el que se representan gráficamente los valores de X y Y para un caso.

Patrón lineal: aquel en el que las coordenadas de un diagrama de dispersión caen en un patrón alargado (en forma de cigarro) que se aproxima a la forma de una recta.

Elaboración del diagrama de dispersión

La tabla 14-1 contiene las estaturas (X) y los pesos (Y) de 16 estudiantes varones del último año de preparatoria. La figura 14-1 muestra el diagrama de dispersión de la estatura (X) como predictor del peso (Y). (Por razones de claridad, no se dibuja la cuadrícula en el diagrama de dispersión; simplemente imagina que allí se encuentra.) Las coordenadas de los estudiantes 2 y 16 se encuentran resaltadas en la tabla y en la figura. Un diagrama de dispersión se asemeja a un histograma; en ambos el eje horizontal muestra los valores de la variable de intervalo/razón X, que recibe el nombre de eje X. No obstante, los ejes verticales de ambos gráficos son diferentes. Recuerda que en el caso de un histograma, el eje vertical representa la frecuencia de ocurrencias (f) de un valor dado de X. *Pero en el caso de un diagrama de dispersión, el eje vertical representa los valores de una segunda variable de intervalo/razón,* Y, la cual recibe el nombre de eje Y. Por tradición, X representa a la variable predictora (o independiente); y Y, a la variable dependiente, la que deseamos explicar.

Las coordenadas se representan gráficamente localizando una puntuación X sobre el eje X y una puntuación Y en el eje Y, colocando posteriormente la coordenada en la cuadrícula. Por ejemplo, en el caso del estudiante 16, la coordenada X Y es $X = 72$ pulgadas, $Y = 176$ libras, o simplemente (72, 176), lo cual se representa gráficamente desplazándose 72 pulgadas sobre el eje XY, posteriormente, ascendiendo 176 libras por el eje Y.

Identificación de un patrón lineal

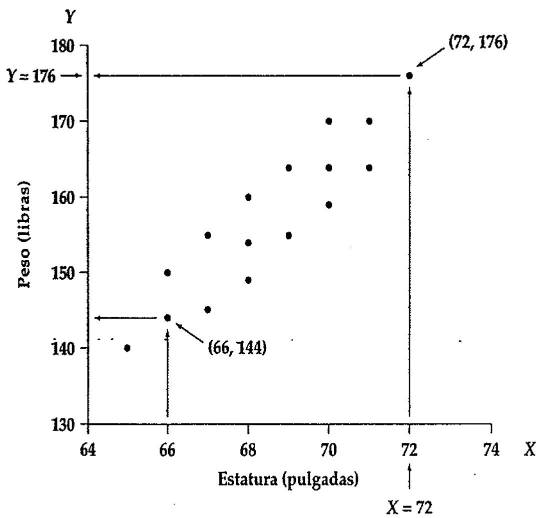
En un diagrama de dispersión, una relación lineal se encuentra presente si las coordenadas forman un patrón alargado, en forma de cigarro y asciende o desciende. Los datos en la figura 14-1 parecen ajustarse a un patrón con pendiente ascendente. Para obtener un sentido de proporción de las relaciones lineales, observa los números reales de la tabla 14-1, en la cual

TABLA 14-1 | Estaturas y pesos de 16 estudiantes varones del último año de preparatoria

Estudiante	Estatura (pulgadas) (X)	Peso (libras) (Y)
1	65	140
2	**66**	**144**
3	66	150
4	67	145
5	67	155
6	68	149
7	68	154
8	68	160
9	69	155
10	69	164
11	70	159
12	70	164
13	70	170
14	71	164
15	71	170
16	**72**	**176**

FIGURA 14-1

Diagrama de dispersión de la estatura (X) como predictora del peso (Y)



los valores de X y Y se ordenan de estaturas menores a mayores. Observa que conforme nos movemos en la columna de las estaturas en la tabla, los pesos también tienden a incrementarse. Esta es una indicación de que los tamaños de X y Y cambian sistemáticamente; es decir que aumentan juntos. En el diagrama de dispersión de la figura 14-1, este cambio resulta evidente por la constante inclinación ascendente de las coordenadas de izquierda a derecha. Las coordenadas se ajustan bastante bien a un patrón lineal alargado. Conforme se incrementa la estatura en el eje X, las puntuaciones de Y (peso) también se incrementan en el eje Y. Cuando las dos variables aumentan simultáneamente, hablamos de una *relación positiva.*

> **Correlación positiva** Un incremento en X se relaciona con un incremento en Y (cuando X se incrementa, Y tiene la tendencia a incrementarse).

El diagrama de dispersión de la figura 14-2 presenta una relación negativa entre dos variables de intervalo/razón. Las coordenadas se ajustan a un patrón lineal, aunque con la caída constante de las posiciones de las coordenadas de izquierda a derecha. Un incremento en la tensión del cuidador se relaciona con una disminución en su satisfacción relacionada con la comunicación con el médico. Conforme las puntuaciones X se elevan, las puntuaciones Y se reducen.

> **Correlación negativa** Un incremento en X se relaciona con una reducción en Y (conforme se incrementa X, Y tiene una tendencia a disminuir).

Finalmente, ¿qué esperamos de un diagrama de dispersión si dos variables no se encuentran relacionadas?; es decir, ¿qué sucedería si X no resultara un buen predictor de Y? Este hecho se ilustra en la figura 14-3, en la cual X representa el nivel educativo de un cuidador y Y, el grado al que un paciente puede funcionar sin la ayuda del cuidador. El diagrama de

FIGURA 14-2

Diagrama de dispersión de la relación negativa entre la satisfacción del cuidador con respecto a la comunicación con el médico regresionada a partir de la carga del cuidador: conforme se incrementa la carga del cuidador, se reduce la satisfacción del cuidador en lo que se refiere a la comunicación con el médico

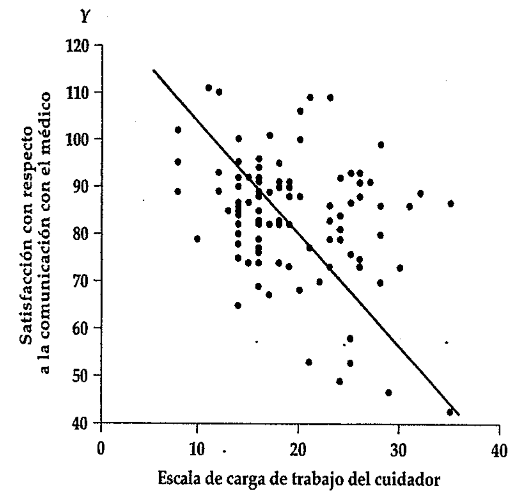
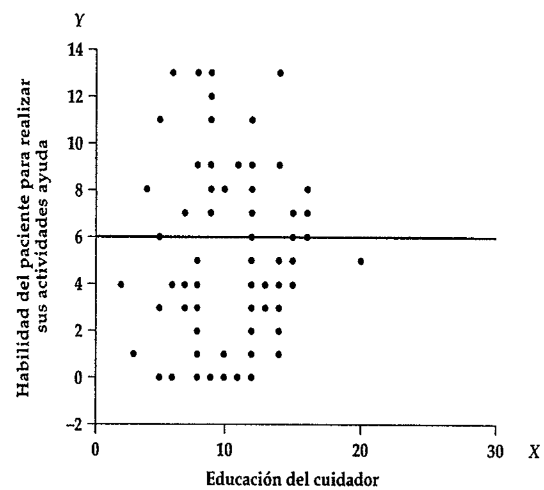


FIGURA 14-3

Diagrama de dispersión de la vida del paciente para desempeñarse sin ayuda regresionada a partir de la educación del cuidador (que muestra que no existe relación entre las dos variables): el nivel educativo del cuidador no predice la vida del paciente para realizar sus actividades independientemente



dispersión no revela ninguna relación entre X y Y. El patrón de coordenadas carece de una forma alargada e inclinada, en forma de cigarro. Cuando no existe una correlación entre X y Y, la media de Y se mantiene como la mejor estimación de Y para todos los valores de X. En otras palabras, conocer el nivel educativo de un cuidador no ofrece ninguna información relacionada con la forma de superar una mejor estimación de la habilidad de un paciente para desempeñar sus actividades sin ayuda. La condición médica del paciente no tiene nada que ver con el nivel educativo del cuidador.

> **Ausencia de correlación** Un incremento en X no se relaciona con las puntuaciones de Y (cuando aumenta X, las puntuaciones de Y varían de forma aleatoria).

Uso de la ecuación de regresión lineal para medir los efectos de X sobre Y

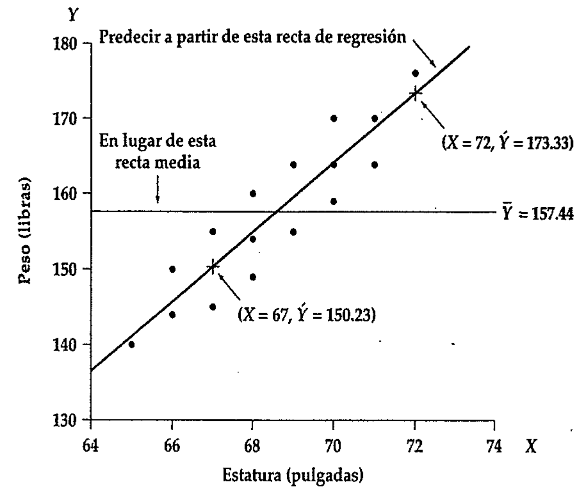
Cuando encontramos un patrón lineal en las coordenadas de un diagrama de dispersión, esto indica que las puntuaciones de una variable dependiente Y siguen las de una variable independiente X de forma ascendente o descendente. El conocimiento de este patrón predecible nos prepara para ajustar las estimaciones de Y. Por ejemplo, en la figura 14-1 parece que existe una relación positiva entre la estatura (X) y el peso (Y). De esta manera, podemos proporcionar una estimación alta del peso en caso de que la estatura sea alta y una estimación baja de peso en caso de que la estatura sea baja. De hecho, la correlación lineal y el análisis de regresión permiten realizar ajustes matemáticos precisos en estas estimaciones.

¿Cómo podemos llevar a cabo ajustes que incrementen la precisión? Si el diagrama de dispersión muestra un patrón lineal de forma alargada en las coordenadas, se traza una recta para *ajustar* el patrón de coordenadas. Esta recta es *aquella que cae lo más cerca posible de cada coordenada*, y recibe el nombre de *recta de mejor ajuste* o, técnicamente, *recta de regresión. Regresión* significa 'regresar' o 'gravitar' en torno a un punto, o tener una tendencia a desplazarse en determinada dirección. Por ejemplo, haremos la predicción de que las personas de mayor estatura tienden a desplazarse en la dirección de los pesos mayores. Se acostumbra decir que la variable dependiente (Y) *se regresiona* a partir de la variable independiente (X). En este ejemplo, se dice que el peso se regresiona a partir de la estatura.

> **Recta de regresión** Recta de mejor ajuste trazada a través de las coordenadas X, Y de un diagrama de dispersión con dos variables de intervalo/razón.

La figura 14-4 representa la figura 14-1 con la recta de regresión (posteriormente analizaremos la forma de determinar la localización precisa de esta recta). Una vez que hemos ubicado la recta, podemos utilizar las coordenadas localizadas sobre ella para identificar la *mejor estimación* del peso (Y) para cualquier estatura (X). En la figura 14-4, se resaltan dos coordenadas de la recta de regresión (67, 150.23) y (72, 173.33) para mostrar la forma en que X (estatura) permite aproximar Y (peso). La mejor estimación del peso de un estudiante del último año de 67 pulgadas de estatura es de 150.23 libras; otra estimación para uno de 72 pulgadas de estatura es de 173.33 libras. De hecho, esta recta se utiliza para calcular el peso de individuos de cualquier estatura. Nos desplazamos hacia un valor de X sobre el eje X y trazamos una recta perpendicular hacia la recta de regresión. En seguida trazamos una recta desde allí hasta el eje Y; nuestra mejor estimación del peso corresponde al punto en el que esta recta corta al eje Y. Dichos valores de X y Y serán las coordenadas que se localizan sobre la recta de regresión. Observa con cuidado que a esta estimación del peso le llamamos *Y predicha*, y se utiliza el símbolo $\acute{Y}$ para distinguirla del peso real observado de $Y = 164$ libras.

FIGURA 14-4
Ilustración de la recta de regresión para la estatura como predictor del peso; es decir, el peso regresionado a partir de la estatura



La $\acute{Y}$ constituye una estimación *susceptible de conocerse —sobre la base del conocimiento de la relación precisa entre la estatura y el peso—*.

El beneficio clave del análisis de regresión consiste en la posibilidad de mejorar las aproximaciones de Y en una población, utilizando la recta de regresión en lugar de presentar solamente la media muestral de Y. Recuerda que si sólo conocemos la media de Y, ésta constituye el mejor resultado que podemos obtener para predecir el peso. En la figura 14-4, observa que también se ha trazado una recta que representa la media de Y. Cuando no se conoce la estatura de una persona, siempre se predice en la dirección de esta recta media de peso. Una vez que se conoce la existencia de una relación con la estatura, se puede predecir un valor de Y en la recta de regresión —$\acute{Y}$, que corresponde a la puntuación X de un individuo—. En esencia, lo que hace el análisis de regresión lineal es permitirnos hacer predicciones a partir de la recta de regresión, en lugar de hacerlo a partir de la recta plana que representa la media de Y. Al hacerlo de esta manera, las estimaciones resultantes se encuentran más cerca de los valores reales de Y. Esto se puede observar en el diagrama de dispersión (figura 14-4). La recta de regresión se diseña de forma intencional para acercarse tanto como sea posible a las coordenadas reales. Ésta es la recta de mejor ajuste.

> **Beneficio del análisis de regresión bivariada para la superación de las mejores estimaciones de una variable dependiente Y**
>
> Nos permite utilizar una coordenada X, $\acute{Y}$ de la recta de regresión como predicción de Y (en lugar de informar simplemente la media de Y).
>
> $\acute{Y}$ = valor de Y predicho por la recta de regresión ($\acute{Y}$ representa una estimación de Y de conocerse sobre la base del conocimiento de que Y se relaciona con X).

Por ejemplo, si se nos pide calcular el peso de un estudiante del último año de preparatoria *sin* conocer su estatura, nuestra mejor aproximación consistiría en estimar 157.44 libras —la media muestral total de Y calculada con los datos de la tabla 14-1—. Sin embargo, si se sabe que la estatura y el peso se encuentran relacionados de forma *lineal*, podríamos trazar una recta a través de las coordenadas en el diagrama de dispersión y aproximar un peso (Y), determinando su coordenada X, $\hat{Y}$ sobre la recta de regresión. Por ejemplo, si dicho estudiante tiene un estatura de 72 pulgadas, calcularemos su peso en 173.33 libras. Al hacerlo, agregamos 15.89 libras a la media de 157.44 para explicar su estatura superior al promedio.

Otra forma de mejorar la predicción consiste en verla a la luz de la reducción proporcional del error (RPE) de predicción. Supongamos que el estudiantes de 72 pulgadas de estatura se llama Christopher. Sin conocer su estatura, habríamos predicho que su peso correspondía a la media de 157.44. Posteriormente encontramos que Christopher representa el caso 16 en la tabla 14-1, cuyo peso real es de 176 libras. La estimación *de la media* subestima su verdadero peso por 18.56 libras (recuerda que, de acuerdo con el capítulo seis, esta cantidad constituye su puntuación de desviación).

Puntuación de desviación de Christopher = error en la estimación de la media

$$= Y_{(Christopher)} - \overline{Y} = 176 - 157.44 = 18.56 \text{ libras}$$

Supongamos que tenemos la oportunidad de predecir nuevamente su peso, pero conociendo la relación entre estatura y peso. En lugar de predecir a partir de la media, se predice *a partir de la recta de regresión* (figura 14-4). La mejor aproximación de un estudiante de 72 pulgadas de estatura es la puntuación $\hat{Y}$ de 173.33. Utilizando este dato como una aproximación susceptible de conocerse el peso de Christopher, se genera un error mucho menor, de sólo 2.67 libras, comparadas con el error de 18.56 libras cuando se predice hacia la media:

Error al calcular el peso de Christopher a partir de la recta de regresión

$$= Y_{(Christopher)} = \hat{Y}_{(X=72)} = 176 - 173.33 = 2.67 \text{ libras}$$

Hemos mejorado esta estimación de peso (Y) por una cantidad considerable, tomando en cuenta el hecho de que el peso se encuentra correlacionado con la estatura (X). Como observamos en el diagrama de dispersión de la figura 14-4, el error en las mediciones de la muestra (y población) entera se reduce mucho si las mediciones se llevan a cabo partir de la recta de regresión en lugar de la línea media.

En suma, cuando se determina la existencia de una relación entre dos variables de intervalo/razón, se deben realizar las mejores estimaciones utilizando la recta de regresión en lugar de la media. Las estimaciones de Y susceptibles de conocerse —las que toman en cuenta a X— reducen los errores en las predicciones. Ésta constituye la esencia de los análisis de correlación y regresión.

Coeficiente de correlación bivariada *r* de Pearson

En un diagrama de dispersión, cuanto más estrechamente se acumulen los datos de las coordenadas en torno a la recta de regresión, mayor será la correlación entre X y Y, y con mayor precisión se aproximará Y dado cualquier valor de X. La r de Pearson es un coeficiente de

correlación muy comúnmente utilizado que mide la estrechez del ajuste de las coordenadas X, Y en torno a la recta de regresión. La fórmula para la r de Pearson es la siguiente:

Cálculo del coeficiente de correlación bivariada *r* de Pearson

$$r = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{\sqrt{\Sigma(X - \overline{X})^2 \, \Sigma(Y - \overline{Y})^2}}$$

donde

r = coeficiente de correlación de Pearson
X = variable independiente de intervalo/razón
Y = variable dependiente de intervalo/razón
$\overline{X}$ = media de la variable independiente, X
$\overline{Y}$ = media de la variable dependiente, Y

Qué mide la *r* de Pearson

La estrechez del ajuste de las coordenadas X, Y con respecto a la recta de regresión.
El grado al que las desviaciones de las puntuaciones de las medias de X y Y tienden a fluctuar conjuntamente.

Hoja de cálculo de computadora para calcular los estadísticos de correlación y regresión bivariadas

Para calcular los estadísticos de este capítulo, se requieren los siguientes estadísticos y sumas (repasa el capítulo 5, tabla 5-1, en la que se utilizaron algunos de estos estadísticos para calcular la desviación estándar).

$$n, \Sigma X, \Sigma Y, \overline{X}, \overline{Y}, \Sigma(X - \overline{X}), \Sigma(Y - \overline{Y}), \Sigma(X - \overline{X})(Y - \overline{Y}),$$
$$\Sigma(X - \overline{X})^2, \Sigma(Y - \overline{Y})^2$$

Con los datos de las estaturas y los pesos de la tabla 14-1, podemos calcular rápidamente estos elementos de las ecuaciones con la hoja de cálculo de la tabla 14-2. Primero calculamos las medias de X y Y:

$$\overline{X} = \frac{\Sigma X}{n} = \frac{1\,097}{16} = 68.56 \text{ pulgadas}$$

$$\overline{Y} = \frac{\Sigma Y}{n} = \frac{2\,519}{16} = 157.44 \text{ libras}$$

Estas medias se utilizan en seguida para calcular la sumas de las puntuaciones de las desviaciones para X y Y (es decir, $\Sigma(X-\bar{X})$ y $\Sigma(Y-\bar{Y})$, columnas A y B de la tabla 14-2). Recuerda que las puntuaciones de las desviaciones para una variable deben sumar cero con cierto error de redondeo. Estas puntuaciones de las desviaciones se elevan al cuadrado y se suman para obtener las sumas de los cuadrados (o *variación*) para X y Y (columnas D y E de la tabla 14-2). La suma para la columna C de la tabla 14- 2 recibe el nombre de *covariación*. En cada caso, la puntuación de la desviación para X se multiplica por la puntuación de la desviación para Y y los resultados se suman.

Covariación de X y Y

$$\Sigma(X-\bar{X})(Y-\bar{Y})$$

La suma de las puntuaciones de las desviaciones de X multiplicada por las puntuaciones de las desviaciones de Y.

TABLA 14-2 | Hoja de cálculo para calcular los estadísticos de correlación y regresión: alturas y pesos de 16 hombres del último año de preparatoria

	Especificaciones		Cálculos				
	Altura (pulgadas)	Peso (libras)	A	B	C	D	E
Estudiante	X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
1	65	140	−3.56	−17.44	62.09	12.67	304.15
2	66	144	−2.56	−13.44	34.41	6.55	180.63
3	66	150	−2.56	−7.44	19.05	6.55	55.35
4	67	145	−1.56	−12.44	19.41	2.43	154.75
5	67	155	−1.56	−2.44	3.81	2.43	5.95
6	68	149	−.56	−8.44	4.73	.31	71.23
7	68	154	−.56	−3.44	1.93	.31	11.83
8	68	160	−.56	2.56	−1.43	.31	6.55
9	69	155	.44	−2.44	−1.07	.19	5.95
10	69	164	.44	6.56	2.89	.19	43.03
11	70	159	1.44	1.56	2.25	2.07	2.43
12	70	164	1.44	6.56	9.45	2.07	43.03
13	70	170	1.44	12.56	18.09	2.07	157.75
14	71	164	2.44	6.56	16.01	5.95	43.03
15	71	170	2.44	12.56	30.65	5.95	157.75
16	72	176	3.44	18.56	63.85	11.83	344.47

$n = 16$ $\quad \Sigma Y = 2\,519$ $\quad \Sigma(Y-\bar{Y}) = -.05$* $\quad \Sigma(X-\bar{X})^2 = 61.88$

$\Sigma X = 1\,097$ $\quad \Sigma(X-\bar{X}) = .04$* $\quad \Sigma(X-\bar{X})(Y-\bar{Y}) = 286.12$ $\quad \Sigma(Y-\bar{Y})^2 = 1\,587.88$

* Las desviaciones no suman cero como consecuencia del error de redondeo.

Los cálculos y sumas en la hoja de cálculo de la tabla 14-2 proporcionan lo que se requiere para calcular la r de Pearson para nuestros datos de las alturas y pesos de los estudiantes del último año de preparatoria:

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2}} = \frac{286.12}{\sqrt{(61.88)\,(1\,587.88)}} = \frac{286.12}{313.46} = .91$$

Características del coeficiente de correlación bivariada r de Pearson

Los valores calculados en la r de Pearson pueden variar de -1.0 a $+1.0$. Por ejemplo, en la figura 14-4, que tiene que ver con la relación entre altura y peso, podemos ver que la recta de regresión se inclina hacia arriba: los valores bajos de X corresponden a los valores bajos de Y, y los valores altos de X corresponden a los valores altos de Y. La pendiente ascendente indica que la dirección de la relación es positiva y que la r de Pearson es un valor positivo. Cuando $r = +1.0$, ésta constituye una correlación positiva perfecta (muy inusual), lo cual quiere decir que X es un predictor perfecto de Y y que conforme X se incrementa, Y se incrementa. En una correlación positiva perfecta, cada coordenada del diagrama de dispersión yace sobre la recta de regresión misma, y la recta se inclina hacia arriba.

Cuando X y Y están correlacionadas en la dirección negativa, la recta de regresión del diagrama de dispersión desciende de tal manera que conforme X aumenta, Y disminuye. El valor de la r de Pearson será negativo. Por ejemplo, a mayor número de personas pobres que viven en una comunidad (X), menor será la tasa de propietarios de casa (Y). Cuando $r = -1.0$, se presenta una correlación negativa perfecta (muy inusual), lo que significa que X representa un predictor perfecto de Y y que, conforme X se incrementa, Y disminuye. Con una correlación negativa perfecta, cada coordenada del diagrama de dispersión yace sobre la recta de regresión inclinada hacia abajo. Las correlaciones perfectas en la dirección positiva o negativa son inusuales, especialmente en las ciencias sociales.

El valor absoluto de la r de Pearson (su tamaño, prescindiendo de su signo) indica la estrechez del ajuste de las coordenadas en torno a la recta de regresión en un diagrama de dispersión. Cuanto mayor es el valor absoluto de la r de Pearson, las coordenadas se encontrarán más próximas a la recta. En el diagrama de la figura 14-4, las coordenadas se ajustan muy estrechamente en torno a la recta de regresión, como se reflejó en la r de Pearson de .91, que resulta considerablemente grande.

Como lo indicamos antes, cuando *no existe correlación entre* X y Y, el conocimiento de los valores de X no mejora las aproximaciones de Y. Es el caso de la figura 14-3. Si no hay correlación, $r = 0$. Otro ejemplo de ausencia de correlación es el peso como predictor del promedio académico; conocer el peso de las personas no nos dice nada en relación con sus promedios académicos.

Las características de la r de Pearson se resumen de la siguiente manera:

Características de coeficientes de correlación r de Pearson

1. Los valores calculados de la r de Pearson pueden oscilar entre -1.0 y $+1$, pasando por cero.
2. A mayor valor absoluto de la r de Pearson, mayor será la estrechez del ajuste de las coordenadas X, Y con respecto a la recta de regresión.

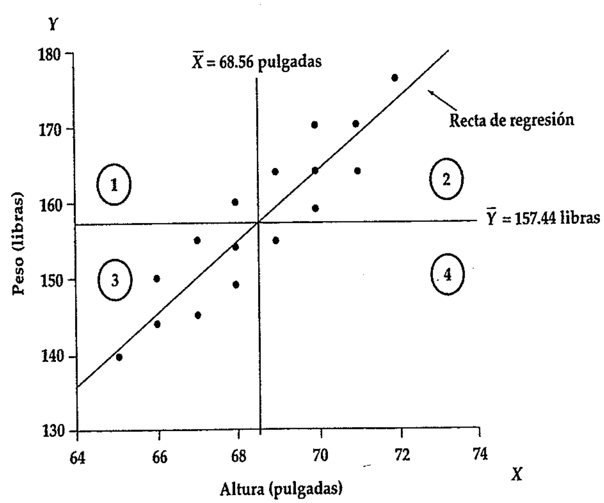
3. Cuando la recta de regresión se inclina hacia arriba, se tiene una correlación positiva: un incremento en el nivel de las puntuaciones X se relaciona con un incremento en el nivel de las puntuaciones Y. La r de Pearson será positiva hasta un valor de +1.00.
4. Cuando la recta de regresión se inclina hacia abajo, se presenta una correlación negativa: un *incremento* en el nivel de las puntuaciones X se relaciona con una *reducción* en el nivel de las puntuaciones Y. La r de Pearson será negativa hasta un valor de −1.00.
5. Cuando la recta de regresión es plana (es decir, que no tiene pendiente), no existe correlación y la r de Pearson = 0. Un *incremento* en el nivel de las puntuaciones X no se relaciona con un cambio en el nivel de las puntuaciones Y. Esto nos indica que el conocimiento de una puntuación X no mejora la estimación de las puntuaciones Y.

Comprensión de la fórmula de r de Pearson

Analicemos la ecuación de r de Pearson y relacionémosla con el diagrama de dispersión para obtener una idea de lo que este estadístico mide. La figura 14-5 muestra el diagrama de dispersión descompuesto en cuadrados o *cuadrantes* circulados con los números 1 a 4. Los cuadrantes identifican áreas por encima y por debajo de las medias de X y Y. El cuadrante 1 incluye las coordenadas de los estudiantes varones del último año de preparatoria cuyo promedio se encuentra debajo de la media X (por debajo de la altura promedio), aunque por encima de la media de Y (por encima del peso promedio). El cuadrante 2 muestra las coordenadas de los estudiantes con puntuaciones por encima de ambas medias X y Y. El cuadrante 3 contiene las coordenadas de aquellos cuyas puntuaciones se encuentran debajo de las medias de X y Y. El cuadrante 4 muestra las coordenadas de los estudiantes por encima de la media de X y debajo de la media de Y. Observa que en esta relación lineal positiva, los cuadrantes 2 y 3 incluyen a la mayoría de las coordenadas. Los estudiantes cuyo promedio de altura (X) es más alto también tienden a tener una puntuación alta de peso (Y) y caen en el cuadrante 2.

FIGURA 14-5

Forma en que las coordenadas de altura ingreso se cargan en los cuadrantes diagonales cuando existe una relación lineal



Los estudiantes con una puntuación baja de altura tienden también a tener una baja puntuación de peso y caen en el cuadrante 3.

Cuando un estudiante cae por encima de la media de X, podemos decir que se desvía en la dirección positiva; es decir que su puntuación de desviación es positiva. Por ejemplo, si un estudiante del último año mide 72 pulgadas de altura, su puntuación de desviación es

$$(X - \overline{X}) = 72 - 68.56 = 3.44 \text{ pulgadas}$$

Un enfoque en las puntuaciones de las desviaciones se hace evidente de la r de Pearson y en la hoja de cálculo en la tabla 14-2. Todos los elementos en la ecuación de la r de Pearson implican puntuaciones de desviación y puntuaciones de desviación al cuadrado (es decir, sumas de cuadrados o variaciones).

Así, ¿cómo mide la correlación bivariada la ecuación de la r de Pearson? Al calcularla, la r de Pearson evalúa el grado en que fluctúan juntas las puntuaciones de desviación de X y Y, o la forma en la que *covarían*. Si existe una relación positiva entre X y Y, entonces esperaríamos que los estudiantes que se desvían de la parte positiva de la altura también se desvíen en lo que se refiere a la parte positiva del peso. Para medir desviaciones de toda una muestra, debemos elevarlas al cuadrado y sumarlas para obtener la variación de la suma de los cuadrados. El denominador de la ecuación de la r de Pearson multiplica la suma total de los cuadrados para X por la suma total de cuadrados para Y. (Resulta necesario sacar la raíz cuadrada para revertir el efecto de que las puntuaciones se elevaron al cuadrado como consecuencia de que la suma de las puntuaciones de desviación de cualquier variable es cero.) Las sumas totales de los cuadrados de la variación total para una variable evalúan el *error*, el grado al que las puntuaciones se apartan de la media. El denominador de la ecuación de la r de Pearson evalúa cuánto del error total tienen X y Y, una con respecto a la otra. Sin embargo, al multiplicar sencillamente estas variaciones de X y Y, este cálculo supone que las puntuaciones de X y Y se encuentran distribuidas aleatoriamente con respecto a sus medias. La suma de estos múltiplos consiste en esencia en un error estándar. El denominador proporciona el error total, suponiendo que no existe una relación entre X y Y. Éste no toma en cuenta el hecho de que las coordenadas X, Y pueden fluctuar juntas en un patrón lineal. En nuestro ejemplo del peso regresionado a partir de la altura, el resultado en el denominador es de 313.46.

Sin embargo, el numerador ecuación de la r de Pearson evalúa cuán adecuadamente fluctúan X y Y en un patrón. Esto implica multiplicar, para cada caso, las puntuaciones de las desviaciones de X por las puntuaciones de desviación de Y para obtener la covariación. Observa las columnas A, B y C de la hoja de cálculo de la tabla 14-2. Las puntuaciones de las desviaciones negativas para X (columna A) tienden a presentarse con puntuaciones de desviación negativas para Y (columna B). Éstas son las coordenadas que caen en el cuadrante 3 de la figura 14-5. Cuando se multiplican dos negativos, el resultado es positivo (columna C). Asimismo, las puntuaciones de desviación positiva para X tienden a acumularse con puntuaciones de desviación positivas para Y (cuadrante 2 de la figura 14-5). El producto de estas puntuaciones también es positivo. Por consiguiente, cuando los cuadrantes 2 y 3 se cargan de puntos como consecuencia de una relación lineal, los resultados para la columna C incluyen muchos números positivos que suman una covariación considerablemente grande; en este caso, 286.12. El numerador toma en cuenta la forma en que las puntuaciones de X y Y se vinculan y, por lo tanto, miden el efecto de correlación de la relación. Cuando este número es grande con respecto al error total del denominador, el resultado es un valor considerable de r. Cuando hay una relación positiva perfecta, el numerador será igual al denominador, lo cual dará como resultado una r de Pearson de 1.00.

Lo que la ecuación hace es calcular en el denominador el error total —desviaciones de las medias de X y Y— ignorando cualquier relación entre X y Y. El numerador evalúa la forma en la que las desviaciones de X y Y fluctúan como consecuencia de su efecto de correlación. Este efecto se presenta cuando existe una relación entre las variables. La razón del efecto de correlación en el numerador con respecto al error total en el denominador es la r de Pearson. En esencia, ésta determina cuántas puntuaciones de desviación para X y Y se pueden explicar por su covarianza, su tendencia a fluctuar en un patrón lineal. En el capítulo 15 analizaremos la fuerza de una relación para una correlación bivariada y los estadísticos de regresión. Mostraremos la forma en la que la r de Pearson se puede aplicar para establecer precisamente qué grado de error en las previsiones de Y se puede reducir conociendo su relación con X. Fundamentalmente, la r de Pearson indica si las desviaciones de X y Y obedecen algún patrón.

Como se indicó, la r de Pearson se puede calcular para dar como resultado una puntuación negativa. Cuando la relación es negativa con pendiente descendente en la recta de regresión, los cuadrantes 1 y 4 se cargarán con casos. En estos cuadrantes, una de las variables posee puntuaciones de desviación positivas y la otra, puntuaciones de desviación negativas. Cuando se multiplican para obtener la covariación, el resultado es una suma alta negativa en el numerador de la ecuación. Esto da como resultado una r negativa.

Cuando no existe una relación entre X y Y, las puntuaciones de desviación serán positivas y negativas de forma aleatoria. Los cuatro cuadrantes tendrán coordenadas dispersas al azar. Los productos positivo y negativo de la covariación (columna C de la tabla 14-2) se eliminarán, lo cual dará como resultado un numerador pequeño. El cálculo de r se aproximará a cero.

Estadísticos de regresión

Ahora que tenemos una idea general de la forma en la que una línea recta en un diagrama de dispersión se puede aplicar para realizar mejores estimaciones de Y cuando conocemos X, analicemos la forma de realizar cálculos matemáticos precisos de estas estimaciones. Por ejemplo, continuemos analizando la correlación entre altura y peso utilizando la hoja de cálculo de la tabla 14-2 y el diagrama de dispersión de la figura 14-4. Ahora que conocemos que la altura y el peso se encuentran relacionados, podemos ajustar los cálculos del peso tomando en cuenta el hecho de que los muchachos altos pesan más que los bajos. Podemos responder la pregunta: en general, ¿cuántas libras extras deben sumarse o restarse a cualquier aproximación del peso para determinada altura? Específicamente, ¿cuáles son las mejores estimaciones del peso en el caso de los estudiantes de último año de preparatoria de diferentes estaturas? Las respuestas se establecen de acuerdo con la siguiente fórmula relacionada con una recta:

Fórmula de la ecuación lineal (o fórmula de la recta de regresión) para la relación entre dos variables de intervalo/razón

$$\hat{Y} = a + bX$$

donde

$\hat{Y}$ = *Y predicha* (una estimación de la variable dependiente Y para un valor dado de la variable independiente X)

a = intersección con Y, el punto en el que la recta de regresión intersecta al eje Y cuando $X = 0$.

b = pendiente de la recta de regresión (denominada coeficiente de regresión)

Para utilizar esta fórmula, comenzamos calculando los valores de a y b. Una vez que a y b se incluyen en la fórmula, cualquier cantidad de pulgadas de altura, o puntuaciones X, se puede sustituir por X en la fórmula. En seguida resolvemos para $\hat{Y}$ — la mejor estimación del peso para dicha cantidad de pulgadas de altura—. Las coordenadas $X, \hat{Y}$ resultantes se ajustarán a la recta que mejor concuerda con el patrón de coordenadas del diagrama de dispersión. Antes de calcular los elementos de esta fórmula, analicemos cada término.

Coeficiente de regresión o pendiente, b

En la ecuación de regresión $\hat{Y} = a + bX$, b representa la *pendiente* de la recta de regresión y recibe el nombre de *coeficiente de regresión*. El coeficiente de regresión nos dice cuántas libras tenemos que sumar a una estimación del peso por cada pulgada adicional de incremento en la lectura.

Coeficiente de regresión, b

b = pendiente de la línea de regresión del diagrama de dispersión

= efecto sobre Y de un cambio de 1 unidad en X

= "elevación dividida entre avance"

donde

Y = una variable dependiente de intervalo/razón

X = una variable independiente de intervalo/razón

¿Qué indica la pendiente? En el caso de nuestros datos relacionados con pesos y alturas de estudiantes del último año de preparatoria, $b = 4.62$ libras por pulgada. Un cambio de 1 unidad en X sería de 1 pulgada. Para interpretar esta pendiente diríamos que "un incremento de 1 pulgada en la altura se relaciona con un incremento de 4.62 libras en el peso".

Cuando corremos montaña arriba o montaña abajo, la pendiente nos indica cuánta elevación (ganancia vertical de b sobre Y) alcanzamos cuando corremos 1 milla (una ganancia horizontal o *avance* de 1 unidad en X). Por ejemplo, si la pendiente de un camino en la ladera de una montaña es de 300 pies por milla, por cada milla que recorremos, nos elevamos 300 pies. Cuanto mayor sea el valor de b, más pronunciada será la pendiente. Si un corredor se eleva 600 pies por milla, corre cuesta arriba por una montaña más inclinada. La figura 14-6 ilustra la forma en la que b indica la inclinación de la pendiente.

Intersección Y, a

La intersección de Y, a, en la ecuación de regresión $\hat{Y} = a + bX$ es el valor de Y cuando $X = 0$. La intersección de Y recibe el nombre de *constante* de la ecuación. Mientras que la ecuación multiplica la pendiente b para diferentes puntuaciones de X, sumamos la misma cantidad (una constante) de a a cada cálculo. De acuerdo con los cálculos que realizaremos más adelante, en el caso de nuestros datos de alturas y pesos, $a = -151.31$ libras. Podemos considerar esta cantidad como punto de partida para sumar o restar cantidades de bX.

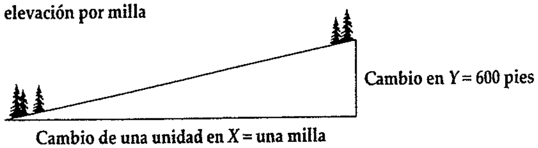
¿Qué hace a? Ésta ancla la recta de regresión al eje Y. Esto es necesario porque puede haber cualquier cantidad de variables independientes (X) que se correlacionan con Y y tienen la misma pendiente, aunque diferentes magnitudes de Y. O puede haber diferentes poblacio-

FIGURA 14-6

Ilustración de la pendiente como la razón de un cambio vertical en Y con respecto a un cambio horizontal de una unidad en X

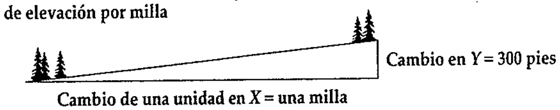
Y = distancia recorrida en millas $\qquad$ X = elevación en pies

A. Pendiente del incremento en la elevación regresionada con respecto a la distancia recorrida: b = 600 pies de elevación por milla



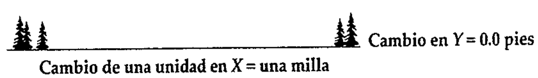
$$b = \text{pendiente} = \frac{\text{elevación}}{\text{recorrido}} = \frac{\text{cambio en } Y}{\text{cambio de una unidad en } X} = \frac{600 \text{ pies}}{1 \text{ milla}} = 600 \text{ pies por milla}$$

B. Pendiente del incremento en la elevación regresionada con respecto a la distancia recorrida: b = 300 pies de elevación por milla



$$b = \text{pendiente} = \frac{\text{elevación}}{\text{recorrido}} = \frac{\text{cambio en } Y}{\text{cambio de una unidad en } X} = \frac{300 \text{ pies}}{1 \text{ milla}} = 300 \text{ pies por milla}$$

C. Pendiente del incremento en la elevación regresionada con respecto a la distancia recorrida en terreno llano (es decir, b = 0 pies *de elevación* por milla)



$$b = \text{pendiente} = \frac{\text{elevación}}{\text{recorrido}} = \frac{\text{cambio en } Y}{\text{cambio de una unidad en } X} = \frac{0 \text{ pies}}{1 \text{ milla}} = 0 \text{ pies por milla}$$

nes con las mismas pendientes y diferentes intersecciones de Y. Si tienen las mismas pendientes, b, estas rectas de regresión serían paralelas entre sí en un diagrama de dispersión y solamente se distinguirían por sus constantes, a. Finalmente, observa que la intersección de Y es con frecuencia un punto hipotético, ya que puede ser que no se presenten casos en los que $X = 0$. Por ejemplo, nadie en una población puede tener una altura de cero. Por consiguiente, es común que a tenga un valor sin sentido como *menos* 159.31 libras.

Intersección de Y, a (constante de la ecuación de regresión)

a = intersección del eje Y de la recta de regresión en un diagrama de dispersión
= punto en el que la recta de regresión cruza el eje Y cuando $X = 0$
= valor predicho de Y cuando $X = 0$ (es decir $\acute{Y}_{(X=0)}$)

donde

Y = variable dependiente de intervalo/razón
X = variable independiente de intervalo/razón

Cálculo de los términos de la fórmula de la línea de regresión

Para determinar la fórmula de la ecuación lineal $\acute{Y} = a + bX$, primero calculamos b de la siguiente manera:

Cálculo del coeficiente de regresión *b*, la pendiente de la recta de regresión

$$b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$$

donde

b = coeficiente de regresión (pendiente de la recta de regresión)
X = variable independiente de intervalo/razón
Y = variable dependiente de intervalo/razón
$\bar{X}$ = media de la variable independiente, X
$\bar{Y}$ = media de la variable dependiente, Y

En seguida calculamos a sustituyendo el valor calculado de b en la ecuación $\acute{Y} = a + bX$. Sin embargo, para despejar a, debemos estipular valores conocidos de X y Y. Como resultado, en cualquier recta de regresión, las coordenadas de las medias de X y Y caerán en la recta. Por consiguiente, podemos sustituir las medias de X y Y, junto con b y despejar a.

Cálculo de la intersección de Y, *a*

$$a = \bar{Y} - b\bar{X}$$

donde

a = intersección Y
$\bar{Y}$ = media de la variable dependiente
b = coeficiente de regresión (pendiente de la recta de regresión)
$\bar{X}$ = media de la variable independiente

Finalmente, incluimos a y b en la fórmula de la recta de regresión:

$$\acute{Y} = a + bX$$

Ahora calculemos los estadísticos de la recta de regresión para el peso regresionado a partir de la estatura (figura 14-4). Tomando las sumas de la hoja de cálculo de la tabla 14-2, primero calculamos b:

$$b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2} = \frac{286.12}{61.88} = 4.62 \text{ libras por pulgada}$$

En seguida utilizamos $\overline{X}$, $\overline{Y}$ y b para calcular a:

$$a = \overline{Y} - b\overline{X} = 157.44 - (4.62)(68.56) = -159.31 \text{ onzas}$$

Finalmente, especificamos la ecuación precisa de la recta de regresión que se ajusta al diagrama de dispersión del peso regresionado a partir de la altura, sustituyendo los valores calculados de a y b (figura 14-4):

$$\acute{Y} = a + bX = -159.31 + (4.62)X$$

Esta ecuación de regresión ahora se puede utilizar para calcular $\acute{Y}$, la *mejor estimación* de Y (peso) para cualquier valor de X (altura). Insertamos unas cuantas puntuaciones X en la ecuación y resolvemos $\acute{Y}$ para obtener los resultados de la tabla 14-3.

Trazamos la recta de regresión marcando dos o más de estas coordenadas X, $\acute{Y}$ sobre el diagrama de dispersión y perfilando la recta entre ellas. Esto se muestra en la figura 14-4, donde las coordenadas (67, 150.23) y (72, 173.33) se utilizan como puntos de referencia para la recta. Se verifica nuevamente la exactitud de la recta determinando en la figura 14-4 las otras coordenadas X, $\acute{Y}$ calculadas en la tabla 14-3. Estas coordenadas caerán en una recta. Observa también que, aunque la intersección Y es *menos* 159.31 libras, la recta de regresión cruza el eje Y en aproximadamente 140 libras. Esta discrepancia se debe al hecho de que el punto donde intersectan los ejes de X y Y, no es cero. Más bien, los ejes, como se representan, comienzan en aproximadamente 65 pulgadas y 130 libras, respectivamente. Esto no constituye un error. Los ejes del diagrama han sido *truncados* (o acortados) para lograr claridad en la presentación.

TABLA 14-3 | Coordenadas X, $\acute{Y}$: mejores estimaciones de los pesos ($\acute{Y}$) de la población de hombres de último año de preparatoria basada en las alturas (X)

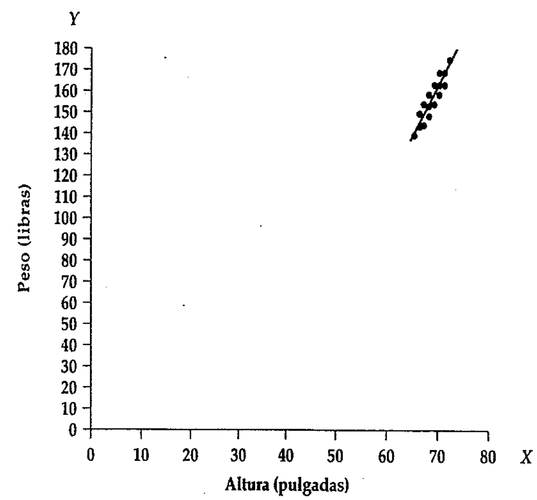
Especificaciones	Cálculos		
X (altura en pulgadas)	$\acute{Y}$ (= $a + bX$) (mejor estimación del peso en libras)		Puntos importantes relacionados con las características de la ecuación lineal
0	−159.31	←	La intersección de Y a veces constituye una abstracción de cálculo; quizá no exista en realidad.
65	140.99		
66	145.61		
67	150.23		
68	154.85		
68.56	157.44	←	Las coordenadas ($\overline{X}$, $\overline{Y}$) siempre se ubicarán en la recta de regresión.
69	159.47		
70	164.09		
		> ←	La diferencia de peso (Y) entre dos alturas (X)
71	168.71		que se encuentran separadas 1 pulgada
		> ←	constituye la pendiente, b = 4.62 libras.
72	173.33		
73	177.95		

¿Por qué truncar los ejes del diagrama de dispersión? A veces las variables de intervalo/razón de un diagrama de dispersión tienen diversos valores que carecen de sentido. En nuestra ilustración de las variables de altura y peso, ninguno de los estudiantes de último año de preparatoria tiene, digamos, 19 pulgadas de altura o pesa solamente 30 libras. En el diagrama de dispersión de peso regresionado a partir de la altura que truncamos, o acortamos, los ejes para cada variable para formar el diagrama de dispersión parecen equilibrados en las páginas. La figura 14-7 muestra el diagrama de dispersión de los datos de la figura 14-4 con los ejes sin truncar. El valor debería resultar obvio cuando las figuras 14-4 y 14-7 se comparan.

Para la mente particularmente inquisitiva: relación matemática entre el coeficiente de correlación r de Pearson y el coeficiente de regresión, b

Observa que existen similitudes en las fórmulas para r y b. En particular, los numeradores son los mismos. Esto significa que los signos de r y b siempre serán iguales. También observa que las dos ecuaciones incluyen sumas de cuadrados, los mismos cálculos que se realizaron al calcular las desviaciones estándar (capítulo 5). Como resultado, la r de Pearson, como b, es una pendiente de una recta de regresión. La r de Pearson es una pendiente estandarizada que va con el diagrama de puntuaciones estandarizadas para X y Y (puntuaciones Z, capítulo 5). Recuerda que una puntuación Z es una puntuación de desviación dividida entre la desviación estándar y su unidad de medida *es* la desviación estándar. Si asignamos coordenadas a las puntuaciones estandarizadas de X y Y (coordenadas Z_x, Z_y), r sería la pendiente de esta recta. El coeficiente de regresión, b, es una pendiente no estandarizada, lo cual significa que en el diagrama de dispersión la recta de regresión se construye alrededor de las unidades de medida de las puntuación en bruto de las variables. En el caso de los estudiantes de último año de preparatoria, estas unidades de medida son *pulgadas* para la altura (X), y *libras* para el peso (Y). La r de Pearson es una pendiente estandarizada por el hecho de que presenta la

FIGURA 14-7
Diagrama de dispersión con los ejes sin truncar: la altura como predictor del peso



pendiente en unidades de medida de desviación estándar. Comparemos las interpretaciones de estas pendientes no estandarizada y estandarizada en el caso de nuestro ejemplo relacionado con pesos y alturas:

Pendiente *no estandarizada*, $b = 4.62$ libras por pulgada: En la recta de regresión del diagrama de dispersión de alturas y pesos, 1 pulgada de incremento en altura se relaciona con 4.62 libras de incremento en peso.

Pendiente *estandarizada*, $r = .91$ desviaciones estándar de libras de peso por 1 desviación estándar de pulgadas de altura: en la recta de regresión del diagrama de dispersión de puntuaciones Z de alturas (Z_x) y pesos (Z_y), 1 *desviación estándar* de incremento en la altura se relaciona con 0.91 *desviaciones estándar* de incremento en el peso.

Las pendientes estandarizadas son importantes en la regresión múltiple, en el caso en el que hay dos o más variables independientes que predicen Y. Estas variables independientes probablemente tendrán diferentes unidades de medida de puntuaciones en bruto. Con el fin de comparar las pendientes, con frecuencia debemos estandarizar las unidades de medida de todas las variables. Más información relacionada con pendientes estandarizadas y la regresión múltiple se puede localizar en las *Extensiones de los capítulos 14 y 15* en el sitio web *The Statistical Imagination* ubicado en www.mhhe.com/ritchey2.

Cálculos paso por paso de los estadísticos de correlación bivariada y regresión

De acuerdo con nuestro ejemplo relativo a si existe una correlación entre la altura y el peso de los estudiantes del último año de preparatoria (datos de la tabla 14-1):

1. Decide qué variable es la independiente y qué variable es la dependiente y traza el diagrama de dispersión (figura 14-1).
2. Observa el diagrama de dispersión para buscar un patrón lineal en forma alargada en las coordenadas. Parece haber un patrón lineal.
3. Organiza la hoja de cálculo (tabla 14-2).
4. Calcula las medias de X y Y:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{1\,097}{16} = 68.56 \text{ pulgadas}$$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{2\,519}{16} = 157.44 \text{ libras}$$

5. Completa las columnas A y E en la hoja de cálculo: *a*) calcula las puntuaciones de desviación de X y Y (columnas A y B) y confirma si las sumas son iguales a cero dentro del error de redondeo; *b*) calcula la covariación entre X y Y (columna C); *c*) calcula la suma de los cuadrados de X y Y (columnas D y E).

6. Calcula el coeficiente de correlación r de Pearson:

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\,\Sigma(Y-\bar{Y})^2}} = \frac{286.12}{\sqrt{(61.88)\,(1\,587.88)}} = \frac{286.12}{313.46} = .91$$

7. Calcula el coeficiente de regresión (pendiente), b:

$$b = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} = \frac{286.12}{61.88} = 4.62 \text{ libras por pulgada}$$

8. Calcula la intersección de Y, a:

$$a = \bar{Y} - b\bar{X} = 157.44 - (4.62)(68.56) = -159.31 \text{ libras}$$

9. Especifica la ecuación de regresión:

$$\acute{Y} = a + bX = -159.31 + (4.62)X$$

10. Calcula las mejores estimaciones de Y incluyendo algunos valores de X en la ecuación de regresión y resolviendo para Y. Por ejemplo, para $X = 67$ pulgadas y $X = 72$ pulgadas:

$$\acute{Y} = a + bX = -159.31 + (4.62)(67) = 150.23 \text{ libras}$$

$$\acute{Y} = a + bX = -159.31 + (4.62)(72) = 173.33 \text{ libras}$$

11. Utiliza estas estimaciones (coordenadas X, $\acute{Y}$) para trazar la recta de regresión en el diagrama de dispersión, como lo indica la figura 14-4.

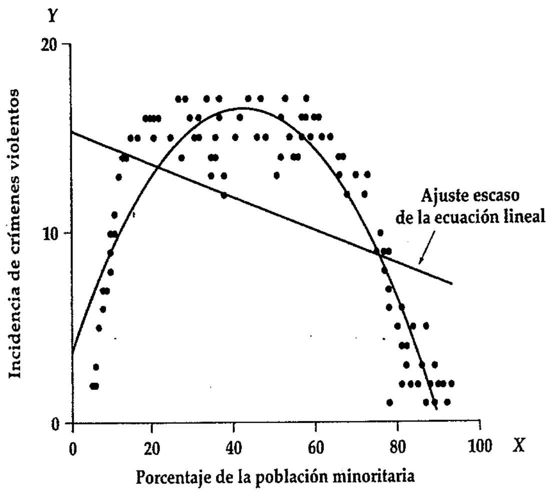
Insensatez y falacias estadísticas: el fracaso para observar un diagrama de dispersión antes de calcular la *r* de Pearson

Las ecuaciones lineales sólo funcionan con un patrón lineal en los diagramas de dispersión

Ansioso de ver sus resultados, un investigador puede sentirse tentado a omitir la tarea de trazar y observar diagramas de dispersión. Incluso con la computadora, esta tarea implica cierto tiempo. Sin embargo, constituye un error proceder sin observar un diagrama de dispersión. Los estadísticos de regresión lineal bivariada se basan en predecir a partir de la recta generada mediante la fórmula $\acute{Y} = a + bX$. Si las coordenadas no se ajustan en torno a la recta, las predicciones de Y basadas en X (es decir las $\acute{Y}$) no se encontrarán cerca de los valores observados de Y. Dicho sencillamente, la recta no se ajustará al patrón de coordenadas; por lo tanto, las predicciones realizadas utilizando dicha recta mal colocada serán erróneas. Los estadísticos de la regresión lineal no son adecuados en el caso de relaciones no lineales (o curvilíneas), tales como la que aparece en la figura 14-8. En la figura 14-8, cada punto (coordenada) representa un vecindario (utilizando datos imaginarios). La posición en el eje X indica el porcentaje de la población del vecindario de estatus minoritario (es decir, clasi-

FIGURA 14-8

Diagrama de dispersión de la incidencia de crímenes violentos contra miembros de minorías de vecindarios regresionados a partir del porcentaje de la población minoritaria en el vecindario (muestra una relación no lineal)



ficaciones raciales o étnicas de afroamericano, asiático-americano, norteamericano nativo e hispano). La posición en el eje Y representa el número de crímenes violentos que se presentan en la vecindad durante cierto periodo. Este diagrama curvilíneo nos dice que los crímenes violentos *no* son frecuentes cuando el tamaño de la población minoritaria es muy bajo o muy alto. Cuando la población minoritaria es muy baja, ésta no representa ninguna amenaza para la población blanca mayoritaria. Cuando la población minoritaria es muy alta, los blancos embargados de odio se sienten superados en número y temerosos de actuar de acuerdo con sus sentimientos. Los crímenes violentos alcanzan su nivel más alto cuando las poblaciones minoritaria y blanca son aproximadamente iguales y *luchan por el control* del vecindario. La forma de esta curva es una parábola invertida, y su ecuación —no la ecuación de una recta— se utilizarían para calcular el número de crímenes violentos sobre la base del tamaño de la población minoritaria. La ecuación de una parábola invertida es $\acute{Y} = [-c_1(X = c_2)^2] + c_3$, donde las c representan constantes que indican la amplitud de la parábola y la localización de su vértice (es decir, su cúspide). Si la ecuación lineal se utiliza de forma equivocada, la recta de la regresión lineal resultante (indicada en la figura 14-8) constituye un ajuste inadecuado con respecto al patrón de las coordenadas. Las estimaciones realizadas a partir de esta recta serán erróneas.

El análisis de las relaciones curvilíneas rebasa el alcance de este libro. Si las coordenadas de un diagrama de dispersión no se ajustan a un patrón lineal en forma alargada, los estadísticos de este capítulo no aplican. Por fortuna, el patrón lineal entre variables de intervalo/razón es extremadamente común. Sin embargo, es importante observar un diagrama de dispersión para identificar patrones no lineales, como el de la figura 14-8.

Coordenadas de valores extremos y la atenuación e inflación de los coeficientes de correlación

Observa que utilizamos las medias de X y Y para calcular el coeficiente de correlación r y la ecuación de la recta de regresión. Todos estos estadísticos se *basan en la media.* Como lo

indicamos en el capítulo 4, la media es susceptible de distorsionarse por las puntuaciones extremas, en una distribución de puntuaciones de la variable. Asimismo, en el caso de las coordenadas de un diagrama de dispersión, unas cuantas coordenadas extremas —*aquellas que caen fuera del patrón general del diagrama de dispersión*— pueden distorsionar los coeficientes de correlación y regresión. Si éstas son significativas, estas distorsiones pueden provocar que la recta de regresión no se ajuste a los datos del diagrama de dispersión. Las coordenadas extremas pueden debilitar o *atenuar* los coeficientes de correlación y regresión calculados con ellas.

La **atenuación de la correlación** es *el debilitamiento o reducción de los coeficientes de correlación y regresión.* Una correlación atenuada generará una r de Pearson pequeña y, por lo tanto, provocará que concluyamos que una relación entre dos variables de intervalo/razón no existe cuando en realidad sí existe o que su tamaño es menor que el real.

> **Atenuación de la correlación** Debilitamiento o reducción de los coeficientes de correlación y regresión (a menudo como consecuencia de la presencia de coordenadas de valores extremos).

El siguiente diagrama de dispersión ilustra la atenuación que resulta de una coordenada extrema en el caso del peso y la altura de los estudiantes del último año de preparatoria. Supongamos que se incluye sin ninguna intención un miembro del equipo de baloncesto en la muestra. Este jugador es alto y delgado, mide 73 pulgadas de estatura y pesa 146 libras. La figura 14-9 muestra el diagrama de dispersión de los nuevos datos. Las coordenadas X, Y (73, 146) del jugador de baloncesto constituyen una coordenada extrema, y las rectas de regresión se presentan con y sin éstas. Las diferentes ubicaciones de las rectas de regresión revelan que la adición de este dato extremo único aparta la recta de regresión del patrón lineal encontrado en el resto de los casos. Esta recta de regresión atenuada no resulta útil

FIGURA 14-9

Ilustración de la atenuación que resulta de una coordenada extrema: peso (Y) en retroceso en alto (X)

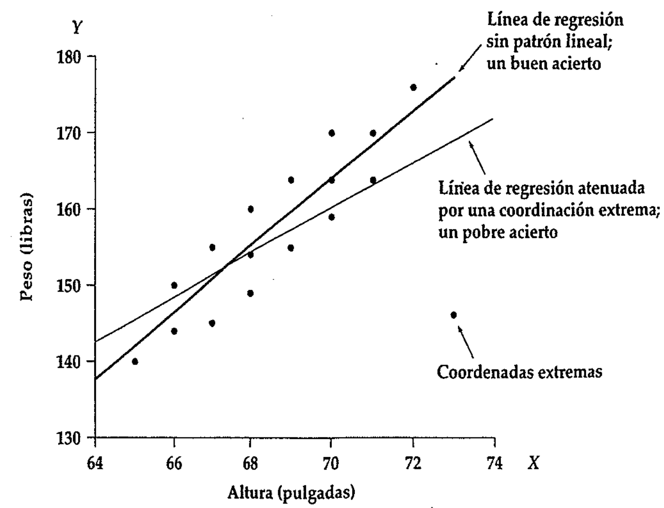
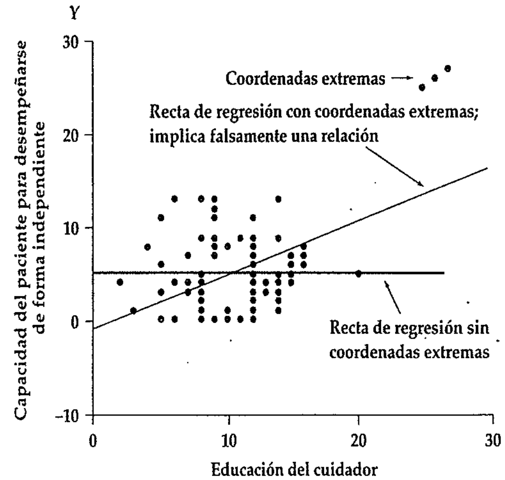


FIGURA 14-10

Ilustración de la inflación de la *r* de Pearson como resultado de coordenadas extremas. Capacidad del paciente para desempeñarse de forma independiente (*Y*) un retroceso de la educación del cuidador (*X*) (compare con la figura 14-3)



para predecir el peso a partir de la altura. Además, la añadidura de esta coordenada extrema única reduce o *atenúa* el coeficiente de correlación del valor original de 0.91 a solamente 0.64 (los cálculos no se incluyen). El potencial de atenuación de la correlación constituye otra razón importante para observar el diagrama de dispersión antes de proceder con el análisis de correlación.

Así como una coordenada extrema puede debilitar los coeficientes, también puede inflarlos. En la figura 14-3 vemos que no existía relación entre la educación del cuidador y la capacidad del paciente para desempeñar su función sin ayuda. Ahora supongamos que la muestra incluía las tres coordenadas extremas de la figura 14-10. Estas coordenadas extremas jalan la recta de regresión calculada hacia arriba y le imprimen una pendiente positiva. Además, la *r* de Pearson se incrementa de cero (no existe relación) a 0.46 (es decir, una relación positiva; no se incluyen los cálculos). Las coordenadas extremas inflan los coeficientes de correlación y regresión, y conducen a una conclusión falsa relativa al hecho de que existe una relación cuando, de hecho no existe.

¿Cómo podemos evitar la atenuación o inflación de los coeficientes? Con frecuencia una o dos de las coordenadas extremas se pueden explicar por las circunstancias y pueden excluirse con justificación del análisis. En el caso de la atenuación de la figura 14-9, podríamos eliminar al jugador de baloncesto y reportar la correlación más fuerte de $r = .91$. Nuestro reporte diría que, salvo en el caso particular de los estudiantes de último grado altos y delgados, como los jugadores de baloncesto, la altura constituye un predictor fuerte del peso en el caso de los estudiantes del último año de preparatoria. Asimismo, en el caso de los coeficientes inflados del ejemplo del cuidador (figura 14-10) podríamos afirmar que las coordenadas extremas son casos aleatorios que no se ajustan al patrón general. En cualquier figura, el diagrama de dispersión revela que la utilización de rectas de regresión basadas en estos datos extremos no proporcionaría buenas estimaciones.

En las circunstancias de investigación de la vida diaria existe mucho potencial para la atenuación o inflación de la correlación, especialmente si el investigador no tiene idea de la conformación de los datos y la población de origen. Por ejemplo, en un estudio de obreros fabriles sobre la relación entre el tiempo de laborar en la empresa (*X*) y el nivel de pago en el tabulador (*Y*), resultaría un error integrar en la misma muestra a los gerentes (cuya paga es elevada) y a los obreros de la línea de montaje. Los gerentes que han laborado en la empresa por poco tiempo quizá tengan sueldos muy altos, y sus coordenadas se ubicarán en la parte superior izquierda del diagrama de dispersión.

Otro ejemplo sería el de un estudio de los estados y la relación entre el tamaño de las poblaciones (*X*) y las tasas de divorcio (*Y*). La inclusión de Nevada con su exorbitante tasa de divorcios distorsionaría mucho los resultados. La mayoría de los divorcios de Nevada se presentan entre personas provenientes de otros estados. Como consecuencia de este hecho comúnmente conocido, podría excluirse a Nevada y los resultados podrían reportarse con la anotación "con excepción de Nevada".

Finalmente, el potencial de atenuación o inflación de los coeficientes de correlación y regresión subraya la conveniencia de emplear una muestra grande. Como en el caso de cualquier estadístico basado en la media, cuantas más oportunidades de muestreo haya para compensar el efecto de las coordenadas extremas, más débiles serán los efectos sobre las predicciones basadas en el conocimiento de *X*. En otras palabras, es deseable contar con muchos grados de libertad.

RESUMEN

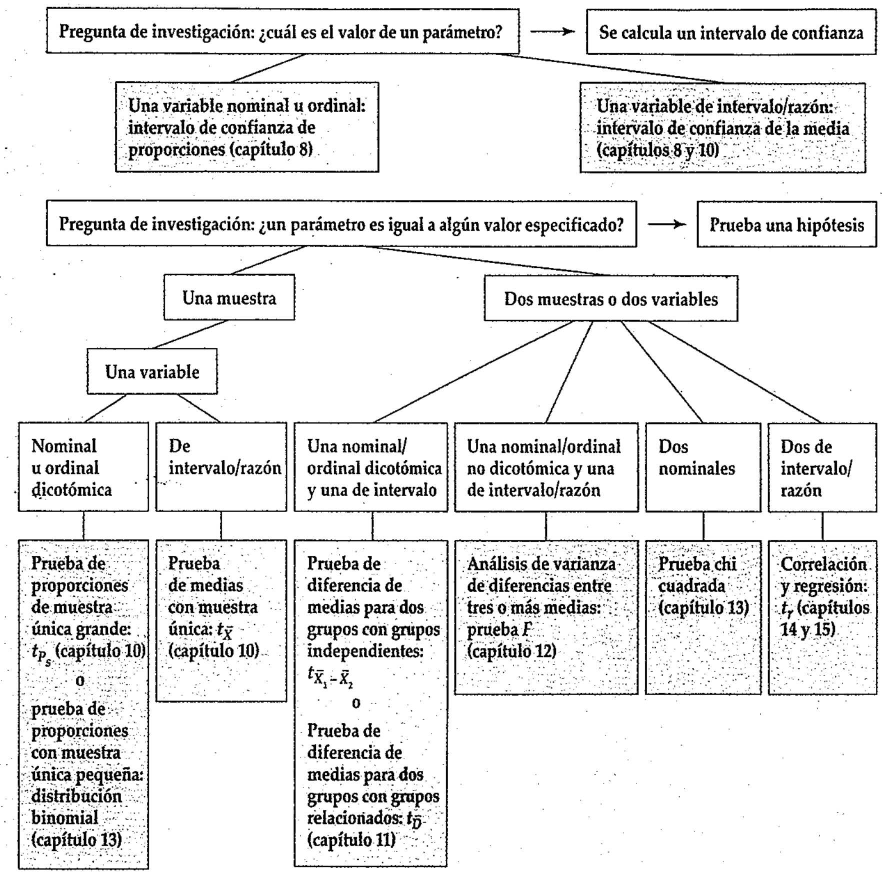
1. Una correlación bivariada es un cambio sistemático en las puntuaciones de dos variables de intervalo/razón.
2. El análisis de correlación y regresión lineal bivariadas simple es el nombre del procedimiento para mejorar las mejores estimaciones de una variable dependiente (*Y*) tomando en cuenta su relación con una variable independiente (*X*). Este procedimiento aplica la fórmula para una recta con el fin de mejorar las mejores estimaciones de *Y* para todos los valores de *X*.
3. Un diagrama de dispersión es una rejilla de dos dimensiones de las coordenadas de dos variables de intervalo/razón, *X* y *Y*. Una coordenada es un punto localizado sobre el diagrama de dispersión en el que los valores de *X* y *Y* se grafican para un caso.
4. Los estadísticos de correlación y regresión lineal aplican solamente a los diagramas de dispersión con coordenadas que se ajustan a un patrón lineal —un patrón alargado (en forma de cigarro) que se aproxima a la forma de una recta.
5. La recta de regresión es la recta de mejor ajuste trazada por medio de las coordenadas *X*, *Y* de un diagrama de dispersión de dos variables de intervalo/razón. La fórmula de una recta para calcular *Y* es $\acute{Y} = a + bX$.
6. El beneficio del análisis de regresión consiste en la habilidad para mejorar las estimaciones de *Y* en una población utilizando la recta de regresión y reportando una $\acute{Y}$ en lugar de reportar simplemente la media muestral de *Y*.
7. A veces es recomendable truncar los ejes de un diagrama de dispersión al momento de detallarlos.

8. La r de Pearson es un coeficiente de correlación bivariada que se utiliza ampliamente, y que mide la estrechez del ajuste de las coordenadas X, Y en torno a la recta de regresión. Éste evalúa el grado en que las puntuaciones de desviación para X y Y varían juntas. Su fórmula incluye puntuaciones de desviación y sumas de cuadrados.
9. Las características de la r de Pearson son las siguientes: *a*) los valores calculados de la r de Pearson pueden variar de –1 a +1; *b*) cuanto mayor sea el valor absoluto de la r de Pearson, más estrecho será el ajuste de las coordenadas X, Y en torno a la recta de regresión; *c*) cuando la recta de regresión tiene una pendiente ascendente, tenemos una correlación positiva. La r de Pearson será positiva hasta un valor de +1.00. *d*) Cuando la recta de regresión tiene una pendiente descendente, tenemos una correlación negativa. La r de Pearson será negativa hasta un valor de –1.00. *e*) Cuando la recta de regresión es plana, no existe correlación y la r de Pearson = 0.
10. Los coeficientes y símbolos de la fórmula de la recta de regresión son los siguientes: $\acute{Y} = a + bX$: *a*) $\acute{Y} = Y$ predicha (una estimación de la variable dependiente Y calculada para un valor calculado de la variable dependiente X); *b*) b = pendiente de la recta de regresión (denominada *coeficiente de regresión*). Ésta expresa la pendiente en el sentido de ascender o descender por una montaña. Responde a la pregunta: ¿cuánto se eleva la recta por cada unidad de recorrido de X? *c*) a = intersección de Y, punto en el que la recta de regresión cruza el eje Y cuando $X = 0$.
11. La pendiente no estandarizada, b, se expresa en unidades de medida de puntuación en bruto. La r de Pearson se puede interpretar como una pendiente estandarizada expresada en unidades de medida de desviación estándar.
12. Observa de cerca los diagramas de dispersión para determinar si una relación lineal es evidente en el patrón de las coordenadas. Verifica al existencia de relaciones no lineales o curvilíneas y de coordenadas extremas.
13. La atenuación de la correlación es el debilitamiento o reducción de los coeficientes de correlación y regresión debida a las coordenadas extremas u otras particularidades en los datos.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones de los capítulos 14 y 15 del material del texto disponibles en el sitio web *The Statistical Imagination*, www.mhhe.com/ritchey2 incluyen una introducción a la correlación y regresión múltiples, situación en la que una variable dependiente se regresiona a partir de dos o más variables independientes.

PROCEDIMIENTOS ESTADÍSTICOS CUBIERTOS HASTA AQUÍ



FÓRMULAS PARA EL CAPÍTULO 14

Para completar la hoja de cálculo (tabla 14-2):

$$\bar{X} = \frac{\Sigma X}{n} \qquad \bar{Y} = \frac{\Sigma Y}{n}$$

Coeficiente de correlación r de Pearson:

$$r = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma (X - \bar{X})^2 \, \Sigma (Y - \bar{Y})^2}}$$

Coeficiente de regresión o pendiente, b:

$$b = \frac{\Sigma(X - \overline{X})(Y - \overline{Y})}{\Sigma(X - \overline{X})^2}$$

Intersección de Y, a:

$$a = \overline{Y} - b\overline{X}$$

Ecuación de regresión:

$$\acute{Y} = a + bX$$

PREGUNTAS PARA EL CAPÍTULO 14

1. Con la ayuda de un diagrama de dispersión trazado a mano y la ecuación de la recta de regresión $\acute{Y} = a + bX$, explica la idea central detrás del análisis de regresión.
2. Con la ayuda de diagramas de dispersión trazados a mano, ilustra la razón por la que el análisis de correlación y regresión lineal se aplica solamente cuando surge un patrón lineal en forma de cigarro a partir de las coordenadas.
3. Con la ayuda de diagramas de dispersión trazados a mano, ilustra el patrón de coordenadas en el caso de relaciones positivas, negativas y la ausencia de relación entre X y Y.
4. Con la ayuda de diagramas de dispersión trazados a mano, ilustra por qué a veces es necesario truncar los ejes de un diagrama de dispersión.
5. ¿Qué mide el coeficiente de correlación r de Pearson?
6. ¿Qué mide el coeficiente de regresión b?
7. En la recta de regresión, la intersección Y, a, es el valor de $\acute{Y}$ en el que X = ________.
8. En una relación lineal entre dos variables de intervalo/razón, ¿cuál es la coordenada X, Y que siempre caerá sobre la recta de regresión?
9. Relaciona las siguientes expresiones en lo que se refiere a la dirección de las correlaciones:

   a) Una correlación positiva ________ La recta de regresión no tiene pendiente; $r = 0$ y $b = 0$.

   b) No hay correlación ________ La pendiente de la recta de regresión es descendente; r y b tienen un signo negativo; un incremento en X se relaciona con una reducción en Y.

   c) Correlación negativa ________ La pendiente de la recta de regresión es ascendente; r y b tienen un signo positivo; un incremento en X se relaciona con un incremento en Y.
10. La atenuación o inflación del cálculo de los coeficientes de correlación y regresión puede ser resultado de la presencia de __________ en el diagrama de dispersión.

EJERCICIOS PARA EL CAPÍTULO 14

Conjunto de problemas 14A

Muestra todas las sumas y las fórmulas.

14A-1. Una instructora con 16 estudiantes aplica un examen parcial y un examen final. Se encuentra interesada en saber si la calificación del examen parcial constituye un buen predictor de la calificación del examen final. Sus datos aparecen en la tabla.

a) Decide cuál es la variable independiente y cuál es la dependiente y traza el diagrama de dispersión.
b) Calcula los estadísticos de la recta de regresión y traza la recta.
c) Calcula el coeficiente de correlación bivariada r de Pearson.
d) ¿Parece haber alguna relación lineal en el patrón de las coordenadas de las calificaciones de la prueba?

Calificación del examen parcial	Calificación del examen final
78	83
91	82
95	91
74	81
87	85
83	87
89	83
92	97
94	98
58	66
71	79
76	84
87	91
91	92
77	75
85	89

14A-2. Los siguientes datos provienen de una muestra de estudiantes de nuevo ingreso (de primero y segundo año) en la universidad de Tuffstuff.

a) Traza el diagrama de dispersión del promedio (Y) regresionado a partir de la puntuación de comprensión de lectura (X).
b) *Sin llevar a cabo ningún cálculo*, traza lo que parecería ser la recta a través del conjunto de coordenadas.
c) *Sin llevar a cabo ningún cálculo*, calcula a y b en la ecuación de regresión $\acute{Y} = a + bX$.
d) Ahora calcula los estadísticos de regresión para la relación de la lectura de comprensión con respecto al promedio y traza la recta de regresión.
e) Compara los estadísticos que calculaste con los que obtuviste en el inciso *c* con el fin de que te asegures de que comprendes lo que cada estadístico significa.

Estudiante	Promedio	Puntuación de lectura de comprensión	Horas de estudio por clase
1	0.90	2	2.00
2	1.12	3	1.25
3	1.46	2	2.50
4	1.93	4	5.00
5	2.00	3	3.00
6	2.16	3	4.15
7	2.18	2	4.72
8	2.21	3	5.75
9	2.33	4	6.13
10	2.39	6	4.75
11	2.46	5	4.25
12	2.54	6	4.84
13	2.68	5	5.75
14	2.73	4	6.00
15	2.85	7	5.89
16	2.87	6	6.37
17	2.93	8	6.50
18	2.99	7	7.00
19	3.04	9	6.00
20	3.14	8	6.93
21	3.22	9	7.16
22	3.27	9	6.94
23	3.28	8	8.10

14A-3. Como empleado de una agencia protectora de la salud del consumidor tú realizas un seguimiento de las organizaciones para el mantenimiento de la salud (OMS). Se ha criticado a estas organizaciones por retirar rápidamente a la gente de los hospitales con el fin de ahorrar dinero y elevar las ganancias. Tú analizas los procedimientos de angioplastías llevados a cabo en hospitales, propiedad de 12 de dichas organizaciones. Los datos imaginarios indican las reducciones promedio en el tiempo de estancia en días y el incremento porcentual en utilidades para las OMS durante el año pasado.

a) Traza el diagrama del incremento porcentual de las ganancias (Y) regresionado a partir de la reducción en el tiempo de estancia (X).
b) Calcula los estadísticos de regresión y traza la recta de regresión.
c) Calcula el coeficiente de correlación bivariada r de Pearson.
d) ¿Parece existir una relación lineal en el patrón de las coordenadas?

Reducción promedio del tiempo de estancia (días)	Incremento porcentual en las ganancias
1.3	2.9
2.2	15.3
2.3	13.9
1.7	6.3
1.9	10.4
1.0	0.8
2.1	15.3
1.6	13.1
2.2	8.9
1.4	6.8
1.3	11.5
1.9	11.4

14A-4. La medida del nivel de pobreza de un condado es el porcentaje de niños que satisfacen los requisitos para recibir almuerzo gratuito en la escuela. La siguiente tabla incluye la variable *Porcentaje de alumnos que reciben almuerzo gratuito* (X), así como una medida compuesta de la forma en que fueron calificados los alumnos del condado en el examen SAT de aptitud académica para ingresar a la universidad (Y).

a) Traza el diagrama de dispersión de Y regresionada a partir de X.
b) Calcula el coeficiente de correlación bivariada r de Pearson.
c) Observa las coordenadas extremas en el diagrama de dispersión e identifícalas en la tabla de datos.
d) Vuelve a calcular la r de Pearson sin estas coordenadas extremas
e) Realiza algún comentario sobre el efecto de las coordenadas extremas sobre el cálculo del coeficiente de correlación r de Pearson.

Calificación en el SAT	Porcentaje de alumnos que reciben almuerzo gratuito
68	34
71	15
56	37
61	33
39	75
65	73
47	59
57	40
43	79
69	69
54	59
55	38
70	17
63	36

Conjunto de problemas 14B

Muestra todas las sumas y las fórmulas.

14B-1. El ACT es un examen estandarizado para evaluar si un estudiante reúne los requisitos para ingresar a la universidad. Una asesora de admisiones de la universidad a la que tú asistes está interesada en saber si las puntuaciones del ACT constituyen un buen predictor del promedio de calificaciones (GPA) en la universidad. Recolecta una muestra de 14 estudiantes universitarios de tercer año. Sus datos aparecen en la tabla.

a) Decide cuál es la variable independiente y cuál es la variable dependiente y traza el diagrama de dispersión.
b) Calcula los estadísticos de la recta de regresión y traza la recta.
c) Calcula el coeficiente de correlación bivariada *r* de Pearson.
d) ¿Parece haber alguna relación lineal en el patrón de puntuación del ACT y el promedio de calificaciones?

Puntuación del ACT	Promedio de calificaciones
24	3.10
29	3.68
31	3.82
21	3.00
17	2.78
21	3.23
24	2.86
23	3.07
25	3.15
21	2.87
28	3.65
27	3.41
23	2.96
21	3.10
18	2.75

14B-2. Tú estás interesado en los usos intergeneracionales de los tipos de tecnología de grabación de sonido. Los siguientes datos imaginarios provienen de una muestra de 20 adultos.

a) Traza el diagrama de dispersión de los discos compactos que se poseen (*Y*) regresionados a partir de la edad (*X*).
b) *Sin llevar a cabo ningún cálculo*, traza lo que parecería ser la recta a través del conjunto de coordenadas.
c) *Sin llevar a cabo ningún cálculo*, calcula *a* y *b* en la ecuación de regresión $\hat{Y} = a + bX$.

d) Calcula los estadísticos de regresión para la relación entre el número de discos compactos que se poseen y la edad, y traza la recta de regresión (para este problema no se emplearán los datos de discos de vinil).
e) Compara los estadísticos que calculaste con los que estimaste en el inciso *c*) con el fin de que te asegures de que comprendes lo que cada estadístico significa.

Adulto	Edad (*X*)	Discos compactos poseídos (*Y*)	Discos de vinil poseídos
1	36	12	16
2	51	6	25
3	32	15	18
4	34	10	20
5	28	15	16
6	27	18	15
7	44	10	19
8	45	8	24
9	29	18	23
10	50	7	27
11	49	9	26
12	37	10	20
13	47	9	23
14	44	12	20
15	41	9	22
16	32	15	21
17	51	6	25
18	33	18	23
19	28	16	18
20	44	12	23

14B-3. El capital social constituye una medida de las relaciones sociales que la gente emplea como recurso para beneficiarse. Como investigador, tú analizas si los individuos con un capital social superior tienen más gente en quién apoyarse y, por consiguiente, tienen una mayor satisfacción con la vida. Tú utilizas una escala de capital social que varía de 10 a 25 y una escala de satisfacción con la vida que varía de 20 a 40.

a) Traza el diagrama de dispersión de satisfacción en la vida (*Y*) regresionada a partir del capital social (*X*).
b) Calcula los estadísticos de regresión y traza la recta de regresión.
c) Calcula el coeficiente de correlación *r* de Pearson.
d) ¿Parece existir una relación lineal en el patrón de coordenadas?

Capital social	Satisfacción con la vida
15	25
12	24
23	31
24	34
19	28
21	31
17	25
14	21
21	35
19	26
20	28
15	23
12	20
17	27
16	24
22	31

14B-4. Los estudios muestran que en los distritos escolares existe una correlación entre las borracheras en la preparatoria y las borracheras en las escuelas secundarias (Guilamo-Ramos, Jaccard, Turrisi y Johansson, 2005). Con el fin de replicar el estudio, tú entrevistas a estudiantes de escuelas de 18 distritos en relación con su conducta de consumo de bebidas alcohólicas. Tú mides las borracheras en la escuela como el porcentaje de estudiantes que indican que consumen cinco o más bebidas en una salida por lo menos una vez durante los pasados 12 meses. En seguida calculas el porcentaje de estudiantes en la escuela que lo hicieron.

a) Decide cuál es la variable independiente y cuál es la variable dependiente y traza el diagrama de dispersión de *Y* regresionada a partir de *X*.
b) Calcula los estadísticos de la recta de regresión y traza la recta.
c) Nota las coordenadas extremas del diagrama de dispersión. Identifícalas en la hoja de cálculo.
d) Vuelve a calcular la *r* de Pearson sin estas coordenadas extremas.
e) Comenta sobre el efecto de las coordenadas extremas en los cálculos del coeficiente de correlación *r* de Pearson.

Distritos con la escuela	Porcentaje de borracheras en la preparatoria	Porcentaje de borracheras en la secundaria
1	16	13
2	15	11
3	19	7
4	18	15
5	15	8
6	18	12
7	14	9
8	12	7
9	10	6
10	19	16
11	12	14
12	16	12
13	15	9
14	13	7
15	14	11
16	20	16
17	14	10
18	10	16

Conjunto de problemas 14C

Muestra todas las sumas y las fórmulas.

14C-1. Como director de personal de una compañía, es tu responsabilidad garantizar la equidad basada en el mérito de los niveles salariales. En otras palabras, el nivel salarial debería ser aproximadamente paralelo a los méritos educativos. En seguida aparecen los niveles salariales y los años de educación de 15 empleados que han laborado en la compañía cinco años. Utiliza las cifras donde los salarios han sido redondeados a miles de dólares.

a) Decide cuál es la variable independiente y cuál es la variable dependiente, y traza el diagrama de dispersión.
b) Calcula los estadísticos de la recta de regresión y traza la recta de regresión. Para facilitar las operaciones, calcula el salario en miles de dólares (por ejemplo, $22 500 = 22.5 miles de dólares).
c) Calcula el coeficiente de correlación bivariada *r* de Pearson.
d) ¿Parece que existe una relación lineal en el patrón de coordenadas?

Años de educación	Salario	Salario en miles de dólares
12	$22 500	22.5
12	17 900	17.9
11	16 500	16.5
16	29 600	29.6
16	34 500	34.5
18	42 600	42.6
17	45 800	45.8
16	24 000	24.0
12	22 300	22.3
10	14 000	14.0
12	13 700	13.7
19	54 000	54.0
18	34 000	34.0
14	25 000	25.0
13	21 400	21.4

14C-2. Supongamos que estudias el desempeño en la universidad, medido por el promedio de calificaciones (GPA). Tú utilizas una medida innovadora de horas de estudio por clase, equipando a los estudiantes con un contador electrónico del tiempo que se enciende y se apaga conforme los estudiantes estudian. En la siguiente tabla aparecen los datos que obtuviste.

a) Traza el diagrama de dispersión del promedio de calificaciones (Y) regresionado a partir de las horas de estudio por clase (X).
b) *Sin llevar a cabo ningún cálculo*, traza lo que parece ser la recta de mejor ajuste a través del conjunto de coordenadas.
c) *Sin llevar a cabo ningún cálculo*, calcula a y b en la ecuación de regresión $\hat{Y} = a + bX$.
d) Ahora calcula los estadísticos de regresión de la relación de horas de estudio con respecto al promedio de calificaciones y traza la recta de regresión.
e) Compara los estadísticos que calculaste con los que obtuviste en el inciso *c*) para asegurar de que entiendes lo que significa cada estadístico.

Estudiante	GPA	Calificación de lectura de comprensión	Horas de estudio por clase
1	0.90	2	2.00
2	1.12	3	1.25
3	1.46	2	2.50
4	1.93	4	5.00
5	2.00	3	3.00
6	2.16	3	4.15
7	2.18	2	4.72
8	2.21	3	5.75
9	2.33	4	6.13
10	2.39	6	4.75
11	2.46	5	4.25
12	2.54	6	4.84
13	2.68	5	5.75
14	2.73	4	6.00
15	2.85	7	5.89
16	2.87	6	6.37
17	2.93	8	6.50
18	2.99	7	7.00
19	3.04	9	6.00
20	3.14	8	6.93
21	3.22	9	7.16
22	3.27	9	6.94
23	3.28	8	8.10

14C-3. Los días de incapacidad se definen como el número de días que los individuos no pueden llevar a cabo sus actividades normales como consecuencia de una enfermedad o una lesión. Los días de incapacidad se relacionan con los riesgos en el trabajo y ambientes domésticos, los cuales, a su vez, se relacionan estrechamente con los niveles de ingreso. Los siguientes datos indican el patrón de la relación entre los días de incapacidad al año y el ingreso familiar (en miles de dólares).

a) Traza el diagrama de dispersión del ingreso familiar total (Y) regresionado a partir de los días de incapacidad (X).
b) Calcula los estadísticos de regresión y traza la recta de regresión.
c) Calcula el coeficiente de correlación r de Pearson.
d) ¿Parece que existe alguna relación lineal en el patrón de las coordenadas?

Ingreso familiar	Días de incapacidad
5	27
15	19
28	14
40	10
6	29
14	21
26	13
37	6

14C-4. Tú te encuentras estudiando el desarrollo de la habilidad de pianistas de 14 años de edad en un club y te preguntas si el tiempo de pertenencia al club (X) se relaciona con la cantidad de premios ganados (Y) en recitales de competencia.

a) Traza el diagrama de dispersión de Y regresionada a partir de X.
b) Calcula el coeficiente de correlación bivariada r de Pearson.
c) Nota las coordenadas extremas en el diagrama de dispersión. Identifícalas en la hoja de cálculo.
d) Vuelve a calcular la r de Pearson sin estas coordenadas extremas.
e) Comenta sobre el efecto de una coordenada extrema del coeficiente de correlación r de Pearson.

Años en el club de música	Premios y trofeos de mérito
4	5
6	6
2	1
3	4
2	7
1	2
3	3
5	4
4	4

Conjunto de problemas 14D

Muestra todas las sumas y las fórmulas.

14D-1. Se ha observado que el éxito educativo logrado a lo largo de los años de escolaridad es benéfico por el hecho de que abre oportunidades de empleo e ingresos. En el caso de los siguientes datos, analiza la relación entre los años de escolaridad y el ingreso.

a) Decide cuál es la variable independiente y cuál es la variable dependiente, y traza el diagrama de dispersión.
b) Calcula los estadísticos de la recta de regresión y traza la recta de regresión.
c) Calcula el coeficiente de correlación bivariada r de Pearson.
d) ¿Parece que existe una relación lineal en el patrón educación e ingresos?

Años de escolaridad	Ingreso (en miles)
12	29
16	45
15	37
16	39
18	51
15	35
12	31
12	33
11	27
15	40
14	36
16	45
15	38
20	55
12	31
16	45
10	27

14D-2. Ante la curiosidad relacionada con el sonido de los discos en vinil y el hecho de que la gente de la generación de tus padres todavía las escuche, tú reúnes los siguientes datos tomados de una muestra de 20 adultos.

a) Traza el diagrama de dispersión de los discos de vinil que se poseen (Y) regresionado a partir de la edad (X).
b) *Sin llevar a cabo ningún cálculo*, traza lo que parecería ser la recta a través del conjunto de coordenadas.
c) *Sin llevar a cabo ningún cálculo*, calcula a y b en la ecuación de regresión $\acute{Y} = a + bX$.

d) Ahora calcula los estadísticos de regresión para la relación del número de discos de vinil poseídos y la edad, y traza la recta de regresión (para este problema no se emplearán los datos de discos compactos).
e) Compara los estadísticos que calculaste con los que obtuviste en el inciso *c)* con el fin de que te asegures de comprender lo que cada estadístico significa.

Adulto	Edad	Discos compactos poseídos	Discos de vinil poseídos
1	36	12	16
2	51	6	25
3	32	15	18
4	34	10	20
5	28	15	16
6	27	18	15
7	44	10	19
8	45	8	24
9	29	18	23
10	50	7	27
11	49	9	26
12	37	10	20
13	47	9	23
14	44	12	20
15	41	9	22
16	32	15	21
17	51	6	25
18	33	18	23
19	28	16	18
20	44	12	23

14D-3. Las investigaciones han encontrado que el estado de salud que uno percibe a menudo constituye un buen predictor del estado real de salud. De acuerdo con los datos ficticios de la siguiente tabla determina si éste es el caso. En el caso de las dos variables, el estado de salud se mide en una escala cuyo rango va de 10 (salud muy deficiente) a 45 (muy buena salud).

a) Traza el diagrama de dispersión del estado real de salud (Y) regresionado a partir del estado de salud percibido (X).
b) Calcula los estadísticos de la recta de regresión y traza la recta de regresión.
c) Calcula el coeficiente de correlación bivariada r de Pearson.
d) ¿Parece que existe una relación lineal en el patrón de coordenadas?

Estado de salud percibido	Estado de salud real
12	10
15	13
17	21
19	22
23	22
25	28
27	32
29	33
31	30
33	29
35	31
37	38
39	39
41	40
43	41
45	44

14D-4. Un estudio reciente concluyó que la exposición a una campaña contra el cigarro predijo tasas más bajas de fumadores jóvenes (Farrelly, Davis, Haviland, Messeri y Healton, 2005). Tú replicas el estudio con datos tomados de 20 países seleccionados aleatoriamente. La exposición a los anuncios contra el cigarro (X) se mide por las horas de transmisión por mes. La prevalencia del tabaquismo (Y) se mide como el porcentaje de jóvenes entre los 13 y 17 años que reportaron haber fumado el mes pasado.

a) Traza el diagrama de dispersión de Y regresionada a partir de X.
b) Calcula el coeficiente de correlación bivariada r de Pearson.
c) Observa las coordenadas extremas en el diagrama de dispersión. Identifícalas en la tabla de datos.
d) Vuelve a calcular la r de Pearson sin estas coordenadas extremas.
e) Comenta sobre el efecto de las coordenadas extremas sobre el cálculo del coeficiente de correlación r de Pearson.

Porcentaje de jóvenes en el país que fuman	Horas de transmisión de anuncios
14	4
12	16
13	18
15	13
20	10
14	16
24	8
26	19
19	9
17	11
15	13
14	17
19	12
18	11
21	9
14	16
23	8
20	10
17	16
13	18

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 14

Si en tu clase se utilizan las aplicaciones opcionales en computadora que acompañan el texto, abre los ejercicios del capítulo 14 localizados en el sitio web de *The Statistical Imagination*, www.mhhe.com/ritchey2. Estos ejercicios le dan énfasis en la comprensión de la relación entre la forma de un diagrama de dispersión y el tamaño de la r de Pearson con instrucciones para generar diagramas de dispersión y estadísticos de correlación y regresión. Además, el apéndice D del texto contiene una vista rápida de las secuencias de los comandos del *SPSS* para los procedimientos cubiertos en este capítulo.

CAPÍTULO

15

Correlación y regresión bivariadas

Parte 2: Prueba de hipótesis y aspectos de una relación

RESUMEN DEL CAPÍTULO

Introducción: prueba de hipótesis y aspectos de una relación entre dos variables de intervalo/razón

En el capítulo 14 introdujimos las ideas básicas que subyacen el análisis de correlación y regresión bivariadas y el cálculo de los coeficientes. En este capítulo utilizaremos dichos estadísticos para probar hipótesis sobre la relación entre dos variables de intervalo/razón.

Para ilustrar la prueba de hipótesis, examinamos la relación entre el nivel de educación y la aceptación de la música *inculta* —estilos relacionados con la protesta política, las clases bajas y las minorías étnicas y raciales (Bryson, 1996)—. Este tipo de música contrasta con la música *culta*, como la música sinfónica clásica, de cámara y la ópera; el swing; las melodías para espectáculos; la música fácil de escuchar; el *soft rock* y la música popular que toca temas sexuales. Los estilos de música que favorece una sociedad son considerados por algunos individuos como símbolos del carácter moral de una sociedad. Una pequeña pero ruidosa minoría considera a la música inculta como una forma de desviación y clara señal de decadencia moral. Aún la clase media en la sociedad norteamericana siente atracción por formas musicales que tienen elementos contraculturales que desafían a los sistemas social,

moral y político dominantes. Algunos padres temen que el heavy metal y la música *rap*, políticamente cargada, corromperían la moral de sus hijos. Otros estilos de música popular —relacionados con las clases bajas y minorías étnicas— suscitan prejuicios clasistas y sentimientos racistas entre algunos. Dichos estilos incluyen el rap, reggae, blues, *rhythm and blues*, *rock'n'roll*, rock alternativo, heavy metal, country, ritmos latinos, jazz contemporáneo y música cristiana contemporánea no tradicional.

¿A qué se deben las diferencias sobre la manera en que la gente evalúa y acepta diferentes formas de música? Entre las variables predictoras normalmente analizadas se encuentran la edad, raza, género, estado civil, opiniones políticas, filiación religiosa y nivel educativo (Bryson, 1996, Peterson y Kern, 1996). Una hipótesis de investigación consiste en que las personas mejor educadas aceptan más —son más tolerantes— los estilos de música inculta.

Organización de los datos para la prueba de hipótesis

Supongamos que investigamos esta hipótesis utilizando los datos recogidos por medio de un cuestionario aplicado a una muestra de 12 madres de estudiantes de preparatoria a quienes se les solicitó que evaluaran diversos estilos musicales. Posteriormente se sumó el número de estilos de música inculta que cada madre tolera (es decir, que considera aceptable para sus hijos adolescentes), y a esta variable se le llamó *tolerancia con respecto a la música inculta*. A mayor puntuación, más favorable será la postura de la madre con respecto a los estilos de música inculta. El nivel educativo de las madres se definió como el número de años de escolaridad formal concluidos.

Sigamos el recuadro de cálculo paso a paso localizado al final del capítulo 14. Comenzamos determinando cuál es la variable independiente y cuál es la variable dependiente. En este caso estamos interesados en explicar la tolerancia con respecto a la música inculta por parte de las madres; por consiguiente, ésta es la variable dependiente, que se designará Y. Hipotetizamos que el nivel educativo es un predictor de la tolerancia con respecto a la música inculta. De esta manera, el nivel educativo es la variable independiente X.

En seguida organizamos los datos en una hoja de cálculo. La tabla 15-1 presenta los datos relativos al nivel educativo y la tolerancia respecto a la música inculta, junto con las sumas que se requieren para calcular los estadísticos de correlación y regresión.

Antes de calcular los estadísticos de regresión lineal, debemos asegurarnos de que dicho procedimiento se aplica a estos datos; es decir, que debemos observar el diagrama de dispersión para determinar si los datos parecen ajustarse a un patrón en forma de cigarro que se aproxima a una recta. La figura 15-1 muestra el diagrama de dispersión de los datos de la tabla 15-1. De hecho, los datos parecen formar un patrón de línea recta. Además, la pendiente es positiva; es decir que las coordenadas del diagrama de dispersión se prolongan hacia arriba de izquierda a derecha. Por lo tanto, procedemos a calcular los estadísticos de correlación y regresión.

Primero calculamos $\bar{X}$ y $\bar{Y}$:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{150}{12} = 12.50 \text{ años}$$

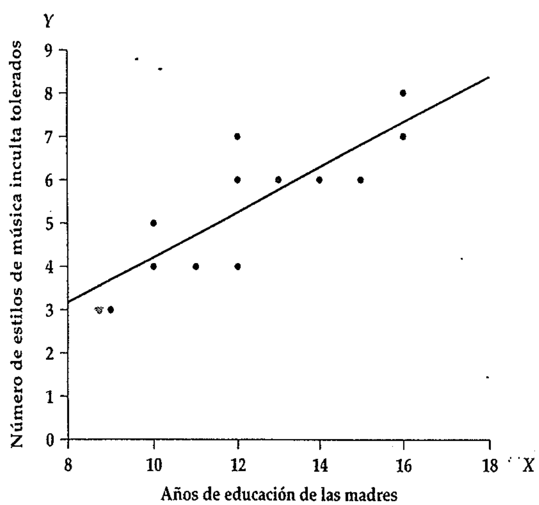
$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{66}{12} = 5.50 \text{ estilos}$$

TABLA 15-1 | Hoja de cálculo para calcular los estadísticos de correlación y de regresión: niveles educativos y tolerancia respecto a la música inculta de 12 madres de estudiantes de preparatoria (X = nivel educativo en años de escolaridad, Y = número de estilos de música inculta tolerados)

Especificaciones			Cálculos				
Madre	**X**	**Y**	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
1	9	3	-3.5	-2.5	8.75	12.25	6.25
2	10	4	-2.5	-1.5	3.75	6.25	2.25
3	10	5	-2.5	-0.5	1.25	6.25	.25
4	11	4	-1.5	-1.5	2.25	2.25	2.25
5	12	4	-0.5	-1.5	.75	.25	2.25
6	12	6	-0.5	0.5	-.25	.25	.25
7	12	7	-0.5	1.5	-.75	.25	2.25
8	13	6	0.5	0.5	.25	.25	.25
9	14	6	1.5	0.5	.75	2.25	.25
10	15	6	2.5	0.5	1.25	6.25	.25
11	16	7	3.5	1.5	5.25	12.25	2.25
12	16	8	3.5	2.5	8.75	12.25	6.25
$n = 12$	$\Sigma X = 150$	$\Sigma Y = 66$	$\Sigma(X-\bar{X}) = 0$	$\Sigma(Y-\bar{Y}) = 0$	$\Sigma(X-\bar{X})(Y-\bar{Y}) = 32$	$\Sigma(X-\bar{X})^2 = 61$	$\Sigma(Y-\bar{Y})^2 = 25$

FIGURA 15-1

Diagrama de dispersión del nivel educativo como predictor de la tolerancia con respecto a la música inculta entre 12 madres de estudiantes de preparatoria



En seguida utilizamos las medias de X y Y para calcular las puntuaciones de desviación y las sumas de cuadrados para completar los cálculos de la tabla 15-1. Al emplear las sumas de cuadrados de la tabla 15-1, entonces podemos calcular el coeficiente de correlación de Pearson r:

$$r = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2\Sigma(Y-\bar{Y})^2}} = \frac{32}{\sqrt{(61)(25)}} = \frac{32}{39.05} = .82$$

En seguida calculamos el coeficiente de regresión, b:

$$b = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} = \frac{32}{61} = .52 \text{ estilo tolerado por año de educación}$$

A continuación utilizamos $\bar{X}$, $\bar{Y}$ y b para calcular a:

$$a = \bar{Y} - b\bar{X} = 5.50 - (.52)(12.50) = -1.00 \text{ estilos}$$

Ahora especificamos la ecuación de la recta de regresión para calcular $\acute{Y}$ (es decir, los valores predichos de Y) sustituyendo los valores calculados de a y b:

$$\acute{Y} = a + bX = -1.00 + (.52)X$$

Finalmente, sustituimos unos cuantos valores de X en la ecuación de regresión y resolvemos para $\acute{Y}$. Es mejor utilizar valores de X bajos, medios y altos.

X	Ý
9	3.68
12	5.24
16	7.32

Por consiguiente, la mejor estimación de la tolerancia a la música inculta de una madre que cuenta con 9 años de educación consiste en la aceptación de 3.68 estilos de música; en el caso de una madre con 12 años de educación, 5.24 estilos de música; etc. Trazamos la recta de regresión en el diagrama de dispersión utilizando las coordenadas resultantes X, $\acute{Y}$ (figura 15-1).

Los seis pasos de la inferencia estadística y los cuatro aspectos de una relación

El coeficiente de correlación r de Pearson es el estadístico que utilizamos para probar la hipótesis de la existencia de una relación entre dos variables de intervalo/razón, una variable independiente X y una variable dependiente Y. Los siguientes criterios deben satisfacerse para que podamos utilizar los estadísticos de correlación y regresión lineal bivariados:

Cuándo probar una hipótesis utilizando el análisis de correlación y regresión bivariados (distribución t, $gl = n - 2$)

En general: se prueba una hipótesis relacionada con el hecho de que existe una relación entre dos variables de intervalo y de razón.

1. Existe una muestra representativa a partir de una sola población.
2. Hay dos variables de intervalo/razón.
3. No hay restricciones en cuanto al tamaño de la muestra, pero, generalmente, cuanto mayor es n, mejor.
4. Un diagrama de dispersión de las coordenadas de las dos variables se ajusta a un patrón lineal.

Existencia de una relación

En nuestro ejemplo de las madres de los estudiantes de preparatoria, utilizamos una muestra de 12 mujeres. Aunque los datos del diagrama de dispersión de la figura 15-1 parecen lineales, es posible que estas coordenadas de la muestra representen escasamente las coordenadas *para la población* de todas las madres de estudiantes de preparatoria. Por supuesto que el patrón lineal de la muestra puede ser simplemente el resultado del error de muestreo. Especialmente con dicha muestra pequeña, extraer una segunda, tercera, cuarta, quinta y sexta muestras producirá diferentes diagramas de dispersión cada ocasión. Algunos parecerán lineales; otros, no.

Cuando utilizamos datos muestrales, entonces debemos plantear la pregunta: ¿en realidad existe una relación entre X y Y *en la población*, o el patrón lineal en esta muestra es resultado del error de muestreo? En términos prácticos, estamos interesados en la pregunta: en el caso de *todas* las mujeres, ¿existe una relación entre el nivel educativo y la tolerancia de la música inculta, o es cuestión de azar el aparente patrón lineal en nuestra muestra? Como en el caso de cualquier prueba de hipótesis, el verdadero interés se centra en el parámetro, la medición sumaria que se aplica a la población.

El coeficiente de correlación r de Pearson nos permite probar una hipótesis para responder esta pregunta. La r de Pearson es un estadístico, una medida de lo ceñido del ajuste de las coordenadas en torno a la recta de regresión *para la muestra. Para la población*, el parámetro correspondiente se representa mediante la letra griega *rho* (ρ). *Rho* es el coeficiente de correlación que se obtendría si el coeficiente de correlación de Pearson se calculara para toda la población. Éste mediría lo ceñido del ajuste de las coordenadas X, Y si se les graficara en el caso de todas las mujeres, no sólo las de la muestra.

Como lo indicamos en el capítulo 14, el coeficiente de correlación de Pearson es igual a cero cuando no existe relación entre X y Y. A partir de esto, podemos formular una hipótesis nula —una que nos diga qué esperar de los cálculos estadísticos en el muestreo repetido cuando este enunciado es verdadero—. Si no existe relación entre el nivel de educación y la tolerancia a la música inculta *en la población de madres*, entonces *rho* es igual a cero y las r de Pearson muestrales serán iguales a cero, con un pequeño error de muestreo. La hipótesis alternativa consiste en que rho *no* es cero. La prueba de hipótesis gira en torno a que la r de Pearson de la muestra es *significativamente diferente* de cero.

Como en el caso de la prueba de hipótesis, el *efecto* de la prueba es la diferencia entre un estadístico de una muestra observada y el parámetro esperado cuando la hipótesis nula es verdadera. En el caso de una hipótesis de correlación, el efecto es la diferencia entre la r de Pearson de la muestra observada y la *rho* esperada de cero. Esta diferencia calcula el valor de r:

$$\text{Efecto de la prueba para una correlación de Pearson} = r - \rho = r - 0 = r$$

La prueba de hipótesis determina si este efecto encontrado en una muestra es real para la población. ¿Es el valor absoluto de r muestral tan grande que nos vemos obligados a creer que ρ no es cero y que este efecto no es simplemente resultado del error de muestreo?

En general, para cualquier hipótesis sobre la relación entre dos variables de intervalo/razón, la hipótesis nula se enuncia de la siguiente manera:

$$H_0\text{: } \rho = 0$$

Es decir que no existe una relación entre X y Y.

El enunciado de la hipótesis alternativa puede ser de dos colas, no direccional (es decir que $\rho \neq 0$; existe una relación), de una cola en la dirección negativa (es decir, $\rho < 0$; *existe* una relación negativa) o de una cola en la dirección positiva (es decir que $\rho > 0$; *existe* una relación positiva).

Existencia de una relación entre dos variables de intervalo/razón

Prueba la hipótesis nula de que no existe relación entre X y Y:

$$H_0\text{: } \rho = 0$$

Es decir que *no* existe relación entre X y Y.

En el paso 2, de los seis pasos de la inferencia estadística, proyectamos el tamaño del error de muestreo describiendo la distribución de muestreo. En este caso, si rho es, de hecho, igual a cero y tomamos repetidamente muestras de tamaño 12 de la población de madres, ¿qué correlaciones muestrales (r) obtendremos? La r de Pearson se centrará en cero aproximadamente como una distribución t normal. El cálculo del error estándar para esta prueba resulta muy engorroso. Por fortuna, el estadístico de la prueba se encuentra diseñado de tal manera que el cálculo del error estándar es innecesario. Sin embargo, podemos decir que el error estándar se encuentra relacionado de forma inversa al tamaño de la muestra. Es decir que cuanto más grande sea el tamaño de la muestra, menor será el error estándar. El estadístico de la prueba para esta prueba de hipótesis es el siguiente:

Fórmula de la prueba *t* para probar la significancia del coeficiente de correlación bivariada *r* de Pearson

$$t_r = r\sqrt{\frac{n-2}{1-r^2}}$$

con $gl = n - 2$,

donde

t_r = prueba t para el coeficiente de correlación r de Pearson

r = coeficiente de correlación r de Pearson calculado sobre una muestra

n = tamaño de la muestra

gl = grados de libertad

Procedamos a probar la hipótesis nula de no relación entre X (nivel educativo) y Y (tolerancia de la música inculta). Continuaremos y completaremos los cuatro aspectos de una relación; en seguida analizaremos los detalles de cada paso y cada aspecto.

Breve lista de verificación de los seis pasos de la inferencia estadística

PREPARACIÓN DE LA PRUEBA

Formula la pregunta de investigación. Traza diagramas conceptuales que describan especificaciones, incluyendo poblaciones y muestras bajo estudio, variables (por ejemplo, $X = ...$, $Y = ...$,) y sus niveles de medición, así como los estadísticos y parámetros dados o calculados. Establece el procedimiento de la prueba estadística adecuado.

SEIS PASOS

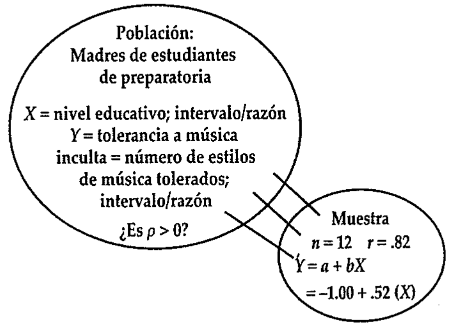
Utilizando el símbolo H para representar la *hipótesis*:

1. Enuncia la H_0 y la H_A y estipula la dirección de la prueba.
2. Describe la distribución muestral.
3. Determina el nivel de significancia (α) y especifica el valor crítico de la prueba.
4. Observa los resultados de la muestra en cuestión y calcula los aspectos de la prueba, el estadístico de la prueba y el valor p.
5. Toma la decisión de rechazo.
6. Interpreta y aplica las mejores estimaciones en términos comunes.

Solución para la prueba de hipótesis de una relación entre dos variables de intervalo o de razón (prueba *t* de la *r* de Pearson

PREPARACIÓN DE LA PRUEBA

Pregunta de investigación: ¿existe una relación entre el nivel educativo y la tolerancia con respecto a la música inculta entre las madres de estudiantes de preparatoria? *Procedimiento estadístico:* prueba t para la significancia del coeficiente de correlación r de Pearson; distribución t. El diagrama de dispersión (figura 15-1) sugiere una relación lineal; los datos y cálculos de la hoja de cálculo se incluyen en la tabla 15-1.



SEIS PASOS

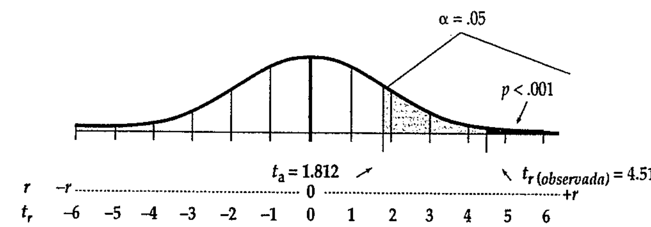
1. H_0: $\rho_{(\text{madres de estudiantes de preparatoria})} = 0$

   Es decir que *no* existe relación entre el nivel educativo y la tolerancia con respecto a la música inculta.

   H_A: $\rho_{(\text{madres de estudiantes de preparatoria})} > 0$

   Es decir que *existe* una relación positiva entre el nivel educativo y la tolerancia a la música inculta. De una cola.

2. *Distribución muestral:* distribución t aproximadamente normal, $gl = n - 2 = 10$. Si H_0 es verdadera y se extraen repetidamente muestras de tamaño 12 de la población de madres de preparatoria, las r de las muestras se centrarán en torno a cero con un error estándar relacionado inversamente con el tamaño de la muestra (es decir, que a mayor tamaño de la muestra, menor el error estándar).



3. *Nivel de significancia:* $\alpha = .05$, de una cola; valor crítico $t_\alpha = 1.812$ [de la tabla de la distribución t (tabla estadística C, apéndice B)]. (Área sombreada sobre la curva.)

4. *Observación:*

   Efecto de la prueba: $= .82$ (es decir, efecto $= r - \rho = r - 0 = .82 - 0 = .82$).

   Estadístico de la prueba:

   $$t_r = r\sqrt{\frac{n-2}{1-r^2}} = .82\sqrt{\frac{10}{1-.67}} = 4.51 \text{ EE}$$

   De la tabla t (tabla estadística C, apéndice B):

   valor p: p [de observar una r tan inusual o más inusual que .82 cuando $\rho = 0$] $< .001$ (el área señalada en la curva en el paso 2).

5. *Decisión de rechazo*: $|t_r| > |t_\alpha|$ (es decir, $4.51 > 1.812$); por lo tanto, $p < \alpha$ (es decir, $p < .05$). Se rechaza H_0 y se acepta H_A en el nivel de confianza del 95%.

6. *Interpreta los resultados* (analizando los cuatro aspectos de una relación) y lleva a cabo las mejores estimaciones (los detalles se analizan en seguida).

   Existencia: existe relación entre el nivel educativo y la tolerancia a la música inculta entre madres de estudiantes de preparatoria; $r = .82$, $p < .001$.

Dirección: positiva. Conforme el nivel educativo aumenta, el número de estilos musicales tolerados tiende a incrementarse.

Fuerza: $r^2 = .82^2 = .6724$; $(100)(.6724) = 67.24\%$; por consiguiente, 67.24% de la variación en el número de estilos musicales tolerados se explica por el conocimiento del nivel educativo.

Aplicaciones prácticas: (interpreta la pendiente de la recta de regresión): $b = .52$ estilos por año de educación; un año de incremento en el nivel educativo relaciona con un .52 de incremento en el número de estilos musicales tolerados.

Mejores estimaciones (utiliza las coordenadas $X, \acute{Y}$ para ilustrar la utilidad del conocimiento sobre la relación entre el número de estilos musicales tolerados):

$$\acute{Y} = a + bX = -1.00 + (.52)X$$

Por consiguiente, la mejor estimación del número de estilos musicales tolerados por madres con un nivel educativo de preparatoria (12 años) es de 5.24 estilos y en el caso de una educación universitaria (16 años) es de 7.32 estilos.

Haremos más comentarios sobre la fuerza, dirección y aplicaciones prácticas a continuación. Primero observemos algunas cuestiones que tienen que ver con esta prueba de hipótesis:

- En el paso 1 utilizamos una prueba de una cola, pero *no* porque hayamos visto una relación positiva en el diagrama de dispersión. El patrón podría haber ocurrido por el azar. Hicimos esto porque incluso antes de que la muestra se extrajera, planteamos la pregunta de investigación relativa al hecho de que las madres *mejor* educadas eran más tolerantes con respecto a los diversos estilos musicales.
- En el paso 2, como lo observamos antes, por razones de simplicidad en los cálculos, utilizamos la prueba estadística t_r. Esto nos evita tener que calcular un error. Sin embargo, una vez que se ha calculado t_r, el error estándar se puede calcular con relativa facilidad. Ahora sabemos que la cantidad r de Pearson de .82 es de 4.51 errores estándares a partir de una rho de cero. Por consiguiente, la magnitud del error estándar es

$$s_{t_r} = \frac{.82}{4.51} = .18$$

donde s_{t_r} es el error estándar calculado de la distribución t para la r de Pearson (calculado porque se basa en la muestra r). Introduce este valor en la curva, y su ajuste resultará evidente. Además, como los estadísticos de la regresión se basan en desviaciones a partir de la media de Y, precisamente como en el caso de los estadísticos del análisis de varianza (ANOVA), podríamos haber utilizado una prueba de razón F. Sin embargo, esto requeriría más cálculos.
- Observa que en el paso 2, en la curva de distribución t, las r muestrales cercanas a cero se presentan con frecuencia cuando la hipótesis nula es verdadera y las r muestrales grandes se presentan con poca frecuencia. Es decir que si, de hecho, rho es cero, los valores absolutos grandes de r son inusuales; por consiguiente, cuando se presenta una r significativamente grande, rechazamos la suposición de que ρ es cero.

- En el paso 2 se pierden 2 grados de libertad porque la r de Pearson se calcula a partir de las medias y las varianzas muestrales de X y Y. El cálculo de cada media da como resultado 1 grado de libertad (véase capítulo 10).
- En lo que se refiere a la existencia de la correlación, esta prueba de hipótesis sencillamente establece si existe una correlación entre X y Y. Si se concluye que esto es verdad (es decir, *si rechazamos* la hipótesis nula que indica que $\rho = 0$), es posible analizar los demás aspectos de la relación. Sin embargo, *si no rechazamos la hipótesis nula, los demás aspectos resultan irrelevantes.* Simplemente concluimos que no existe relación y no mencionamos la dirección, la fuerza ni las aplicaciones prácticas. Ya que rechazamos la hipótesis nula, ahora analizaremos la dirección, la fuerza y las aplicaciones prácticas de una relación para dos variables de intervalo/razón.

Dirección de la relación

La dirección de una relación entre dos variables de intervalo/razón se determina por medio del signo de la r y b, la pendiente de la recta de regresión. Una pendiente ascendente es positiva; una pendiente descendente es negativa. La dirección de la pendiente para una relación realmente lineal y moderadamente fuerte será evidente en el diagrama de dispersión. La dirección de la relación se puede observar directamente en los signos (+ o –) de b y la r de Pearson, los cuales siempre serán iguales. Como parte de nuestra investigación en el paso 6 de la prueba de hipótesis, la dirección de la relación para el nivel educativo y la tolerancia con respecto a la música inculta se revela en el signo de r, con $r = +0.82$. Por lo tanto, la dirección de la relación es positiva y se describe de la siguiente manera:

Positiva: conforme se incrementa el nivel educativo, el número de estilos musicales tolerados tiende a incrementar.

La dirección de una relación entre dos variables de intervalo/razón Observa el signo (+ o –) de la b y la r de Pearson. También observa la pendiente (o inclinación) del patrón de coordenadas en un diagrama de dispersión.

Fuerza de la relación

La correlación del nivel educativo con la tolerancia con respecto a la música inculta indica que una parte de las diferencias en la cantidad de estilos musicales tolerados (Y) por las madres se debe a las diferencias en sus niveles educativos —los efectos de X—. En el capítulo 14 observamos que cuando esas diferencias son sistemáticas y las coordenadas X, Y en un diagrama de dispersión obedecen un patrón lineal, es posible superar las mejores estimaciones de Y determinando la recta de regresión en lugar de la *línea media.* La recta de regresión es aquella que se encuentra tan cerca como se puede de todas las coordenadas X, Y. Al predecir la recta de regresión, reducimos el error en las predicciones.

Apliquemos dichas ideas a la población de madres de estudiantes de preparatoria. De acuerdo con los datos de la tabla 15-1, encontramos que la cantidad media de estilos de música inculta tolerados por estas madres es de 5.50 estilos. Si se nos pidiera calcular dicha cantidad en el caso de cualquier madre *sin más información sobre ella*, nuestra mejor estimación sería de 5.50: una predicción en la dirección de la línea media del diagrama de dispersión. ¿Cómo sería esta información *sin* más información? La madre 12 tolera ocho estilos musicales incultos. Si hubiéramos calculado su puntuación Y hacia la media de 5.50, habríamos errado por 2.50 estilos. Según el capítulo 5, recuerda que esta cantidad —la diferencia entre una puntuación y la media de las puntuaciones— es su puntuación de desviación:

Puntuación de desviación de la madre 12 = error en el cálculo hacia la media

$$= Y_{(madre\ 12)} - \bar{Y} = 8 - 5.50 = 2.50 \text{ estilos musicales}$$

Supongamos ahora que seguimos un curso de acción diferente. En este escenario tenemos la oportunidad de predecir su nivel de tolerancia con respecto a la música inculta sabiendo que el nivel educativo y la tolerancia se encuentran correlacionados. En lugar de predecir en la dirección de la media, predecimos *hacia la recta de regresión* (figura 15-1), que sabemos que se localiza cerca de las coordenadas *X*, *Y* observadas. La mejor estimación de los estilos tolerados para la madre 12, con sus 16 años de educación, ahora se calcula con mayor precisión por medio de la recta de regresión. Sobre la base de sus 16 años de educación, su puntuación $\acute{Y}$ predicha de estilos musicales es de 7.32, se calcula de la siguiente manera:

$$\acute{Y}_{(X=16\ años)} = a + bX = -1.00 + (.52)16 = 7.32 \text{ estilos musicales}$$

Utilizando este resultado como una *estimación posible de conocer* —una coordenada en la recta de regresión— ahora calculamos el nivel de tolerancia de la madre 12 mucho más cerca de su nivel de tolerancia real. El error en esta estimación posible de conocer es de solamente .68 estilos:

Error al calcular el nivel de tolerancia musical de la madre 12 al predecir hacia la recta de regresión $= Y_{(madre\ 12)} - \acute{Y}_{(X=16\ años)} = 8 - 7.32 = .68$ estilos musicales

Esta estimación constituye una mejora considerable con respecto a la que se llevó a cabo en la dirección de la media. El error se redujo de 2.50 a .68 estilos. Para determinar la magnitud de la mejora, calculamos la diferencia entre la estimación posible de conocer ($\acute{Y}$ predicha para 16 años de educación) y la estimación sin más información (es decir, $\bar{Y}$, la media de *Y*):

Mejora en la estimación de *Y* utilizando la información que ésta se encuentra relacionada con $X = \acute{Y}_{(X=16\ años)} - \bar{Y} = 7.32 - 5.50 = 1.82$ estilos musicales

Esta reducción de 1.82 estilos en el error es la parte del error *explicado* por el nivel educativo. Al predecir en la dirección de la línea de regresión explicamos 1.82 de la puntuación de desviación de 2.50 de la madre 12. Atribuimos los 1.82 estilos por encima de la media de 5.50 a su mayor nivel educativo.

La fuerza de la relación entre dos variables de intervalo/razón constituye una medida del grado de mejora en las estimaciones de la muestra (y la población) en conjunto. En el caso de un estadístico sumario que se aplique a toda la muestra, debemos calcular las mejores estimaciones de cada caso y comparar estas estimaciones posibles de conocer con las estimaciones sin más información sobre la base de la media. Una medida de la fuerza total para la muestra debe evaluar el grado del error total —la suma de las puntuaciones de desviación— que se puede eliminar utilizando la recta de regresión en lugar de la línea media como base para predecir la tolerancia con respecto a la música inculta.

Ahora bien, se requiere un paso adicional al llevar a cabo estos cálculos sumarios. No es posible sumar sencillamente las puntuaciones de desviación, ya que siempre suman cero. De esta manera, se requiere que las puntuaciones de desviación se eleven al cuadrado. Recuerda que la suma de las puntuaciones de desviación al cuadrado para una muestra completa recibe el nombre de *variación total*:

$$\text{Variación total en } Y = \Sigma(Y_{(cada\ caso)} - \bar{Y})^2$$

La variación es la cantidad total de desviaciones de la media elevadas al cuadrado *que es necesario explicar.* ¿Por qué tolera una madre dos estilos musicales por encima de la media, mientras otra tolera un estilo por debajo de ésta? Si la relación entre el nivel educativo y la tolerancia de estilos musicales incultos es fuerte, el nivel educativo explicará una buena *proporción de la variación total* de la tolerancia de la música inculta. Una buena parte de la puntuación de desviación de la tolerancia de la música inculta de cada individuo podría atribuirse a su nivel educativo. De esta manera, debemos determinar *qué proporción de la variación total en Y puede explicarse por X*.

Llevemos a cabo el cálculo del cuadrado en el caso de la madre 12. De nuevo hallamos que su puntuación de desviación es de 2.50 estilos. Su desviación elevada al cuadrado es de $(2.50)^2 = 6.25$. ¿Qué proporción de esta desviación al *cuadrado* se explica por la educación de la madre?

Cantidad de desviación al cuadrado de la madre 12 por su nivel educativo

$$= \frac{\text{mejora al cuadrado utilizando la estimación posible de conocer}}{\text{puntuación de desviación al cuadrado}}$$

$$= \frac{(\acute{Y}_{(X=16\ años)} - \bar{Y})^2}{(Y_{(madre\ 12)} - \bar{Y})^2} = \frac{1.82^2}{2.50^2} = .5296$$

Por consiguiente, la proporción de la desviación *al cuadrado* de la madre 12 en lo que se refiere a tolerancia con respecto a la música inculta (*Y*) explicada por sus 16 años de educación (*X*) es de .5296 o 52.96%.

Para obtener la cantidad total de la variación en *Y* explicada por *X* en el caso de toda la muestra, efectuamos estos cálculos para cada individuo de la muestra y sumamos los cálculos. Éste es un proceso engorroso, aunque, afortunadamente, existe un método abreviado para obtener la proporción de la variación total en *Y* explicada por *X*.

Método abreviado para calcular la fuerza de la relación Matemáticamente, la proporción de la variación total en *Y* explicada por *X* puede obtenerse rápidamente elevando al cuadrado el coeficiente de correlación *r* de Pearson. Es decir que el estadístico r^2 permite obtener rápidamente esta proporción y constituye una medida de la fuerza de la relación entre *X* y *Y*.

$$r^2 = \text{proporción de la variación en } Y \text{ explicada por } X$$

$$= \frac{\Sigma(\acute{Y}_{(cada\ caso)} - \bar{Y})^2}{\Sigma(Y_{(cada\ caso)} - \bar{Y})^2}$$

Fuerza de una relación entre dos variables de intervalo o de razón

r^2 = proporción de la variación en *Y* explicada sabiendo que se relaciona con *X*

En el caso de la hipótesis referente al nivel educativo y la tolerancia respecto a la música inculta, $r^2 = (.82)^2 = .6724$; $(100)(.6724) = 67.24\%$. De esta manera 67.24% de la variación del número de estilos musicales tolerados se explica por el conocimiento de los niveles educativos.

Cuando existe una fuerte relación entre dos variables de intervalo/razón, las coordenadas X, Y del diagrama de dispersión se ajustarán estrechamente en torno a la recta de regresión. Cuanto más ceñido sea el ajuste, mayor será el valor de la r de Pearson y, por consiguiente, mayor será el valor de r^2. Y cuando el ajuste es ceñido, las mejores estimaciones hacia la recta de regresión se encuentran cerca de las puntuaciones Y observadas, como sucedió en el caso de la madre 12. El error en las predicciones de Y, tomando en cuenta su relación con X, será pequeño y la reducción proporcional del error (RPE) será grande. Esto significa que las puntuaciones X son buenas y fuertes predicciones de las puntuaciones Y. En otras palabras, la relación es fuerte.

Enfoque en r^2, no en r Al interpretar la fuerza de una relación, nos enfocamos en r^2 más que en r, ya que ambos, el signo de r así como su tamaño, pueden inducir a error. El signo de r no tiene nada que ver con la fuerza de la relación; solamente indica la dirección de la relación. Además, si la relación es positiva o negativa, al elevar r al cuadrado siempre resulta una *proporción de la variación explicada* positiva, ya que esta operación elimina cualquier signo negativo. Con respecto al tamaño de r, elevar al cuadrado revela la verdadera fuerza de una relación —la cantidad de RPE al predecir la variable dependiente Y debida a su correlación con la variable independiente X—.

Observar directamente r (es decir, sin elevarlo al cuadrado) estimula a exagerar la fuerza de la relación, lo cual se refleja en la tabla 15-2. Por ejemplo, la observación directa de $r = .50$ podría conducirnos a la conclusión incorrecta de que estamos *a medio camino* en la reducción de los errores de predicción. De hecho, nos encontramos a un cuarto de camino, porque $r^2 = (.50)^2 = .25$; es decir, sólo 25% de la variación en Y se explica por X. Recuerda, es r^2, no r, la que en realidad indica la fuerza de una relación entre dos variables de intervalo/razón.

Modelo lineal general aplicado al análisis de correlación y regresión El examen de la cantidad de variación explicada debería resultar familiar a quienes han leído el material del capítulo 12 relativo al ANOVA. Ahí utilizamos el modelo lineal general por medio de la descomposición de las puntuaciones Y de una muestra en sus partes explicadas por la media de Y, los efectos de X, y los efectos de otras variables no medidas (el error restante). Se puede aplicar el mismo enfoque al análisis de regresión. Por ejemplo, podemos descomponer la puntuación de tolerancia a la música inculta de la madre 12 en las partes explicadas y las no explicadas por el nivel educativo:

$$\begin{aligned} Y_{(madre\ 12)} &= \text{media de } Y + \text{puntuación de desviación de la madre 12} \\ &= \bar{Y} + (Y_{(madre\ 12)} - \bar{Y}) \\ &= \bar{Y} + (\acute{Y}_{(X=16\ años)} - \bar{Y}) + (Y_{(madre\ 12)} - \acute{Y}_{(X=16\ años)}) \end{aligned}$$

Por lo tanto,

$$\begin{aligned} 8 \text{ estilos} &= 5.50 \text{ estilos} + 2.5 \text{ estilos} \\ &= 5.50 \text{ estilos} + 1.82 \text{ estilos} \quad + .68 \text{ estilos} \\ &= \text{media} + \begin{matrix}\text{cantidad explicada}\\ \text{por 16 años de}\\ \text{educación}\end{matrix} + \begin{matrix}\text{error: cantidad}\\ \text{explicada por otras}\\ \text{variables}\end{matrix} \end{aligned}$$

TABLA 15-2 | Comparación de r con r^2: importancia del enfoque en r^2 para evaluar la fuerza de una relación

		Proporción y porcentaje de variación en Y explicada por X		
r	r^2	p	%	Fuerza de la relación
1.00	1.00	1.00	100	Relación perfecta positiva
.90	.81	.81	81	Muy fuerte positiva
.80	.64	.64	64	
.70	.49	.49	49	Moderadamente fuerte positiva
.60	.36	.36	36	
.50	.25	.25	25	
.40	.16	.16	16	Moderadamente débil positiva
.30	.09	.09	9	
.20	.04	.04	4	
.10	.01	.01	1	Muy débil positiva
.00	**.00**	**.00**	**0**	**No existe relación**
–.10	.01	.01	1	Muy débil negativa
–.20	.04	.04	4	
–.30	.09	.09	9	
–.40	.16	.16	16	Moderadamente débil negativa
–.50	.25	.25	25	
–.60	.36	.36	36	
–.70	.49	.49	49	Moderadamente fuerte negativa
–.80	.64	.64	64	
–.90	.81	.81	81	Muy fuerte negativa
–1.00	1.00	1.00	100	Relación negativa perfecta

Esto ilustra el hecho de que el modelo lineal general funciona siempre que tenemos una variable dependiente de intervalo/razón (Y) con respecto a la que calculamos la media. Las medidas de la fuerza de una relación tanto en el ANOVA como en el análisis de regresión simplemente suman la cantidad de variación explicada en Y y permiten calcular la proporción de la variación total que esta cantidad constituye.

Aplicaciones prácticas de la relación

No olvides que las aplicaciones prácticas de la relación proporcionan una descripción práctica de los resultados —al grado posible en términos comunes—. En el caso de dos variables de intervalo/razón, comenzamos con una descripción general de la forma de calcular el número de estilos musicales tolerados cuando conocemos el nivel educativo. Esta descripción general es simplemente una interpretación del coeficiente de regresión b, la pendiente de la recta de regresión. Recordemos, de acuerdo con el capítulo 14, que

$$\begin{aligned} b &= \text{pendiente de la recta de regresión del diagrama de dispersión} \\ &= \text{efecto sobre } Y \text{ de un cambio de } \textit{una unidad} \text{ en } X \end{aligned}$$

La pendiente de la recta de regresión en el diagrama de dispersión nos indica cuánta elevación se presenta con un recorrido horizontal de una unidad (véase la figura 14.6 en el capítulo 14.6). En el capítulo 14 establecimos el hecho de que la r de Pearson constituye una medida del efecto de X sobre Y. De hecho, r y b son medidas de la pendiente, aunque de manera diferente. (La r de Pearson es una pendiente estandarizada, la pendiente como número de desviaciones estándares. Nos dice cuántas desviaciones estándar de elevación de Y es posible esperar con 1 desviación estándar de recorrido en X.) Esta conexión entre los dos estadísticos también significa que b constituye una medida del efecto de X sobre Y.

En nuestra muestra de madres, una descripción general de la naturaleza de la relación entre el nivel educativo y el número de estilos musicales tolerados se enunciaría de la siguiente manera:

$$b = .52 \text{ estilos por año de educación}$$

Un incremento de un año de nivel educativo se relaciona con un incremento de .52 en la cantidad de estilos de música inculta tolerados. En otras palabras, suma .52 a las estimaciones de estilos musicales tolerados con la adición con cada año de educación. Si la señora Franklin tiene un año de educación más que la señora Curry, estima la tolerancia de la música inculta de la señora Franklin como .52 estilos más arriba.

Después de reportar la pendiente, proporcionamos ejemplos de las *mejores estimaciones* (es decir, las $\acute{Y}$) del número de estilos musicales tolerados para unos cuantos niveles educativos seleccionados (X). Los ejemplos específicos son particularmente significativos para el público en general. Estas mejores estimaciones se calculan a partir de la ecuación de regresión $\acute{Y} = a + bX$. Introducimos estos valores de X y calculamos las $\acute{Y}$. Resulta de utilidad elegir un valor bajo y un valor alto de X o una cantidad de años de educación especialmente significativa, como 12 años (es decir, graduada de preparatoria). Por consiguiente, llevemos a cabo los siguientes cálculos:

$$\acute{Y} = a + bX = -1.00 + (.52)X$$

$$\acute{Y}_{(X=12)} = -1.00 + (.52)12 = 5.24 \text{ estilos}$$

$$\acute{Y}_{(X=16)} = -1.00 + (.52)16 = 7.32 \text{ estilos}$$

Al informar estas mejores estimaciones, establecemos que

$$\acute{Y} = a + bX = -1.00 + (.52)X$$

Por consiguiente, por ejemplo, la mejor estimación del número de estilos musicales tolerados por madres con educación de preparatoria (12 años) es de 5.24 estilos, y entre madres con una educación universitaria (16 años) es de 7.32 estilos.

Aplicaciones prácticas de una relación entre dos variables de intervalo/razón En general, describe la pendiente de la recta de regresión del diagrama de dispersión:

$$b = \text{efecto sobre } Y \text{ de una unidad de cambio en } X$$

Proporciona las mejores estimaciones utilizando la ecuación de la recta de regresión:

$$\acute{Y} = a + bX$$

Sustituye los valores elegidos de X, calcula las $\acute{Y}$ e interprétalas en términos comunes.

Interpretación correcta de los estadísticos de correlación y regresión

Las correlaciones aplican a una población, no a un individuo

Una faceta fundamental de la imaginación estadística consiste en interpretar datos en relación con un todo —una población—, en lugar de hacerlo con respecto a un solo individuo o caso. Las afirmaciones sobre un caso individual son estimaciones débiles porque los individuos son sumamente complejos. Las afirmaciones sobre una población que se basan en una muestra también son estimaciones. La interpretación adecuada de datos requiere un entendimiento claro de la aplicación de las estimaciones. ¿Se aplica una estimación a un individuo o a un grupo al que pertenece el individuo?

Esta distinción entre individuo y grupo es importante en la interpretación de los estadísticos de correlación. La interpretación se encuentra limitada por el hecho de que en la mayoría de los estudios, los datos relacionados con individuos se recolectan una sola vez. *Los datos recogidos en determinado momento para cada persona en una muestra* reciben el nombre de **datos transeccionales**, y la muestra se denomina *muestra transeccional*. Por ejemplo, podríamos estudiar los efectos de un nuevo fármaco prescrito para la presión sanguínea alta durante un año, aplicando un examen a la misma muestra de pacientes cada dos meses.

La interpretación de los estadísticos de regresión difiere en estos dos tipos de datos. En el caso de los datos transeccionales del texto, no pueden llevarse a cabo estimaciones sobre los cambios en el tiempo. La interpretación de la pendiente se aplica a un solo punto en el tiempo. Por ejemplo, cuando afirmamos que el cambio de un año en el nivel educativo se relaciona con un incremento de .52 en los estilos musicales incultos tolerados, no deseamos indicar que una madre necesariamente se volverá más tolerante si recibe un año adicional de educación. Más bien, estamos afirmando que dentro de la muestra y como estimación para la población, .52 estilos constituye la diferencia *en el número de estilos musicales tolerados*, entre dos categorías del nivel educativo con un año de diferencia entre sí. Estamos comparando niveles educativos entre individuos de la muestra en un punto del tiempo.

Esta distinción se ilustra por la correlación *negativa* entre la edad y el nivel educativo. En una muestra transeccional de adultos, un aumento en la edad de los que *han concluido* su educación escolar (personas de 25 años o más edad, por así decirlo) se relaciona con una *disminución* en la educación; es decir que la correlación se presenta en dirección negativa. ¿Significa esto que las personas pierden educación al envejecer? Desde luego que no. Los datos son transeccionales; no rastreamos a la gente en el tiempo. Comparamos diferentes individuos de diferentes edades, como por ejemplo, 30, 40, 50 y 60 años, en un punto del tiempo. La correlación negativa solamente indica que las personas en los grupos de mayor edad cuentan con menos años de educación. Por ejemplo, miembros de la generación de tus abuelos, por lo general sólo recibían un diploma de preparatoria. Una interpretación correcta consistiría en que un incremento en *el nivel de edad dentro de la población* se relaciona con una reducción *del nivel* educativo. Además, supongamos que utilizamos una ecuación de regresión para calcular una mejor estimación del nivel educativo del abuelo Parker. Determinamos que su edad es de 76 años y predecimos un nivel educativo de 10.5 años de escolaridad, lo cual no significa que abandonó la escuela a la mitad de su primer año de preparatoria. Más bien, dicha estimación es la media del nivel educativo de todas las personas de 76 años de edad. La mejor estimación se basa en una puntuación X. Ésta se aplica a la puntuación X de todas las personas de 76 años. Aunque más breve que un análisis multivariado, el cual toma en cuenta una gran cantidad de variables predictoras, esta estimación es adecuada. Pero no debería hacerse mucho en lo que se refiere a estimaciones de Y basadas en X a menos que la correlación sea muy fuerte, es decir, de .90 o más alta. A menos que la correlación sea perfecta, aún habrá error en las estimaciones de Y sobre la base del conocimiento de X.

Interpretación cuidadosa de la pendiente, b

Los análisis de correlación y de regresión tienen mucho en común con el análisis de varianza (ANOVA) (véase capítulo 12). En ambos procedimientos analizamos la variación en torno a la media de una variable dependiente Y de intervalo/razón. Con el ANOVA, analizamos las diferencias en las medias entre tres o más *grupos o categorías*. Por ejemplo, ¿existe alguna diferencia entre los ingresos medios de familias católicas, protestantes y judías?, y, de ser así, ¿cuáles son las mejores estimaciones de las diferencias en la cantidad de dólares en el ingreso? Asimismo, con el análisis de regresión y correlación indirectamente nos enfocamos en la media, ya que las puntuaciones de desviación son distancias a partir de la media. Éste constituye un punto importante: los estadísticos de regresión y correlación se basan en las medias de X y de Y, y estamos haciendo estimaciones en torno a la media de Y. Cuando calculamos que una madre con 12 años de educación tolera 5.24 estilos de música inculta, no estamos afirmando que toda madre con este nivel educativo obtiene una puntuación de 5.24 estilos. Lo que estamos diciendo es que la mejor estimación *de la cantidad media* de estilos musicales tolerados por todas las madres con 12 años de educación es de 5.24 estilos. Justo como en el caso del ANOVA, nos estamos basando en las mejores estimaciones de un individuo sobre la media de su grupo. En el caso de estas madres, el grupo es una puntuación X sobre una medida de intervalo/razón.

Distinción entre la significancia estadística y la significancia práctica

La significancia estadística demuestra que un efecto de la prueba de medición para una muestra es tan grande que indica un efecto real en la población y no es meramente resultado de un error de muestreo aleatorio. Como indicamos en los capítulos 11 y 12, cuando se analizaron las pruebas de diferencia de medias con una muestra grande, una pequeña diferencia podría resultar estadísticamente significativa. Por ejemplo, con una muestra muy grande podemos encontrar una diferencia de \$10 en el sueldo anual medio de hombres y mujeres de una empresa. Con una prueba t, concluimos que la diferencia de \$10 entre las medias muestrales refleja una diferencia real de \$10 en las poblaciones de hombres y mujeres.

Ahora bien, ¿sugieren \$10 una clara discriminación de género? Por 250 días de trabajo al año, una diferencia de \$10 da como resultado 4 centavos al día o menos de un centavo por hora. En términos prácticos, la diferencia carece de significado. La significancia estadística y la significancia práctica, son cuestiones separadas.

En todas las pruebas de hipótesis, un tamaño muestral grande reduce el error estándar de la distribución muestral. Esto vuelve a la prueba más sensible para detectar la significancia estadística aun cuando el efecto de la prueba sea pequeño. Con los análisis de correlación simple y de regresión, la r de Pearson constituye la medida del efecto de la prueba. Con muestras grandes, incluso una r pequeña producirá un estadístico de prueba t grande. Esto, a su vez, producirá un valor p pequeño y llevará a un rechazo de la hipótesis estadística. El tamaño del error estándar —y, por consiguiente, el tamaño del valor p— es influenciado por el tamaño de la muestra. Si todo lo demás permanece igual, cuanto mayor es el tamaño de la muestra, menor es el valor p y más probable que se rechace la hipótesis nula. Esto se ilustra

en el caso de las pruebas t con dos muestras, una muy pequeña y la otra muy grande, en las que la r de Pearson resultó la misma:

Ilustración 1: tamaño de la muestra, $n = 16$; $r = .10$; prueba de una cola, $\alpha = .05$, $gl = n - 2 = 14$.

$$t_r = r\sqrt{\frac{n-2}{1-r^2}} = .10\sqrt{\frac{14}{1-.01}} = .38 \text{ EE}$$

A partir de la tabla de distribución t (tabla estadística C, apéndice B): valor p: p [de observar una r tan inusual o más inusual que .10 cuando $\rho = 0$] $> .05$.

Conclusión: no es estadísticamente significativa.

Ilustración 2: tamaño de la muestra, $n = 2\,002$; $r = 0.10$; prueba de una cola, $\alpha = .05$, $gl = n - 2 = 2\,000$.

$$t_r = r\sqrt{\frac{n-2}{1-r^2}} = .10\sqrt{\frac{2\,000}{1-.01}} = 4.49 \text{ EE}$$

A partir de la tabla de la distribución t (tabla estadística C, apéndice B) se obtiene el valor p: p [de observar una r tan inusual o más inusual que .10 cuando $\rho = 0$] $< .001$.

Conclusión: estadísticamente significativa al nivel de significancia de .05.

Aun cuando las r son las mismas, en las dos ilustraciones se llega a diferentes conclusiones. En la ilustración 2, concluimos que el coeficiente de correlación r de la muestra indica una correlación real en la población.

Ahora bien, ¿significa algo una r de .10? La fuerza de la relación se mide mediante r^2. En el caso de la ilustración 2: $r^2 = (.10)^2 = .01$ (o 1%). De esta manera, 1% de la variación en la variable dependiente Y se explica por el conocimiento de la variable independiente X.

El efecto es real en la población, como lo establecimos con 95% de confianza cuando rechazamos la hipótesis nula de no relación. No obstante, el efecto es tan débil que quizá carece de significado. En cierta forma, una muestra grande puede ser demasiado sensible a efectos pequeños de la prueba. La determinación de la significancia estadística no es la única consideración. La fuerza también debe tomarse en cuenta.

Por supuesto, ya que resultan insensibles incluso a efectos de prueba grandes, las muestras pequeñas presentan un problema distinto. En el caso de una muestra pequeña, podríamos concluir que una r de Pearson grande no es estadísticamente diferente de cero. Si en realidad existe una relación en la población, pero sencillamente no dimos con ella, entonces hemos cometido un error tipo II: al no rechazar una hipótesis estadística falsa. Otra ilustración revela que inclusive una relación aparentemente fuerte puede pasar inadvertida cuando empleamos una muestra pequeña.

Ilustración 3: tamaño de la muestra, $n = 5$; $r = .70$; prueba de una cola, $\alpha = .05$, $gl = n - 2 = 3$.

$$t_r = r\sqrt{\frac{n-2}{1-r^2}} = .70\sqrt{\frac{3}{1-.49}} = 1.69 \text{ EE}$$

A partir de la tabla de la distribución t (tabla estadística C, apéndice B): valor p: [de observar una r tan inusual o más inusual que .70 cuando $\rho = 0$] > .05.

Conclusión: no es estadísticamente significativa.

En esta ilustración, la fuerza de la relación sería $r^2 = (.70)^2 = .49$ (o 49%). De esta manera, parece que el 49% de la variación de la variable dependiente Y se explica por el conocimiento de la variable independiente X, sin embargo, debemos concluir que no existe una relación entre las variables y, dado el pequeño tamaño de la muestra, es todo lo que podemos hacer. Este hecho despierta la sospecha de que hemos cometido un error tipo II. Quizás exista o no la relación, pero con una muestra tan pequeña quizá nunca lo sepamos. Dicha ilustración destaca la importancia de tener una muestra lo suficientemente grande para evitar errores tipo II, en el apartado "Insensatez y falacias estadísticas", del capítulo 13, abordamos este problema como un problema de poder estadístico. Las muestras pequeñas tienen un poder estadístico bajo.

Estas ilustraciones relacionadas con la influencia del tamaño de la muestra en la prueba de hipótesis también destacan la importancia de no considerar el valor p de una prueba estadística como una medida de la fuerza de la relación. Si se encuentra que una correlación es significativa con el nivel de significancia de .001 y una segunda correlación con un nivel de .05, esto no significa necesariamente que la primera correlación sea más fuerte que la segunda. Los valores p de las ilustraciones 1 y 2 son muy diferentes; sin embargo, la fuerza de la relación es la misma. La existencia y la fuerza de la relación son temas independientes.

Presentación tabular: tablas de correlación

Con frecuencia, las correlaciones bivariadas se presentan en tablas como matrices de forma diagonal. La tabla 15-3 constituye un ejemplo. Ésta presenta datos agrupados —datos basados en grupos— para los estudiantes de 76 distritos escolares seleccionados de forma aleatoria. La puntuación media de la prueba de admisión a Stanford (SAT) para el distrito se encuentra correlacionada con el porcentaje del distrito de familias monoparentales, y con el porcentaje del distrito de estudiantes que reciben desayunos escolares gratuitos. La última variable constituye una medida común de pobreza de un área residencial.

La tabla se encuentra diseñada de tal manera que sea posible hacer comparaciones rápidas entre las correlaciones. Primero observa que la correlación de una variable consigo misma es perfecta (es decir, $r = 1.00$). Por supuesto, una variable es un predictor perfecto de sí misma. Segundo, las correlaciones sólo se presentan en una esquina de la tabla, o diagonal, ya que resulta redundante repetirlas en el otro lado. Tercero, los valores p para las r significativas se presentan por medio de asteriscos al pie de la página. Los signos de los coeficientes de correlación proporcionan información relacionada con la dirección de la relación y, al elevar al cuadrado un coeficiente, podemos estimar la proporción y el porcentaje de la variación explicada por una variable independiente elegida.

Los estadísticos de la tabla revelan que la ejecución media de la prueba en un distrito escolar se relaciona moderadamente con el porcentaje de familias monoparentales (r de Pearson = –0.61; $p < .001$). Esta fuerza moderada se determina elevando al cuadrado –0.61 y observando que aproximadamente 36% de la varianza se explica por dicha variable referente a la estructura familiar. El signo negativo nos indica que a mayor porcentaje de familias monoparentales en el distrito menor será la puntuación media SAT.

La puntuación media SAT en un distrito también tiene una fuerte correlación negativa con el porcentaje de estudiantes que reciben desayunos gratuitos: $r = -.91$; $p < .001$; r^2 es igual aproximadamente a 82%. A mayor porcentaje de estudiantes en el distrito que reciben desayunos escolares gratuitos menor será la puntuación media SAT en el distrito.

También vemos que el porcentaje de familias monoparentales se relaciona con la pobreza; es decir, la correlación entre familias monoparentales y desayunos gratuitos es moderada y positiva. $r = .63$; $p < .001$; r^2 es aproximadamente igual a 40%. Como podemos imaginar, una tabla de correlaciones con una gran cantidad de variables contiene mucha información en una sola página.

TABLA 15-3 | Correlaciones bivariadas entre la puntuación media de la Prueba de Admisión a Stanford (SAT), el porcentaje de familias monoparentales y el porcentaje de estudiantes que reciben desayunos gratis para 76 distritos escolares (datos ficticios)

		1	2	3
1	Puntuación media SAT	1.00		
2	Porcentaje de familias monoparentales	–.61***	1.00	
3	Porcentaje de estudiantes que reciben desayunos escolares gratuitos	–.91***	.63***	1.00

*$p < .05$. **$p < .01$. ***$p < .001$. Pruebas de una cola.

Insensatez y falacias estadísticas: la correlación no siempre indica causalidad

Al intentar explicar un fenómeno de interés (nuestra variable dependiente Y), buscamos una variable correlacionada X. Sin embargo, hallar una correlación no necesariamente implica que X sea la causa de Y. La existencia de una correlación sencillamente significa que las puntuaciones de las dos variables varían sistemática y conjuntamente en un patrón predecible. Este descubrimiento en sí mismo no establece ninguna causalidad entre las variables. Muchas correlaciones son espurias.

Una **correlación espuria** es *una correlación conceptualmente falsa, sin sentido o teóricamente sin sentido*, lo cual se ilustra por la correlación entre el consumo de helado (X) y la tasa de violaciones (Y). Cuando aumenta el consumo de helado, la tasa de violaciones se incrementa. Muchas correlaciones espurias son de esta naturaleza. Los incrementos y reducciones simultáneos en las tasas de estas dos variables se explican por una tercera variable —el cambio estacional—. Como resultado, la gente come más helado en época de calor y, por razones que tienen que ver con la disponibilidad de la víctima, ocurren más violaciones durante los meses calurosos del verano. De esta manera, aunque los cambios en las tasas de uno no tienen nada que ver con los cambios en las tasas del otro, el consumo de helado y las tasas de violaciones suben y bajan juntos con el cambio de las estaciones. No obstante, suponer una relación significativa entre las dos conductas carece de sentido.

Correlación espuria Correlación entre dos variables, conceptualmente falsa, sin sentido o teóricamente sin sentido.

Otra correlación espuria involucra la relación entre la tasa de delito entre los barrios de la ciudad y la composición racial de una comunidad. Existe una correlación positiva entre el porcentaje de la población minoritaria (por ejemplo, afroamericanos) que viven en barrios y las tasas de crímenes. Es decir, en una muestra de comunidades, las que tienen un alto porcentaje de afroamericanos tienden a tener tasas altas de delitos. No obstante, ello sugiere que los afroamericanos son más propensos al comportamiento delictivo, y de hecho, los racistas a menudo citan tal estadístico. Esta correlación, sin embargo, resulta espuria. Las tasas de delito son altas en los barrios *pobres* sin tener en cuenta su composición racial, y una parte desproporcionada de los barrios minoritarios son pobres. Es más, la relación entre pobreza y composición racial se debe al racismo, no a la raza biológica. Es decir, ser pobre, no tiene nada que ver con la genética. La herencia racista de Estados Unidos contribuye al hecho de que una parte desproporcionada de afroamericanos viva en la pobreza, lo cual, a su vez, constituye un buen predictor de las tasas de delitos.

¿Cómo le demostramos tal cuestión a un racista? Examina sólo los barrios adinerados, algunos que son predominantemente blancos y algunos que son predominantemente afroamericanos. Sin tener en cuenta el predominio racial, dichos barrios tendrán bajas tasas de delito. Es más, *dentro* de una muestra de barrios adinerados, la correlación entre el porcentaje minoritario y la tasa de delito será cero. De igual forma, en los barrios pobres —sin tener en cuenta la composición racial— las tasas de delito son relativamente altas, y se encuentra una correlación de cero entre la pobreza y la composición racial. Esta perspectiva de enfocarse en un nivel constante para una tercera variable, como el estatus socioeconómico de un barrio, se llama *mantener constante esa variable*. Cuando mantenemos constante un estatus socioeconómico, se eliminan sus efectos en la tasa de delito, y así tal variable ya no produce la relación espuria entre X (porcentaje de población minoritaria) y Y (tasa de delito). El cálculo de los estadísticos de correlación y de regresión para controlar variables adicionales se realiza mediante los *análisis de correlación y regresión múltiple*, un tema que se abarca en el sitio web *The Statistical Imagination*, www.mhhe.com/ritchey2. Estas ilustraciones de los efectos espurios subrayan la importancia de interpretar los hallazgos estadísticos con gran cautela.

Las correlaciones espurias no son inusuales, lo cual, en parte, se debe al hecho de que muchas cosas, como el tamaño de la población de la Tierra, continuamente se incrementan. Cualquier otra cuestión que continuamente se aumente (como el crecimiento de gigantescos hongos en el subsuelo) estará positivamente correlacionada con el tamaño de la población. Asimismo, cualquier otra variable que continuamente disminuya (como el tamaño de las capas del hielo en el continente Antártico o el porcentaje de adultos estadounidenses que fuman cigarros) estará negativamente correlacionada con el tamaño de la población. Recuerda del capítulo 1 que una buena teoría científica implica dos elementos: un sentido de la comprensión y la habilidad para brindar predicciones empíricas. Las correlaciones espurias permiten hacer predicciones pero no ofrecen un sentido de la comprensión del fenómeno en cuestión; de hecho, a menudo confunden una situación.

RESUMEN

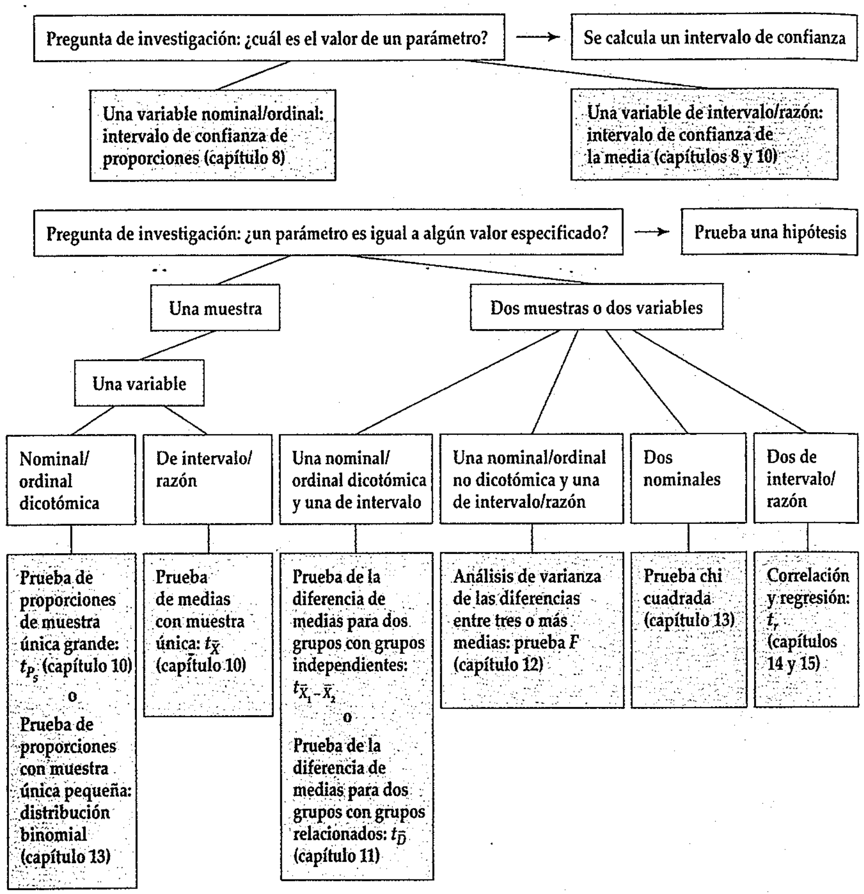
1. La prueba de hipótesis para una relación entre dos variables de intervalo/razón constituye una prueba t del coeficiente de correlación bivariada r de Pearson.
2. Se utiliza esta prueba cuando: *a*) existe una muestra representativa proveniente de una sola población; *b*) hay dos variables de intervalo/razón; *c*) no existen restricciones en el tamaño de la muestra, aunque por lo general, mientras más grande n, mejor; *d*) la observación del diagrama de dispersión de las coordenadas de las dos variables revela un patrón lineal.

3. Prueba la H_0 de que $\rho = 0$; es decir que no existe relación entre X y Y. La letra griega rho (ρ) es el coeficiente de correlación obtenido si la r de Pearson se calculara para toda la población.
4. La distribución muestral es una distribución t para el coeficiente de correlación bivariada r con $gl = n - 2$. Si de hecho, $\rho = 0$, entonces con el muestreo repetido la r de Pearson daría un valor de cero con un error de muestreo. No es necesario calcular un error estándar.
5. El efecto de la prueba es el coeficiente de correlación bivariada r de Pearson, que es la r muestral observada menos la rho hipotética de cero.
6. El estadístico de la prueba es t_r y el valor p se calcula por medio de la tabla de la distribución t, apéndice B, tabla estadística C.
7. La existencia de una relación se establece probando la H_0 de que $\rho = 0$; es decir que no existe una relación entre X y Y. Si la H_0 se rechaza, entonces existe una relación.
8. La dirección se indica por medio del signo de r y b observando la pendiente de la recta de regresión en el diagrama de dispersión. Se revela una relación positiva por medio de una pendiente ascendente, y r y b serán positivos. Una relación negativa se revela por medio de una pendiente descendente, y r y b serán negativos.
9. La fuerza de la relación se determina por medio de la proporción de la variación total de Y explicada por X. Ésta se obtiene rápidamente elevando al cuadrado el coeficiente de correlación bivariada r de Pearson.
10. Las aplicaciones prácticas de los resultados se presentan de la siguiente manera: *a*) Interpreta el coeficiente de regresión, b, la pendiente de la recta de regresión. Establece el efecto sobre Y para una unidad de cambio en X. *b*) Proporciona las mejores estimaciones utilizando la ecuación de la recta de regresión. Incluye los valores elegidos de X, calcula $\acute{Y}$ e interprétalas en términos familiares.
11. Debe tenerse mucho cuidado cuando se interpreten los coeficientes de correlación: *a*) Una correlación se aplica a una población, no a un individuo. Una $\acute{Y}$ es una estimación de la media de Y para todos los individuos de la población con el valor dado de X. *b*) Con muestras grandes particularmente, un coeficiente de correlación estadísticamente significativo puede tener un valor pequeño, y por lo tanto representa una relación muy débil o prácticamente insignificante. *c*) Una correlación entre X y Y no siempre significa que X es causa de Y. Una correlación puede ser espuria; matemáticamente puede ser adecuada, aunque, conceptualmente falsa, carente de sentido o teóricamente sin significado.

EXTENSIONES DEL CAPÍTULO EN EL SITIO WEB *THE STATISTICAL IMAGINATION*

Las extensiones de los capítulos 14 y 15 del material del texto disponibles en el sitio web *The Statistical Imagination*, www.mhhe.com/ritchey2 incluyen una introducción a la correlación y regresión múltiples, situación en la que una variable dependiente es regresionada a partir de dos o más variables independientes.

PROCEDIMIENTOS ESTADÍSTICOS CUBIERTOS HASTA AQUÍ



FÓRMULAS PARA EL CAPÍTULO 15

Para calcular la ecuación de los coeficientes de correlación y regresión, véase capítulo 14. Para completar la hoja de cálculo (tabla 15-1):

$$\bar{X} = \frac{\Sigma X}{n} \qquad \bar{Y} = \frac{\Sigma Y}{n}$$

Coeficiente de correlación de Pearson:

$$r = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma (X - \bar{X})^2 \, \Sigma (Y - \bar{Y})^2}}$$

Coeficiente de regresión o pendiente, b:

$$b = \frac{\Sigma (X - \bar{X})(Y - \bar{Y})}{\Sigma (X - \bar{X})^2}$$

Intersección de Y, a:

$$a = \bar{Y} - b\bar{X}$$

Ecuación de regresión:

$$\acute{Y} = a + bX$$

Estadístico de la prueba:

$$t_r = r\sqrt{\frac{n-2}{1-r^2}}$$

Para describir la fuerza de una relación entre dos variables de intervalo/razón:

$$r^2$$

PREGUNTAS PARA EL CAPÍTULO 15

1. En el caso del coeficiente de correlación r de Pearson, traza un diagrama conceptual que describa una población e inserta los estadísticos y parámetros adecuados. Estipula cómo se establece la hipótesis nula para la prueba de hipótesis.
2. En el caso de una prueba de hipótesis de la relación entre dos variables de intervalo/razón, ¿cuál es la forma de la distribución muestral y cuál es el estadístico de prueba que se utiliza?
3. Al probar una hipótesis entre dos variables de intervalo/razón, la dirección de la prueba se establece en la hipótesis alternativa. En el caso general de Y regresionada a partir de X, muestra la forma en la que se establece la hipótesis alternativa cuando se hipotetiza que la dirección es positiva. Indica la forma del patrón de coordenadas de un diagrama de dispersión cuando éste es el caso. Haz lo mismo en el caso de una relación negativa.
4. Observa las fórmulas del coeficiente de correlación r de Pearson y el coeficiente de regresión b, y explica por qué estos dos coeficientes siempre tienen el mismo signo direccional (+ o –).
5. Explica lo que nos dice r^2.
6. Al determinar la fuerza de una relación entre dos variables de intervalo/razón, ¿cuál es el peligro de confiar solamente en una interpretación del valor absoluto de la r de Pearson? Explica.
7. Responde con una palabra la siguiente pregunta: cuando no se encuentra una relación entre X y Y, ¿qué decimos sobre los demás aspectos de una relación?

8. Relaciona las siguientes afirmaciones en lo que se refiere a la relación entre dos variables de intervalo o de razón:

a) *r* de Pearson	______	Intersección de *Y*; valor de *Y* cuando $X = 0$
b) *a*	______	Proporción de la variación en *Y* explicada por el conocimiento de *X*; una medida de la fuerza de la relación
c) r^2	______	Pendiente de la recta de regresión; efecto sobre *Y* de un cambio de una unidad en *X*; una medida de las aplicaciones prácticas de la relación
d) *b*	______	Valor predicho de *Y*; mejor estimación de *Y* para un valor dado; se utiliza para describir las aplicaciones prácticas de la relación
e) $\hat{Y}$	______	Mide lo ceñido del ajuste de las coordenadas *X*, *Y* en torno a la recta de regresión; se utiliza para describir la existencia y dirección de una relación

9. Explica la diferencia entre datos transeccionales y datos longitudinales.
10. Un investigador determina una relación negativa entre la edad y el conocimiento sobre las computadoras personales utilizando datos transeccionales. ¿Significa esto que conforme la gente envejece, pierde este conocimiento? Explica.
11. Matemáticamente existe una correlación positiva entre el número de calzado y la habilidad para resolver problemas matemáticos complejos. ¿Es esto realmente significativo? Explica.
12. Matemáticamente existe una correlación negativa entre la cantidad de películas producidas en Hollywood cada año y el tamaño de las bosques tropicales del Amazonas. ¿Es esto realmente significativo? Explica.

EJERCICIOS PARA EL CAPÍTULO 15

Conjunto de problemas 15A

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, el diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

15A-1. Interpreta los siguientes coeficientes de correlación bivariada *r* de Pearson en lo que se refiere a la dirección de la relación (datos ficticios).

X	Y	r
a) Ingreso anual	Valor del dólar en inversiones a largo plazo	.69
b) Tasa de alfabetización de la población	Grado en que el gobierno es democrático	.73
c) Edad	Número de películas vistas durante los pasados seis meses	−.45
d) Infracciones de tránsito y accidentes en los últimos tres años	Costos de las primas de seguros de automóviles	.87

15A-2. Tú diriges un estudio de *adictos a la televisión* con adolescentes en riesgo de desarrollar obesidad. Recolectas los datos de una muestra de varones adolescentes, en uno de los 14 grupos experimentales de siete ciudades. Los siguientes datos ficticios indican la cantidad de horas que los individuos ven televisión cada semana, la edad, el peso, la estatura y la cantidad de comidas consumidas la semana pasada.

a) Traza un diagrama de dispersión del peso regresionado a partir de las horas de televisión que se ven por semana.
b) Calcula el coeficiente de correlación *r* de Pearson para estas dos variables, así como la ecuación de regresión, $\hat{Y} = a + bX$.
c) Prueba la hipótesis de que existe una relación entre estas dos variables. Si existe, analiza los demás aspectos de la relación.
d) La edad y la estatura son casi iguales entre los individuos. De hecho, estas variables se *mantuvieron constantes intencionalmente*. ¿Por qué resulta conveniente realizar esto al probar la relación entre la cantidad de tiempo que los individuos ven televisión y sus pesos?

Individuo	Horas de televisión a la semana	Peso (libras)	Edad (años)	Estatura (pulgadas)
1	9	112	13	62
2	14	131	14	65
3	20	171	14	64
4	18	160	13	63
5	16	182	14	64
6	14	165	15	67
7	19	149	14	64
8	12	137	13	65

15A-3. Imagina que tú eres el responsable de las admisiones en cierta universidad. En el caso de una muestra aleatoria de 114 estudiantes de segundo año, tú comparas el promedio actual de la universidad con el promedio de la preparatoria (los datos no se muestran). Un diagrama de dispersión generado en computadora revela una relación lineal entre dichas variables. Los resultados de la computadora indican lo siguiente: *r* de Pearson = .47, $b = .73$ y la intersección con el eje *Y*, $a = .80$. Prueba la hipótesis de que los estudiantes que cursaron satisfactoriamente la preparatoria van bien en la universidad.

15A-4. Las siguientes variables son de intervalo/razón y sus correspondientes estadísticos se calcularon con una muestra de 100 usuarios de internet (datos ficticios). Todos los coeficientes son estadísticamente significativos al nivel de .05.

X = Edad	Y = Puntuación de prestigio profesional	Z = Nivel educativo
$r_{yx} = .24$	$r_{zx} = -.32$	$r_{zy} = .31$
$b_{yx} = 2$	$b_{zx} = -.3$ años	$b_{zy} = 6$

a) ¿Cuál es la relación más fuerte, la que existe entre la edad y la puntuación de prestigio profesional o la que existe entre la edad y el nivel educativo? ¿Por qué razón?
b) Interpreta b_{zx}.
c) Interpreta la dirección de r_{zx}.
d) Organiza los coeficientes de correlación bivariada en una tabla de correlación.

15A-5. Los siguientes datos ficticios corresponden a 12 ciudades seleccionadas aleatoriamente. ¿Existe una relación entre el nivel educativo de una ciudad (es decir, la media de años de educación de los adultos jefes de familia) y las muertes por incendios y quemaduras por cada 100 000 habitantes.

Muertes por incendio/quemaduras por cada 100 000	Nivel medio educativo
2.3	11.7
1.5	12.2
2.1	11.3
2.3	11.1
1.4	12.2
1.9	12.0
1.6	12.3
1.7	12.4
1.5	12.5
1.7	11.6
2.2	10.9
2.1	12.1

15A-6. Utiliza los datos del ejercicio 14A-3 del capítulo 14, sobre las organizaciones para el mantenimiento de la salud (OMS). Completa el ejercicio para obtener un diagrama de dispersión y los estadísticos de correlación y regresión bivariadas. Prueba la hipótesis de que la reducción promedio de tiempo de estancia se encuentra correlacionada con el incremento porcentual de ganancias de las organizaciones para el mantenimiento de la salud durante el año pasado.

Conjunto de problemas 15B

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Por consistencia, redondea los cálculos a dos decimales. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

15B-1. Interpreta los siguientes coeficientes de correlación bivariada con respecto a la dirección de la relación (datos ficticios).

X	Y	r
a) Nivel educativo	Prestigio profesional	.75
b) Suma gastada en la promoción de cinturones de seguridad	Muertes por accidentes de tráfico (en un estado)	–.58
c) Edad	Estado de salud	–.51
d) Suma gastada en un automóvil	Suma gastada en la casa	.81

15B-2. Los siguientes datos ficticios corresponden a un estudio de niños de 10 años de edad, así como de sus madres. Determina si existe una relación entre el peso de la madre y el peso del niño. En el caso de estos datos, realiza lo siguiente:

a) Traza un diagrama de dispersión del peso del niño regresionado a partir del peso de la madre.
b) Calcula el coeficiente de correlación bivariada *r* de Pearson para estas dos variables, así como la ecuación de la recta de regresión: $\hat{Y} = a + bX$.
c) Lleva a cabo la prueba de hipótesis de que existe una relación entre estas dos variables y aborda los aspectos adecuados de la relación.
d) La edad se mantiene constante a través del muestreo. También se *mantuvieron constantes* intencionalmente las calorías diarias consumidas, así como el ejercicio diario. ¿Por qué resulta prudente realizar esto al probar una relación entre el peso de la madre y el peso del niño?

Individuo	Calorías diarias	Peso de la madre	Peso del niño	Ejercicio diario (en minutos)
1	2 206	150	80	24
2	2 246	185	91	23
3	2 211	134	68	23
4	2 203	215	93	22
5	2 229	287	91	22
6	2 223	147	71	23
7	2 241	175	81	24
8	2 233	195	92	24
9	2 219	167	75	22

15B-3. Supongamos que deseas probar la hipótesis de que la gente más joven conoce más de computadoras que la gente mayor. Tú reúnes una muestra aleatoria de 150 adultos y recopilas datos relacionados con su edad y conocimientos computacionales utilizando la escala de conocimientos computacionales (ECC). Un diagrama de dispersión generado por computadora revela una relación lineal entre las variables. Los resultados de la computadora arrojan los siguientes datos: *r* de Pearson $= -0.72$; $b = -0.08$ y la intersección con el eje *Y*, $a = 25.31$. Prueba la hipótesis de que la gente más joven conoce más de computadoras que la gente mayor.

15B-4. Las siguientes variables son variables de nivel de intervalo/razón con los estadísticos adecuados calculados para una muestra de hombres adultos (datos ficticios). Todos los coeficientes son estadísticamente significativos al nivel de .05.

X = Ingreso personal (en miles de dólares)	Y = Ingreso de los padres (en miles de dólares)	Z = Número de parientes incluidos en la familia
$r_{yx} = .51$	$r_{zx} = -.47$	$r_{zy} = -.39$
$b_{yx} = 1.7$	$b_{zx} = -.11$	$b_{zy} = -.08$

a) ¿Cuál es la relación más fuerte, la que se da entre el ingreso personal y el ingreso de los padres, o la que se da entre el número de parientes incluidos en la familia? ¿Por qué?

b) Interpreta b_{zx}.

c) Interpreta la dirección de r_{yx}.

d) Organiza los coeficientes de correlación bivariada en una matriz de correlación.

15B-5. Los siguientes datos ficticios provienen de 10 escuelas de educación media seleccionadas de forma aleatoria. ¿Existe una relación entre las opciones de comida saludable en las cafeterías de las escuelas y la cantidad de estudiantes obesos? Las opciones de comida saludable se midieron mediante la razón de comidas saludables (frutas, verduras, granos enteros, etc.) a los alimentos no saludables (alimentos con mucha grasa, sodio, etcétera).

Opciones de comida saludable	Número de estudiantes obesos por cada 100
2.01	22
.82	25
.67	27
1.50	25
1.90	21
2.14	23
1.01	26
.87	25
.91	24
1.75	23

15B-6. Utiliza los datos del ejercicio 14B-3 del capítulo 14, relacionados con el capital social y la satisfacción con la vida. Completa dicho ejercicio para obtener un diagrama de dispersión y los estadísticos de correlación y regresión bivariadas. Prueba la hipótesis de que los individuos con un alto capital social tienden a tener mayor satisfacción con la vida.

Conjunto de problemas 15C

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Por coherencia, redondea los cálculos a dos decimales. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

15C-1. Interpreta las siguientes *r* de Pearson en lo que se refiere a la dirección de la relación (datos ficticios).

X	Y	r
a) Promedio de preparatoria	Promedio en la universidad	.57
b) Número de condenas	Duración de la sentencia en prisión	.38
c) Nivel socioeconómico del convicto	Duración de la sentencia en prisión	−.71
d) Nivel socioeconómico de un barrio	Porcentaje de graduados de preparatoria que asisten a la universidad	.68

15C-2. Supongamos que la siguiente tabla contiene datos de un estudio de adolescentes en riesgo de desarrollar obesidad. Lleva a cabo lo siguiente:

a) Traza un diagrama de dispersión del peso regresionado a partir de la cantidad de comidas rápidas consumidas la semana pasada.

b) En el caso de estas dos variables, calcula el coeficiente de correlación bivariada de Pearson y la ecuación de regresión: $\hat{Y} = a + bX$.

c) Prueba la hipótesis de que existe una relación entre estas dos variables y analiza los aspectos apropiados de la relación.

d) La edad y la estatura son aproximadamente iguales entre los individuos. De hecho, se *mantuvieron constantes* de manera intencional. ¿Por qué resulta prudente hacer esta prueba en el caso de una relación entre la cantidad de comidas rápidas consumidas la semana pasada y el peso?

Individuo	Horas de televisión a la semana	Edad (años)	Peso (libras)	Altura (pulgadas)	Número de comidas rápidas consumidas la semana pasada
1	9	12	112	65	2
2	14	13	131	66	2
3	20	12	171	66	5
4	18	14	160	66	2
5	16	13	182	67	4
6	14	14	165	66	3
7	19	13	149	67	3
8	12	13	137	66	2

15C-3. Las medidas de religiosidad obtenidas por medio de una encuesta evalúan el grado al que una persona cree, reza, asiste a la iglesia y se rige por la moral establecida por una religión. La variable de satisfacción con la vida constituye una percepción de lo bien que le va en la vida a una persona en la actualidad. Un diagrama de dispersión generado en computadora revela una relación lineal entre estas variables. La información de la computadora revela los siguientes resultados (ficticios): $n =$ 14 adultos; *r* de Pearson = 0.48; $b = 0.78$ y la intersección con el eje Y $a = 6.24$.

Prueba la hipótesis referente al hecho de que las personas religiosas tienden a presentar mayor satisfacción con la vida.

15C-4. Las siguientes variables son de intervalo/razón, y sus correspondientes estadísticos se calcularon con una muestra de 100 usuarios de internet (datos ficticios). Todos los coeficientes son estadísticamente significativos con un nivel de significancia de .05.

a) ¿Cuál es la relación más fuerte, la que existe entre la edad y las horas de conexión a internet, o la que existe entre el conservadurismo político y las horas de conexión a internet? ¿Por qué?

b) Interpreta b_{zx}.

c) Interpreta la dirección de r_{yx}.

d) Organiza los coeficientes de correlación bivariada en una tabla de correlación.

X = Edad	Y = Conservadurismo político	Z = Horas de conexión (a la World Wide Web en internet por mes)
$r_{yx} = .19$	$r_{zx} = -.26$	$r_{zy} = .25$
$b_{yx} = .35$ puntos en la escala de conservadurismo	$b_{zx} = -1.1$ horas	$b_{zy} = .79$ horas

15C-5. Un interesante rasgo psicológico consiste en abrirse a la experiencia: la disposición para probar nuevas cosas y mantener una mente abierta (McCrae, 1996). Otro rasgo que se investiga con frecuencia es el autoritarismo: la creencia general de que es mejor contar con leyes firmes en la sociedad, con padres rigurosos y jefes que mantengan a la gente en orden. ¿Existe alguna relación entre estos dos rasgos en un grupo de adultos elegidos al azar?

Escala de apertura a experimentar	Escala de autoritarismo
9	31
12	20
17	15
16	14
15	12
13	26
10	28
7	34
11	27
9	24
10	22
6	33
7	27
13	22
15	17

15C-6. De acuerdo con el capítulo 14, utiliza los datos del ejercicio 14C-3, relacionados con los días de incapacidad y la pérdida de ingresos. Completa dicho ejercicio para obtener un diagrama de dispersión y estadísticos de correlación y regresión bivariadas. Prueba la hipótesis de que los días de incapacidad al año predicen el ingreso familiar total (en miles de dólares).

Conjunto de problemas 15D

En todas las pruebas de hipótesis, sigue los seis pasos de la inferencia estadística, incluyendo la preparación de la prueba, un diagrama conceptual, las curvas de probabilidad y los aspectos adecuados de una relación. Utiliza $\alpha = .05$, a menos que se estipule otra cosa.

15D-1. Interpreta los siguientes coeficientes de correlación bivariada *r* de Pearson en lo que se refiere a la dirección de la relación (datos ficticios).

X	Y	*r*
a) Altura	Peso	.68
b) Edad	Conciertos a los que se asistió el año pasado	−.72
c) Cobertura media del crimen	Temor al crimen	.56
d) Tasa de pobreza en el vecindario	Tasa de crímenes en el vecindario	.79

15D-2. Tú llevas a cabo un estudio para determinar si la asistencia a clase predice el promedio de calificaciones (GPA). Integras una muestra de estudiantes universitarios y reúnes los siguientes datos sobre sus promedios de calificaciones y la cantidad de clases a las que faltaron el año pasado. Determina si los estudiantes que perdieron una gran cantidad de clases tienden a tener promedios más bajos.

a) Traza un diagrama de dispersión del GPA regresionado a partir de las clases a las que se faltó.

b) En el caso de estas dos variables calcula el coeficiente de correlación bivariada *r* de Pearson y la ecuación de regresión $\hat{Y} = a + bX$.

c) Prueba la hipótesis de que existe una relación entre estas dos variables y analiza los aspectos más apropiados de la relación.

d) Las puntuaciones del CI y del SAT son aproximadamente iguales entre los individuos. De hecho, estas variables se *mantuvieron constantes* de forma intencional. ¿Por qué resulta prudente hacer esto al probar la existencia de una relación entre el número de clases a las que se faltó y el GPA?

Individuo	GPA	Clases a las que se faltó	CI	SAT
1	3.12	5	114	1 014
2	2.42	9	110	996
3	1.95	16	101	1 012
4	3.89	4	112	978
5	2.56	7	108	994
6	2.34	9	119	1 025
7	1.84	14	114	1 017
8	3.64	4	106	980
9	3.91	3	116	1 020
10	2.49	8	115	989

15D-3. La carga de pacientes de una enfermera se relaciona con la satisfacción en el trabajo (Aiken, Clarke, Sloane, Sochalski y Silber, 2002). En otras palabras, cuantos menos pacientes haya bajo el cuidado de una enfermera, mayor será la satisfacción en el trabajo. Tú deseas confirmar esta conclusión con un estudio propio. Mides la carga de pacientes como razón de enfermera-paciente, la cantidad de enfermeras por cada 100 pacientes. La satisfacción en el trabajo se evalúa con la escala *Me gusta la enfermería.* Un diagrama de dispersión generado por computadora revela una relación lineal entre las siguientes variables. La computadora arroja los siguientes resultados (ficticios): $n = 210$ enfermeras; r de Pearson $= .79$; $b = .44$ e intersección con Y, $a = 11.53$. Prueba la hipótesis de que la razón paciente-enfermera se relaciona con la satisfacción en el trabajo.

15D-4. Las siguientes variables son de nivel de intervalo/razón con estadísticos apropiados calculados para una muestra de 442 madres que trabajan (datos ficticios). Todos los coeficientes son estadísticamente significativos con un nivel de confianza de .05.

a) ¿Cuál es la relación más fuerte, la que existe entre la edad y las horas de conexión a internet o la que existe entre el conservadurismo político y las horas de conexión a internet? ¿Por qué?
b) Interpreta *bzx*.
c) Interpreta la dirección de r_{zx}.
d) Organiza los coeficientes de correlación bivariada en una tabla de correlación.

X = Sucesos estresantes en la vida (el año pasado)	Y = Depresión	Z = Satisfacción en la vida
$r_{yx} = .46$	$r_{zx} = -.51$	$r_{zy} = -.39$
$b_{yx} = -1.3$ puntos de la escala	$b_{zx} = -1.9$ puntos de la escala	$b_{zy} = .81$ puntos de la escala

15D-5. Tú te encuentras investigando sobre una diversidad de relaciones con el control psicológico, el grado de control que los individuos creen tener sobre sus vidas y lo que les sucede. Por medio de diversas escalas de medición para encuestas, obtienes los siguientes datos para un grupo de adultos elegidos al azar. Determina si existe una relación entre el dominio y la creencia de que votar es importante.

Escala de dominio	Escala referente a la importancia del voto
10	12
25	30
33	34
31	41
15	13
19	28
22	27
29	24
24	21
25	23
13	15
17	14
21	20
28	23
30	29

15D-6. Utilice datos del ejercicio 14D-1 del capítulo 14. Complete el ejercicio para obtener el diagrama de dispersión, la correlación bivariada y la regresión estadística. Pruebe la hipótesis de que hay una relación entre los años de escolaridad y el ingreso.

APLICACIONES OPCIONALES DE COMPUTADORA PARA EL CAPÍTULO 15

Si en tu clase se utilizan las aplicaciones opcionales en computadora que acompañan el texto, abre los ejercicios del capítulo 15 localizados en el sitio web de *The Statistical Imagination*, www.mhhe.com/ritchey2. Estos ejercicios involucran el uso del *SPSS* para Windows para calcular los estadísticos para la prueba de hipótesis y abarcan los cuatro aspectos de las relaciones para dos variables de intervalo/razón. Además, el apéndice D del texto contiene una vista rápida de las secuencias de comandos del *SPSS* para los procedimientos cubiertos en este capítulo.

A APÉNDICE

Repaso de las operaciones matemáticas básicas

Este apéndice no proporciona una lección minuciosa sobre cálculos matemáticos. Tan sólo ofrece ejemplos de repaso para ayudar a los estudiantes a recordar las operaciones matemáticas básicas que se encuentran por lo común en los cálculos estadísticos. Después de cada sección de repaso, se incluyen problemas; las respuestas aparecen al final de este apéndice.

SÍMBOLOS Y TÉRMINOS MATEMÁTICOS BÁSICOS

$\pm$	Más menos	$\div$ o /	Dividido entre
$<$	Menor que	$\times$ o $\cdot$ o $(\cdot\cdot\cdot)$	Multiplicado por
$>$	Mayor que	$\leq$	Menor o igual que
$\geq$	Mayor o igual que	∞	Infinito
$\approx$	Aproximadamente igual	$\neq$	Desigual
$\lvert\ \rvert$	Valor absoluto de (ejemplo, $\lvert -15 \rvert = 15$)	$\sqrt{\ }$	Raíz cuadrada

Una *suma* es la respuesta a un problema de adición o de sustracción.

Un *producto* es la respuesta a un problema de multiplicación.

Un *cociente* es la respuesta a un problema de división.

ORDEN DE LAS OPERACIONES MATEMÁTICAS

Al calcular las partes de una fórmula, sigue estas reglas:

1. Trabaja dentro de cada paréntesis antes de seguir fuera de ellos. Si conjuntos de paréntesis están dentro de corchetes, trabaja dentro de cada corchete antes de salirte.
2. Los términos son partes de una ecuación separados por signos de suma y resta. Antes de sumar y restar términos, completa cualquier multiplicación y división dentro de cada término.
3. En problemas de división, completa todos los cálculos arriba y abajo del signo de división antes de dividir.
4. Trata al signo de un radical (es decir, el signo de la raíz cuadrada) como un conjunto grande de corchetes; es decir, completa todos los cálculos bajo el signo del radical antes de calcular la raíz cuadrada.

EXPONENTES: ELEVAR AL CUADRADO Y RAÍCES CUADRADAS

Exponentes: Elevación a una potencia Un exponente es el número de veces que un número base se multiplica por sí mismo. Cuando realizamos tales cálculos, decimos que "ele-

vamos el número base a una potencia". "Elevar al cuadrado" un número base consiste en elevarlo a la segunda potencia. Por ejemplo, 4 a la segunda potencia es 4^2 (es decir, 4 al cuadrado), o 4 multiplicado por sí mismo:

$$4^2 = (4)\,(4) = 16$$

Un número base elevado a la primera potencia es el propio número base, y no nos molestamos en colocar el exponente 1 en la notación: $4^1 = 4$. En general, para cualquier número a: $a^1 = a$.

Cualquier número elevado al exponente cero es igual a 1: $4^0 = 1$, $7^0 = 1$. En general, $a^0 = 1$.

Un número base elevado a la tercera potencia se "eleva al cubo". Así, 4 elevado al cubo es

$$4^3 = (4)\,(4)\,(4) = 64$$

De igual forma, 5 elevado a la quinta potencia es

$$5^5 = (5)\,(5)\,(5)\,(5)\,(5) = 3\,125$$

Note que 5^5 es igual a 5 al cuadrado por 5 al cubo, o 5 a la cuarta potencia por 5. Es decir, cuando un número base se eleva a una potencia y se multiplica por el mismo número base elevado a una potencia, el resultado es igual al número base elevado a la suma de los exponentes:

$$5^5 = (5^2)\,(5^3) = (25)\,(125) = 3\,125 \quad \text{o} \quad 5^5 = (5^1)\,(5^4) = (5)\,(625) = 3\,125$$

Esto se llama la regla multiplicativa para exponentes. En general, establece que para cualesquier enteros positivos, m y n,

$$(a^m)\,(a^n) = a^{m+n}$$

Existen varias *reglas generales de potencias para los exponentes.* Para cualesquier enteros positivos m y n,

$$(a^m)^n = a^{mn}; \text{ por ejemplo, } (6^2)^3 = 6^6 = 46\,656$$
$$(ab)^m = a^m b^m; \text{ por ejemplo, } (2 \cdot 3)^2 = (2^2)\,(3^2) = 4 \cdot 9 = 36$$
$$(a/b)^m = \frac{a^m}{b^m}; \text{ por ejemplo, } (4/2)^3 = \frac{4^3}{2^3} = \frac{64}{8} = 8$$

La regla del cociente para exponentes establece que

$$\frac{a^m}{a^n} = a^{m-n}; \text{ por ejemplo, } \frac{3^4}{3^2} = 3^{(4-2)} = 3^2 = 9$$

PROBLEMAS DE REPASO

1. $7^2 =$
2. $21^2 =$
3. $21^3 =$
4. $19^2 =$
5. $9^4 =$
6. $10^3 =$
7. $(2^2)(2^5) =$
8. $(3^2)^4 =$
9. $100^2 \div 10^2 =$
10. $10^4 \div 10^3 =$

Raíces cuadradas Obtener la raíz cuadrada es lo inverso a elevar al cuadrado. Por tanto, la raíz cuadrada de 16 es 4:

$$\sqrt{16} = 4$$

Para verificar la exactitud al sacar una raíz cuadrada, eleva al cuadrado el resultado para ver si reaparece el número bajo el radical. De igual forma, cuando un número se eleva al cuadrado, saca la raíz cuadrada del resultado para ver si reaparece el número base. Esto significa que la raíz cuadrada de un número elevada al cuadrado es igual al número:

$$\sqrt{16} = \sqrt{4^2} = 4$$

En general, $\sqrt{a^2} = a$, donde a es algún número base.

PROBLEMAS DE REPASO

Observa el patrón de las respuestas para los ejercicios 11 al 16.

11. $\sqrt{8.1} =$ **12.** $\sqrt{81} =$

13. $\sqrt{810} =$ **14.** $\sqrt{8\,100} =$

15. $\sqrt{81\,000} =$ **16.** $\sqrt{810\,000} =$

17. $\sqrt{7.568} =$ **18.** $\sqrt{100} =$

19. $\sqrt{10^2} =$ **20.** $\sqrt{41^2}$

NOTACIÓN DE LA SUMATORIA

La letra griega sigma (Σ) sirve para indicar "suma de". Debe tenerse cuidado en sumar los términos apropiados de una ecuación.

Ilustración: La variable X = edad. John tiene 32 años, Kirk 40 y Jim 24.

Ejemplo 1: su edad "media" o promedio es "la suma de X" dividida entre 3:

$$\bar{X} = \frac{\Sigma X}{3} = \frac{(32 + 40 + 24)}{3} = 32 \text{ años}$$

donde el símbolo $\bar{X}$ se lee "media" (véase el capítulo 4).

Ejemplo 2: para calcular la *suma de las edades elevadas al cuadrado*, primero eleva al cuadrado cada puntuación X y después suma:

$$\Sigma X^2 = (32^2 + 40^2 + 24^2) = 1\,024 + 1\,600 + 576 = 3\,200 \text{ años cuadrados}$$

Ejemplo 3: para calcular la *suma de las edades* al cuadrado, primero suma las edades y después eleva al cuadrado el resultado:

$$(\Sigma X)^2 = 96^2 = 9\,216 \text{ años cuadrados}$$

Las sumatorias se facilitan cuando los datos están organizados en hojas de cálculo con las variables colocadas en las columnas y los casos ordenados en las filas. La siguiente hoja de cálculo indica edades (X) y pesos (Y) para cuatro casos o sujetos. Usa esta hoja de cálculo para repasar los problemas del 21 al 26.

Caso/sujeto	X (edad)	X^2	Y (peso)	Y^2
John	32	1 024	169	28 561
Kirk	40	1 600	191	36 481
Jim	24	576	157	24 649
Carl	44	1 936	212	44 944

PROBLEMAS DE REPASO

21. $\Sigma X =$ **22.** $\Sigma X^2 =$ **23.** $(\Sigma X)^2 =$ **24.** $\Sigma Y =$

25. $\Sigma Y^2 =$ **26.** $(\Sigma Y)^2 =$

FRACCIONES Y COMUNES DENOMINADORES

Antes de sumar o dividir fracciones, éstas deben tener el mismo denominador o un común denominador. Una forma sencilla para obtener un común denominador consiste en multiplicar los denominadores. (Este método funciona, aunque no siempre produce el "mínimo común denominador".) Una vez que se encuentra un común denominador, los numeradores se suman o se restan.

Ejemplo:

$$\frac{1}{2} + \frac{1}{3} = \frac{3}{6} + \frac{2}{6} = \frac{5}{6}$$

Ejemplo:

$$\frac{1}{2} - \frac{1}{3} = \frac{3}{6} - \frac{2}{6} = \frac{1}{6}$$

La multiplicación de fracciones es directa: tan sólo se multiplican los numeradores y denominadores. De nuevo, la respuesta a un problema de multiplicación se llama *producto.*

Ejemplo 1: el producto de un medio por un tercio es un sexto.

$$\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

Ejemplo 2:

$$\frac{3}{52} \cdot \frac{4}{51} = \frac{12}{2\,652}$$

En general, para multiplicar fracciones,

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}$$

Al dividir fracciones, nos referimos a la respuesta como *cociente.* En general,

$$\frac{a}{b} \div \frac{c}{d} = \frac{a \cdot c}{b \cdot c}$$

Por ejemplo, el cociente de tres cuartos entre dos tercios es un entero y un octavo:

$$\frac{3}{4} \div \frac{2}{3} = \frac{3 \cdot 3}{4 \cdot 2} = \frac{9}{8} = 1\tfrac{1}{8}$$

Una manera más fácil de tratar con fracciones consiste en transformarlas en números decimales, lo cual se ilustra a continuación.

PROBLEMAS DE REPASO

27. $\frac{5}{8} + \frac{3}{12} =$ **28.** $\frac{13}{22} - \frac{7}{49} =$

29. $\frac{6}{11} + \frac{2}{3} - \frac{1}{8} =$ **30.** $\frac{10}{12} \cdot \frac{4}{5} =$

31. $\frac{13}{25} \cdot \frac{6}{14} \cdot \frac{2}{5}$ **32.** $\frac{16}{18} \div \frac{5}{6} =$

DECIMALES Y POSICIONES DECIMALES

Deci- significa 10, y el sistema numérico decimal se basa en múltiplos de 10. La figura A-1 estipula las posiciones decimales. Cuando un número se multiplica por un múltiplo de 10, el punto decimal simplemente se recorre el número de lugares apropiado hacia la derecha. Cuando un número se divide entre un múltiplo de 10, el punto decimal se recorre hacia la izquierda. La posición del lugar decimal también puede concebirse en función de elevar 10 a una potencia o a un exponente de número entero. Por ejemplo,

$$10^1 = 10, \quad 10^2 = 100, \quad 10^3 = 1\,000, \quad 10^4 = 10\,000, \quad \text{etcétera}$$

Al multiplicar por 10, mueve el punto decimal un lugar a la derecha; si es por 100, dos lugares, y así sucesivamente:

$$(98.49)\,(1\,000) = 98\,490 \qquad (.3587)\,(100) = 35.87$$

Los exponentes negativos implican división. Así, $10^{-1} = .1$, $10^{-2} = .01$, $10^{-3} = .001$, $10^{-4} = .0001$, etc. Al dividir entre 10, mueve el punto decimal un lugar a la izquierda; si es entre 100, dos lugares, y así sucesivamente:

$$(45.91)/100 = .4591$$

$$(.0083)/1\,000 = .0000083$$

FIGURA A-1
Posición de los lugares decimales

Lugares decimales														
X	X	X	X	X	X	X	•	X	X	X	X	X	X	X
Millones	Cientos de millares	Decenas de millares	Millares	Centenas	Decenas	Unidades (entero)	Punto decimal	Décimos	Centésimos	Milésimos	Diezmilésimos	Cienmilésimos	Millonésimos	Diezmillonésimos

PROBLEMAS DE REPASO

Redondea las respuestas a cuatro lugares decimales.

33. $29.869/1\,000 =$ **34.** $(.0388)(10\,000) =$ **35.** $4/1\,000 =$

36. $(1.957)(100) =$ **37.** $(3.503)(10^{-3}) =$ **38.** $(3.503)(10^{3}) =$

Como se expuso en el capítulo 1, es más fácil lidiar con las fracciones en forma decimal. Una fracción se "decimaliza" tan sólo dividiendo el numerador entre el denominador. Para transformar este cociente en un porcentaje, mueve el punto decimal dos lugares a la derecha, lo cual simplemente es cuestión de multiplicar por 100.

Debe tenerse cuidado al utilizar varios lugares decimales en un solo problema. Prueba tu habilidad para manejar los lugares decimales con los siguientes problemas de repaso.

PROBLEMAS DE REPASO

39. $(.15)(4) =$ **40.** $(2.0)(.3) =$ **41.** $(.024 - .03) =$

42. $(.05 + 235.44) =$ **43.** $(34.076 - 6.3) =$ **44.** $(4.141 - .09) =$

LA RELACIÓN DEL NUMERADOR CON EL DENOMINADOR

Dimensionar el tamaño relativo de fracciones, porcentajes y proporciones resulta importante en estadística, porque todo procedimiento estadístico implica fracciones. *Porcentaje* significa "por cien". Si el término *porcentaje* no te transmite un sentido de proporción, toma 100 monedas y lánzalas sobre la cama. Después calcula el porcentaje de las caras simplemente contando el número de ellas.

Las proporciones matemáticas se derivan de las fracciones, y entender la dinámica de las fracciones nos ayuda a tomar sentido de la estadística. Analicemos esta dinámica comparando los tamaños relativos del numerador y el denominador de fracciones y veamos cómo afectan estos tamaños a los cocientes. Dicho análisis revela lo siguiente:

1. Cuando el numerador es pequeño comparado con el denominador, el cociente será pequeño. Como un ejemplo, compara 1/567 con 439/567, dividiendo cada una de estas fracciones para obtener sus proporciones en forma decimal.
2. Cuando el numerador es por lo menos de la mitad del tamaño del denominador, el cociente estará arriba del 50% y ello constituye una mayoría. Por ejemplo, en la recta final de una elección para presidenta del club femenino, si Nancy consigue 51 de los 100 votos, ella gana.
3. Cuando el numerador es casi tan grande como el denominador, el cociente estará cerca de una proporción de 1.0 y un porcentaje de 100%. Por ejemplo, si 222 de 236 estudiantes aprueban un curso,

$$p\text{ [de estudiantes que aprobaron un curso]} = \frac{\#\text{ aprobados}}{\text{tamaño total de la clase}} = \frac{222}{236} = .9407$$

$$\%\text{ [de estudiantes que aprobaron un curso]} = (p)\,(100) = 94.07\%$$

Es decir, por cada 100 estudiantes, aproximadamente 94 aprobaron.

4. Cuando el numerador es más grande que el denominador, el cociente será mayor que 1.
5. En resumen, mientras mayor sea el numerador respecto al denominador, mayor será el cociente.

SOLUCIONES ALGEBRAICAS BÁSICAS

El álgebra implica el uso de símbolos, como letras, para representar un caso matemático general. Se sustituyen números específicos por símbolos para llegar a una respuesta específica. Por ejemplo, es posible definir X como altura y Y como peso, y entender que el peso es una función de la altura: cuanto más alta sea una persona, más alto tenderá a ser su peso. Para estimar el peso para una altura dada, sustituimos valores de X en una ecuación apropiada y calculamos Y. Por ejemplo,

$$Y = -159.31 + (4.62)X$$

Si X es 68 pulgadas, la mejor estimación del peso de esa persona es 154.85 libras:

$$Y = -159.31 + (4.62)\,(68) = 154.85\text{ libras}$$

REGLAS BÁSICAS DE ARITMÉTICA Y ÁLGEBRA

Al multiplicar o dividir dos números, el producto o cociente es positivo si ambos números tienen el mismo signo, pero es negativo si los dos números tienen signos diferentes:

$$(4)\,(3) = 12 \qquad (-4)\,(3) = -12$$

$$(-4)\,(-3) = 12 \qquad (4)\,(-3) = -12$$

$$\frac{4}{3} = 1.25 \qquad \frac{-4}{3} = -1.25 \qquad \frac{4}{-3} = -1.25$$

En general, cualquier número multiplicado por cero es cero. Cualquier número multiplicado por 1 es ese número. Cero dividido entre cualquier número es cero. La división entre cero no se permite (porque produciría un cociente indefinido). Por tanto,

$$0 \cdot a = 0;\text{ por ejemplo, } 0 \cdot 25 = 0 \text{ y } (0)\,(4\,500) = 0$$
$$1 \cdot a = a;\text{ por ejemplo, } 1 \cdot 25 = 25 \text{ y } (1)\,(4\,500) = 4\,500$$
$$0 \div a = 0;\text{ por ejemplo, } 0 \div 25 = 0 \text{ y } (0/4\,500) = 0$$
$$a \div 0 = \text{indefinido (porque algo no puede dividirse entre nada)}$$

Cualquier número dividido entre sí mismo es igual a 1. En general,

$$\frac{a}{a} = 1;\text{ por ejemplo } \frac{34}{34} = 1 \text{ y } \frac{3}{3} = 1$$

El *inverso multiplicativo* de un número es igual a 1 dividido entre ese número. Un número multiplicado por su inverso multiplicativo es igual a 1:

$$a \cdot \frac{1}{a} = 1;\text{ por ejemplo, } 3 \cdot \frac{1}{3} = 1$$

Los términos pueden simplificarse usando las *propiedades conmutativa y distributiva*. En general,

$$(a + b) = (b + a);\text{ por ejemplo, } (2 + 3) = (3 + 2) = 5$$
$$(ab) = (ba);\text{ por ejemplo, } (2 \cdot 3) = (3 \cdot 2) = 6$$
$$a(b + c) = ab + ac;\text{ por ejemplo, } 6(2 + 3) = 6(2) + 6(3) = 30$$
$$a(b - c) = ab - ac;\text{ por ejemplo, } 6(3 - 2) = 6(3) - 6(2) = 6$$

Consistentes con las propiedades y las reglas algebraicas, las ecuaciones exponenciales pueden "desarrollarse" en un conjunto de términos. Por ejemplo,

$$(a + b)^3 = (a + b)(a + b)(a + b) = a^3 + 3a^2b + 3ab^2 + b^3$$

Existe un método simple para desarrollar tales ecuaciones cuando sólo tienen dos números base, a y b (véase "La ecuación de la distribución binomial" en el capítulo 13).

CÁLCULO DE UNA CANTIDAD DESCONOCIDA

Un problema matemático común implica resolver una ecuación para una variable o función desconocidas. Por ejemplo, podrían decirnos que una cantidad X es tal que 3 veces esa cantidad más 100 es igual a 100 menos 6 veces la cantidad más 2 veces la cantidad. En forma simbólica,

$$3X + 30 = 100 - 6X + 2X$$

Nos piden que determinemos la cantidad; es decir, se nos pide "calcular X". Estos tipos de soluciones dependen de la *equivalencia matemática*. Es decir, el signo de igual debe respetarse para que las cantidades en ambos lados del signo igual permanezcan idénticas. Cualquier operación matemática realizada en un lado de la ecuación debe realizarse en el otro lado para mantener la equivalencia matemática. Se llega a las soluciones combinando los términos semejantes y realizando operaciones matemáticas en cada lado de la ecuación. Para calcular X, elegimos operaciones matemáticas que "aíslen" X en un lado de la ecuación. Por tanto,

Empezamos con:	$3X + 30 = 100 - 6X + 2X$
Simplifica combinando los términos semejantes para obtener:	$3X + 30 = 100 - 4X$
Suma $4X$ a ambos lados de la ecuación:	$3X + 30 + 4X = 100 - 4X + 4X$
Para obtener:	$7X + 30 = 100$
Resta 30 de ambos lados:	$7X + 30 - 30 = 100 - 30$
Para obtener:	$7X = 70$
Divide ambos lados entre 7:	$\frac{7X}{7} = \frac{70}{7}$
Para obtener:	$X = 10$

Verifica la exactitud de la respuesta sustituyendo 10 por X en la ecuación original:

	$3X + 30 = 100 - 6X + 2X$
Sustituyendo X por 10:	$3(10) + 30 = 100 - 6(10) + 2(10)$
Resuelve:	$60 = 60$

Por tanto, sabemos que 10 es una solución correcta porque su sustitución por X mantiene la equivalencia matemática.

PROBLEMAS DE REPASO

45. Especificaciones: $Y = a + bX$, $a = 17$, $b = 5$ y $X = -2$. Calcula Y.

46. Si $a = 3$, resuelve la ecuación $(4a^2)(6a^3) =$

47. Especificaciones: $8 - 3X = 4X - 6$. Resuelve X.

48. Especificaciones: $7X - 5X - 4 = 4X - 10$. Resuelve X.

49. $a = \frac{b}{c}$. Resuelve c.

50. Desarrolla la ecuación $(a + b)^2 =$

RESPUESTAS

1. 49 **2.** 441 **3.** 9 261 **4.** 361 **5.** 6 561 **6.** 1 000 **7.** 128 **8.** 6 561 **9.** 100 **10.** 10 **11.** 2.85 **12.** 9 **13.** 28.46 **14.** 90 **15.** 284.60 **16.** 900 **17.** 2.75 **18.** 10 **19.** 10 **20.** 41 **21.** 140 años **22.** 5 136 años cuadrados **23.** 19 600 años cuadrados **24.** 729 libras **25.** 134 635 libras cuadradas **26.** 531 441 libras cuadradas **27.** 21/24 = .8750 **28.** 483/1 078 = .4480 **29.** 287/264 = 1.0871 **30.** 2/3 = .6667 **31.** 156/1 750 = .0891 **32.** 96/90 = 1.0667 **33.** .0299 **34.** 388 **35.** .0040 **36.** 195.7 **37.** .0035 **38.** 3 503 **39.** .60 **40.** .6 **41.** –.006 **42.** 235.49 **43.** 27.776 **44.** 4.051 **45.** $Y = 7$ **46.** 5 832 **47.** $X = 2$ **48.** $X = 3$ **49.** $c = b/a$ **50.** $a^2 + 2ab + b^2$

APÉNDICE B

Tablas estadísticas de probabilidad

TABLA ESTADÍSTICA A Tabla de números aleatorios

9	5	7	3	4	3	9	3	1	1	1	5	6	7	8	2	9	3	5	3	2	5	0	1	4	8	2	4	3	7	3	2	4	2	8
5	7	9	4	6	5	7	5	9	3	6	5	8	5	3	7	2	4	9	3	7	7	5	2	6	1	5	1	6	5	2	6	0	9	4
3	9	9	2	6	8	1	4	8	1	8	7	1	8	3	5	0	7	9	6	0	8	2	2	5	1	8	7	7	3	7	3	1	3	7
3	6	1	4	3	4	8	4	9	7	0	3	1	8	3	6	8	0	9	2	4	9	5	3	8	9	2	9	6	7	3	7	5	1	1
3	2	7	6	1	8	7	3	2	2	2	9	8	4	0	5	4	0	5	7	3	3	2	7	2	8	4	7	0	5	9	3	8	1	1
5	8	3	7	5	7	2	0	1	9	4	9	4	1	8	7	0	4	4	9	1	3	5	9	3	1	1	8	5	0	1	7	8	2	8
6	6	7	9	7	5	5	9	0	8	7	4	1	1	1	6	4	9	3	5	1	0	0	6	0	2	7	9	6	2	5	9	8	3	7
1	0	9	6	5	4	0	7	6	1	2	1	6	0	2	1	5	8	4	1	3	5	0	7	6	9	0	8	1	6	0	8	2	5	9
2	0	8	5	1	0	4	4	0	7	6	2	2	5	5	9	4	6	6	7	4	0	2	5	9	3	5	8	3	9	5	7	6	7	0
0	4	2	5	8	3	4	3	3	5	3	6	3	1	8	7	9	1	1	7	5	6	8	5	9	3	9	3	8	7	7	9	5	2	6
0	5	7	5	5	2	6	0	4	5	0	9	8	6	2	2	3	2	0	8	5	6	4	8	9	7	9	7	6	5	6	2	0	9	5
2	3	8	8	4	9	4	2	2	0	0	8	2	0	4	1	8	6	3	9	6	0	7	2	4	7	7	9	5	9	9	6	9	2	8
1	4	0	2	8	2	5	7	0	2	2	5	6	9	6	5	9	3	2	9	8	1	1	7	8	5	1	2	1	8	9	0	6	5	4
1	6	8	2	3	7	6	4	0	5	9	3	4	7	8	9	0	4	9	0	8	3	4	3	6	6	1	1	4	7	0	8	7	9	6
3	6	9	5	8	1	1	5	3	1	1	3	0	8	4	7	8	5	7	7	0	2	2	4	0	2	1	8	2	1	9	3	1	8	2
1	2	2	3	9	4	2	4	0	1	8	1	0	2	1	1	7	3	6	8	9	8	3	8	7	8	0	2	4	9	0	4	8	8	4
9	3	4	4	9	2	2	5	9	3	3	4	8	0	9	0	6	2	2	4	5	7	6	2	8	0	7	9	2	4	1	5	0	4	7
9	4	5	9	6	9	2	4	1	8	1	3	2	9	7	6	3	0	0	5	6	8	9	1	4	4	2	6	0	8	1	5	3	1	5
1	4	6	4	5	4	2	4	3	6	7	1	4	3	9	4	4	0	5	7	4	0	3	9	0	7	9	5	3	6	6	0	1	6	7
6	8	0	5	5	2	3	5	8	6	4	4	1	0	7	5	5	7	0	3	1	9	6	9	8	6	1	3	4	4	6	9	5	7	9
2	1	7	4	7	0	5	3	6	7	8	6	5	0	6	4	7	6	0	6	2	5	2	7	1	8	3	6	1	8	7	1	2	1	5
3	9	1	8	7	0	5	6	6	7	3	3	0	4	0	2	4	3	5	7	7	3	7	7	1	8	4	2	4	0	9	4	8	3	0
4	8	2	8	4	0	2	7	9	8	6	7	1	0	4	6	1	6	9	7	0	9	1	7	6	6	4	3	8	6	5	9	6	0	9
5	1	9	8	4	8	4	6	4	1	4	0	1	8	2	3	7	5	3	0	7	9	5	1	3	8	0	1	6	8	9	2	0	6	5
3	7	4	9	7	9	4	4	7	2	3	9	3	8	4	9	2	2	1	8	5	9	7	7	1	0	6	0	2	0	7	8	0	5	6
5	9	1	1	0	6	5	5	0	4	7	8	6	8	7	8	8	6	3	1	3	1	4	6	0	8	5	5	2	8	9	7	4	6	0
2	2	6	4	9	7	8	8	4	3	6	9	6	3	7	6	3	1	2	5	4	3	0	5	1	1	5	6	6	6	1	2	9	4	1
6	5	2	2	5	3	4	3	3	9	1	3	4	2	2	0	7	5	1	9	2	7	9	5	6	0	9	4	9	4	6	0	3	3	3
1	7	1	4	3	3	3	4	8	7	8	2	2	8	3	9	3	2	4	1	6	5	1	6	5	7	5	5	9	9	5	6	1	4	5
9	6	7	5	3	4	9	2	2	8	1	5	0	1	9	3	8	4	5	4	6	3	4	3	4	4	5	8	4	2	9	3	3	7	3
5	0	4	1	9	6	3	9	2	3	2	5	7	8	9	2	8	0	6	3	6	0	6	7	4	6	3	8	0	8	9	4	2	9	5
9	5	6	8	8	4	0	2	8	9	3	9	7	5	3	9	1	7	3	9	3	2	7	8	3	4	7	3	0	8	0	5	1	4	2
1	9	4	6	0	7	1	9	6	6	2	9	8	0	4	8	7	4	6	4	7	4	8	2	9	6	1	5	0	8	0	8	3	6	7
9	3	0	9	9	7	0	5	5	7	3	4	5	5	0	3	0	2	8	4	7	4	6	2	1	9	4	1	9	6	2	8	3	3	4
6	7	6	6	1	2	7	5	4	5	6	7	5	9	5	3	4	2	4	0	1	0	3	8	7	5	9	5	2	3	7	1	0	0	0
3	8	7	5	1	2	2	0	1	3	5	2	2	5	5	0	4	8	0	8	9	7	6	4	3	7	9	5	0	2	9	6	7	2	1
3	9	3	2	4	1	1	1	6	5	7	4	2	6	5	0	0	2	2	0	0	1	5	0	2	1	6	8	6	7	9	7	3	2	5

TABLA ESTADÍSTICA B Tabla de la distribución normal

Áreas bajo la curva normal (columnas B y C) para puntuaciones Z estipuladas (columna A)

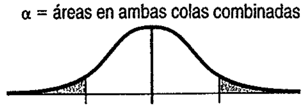
(A) +Z o −Z	(B)	(C)
0.00	.0000	.5000
0.01	.0040	.4960
0.02	.0080	.4920
0.03	.0120	.4880
0.04	.0160	.4840
0.05	.0199	.4801
0.06	.0239	.4761
0.07	.0279	.4721
0.08	.0319	.4681
0.09	.0359	.4641
0.10	.0398	.4602
0.11	.0438	.4562
0.12	.0478	.4522
0.13	.0517	.4483
0.14	.0557	.4443
0.15	.0596	.4404
0.16	.0636	.4364
0.17	.0675	.4325
0.18	.0714	.4286
0.19	.0753	.4247
0.20	.0793	.4207
0.21	.0832	.4168
0.22	.0871	.4129
0.23	.0910	.4090
0.24	.0948	.4052
0.25	.0987	.4013
0.26	.1026	.3974
0.27	.1064	.3936
0.28	.1103	.3897
0.29	.1141	.3859
0.30	.1179	.3821
0.31	.1217	.3783
0.32	.1255	.3745
0.33	.1293	.3707
0.34	.1331	.3669
0.35	.1368	.3632
0.36	.1406	.3594
0.37	.1443	.3557
0.38	.1480	.3520
0.39	.1517	.3483
0.40	.1554	.3446
0.41	.1591	.3409
0.42	.1628	.3372
0.43	.1664	.3336
0.44	.1700	.3300
0.45	.1736	.3264
0.46	.1772	.3228
0.47	.1808	.3192
0.48	.1844	.3156
0.49	.1879	.3121
0.50	.1915	.3085
0.51	.1950	.3050
0.52	.1985	.3015
0.53	.2019	.2981
0.54	.2054	.2946
0.55	.2088	.2912
0.56	.2123	.2877
0.57	.2157	.2843
0.58	.2190	.2810
0.59	.2224	.2776
0.60	.2257	.2743
0.61	.2291	.2709
0.62	.2324	.2676
0.63	.2357	.2643
0.64	.2389	.2611
0.65	.2422	.2578
0.66	.2454	.2546
0.67	.2486	.2514
0.68	.2517	.2483
0.69	.2549	.2451
0.70	.2580	.2420
0.71	.2611	.2389
0.72	.2642	.2358
0.73	.2673	.2327
0.74	.2704	.2296
0.75	.2734	.2266
0.76	.2764	.2236
0.77	.2794	.2206
0.78	.2823	.2177
0.79	.2852	.2148
0.80	.2881	.2119
0.81	.2910	.2090
0.82	.2939	.2061
0.83	.2967	.2033
0.84	.2995	.2005
0.85	.3023	.1977
0.86	.3051	.1949
0.87	.3078	.1922
0.88	.3106	.1894
0.89	.3133	.1867
0.90	.3159	.1841
0.91	.3186	.1814
0.92	.3212	.1788
0.93	.3238	.1762
0.94	.3264	.1739
0.95	.3289	.1711
0.96	.3315	.1685
0.97	.3340	.1660
0.98	.3365	.1635
0.99	.3389	.1611
1.00	.3413	.1587
1.01	.3438	.1562
1.02	.3461	.1539
1.03	.3485	.1515
1.04	.3508	.1492
1.05	.3531	.1469
1.06	.3554	.1446
1.07	.3577	.1423
1.08	.3599	.1401
1.09	.3621	.1379
1.10	.3643	.1357
1.11	.3665	.1335
1.12	.3686	.1314
1.13	.3708	.1292
1.14	.3729	.1271
1.15	.3749	.1251
1.16	.3770	.1230
1.17	.3790	.1210
1.18	.3810	.1190
1.19	.3830	.1170
1.20	.3849	.1151
1.21	.3869	.1131
1.22	.3888	.1112
1.23	.3907	.1093
1.24	.3925	.1075
1.25	.3944	.1056
1.26	.3962	.1038
1.27	.3980	.1020
1.28	.3997	.1003
1.29	.4015	.0985
1.30	.4032	.0968
1.31	.4049	.0951
1.32	.4066	.0934
1.33	.4082	.0918
1.34	.4099	.0901
1.35	.4115	.0885
1.36	.4131	.0869
1.37	.4147	.0853
1.38	.4162	.0838
1.39	.4177	.0823
1.40	.4192	.0808
1.41	.4207	.0793
1.42	.4222	.0778
1.43	.4236	.0764
1.44	.4251	.0749
1.45	.4265	.0735
1.46	.4279	.0721
1.47	.4292	.0708
1.48	.4306	.0694
1.49	.4319	.0681
1.50	.4332	.0668
1.51	.4345	.0655
1.52	.4357	.0643
1.53	.4370	.0630
1.54	.4382	.0618
1.55	.4394	.0606
1.56	.4406	.0594
1.57	.4418	.0582
1.58	.4429	.0571
1.59	.4441	.0559
1.60	.4452	.0548
1.61	.4463	.0537
1.62	.4474	.0526
1.63	.4484	.0516
1.64	.4495	.0505
1.65	.4505	.0495
1.66	.4515	.0485
1.67	.4525	.0475
1.68	.4535	.0465
1.69	.4545	.0455
1.70	.4554	.0446
1.71	.4564	.0436
1.72	.4573	.0427
1.73	.4582	.0418
1.74	.4591	.0409
1.75	.4599	.0401
1.76	.4608	.0392
1.77	.4616	.0384
1.78	.4625	.0375
1.79	.4633	.0367
1.80	.4641	.0359
1.81	.4649	.0351
1.82	.4656	.0344
1.83	.4664	.0336
1.84	.4671	.0329
1.85	.4678	.0322
1.86	.4686	.0314
1.87	.4693	.0307
1.88	.4699	.0301
1.89	.4706	.0294
1.90	.4713	.0287
1.91	.4719	.0281
1.92	.4726	.0274
1.93	.4732	.0268
1.94	.4738	.0262
1.95	.4744	.0256
1.96	.4750	.0250
1.97	.4756	.0244
1.98	.4761	.0239
1.99	.4767	.0233
2.00	.4772	.0228
2.01	.4778	.0222
2.02	.4783	.0217
2.03	.4788	.0212
2.04	.4793	.0207
2.05	.4798	.0202
2.06	.4803	.0197
2.07	.4808	.0192
2.08	.4812	.0188
2.09	.4817	.0183
2.10	.4821	.0179
2.11	.4826	.0174
2.12	.4830	.0170
2.13	.4834	.0166
2.14	.4838	.0162
2.15	.4842	.0158
2.16	.4846	.0154
2.17	.4850	.0150
2.18	.4854	.0146
2.19	.4857	.0143
2.20	.4861	.0139
2.21	.4864	.0136
2.22	.4868	.0132
2.23	.4871	.0129
2.24	.4875	.0125
2.25	.4878	.0122
2.26	.4881	.0119
2.27	.4884	.0116
2.28	.4887	.0113
2.29	.4890	.0110
2.30	.4893	.0107
2.31	.4896	.0104
2.32	.4898	.0102
2.33	.4901	.0099
2.34	.4904	.0096
2.35	.4906	.0094
2.36	.4909	.0091
2.37	.4911	.0089
2.38	.4913	.0087
2.39	.4916	.0084
2.40	.4918	.0082
2.41	.4920	.0080
2.42	.4922	.0078
2.43	.4925	.0075
2.44	.4927	.0073
2.45	.4929	.0071
2.46	.4931	.0069
2.47	.4932	.0068
2.48	.4934	.0066
2.49	.4936	.0064
2.50	.4938	.0062
2.51	.4940	.0060
2.52	.4941	.0059
2.53	.4943	.0057
2.54	.4945	.0055
2.55	.4946	.0054
2.56	.4948	.0052
2.57	.4949	.0051
2.58	.4951	.0049
2.59	.4952	.0048
2.60	.4953	.0047
2.61	.4955	.0045
2.62	.4956	.0044
2.63	.4957	.0043
2.64	.4959	.0041
2.65	.4960	.0040
2.66	.4961	.0039
2.67	.4962	.0038
2.68	.4963	.0037
2.69	.4964	.0036
2.70	.4965	.0035
2.71	.4966	.0034
2.72	.4967	.0033
2.73	.4968	.0032
2.74	.4969	.0031
2.75	.4970	.0030
2.76	.4971	.0029
2.77	.4972	.0028
2.78	.4973	.0027
2.79	.4974	.0026
2.80	.4974	.0026
2.81	.4975	.0025
2.82	.4976	.0024
2.83	.4977	.0023
2.84	.4977	.0023
2.85	.4978	.0022
2.86	.4979	.0021
2.87	.4979	.0021
2.88	.4980	.0020
2.89	.4981	.0019
2.90	.4981	.0019
2.91	.4982	.0018
2.92	.4982	.0018
2.93	.4983	.0017
2.94	.4984	.0016
2.95	.4984	.0016
2.96	.4985	.0015
2.97	.4985	.0015
2.98	.4986	.0014
2.99	.4986	.0014
3.00	.4987	.0013
3.01	.4987	.0013
3.02	.4987	.0013
3.03	.4988	.0012
3.04	.4988	.0012
3.05	.4989	.0011
3.06	.4989	.0011
3.07	.4989	.0011
3.08	.4990	.0010
3.09	.4990	.0010
3.10	.4990	.0010
3.11	.4991	.0009
3.12	.4991	.0009
3.13	.4991	.0009
3.14	.4992	.0008
3.15	.4992	.0008
3.16	.4992	.0008
3.17	.4992	.0008
3.18	.4993	.0007
3.19	.4993	.0007
3.20	.4993	.0007
3.21	.4993	.0007
3.22	.4994	.0006
3.23	.4994	.0006
3.24	.4994	.0006
3.25	.4994	.0006
3.30	.4995	.0005
3.35	.4996	.0004
3.40	.4997	.0003
3.45	.4997	.0003
3.50	.4998	.0002
3.60	.4998	.0002
3.70	.4999	.0001
3.50	.4999	.0001
3.90	.49995	.00005
4.0	.49997	.00003

Fuente: Tabla III, página 45 de *Statistical Tables for Biological, Agricultural and Medical Research.* Copyright 1963. R.A. Fisher y S. Yates. Reimpreso con autorización de Pearson Education Limited.

TABLA ESTADÍSTICA C Tabla de la distribución t

Valores críticos de $t(t_\alpha)$ para los niveles de significancia y los grados de libertad especificados

α = áreas en ambas colas combinadas



Prueba de dos colas o no direccional
Nivel de significancia

df	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$
1	12.706	63.657	636.62
2	4.303	9.925	31.598
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.335	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587
11	2.201	3.105	4.437
12	2.179	3.055	4.318
13	2.160	3.012	4.221
14	2.145	2.977	4.140
15	2.131	2.947	4.073
16	2.120	2.921	4.015
17	2.110	2.898	3.965
18	2.101	2.878	3.922
19	2.093	2.861	3.883
20	2.086	2.845	3.850
21	2.080	2.831	3.819
22	2.074	2.819	3.792
23	2.069	2.807	3.767
24	2.064	2.797	3.745
25	2.060	2.787	3.725
26	2.056	2.779	3.707
27	2.052	2.771	3.690
28	2.048	2.763	3.674
29	2.045	2.756	3.659
30	2.042	2.750	3.646
40	2.021	2.704	3.551
60	2.000	2.660	3.460
120	1.980	2.617	3.373
∞	1.96	2.58	3.30

α = áreas en una cola de la curva

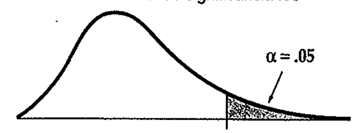
Prueba de una cola o direccional
Nivel de significancia

df	$\alpha = .05$	$\alpha = .01$	$\alpha = .001$
1	6.314	31.821	318.31
2	2.920	6.965	22.326
3	2.353	4.541	10.213
4	2.132	3.747	7.173
5	2.015	3.365	5.893
6	1.943	3.143	5.208
7	1.895	2.998	4.785
8	1.860	2.896	4.501
9	1.833	2.821	4.297
10	1.812	2.764	4.144
11	1.796	2.718	4.025
12	1.782	2.681	3.930
13	1.771	2.650	3.852
14	1.761	2.624	3.787
15	1.753	2.602	3.733
16	1.746	2.583	3.686
17	1.740	2.567	3.646
18	1.734	2.552	3.610
19	1.729	2.539	3.579
20	1.725	2.528	3.552
21	1.721	2.518	3.527
22	1.717	2.508	3.505
23	1.714	2.500	3.485
24	1.711	2.492	3.467
25	1.706	2.485	3.450
26	1.705	2.479	3.435
27	1.703	2.473	3.421
28	1.701	2.467	3.408
29	1.699	2.462	3.396
30	1.697	2.457	3.385
40	1.684	2.423	3.307
60	1.671	2.390	3.232
120	1.658	2.358	3.160
∞	1.64	2.33	3.096

Fuente: Tabla III, página 46 de *Statistical Tables for Biological, Agricultural and Medical Research.* Copyright 1963. R.A. Fisher y S. Yates. Reimpreso con autorización de Pearson Education Limited.

TABLA ESTADÍSTICA D

Valores críticos de la distribución de la razón F al nivel de significancia .05



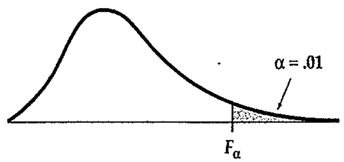
gl para el denominador — *gl* para el numerador $\alpha = .05$

	1	2	3	4	5	6	8	12
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28
21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.25
22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.23
23	4.28	3.42	3.03	2.80	2.64	2.53	2.38	2.20
24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.18
25	4.24	3.38	2.99	2.76	2.60	2.49	2.34	2.16
26	4.22	3.37	2.98	2.74	2.59	2.47	2.32	2.15
27	4.21	3.35	2.96	2.73	2.57	2.46	2.30	2.13
28	4.20	3.34	2.95	2.71	2.56	2.44	2.29	2.12
29	4.18	3.33	2.93	2.70	2.54	2.43	2.28	2.10
30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.09
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.00
60	4.00	3.15	2.76	2.52	2.37	2.25	2.10	1.92
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.83
∞	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.75

Fuente: Tomado de la tabla 18 de Pearson y Hartley (1976: 171), *Biometrika Tables for Statisticians,* volumen 1, Londres: Biometrika Trust. Con autorización de Oxford University Press.

TABLA ESTADÍSTICA E

Valores críticos de la distribución de la razón F al nivel de significancia .01



gl para el denominador — *gl* para el numerador α = .01

	1	2	3	4	5	6	8	12
1	4 052	4 999.5	5 403	5 625	5 764	5 859	5 981	6 106
2	98.49	90.01	99.17	99.25	99.30	99.33	99.36	99.42
3	34.12	30.81	29.46	28.71	28.24	27.91	27.49	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.27	9.89
6	13.74	10.92	9.78	9.15	8.75	8.47	8.10	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.47
8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.71
11	9.65	7.20	6.22	5.67	5.32	5.07	4.74	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.16
13	9.07	6.70	5.74	5.20	4.86	4.62	4.30	3.96
14	8.86	6.51	5.56	5.03	4.69	4.46	4.14	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.55
17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.45
18	8.28	6.01	5.09	4.58	4.25	4.01	3.71	3.37
19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.30
20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.23
21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.17
22	7.94	5.72	4.82	4.31	3.99	3.76	3.45	3.12
23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.07
24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.03
25	7.77	5.57	4.68	4.18	3.86	3.63	3.32	2.99
26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	2.96
27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	2.93
28	7.64	5.45	4.57	4.01	3.75	3.53	3.23	2.90
29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	2.87
30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.84
40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.66
60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.50
120	6.85	4.79	3.95	6.48	3.17	2.96	2.66	2.34
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.18

Fuente: Tomado de la tabla 18 de Pearson y Hartley (1976: 173), *Biometrika Tables for Statisticians*, volumen 1, Londres: Biometrika Trust. Con autorización de Oxford University Press.

TABLA ESTADÍSTICA F

Valores q de las pruebas de rango a los niveles de significancia .05 y .01

gl para VCM_d $(n-k)$	Nivel de significancia	k = número de medias grupales comparadas									
	α	2	3	4	5	6	7	8	9	10	11
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30
7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61
	.01	4.36	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13
12	.05	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79
14	.05	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46
17	.05	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01
	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02
30	.05	2.89	3.49	9.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82
	.01	3.82	4.37	4.70	4.33	5.11	5.26	5.39	5.50	5.60	5.69
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53
120	.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23

Fuente: Tomado de la tabla 29 de Pearson y Hartley (1976: 192-193), *Biometrika Tables for Statisticians*, volumen 1, Londres: Biometrika Trust. Con autorización de Oxford University Press.

TABLA ESTADÍSTICA G

Valores críticos de la distribución de la chi cuadrada

	Nivel de significancia			
gl	χ^2 crítica, $\alpha = .10$	χ^2 crítica, $\alpha = .05$	χ^2 crítica, $\alpha = .01$	χ^2 crítica, $\alpha = .001$
1	2.71	3.84	6.64	10.83
2	4.50	5.99	9.21	13.82
3	6.25	7.81	11.34	16.27
4	7.78	9.49	13.28	18.47
5	9.24	11.07	15.09	20.52
6	10.64	12.59	16.81	22.46
7	12.02	14.07	18.48	24.32
8	13.36	15.51	20.09	26.12
9	14.68	16.92	21.67	27.88
10	15.99	18.31	23.21	29.59
11	17.28	19.68	24.72	31.26
12	18.55	21.03	26.22	32.91
13	19.81	22.36	27.69	34.53
14	21.06	23.68	29.14	36.12
15	22.31	25.00	30.58	37.70
16	23.54	26.30	32.00	39.25
17	24.77	27.59	33.41	40.79
18	25.99	28.87	34.80	42.31
19	27.20	30.14	36.19	43.82
20	28.41	31.41	37.57	45.32
21	29.62	32.67	38.93	46.80
22	30.81	33.92	40.29	48.27
23	32.01	35.17	41.64	49.73
24	33.20	36.42	42.98	51.18
25	34.38	37.65	44.31	52.62
26	35.56	38.88	45.64	54.05
27	36.74	40.11	45.96	55.48
28	37.92	41.34	48.28	56.89
29	39.09	42.56	49.59	58.30
30	40.25	43.77	50.89	59.70
40	51.80	55.76	63.69	73.40
50	63.17	67.50	75.15	86.66
60	74.40	79.08	88.38	99.61
70	85.53	90.53	100.42	112.32

Fuente: Tomado de la tabla 8 de Pearson y Hartley (1976: 137), *Biometrika Tables for Statisticians,* volumen 1, Londres: Biometrika Trust. Con autorización de Oxford University Press.

APÉNDICE C

Respuestas a ejercicios seleccionados de los capítulos

Éstas son respuestas parciales. Asegúrate de mostrar todos los procedimientos, incluyendo fórmulas y curvas. Ve las respuestas de otros conjuntos de problemas para un capítulo para obtener una guía adicional.

CAPÍTULO 1

Conjunto de problemas 1A

1A-1. *a*) 13.64% *e*) $\frac{2\,321}{10\,000}$

1A-3. 56 227 presos de baja seguridad; 15 691 presos de alta seguridad
1A-5. .8947, 89.47% 1A-7. 30 081 por 100 000 nunca casados; 2 464 por 100 000 separados; 9 490 por 100 000 divorciados

Conjunto de problemas 1B

1B-2. Byron comenzará; tiene un porcentaje de lanzamiento de 43.75% comparado con 35.42% de David 1B-4. .0621 edades 15-24 años; .2219 edades 35-44 años; .1381 edades 55-64 años 1B-6. Hombres: edades 21-30, $p = .0992$, 5 asistirán a las Olimpiadas; edades 41-50, $p = .3421$, 17 asistirán a las Olimpiadas; Mujeres: edades 21-30, $p = .1307$, 7 asistirán a las Olimpiadas; edades 41-50, $p = .2876$, 14 asistirán a las Olimpiadas 1B-8. $p = .2441$, 24.41% sin cursos profesionales; $p = .2142$, 21.42% con escuela técnica

Conjunto de problemas 1C

1C-1. *a*) $\frac{6\,046}{10\,000}$, $p = .6046$ *d*) $p = .2136$, 21.36%

1C-3. *a*) 7 764 crímenes contra personas *c*) 4 340 actos de intimidación
1C-5. $p = .7859$, 78.59% 1C-7. Anderson, Indiana: 3 047 por 100 000 de población; Duluth, Minnesota: 1 921 por 100 000 de población

Conjunto de problemas 1D

1D-2. Alabama ($p = .1304$). 1D-4. Anne. 1D-6. p[Departamento de Personal de Los Ángeles a atender] = .1558; número a atender = 6. 1D-8. *a*) Mujeres ($p = .1423$); *b*) Mujeres; .7297 regresó a trabajar.

CAPÍTULO 2

Conjunto de problemas 2A

2A-1. *a*) Nivel de razón *c*) Nivel de intervalo *g*) Nivel nominal
2A-3. *a*) No inclusiva. No hay lugar para calificar una respuesta de Cirugía General, Pediatría, etc. Para mejorarla: agrega una categoría "otros" *c*) No inclusiva; no hay lugar para calificar ingresos de $41 000 a $55 000. Para mejorarla, cambia la categoría $26 000-$40 000 a $26 000-$50 000. No exclusiva; el ingreso de $100 000 podría incluirse en dos categorías. Para mejorarla, cambia la categoría $100 000-$150 000 a $101 000-$150 000 2A-5. *a*) .0585-.0595 *d*) 5 350-5 450 *f*) 3 años a 3 años, 364 días 2A-7. *a*)

Número de vehículos registrados	*f*	*f* proporcional	Porcentaje (%) *f*	Porcentaje acumulado (%) *f*
0	1	.0500	5.00	5.00
1	3	.1500	15.00	20.00
2	8	.4000	40.00	60.00
3	3	.1500	15.00	75.00
4	5	.2500	25.00	100.00
Totales	20	1.0000	100.00	

b) El percentil 75o. Este hogar tiene tantos vehículos o más que 75% de todos los hogares.

Conjunto de problemas 2B

2B-2. *a*) Nivel de razón *c*) Nivel ordinal *f*) Nivel de intervalo
2B-4. *a*) 28.3 *d*) 25.638 *g*) 30 2B-6. 1.46:1
2B-8. *a*) $p = .7857$, 79%. Por consiguiente, el rango percentilar de Jennifer es 79. Ella obtuvo una calificación igual o mayor que el 79% de sus compañeros de clases.

Conjunto de problemas 2C

2C-1. *a*) Nivel de razón *c*) Nivel nominal *d*) Nivel de intervalo
2C-3. *a*) No inclusiva; no hay respuestas para individuos con edades menores que 25. Para mejorarla, agrega esta categoría de respuesta. No exclusiva; los individuos con edad de 74 pueden entrar en cualquiera de las últimas dos categorías de respuesta. Para mejorarla, crea categorías de respuesta que no se superpongan (por ejemplo, 65-74, 75 o más, etcétera).

c) No exclusiva; individuos con 9 o 15 años de educación pueden caer en más de una categoría de respuesta. Para mejorarla, crea categorías de respuesta que no se superpongan (es decir, menos que 9 años, más que 15 años, etc.)
2C-5. *b*) .023805-.023815 *e*) 6.5-7.5 años *g*) 7.5-8.5
2C-7. *a*)

Puntuaciones en la escala CES-D	*f*	*f* proporcional	Porcentaje (%) *f*	Porcentaje acumulado (%) *f*
2	3	.1500	15.00	15.00
3	3	.1500	15.00	30.00
4	5	.2500	25.00	55.00
5	4	.2000	20.00	75.00
6	3	.1500	15.00	90.00
7	2	.1000	10.00	100.00
Totales	20	1.0000	100.00	

b) El percentil 90o. Este trabajador obtuvo una puntuación tan alta o mayor que 90% de los trabajadores que participaron en este estudio.

Conjunto de problemas 2D

2D-2. *a*) Nivel ordinal *c*) Nivel nominal *e*) Nivel de intervalo
2D-4. *a*) 5.5 *d*. 400 *f*) 500 000 2D-6 2.82:1
2D-8. $p = .8500$, 85%. Por consiguiente, el rango percentilar de Jeff es 85. Obtuvo una puntuación igual o mayor que 85% de sus compañeros de clases.

CAPÍTULO 3

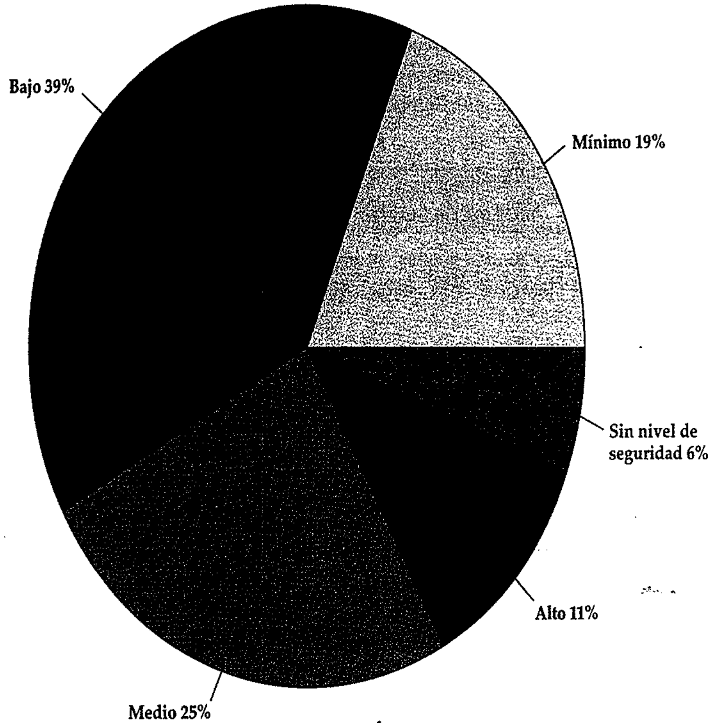
Conjunto de problemas 3A

3A-1.

Nivel de seguridad	Porcentaje (%)	*p*	$p \times 360$
Mínimo	19.4	.194	70
Bajo	38.9	.389	140
Medio	24.8	.248	89
Alto	10.7	.107	39
Sin nivel de seguridad	6.1	.061	22
Totales	99.9*	.999*	360°

*Los totales pueden no sumar 100% o 1.000 debido al error de redondeo.

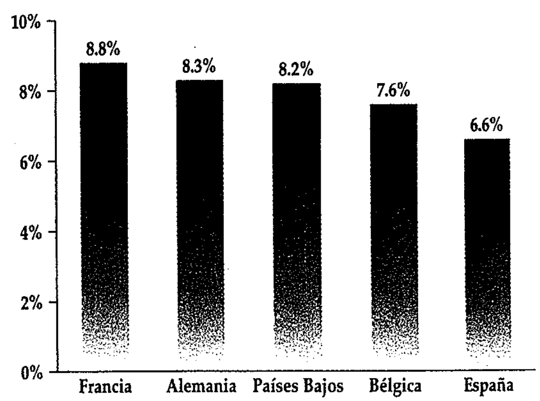
GRÁFICO 3A-1
Prisioneros en Estados Unidos por nivel de seguridad, 2003



Nota: Los porcentajes se redondearon para simplificar la presentación.

3A-2. Gráfico de barras:

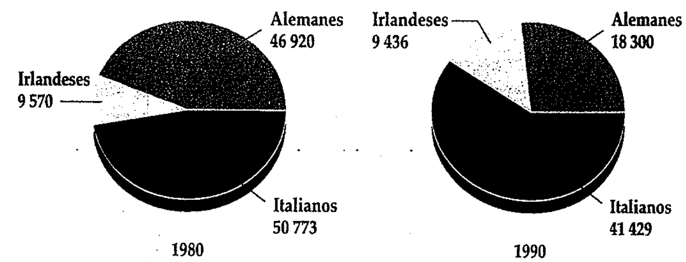
GRÁFICO 3A-2
Porcentaje del producto interno bruto gastado en cuidado de la salud en los países europeos seleccionados



3A-3. Gráfico de barras agrupadas: véase el gráfico 3C-3 más adelante para el trazado básico. Comentario: parece que las mujeres en Estados Unidos poseen más diplomas de bachillerato y títulos de escuelas de formación profesional, mientras los hombres parecen poseer más títulos de licenciatura y profesionales.

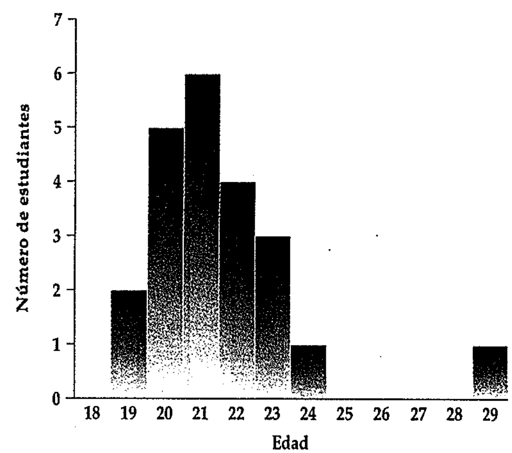
3A-4. *a*) Gráficos de pastel:

GRÁFICO 3A-4a
Composición de barrios étnicos blancos en la ciudad de Nueva York, 1980 y 1990



3A-5. *a*) Histograma de frecuencias:

GRÁFICO 3A-5a
Edades de los estudiantes en el equipo de debate de la universidad

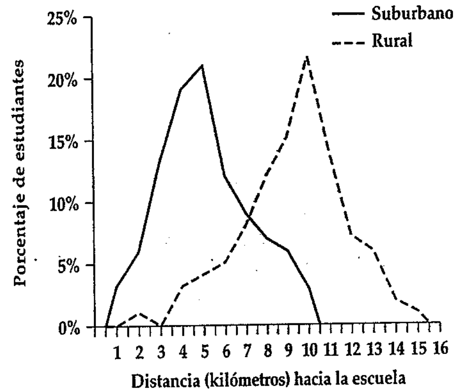


b) Véase el ejercicio 3B-6 para obtener sugerencias. *c*) Probablemente el histograma. Con un tamaño de la muestra pequeño, en el histograma es más fácil asociar la frecuencia (número de estudiantes) con cada edad. *d*) El estudiante con 29 años de edad es un valor extremo.

3A-6. *a*) Polígonos de frecuencias superpuestos (gráficos de línea):

GRÁFICO 3A-6a

Una comparación de las distancias recorridas hacia la escuela por estudiantes suburbanos y rurales

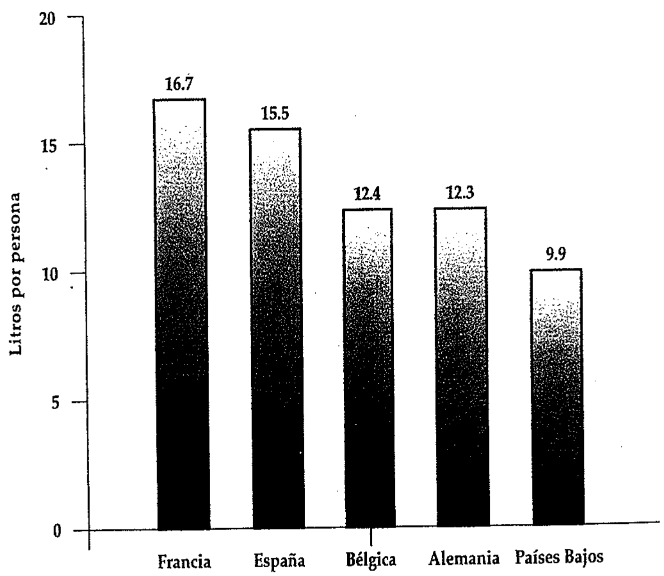


Conjunto de problemas 3B

3B-1. Gráfico de pastel: véase el gráfico de pastel del ejercicio 3A-1 en la página 606 para el trazado básico.
3B-2. Gráfico de barras:

GRÁFICO 3B-2

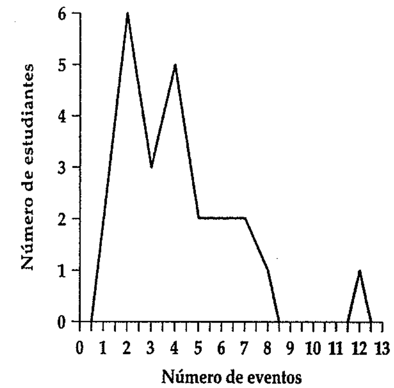
Litros de alcohol consumidos por adultos mayores de 24 años en 1990 en los países europeos seleccionados



Comentario: el consumo promedio de alcohol es bastante más alto en Francia y España.
3B-4. *a*) Gráficos de pastel: véase el gráfico 3A-4a en la página 607 para el trazado básico. *b*) Véase el gráfico 3C-3a en la página 610 para el trazado básico. *c*) Aunque las investigaciones de patrones y contrabando por lo general declinaron entre 1992 y 2002 dentro del Programa de Investigaciones INS, el número total de investigaciones criminales se incrementó de manera significativa. El gráfico de barras agrupadas es mejor que los gráficos de pastel para describir estos cambios. El gráfico de barras no sólo transmite una comparación de los años sino también los cambios proporcionales generales entre tipos diferentes de investigaciones INS. Para describir del mismo modo estos cambios, los gráficos de pastel tendrían que ser de tamaños diferentes con base en la proporción de las investigaciones totales emprendidas en un año determinado. Dos gráficos de pastel de tamaños diferentes parecerían extraños.
3B-5. *b*) Polígono de frecuencia (gráfico de línea):

GRÁFICO 3B-5b

Participación en eventos del campus



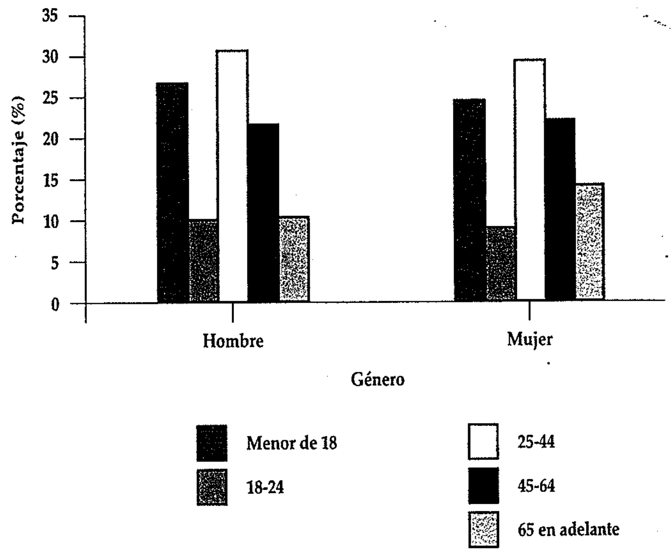
3B-6. *a*) Polígonos de frecuencia superpuestos (gráficos de línea); véase el gráfico 3A-6a en la página 608 para el trazado básico. En vista que los tamaños de la muestra difieren, grafica las frecuencias del porcentaje.
b) Los adictos a la cocaína tienden a ser más jóvenes que los adictos al alcohol.

Conjunto de problemas 3C

3C-1. *a*) Gráfico de pastel; véase el gráfico 3A-1 en la página 606 para el trazado básico.

3C-3. *a*) Gráfico de barras agrupadas:

GRÁFICO 3C-3a
Distribuciones de edades de hombres y mujeres en Estados Unidos, 2000



b) Comentario: hay porcentajes similares de hombres y mujeres excepto en el grupo de edad de 65 años en adelante donde hay más mujeres (lo que refleja el hecho que las mujeres tienden a vivir más que los hombres).

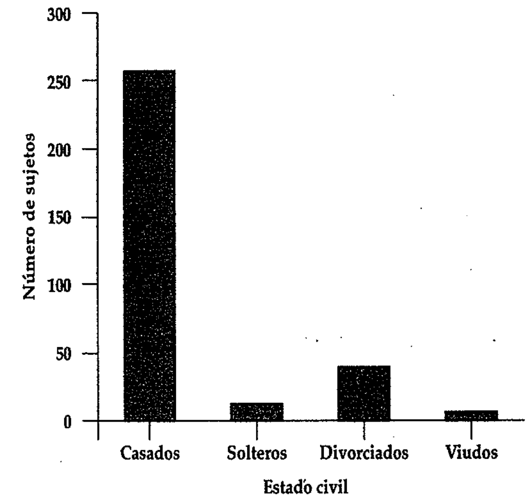
3C-5. *a*) Histogramas de frecuencias: véase el gráfico 3A-5a en la página 607 para el trazado básico. *b*) Véase la respuesta parcial para el ejercicio 3B-5b en la página 609 para obtener sugerencias. *c*) Probablemente el histograma. Con un tamaño de muestra pequeño como éste y el rango pequeño de las puntuaciones, en el histograma es más fácil asociar la frecuencia (número de estudiantes) con cada edad. *d*) El estudiante con 28 años de edad es un valor extremo.

3C-6. *a*) Polígonos de frecuencias superpuestos (gráficos de línea): véase el gráfico 3A-6a en la página 108 para el trazado básico. En vista que difieren los tamaños de la muestra, grafica las frecuencias de los porcentajes.

Conjunto de problemas 3D

3D-2. *a*) Gráfico de barras:

GRÁFICO 3D-2a
Estado civil de los adultos de la República Checa



Comentario: la abrumadora mayoría de participantes en el estudio en la República Checa están casados.

3D-3. *a*) Gráfico de barras agrupadas: véase el gráfico 3C-3a en la página 610 para el trazado básico.

3D-4. *a*) Gráficos de pastel: véase el gráfico 3A-4a en la página 607 para el trazado básico. *b*) Véase el gráfico 3C-3a en la página 610 para el trazado básico. *c*) En general, ambos tipos de gráficos indican un crecimiento considerable de la clase media alta al igual que una disminución marcada en la clase trabajadora. Sin embargo, el gráfico de barras agrupadas es más apropiado para describir este fenómeno, ya que es capaz de mostrar cambios más sutiles en la composición de la clase social durante los dos años seleccionados.

3D-6. *a*) Polígono de frecuencias superpuestas; véase el gráfico 3A-6a en la página 608 para el trazado básico.
b) El centro I parece estar compuesto por jubilados más jóvenes, mientras el centro II parece estar conformado por jubilados que son un poco mayores.

CAPÍTULO 4

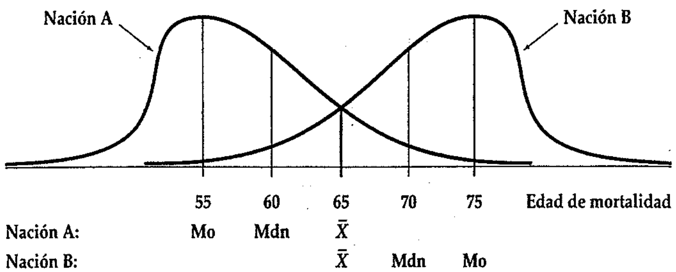
Conjunto de problemas 4A

4A-1. Hoja de trabajo con los cálculos:

X	*X (cont.)*	*X (rango)*	*X (rango, cont.)*
...	...	...	...

Moda = 19 años; Mediana = 19 años; Media = 18.9 años

4A-3. *a*) Curvas de frecuencias:



b) La nación B parece mejor. Su sesgo hacia la izquierda revela que menos personas mueren a edades más tempranas.
4A-5. *a*) Mediana = \$11 300; Media = \$14 160 *d*) Media ajustada = \$10 450
4A-7. Hoja de cálculo:
X = Tiempos de reacción del pollo

X (rango)
.88
...
...
1.45

b) Mediana = .96 segundos; Moda = .93 segundos *c*) Sesgo a la derecha

Conjunto de problemas 4B

4B-2. *a*) Mediana = 625 puntos; Media = 631 puntos 4B-4. *b*) Mediana = 2.2 puntos GPA; Media = 2.39 puntos GPA *e*) Media ajustada = 2.13 puntos GPA
4B-6. Media = 34.82 años 4B-8. Altura: sesgo a la derecha; Presupuesto de comestibles: sesgo a la derecha

Conjunto de problemas 4C

4C-1. Moda = 70 pulgadas; Mediana = 69.5 pulgadas; Media = 68.7 años
4C-3. *a*) Curvas de frecuencias: véase la respuesta parcial al ejercicio 4A-3 anterior para el trazado básico. *b*) La clase B parece haber obtenido mejores puntuaciones en este examen. Su sesgo a la izquierda revela más estudiantes que obtuvieron calificaciones más altas en el examen.
4C-5. Y = calificación del empleado, hoja de cálculo:

Y	Y (rango)
8	3
...	...
...	...
7	9

b) Media = 7.2 puntos; Mediana = 8 puntos *e*) Media ajustada = 7.75 puntos
4C-7. *b*) Media = 4.98 segundos; Mediana = 4.9 segundos; Moda = 4.8 segundos
c) Se aproxima a la normalidad, con un ligero sesgo a la derecha

Conjunto de problemas 4D

4D-2. Mediana = 3.6 pulgadas; Media = 3.86 pulgadas 4D-4. *a*) Media = 82.9 estudiantes; Mediana = 79 estudiantes *e*) Media ajustada = 77.5 estudiantes 4D-6. La puntuación GRE media para los 93 solicitantes es 1 181. 4D-8. Niveles de colesterol: Media = 182; Mediana = 207; Moda = 219; sesgo a la izquierda

CAPÍTULO 5

Conjunto de problemas 5A

5A-1. Respuesta parcial:

Suma de cuadrados	n	Varianza	Desviación estándar
893.49	30	30.81	5.55
43 128.90	347	124.65	11.16

5A-3. *b*) $\bar{X} = 7.86$ pacientes *d*) $s_X = 3.18$ visitas 5A-5. *b*) $\bar{X} = 20.40$ años *c*) $s_x = 1.42$ años 5A-7. Para $X = 128$, $Z_x = -2.28$ SD

Conjunto de problemas 5B

5B-2. *b*) $\bar{X} = \$2\,198.43$ *d*) $s_X = \$356.53$ 5B-4. *b*) $\bar{X} = 3.59$ puntos en la escala de Richter; $s_X = 1.73$ puntos en la escala de Richter 5B-6. Texas. (De hecho, el índice alto para Texas sesga la distribución.)

Conjunto de problemas 5C

5C-1. Respuesta parcial:

Suma de cuadrados	n	Varianza	Desviación estándar
975.46	29	34.82	5.90
74 828.25	526	142.53	11.94

5C-3. *b*) $\bar{X} = 9.29$ contactos de caso *d*) $s_X = .54$ contactos de caso 5C-5. *b*) $\bar{X} = 74.50$ años *c*) $s_X = 2.98$ años 5C-7. Para $X = 42$, $Z_x = 1.22$ SD

Conjunto de problemas 5D

5D-2. *b*) $\bar{X} = 60.14$ años *d*) $s_X = 3.29$ años 5D-4. *d*) $\bar{X} = 170$ libras; $s_X = 19.36$ libras 5D-6. *b*) El área 2 destaca como la que tiene un índice relativamente alto de homicidios.

CAPÍTULO 6

Conjunto de problemas 6A

6A-1. *a*) .1667 *b*) .3333 *c*) .0046 6A-3. *a*) .5000 *b*) .2500 *c*) .1250 6A-5. *a*) .0228 *b*) 26 759 *c*) .6687 6A-7. 95% de las puntuaciones tienen un rango de 20.93 a 33.47 puntos de escala CESD

Conjunto de problemas 6B

6B-2. *a*) .0628 *b*) .0625 *c*) .0316 6B-4. *a*) .0769 *b*) .1538 *c*) .3077 *d*) .0090
6B-6. *a*) .8413 *b*) .0359 *c*) 31.3 puntos CESD
6B-8. *a*) .9878; Jennifer obtuvo mejores puntuaciones que 98% de los que respondieron la prueba. (Si el percentil se redondea, debe redondearse hacia abajo. Sería incorrecto decir que ella obtuvo mejores puntuaciones que el 99%.)

Conjunto de problemas 6C

6C-1. *a*) .1667 *b*) .3333 *c*) .0046 6C-3. *a*. .5000 *b*) .2500 *c*) .1250
6C-5. *a*) .0228 *b*) 135 786 *c*) .6687 6C-7. *b*) Los trabajadores sociales menos satisfechos fueron el 2.5% que obtuvo puntuaciones de alrededor de 19 y menores. Los trabajadores sociales más satisfechos fueron el 2.5% que obtuvo puntuaciones de alrededor de 28 y superiores.

Conjunto de problemas 6D

6D-2. *a*) .0333 *b*) .0460 *c*) .0657 6D-4. *a*) .0769 *b*) .1538 *c*) .3077 *d*) .0090 6D-6. *a*) 168 trabajadores *b*) .0294 *c*) 23.2 o más
6D-8. .9925; el rango percentilar de Caroline en el GRE es 99. Es decir, Caroline obtuvo mejores puntuaciones que el 99% de los que tomaron la prueba en esta muestra.

CAPÍTULO 7

Conjunto de problemas 7A

7A-1. .22 7A-3. *c*) El histograma deberá parecer como uno con un pico en medio, centrado alrededor del resultado de 5 caras. *d*) *p*[5 caras] = .25
7A-7. *b*) 50

Conjunto de problemas 7B

7B-2. *a*) .25 *b*) .08 *c*) .02 *d*) .10 *e*) .02 7B-6. *a*) $P_u = .5$ *b*) La estimación de σ_{p_s} deberá caer entre .06 y .08, dependiendo del tamaño de la cuchara y sus muestras.

Conjunto de problemas 7C

7C-1. .19 7C-3. *c*) El histograma deberá parecer uno con un pico en medio, centrado alrededor del resultado de 5 caras. *d*) *p*[8] = .05
7C-7. *b*) 13

Conjunto de problemas 7D

7D-2. *a*) .1 *b*) .09 *c*) .03 *d*) .10 *e*) .03 7D-6. $P_u = .5$ *b*) La estimación de σ_{p_s} deberá caer entre .06 y .08, dependiendo del tamaño de la cuchara y sus muestras.

CAPÍTULO 8

Conjunto de problemas 8A

8A-1. 55.73 a 58.27 años 8A-3. $42 725.87 a $44 416.13
8A-5. 46.67% a 63.33% No puede garantizarse que se apruebe el proyecto de ley.
8A-7. 99%: 2.58; .0158; .0408; .4592; .5408; .0816 *a*) El ancho del intervalo de confianza aumenta conforme se incrementa el nivel de confianza. Un nivel más alto de confianza requiere un nivel menor de precisión. *b*) El nivel de confianza inferior se resta del nivel de confianza superior.

Conjunto de problemas 8B

8B-2. 172.74 a 175.26 libras 8B-4. No. Su apoyo podría ser tan bajo como 40.67%. 8B-6. Necesitará una muestra de 1 508 personas.

Conjunto de problemas 8C

8C-1. 14.8 a 15.2 años 8C-3. 34.78 a 37.82 años
8C-5. 30.38% a 41.62% 8C-7. 99%: 2.58; .0209; .0539; .3461; .4539; .1078 *a*) El ancho del intervalo de confianza aumenta conforme se incrementa el nivel de confianza. Un nivel más alto de confianza requiere un nivel menor de precisión. *b*) El nivel de confianza inferior se resta del nivel de confianza superior.

Conjunto de problemas 8D

8D-2. 2.86 a 2.94 puntos 8D-4. Sí. Su apoyo es probable de al menos 56% de los votantes registrados. 8D-6. Necesitaría una muestra de 3 393 inversionistas.

CAPÍTULO 9

Conjunto de problemas 9A

9A-1. *a*) Si H_0 es verdadero, que la edad media de los estudiantes en el campus es 21 años, y si muestreamos de manera repetida la población del campus, las medias muestrales para la edad se centrarán en 21 años. *b*) Si H_0 es verdadero, que el porcentaje de mujeres miembros de consejos corporativos es 20%, y si muestreamos de manera repetida a las compañías Fortune 500, las proporciones muestrales de mujeres miembros de consejos corporativos se centrará en .20.
9A-3. *a*) De una cola, dirección positiva. Use el signo mayor que (>) para indicar *más de* 50% 9A-5. *a*) Se rechaza. *e*) No se rechaza 9A-7. Sí. La edad promedio de los hogares Clarksdale parece ser más de 15 años.

Conjunto de problemas 9B

9B-2. *a*) Hipótesis alternativa. La pregunta de investigación examina si la velocidad es *mayor que* 70. La hipótesis nula se plantea como *igual a* 70 millas por hora.

b) Hipótesis nula. La pregunta de investigación examina si el peso promedio es *igual a* 224 libras. La hipótesis alternativa se plantea como *no igual a* 224 libras.
9B-4. *a*) La hipótesis nula se plantea como igual, de modo que podemos establecer una distribución muestral. El asunto del ingreso más bajo para las mujeres se aborda en la hipótesis alternativa. *b*) Ella usará una prueba de una cola debido a que está planteando la hipótesis de que el ingreso de las mujeres es *menor que* el de los hombres. 9B-6. El índice de masa corporal (IMC) promedio entre los estudiantes de la secundaria Jackson parece ser mayor que 25.
9B-8. *a*) 25.25 kg por metros cuadrados

Conjunto de problemas 9C

9C-1. *a*) Si H_0 es verdadera, que la edad media de los empleados es 32 años, y si muestreamos de manera repetida la población de empleados, las medias muestrales para la edad se centrarán en 32 años. *b*) Si H_0 es verdadera, que el peso medio de los jugadores del equipo es 207 libras, y si muestreamos de manera repetida la población de jugadores, las medias muestrales para el peso se centrarán en 207 libras. 9C-3. *a*) Una cola, dirección positiva. Usa el signo mayor que (>) para indicar *más de* 60%. 9C-5. *a*) No se rechaza. *d*) Se rechaza. 9C-7. El promedio de calificaciones (GPA) de los estudiantes de la universidad estatal parece ser menor que un promedio de B.

Conjunto de problemas 9D

9D-2. *a*) Hipótesis alternativa. La pregunta de investigación examina si los dados *no* son legítimos (es decir, están cargados). La hipótesis nula se plantea como: los dados son legítimos, debido a que podemos predecir cómo caerán los dados a la larga y, por consiguiente, predecir una distribución muestral. *b*) Hipótesis alternativa. La pregunta de investigación examina si el índice de deserción escolar entre estudiantes de primer año de la carrera de química es *mayor que* el índice de deserciones entre otras disciplinas. La hipótesis nula se plantea como que los índices son *iguales*.
9D-4. *a*) La hipótesis nula se plantea como igual, así que podemos establecer una distribución muestral. La cuestión de la estima ocupacional para las mujeres es abordada en la hipótesis alternativa. *b*) El investigador usará una prueba de una cola debido a que está planteando la hipótesis de que los niveles de estima ocupacional entre mujeres son *menores que* aquellos entre los hombres. 9D-6. Sí. La edad promedio de las personas sin hogar ahora parece ser menor que 40 años. 9D-8. *a*) 38.71 años.

CAPÍTULO 10

Conjunto de problemas 10A

10A-1. *a*) −3.4 años *b*) −.14 10A-3. La proporción de las mujeres que estudian la carrera de sociología en una universidad en la actualidad es significativamente diferente de .47.
10A-5. Los pacientes en el estudio nuevo perdieron significativamente menos peso que los participantes en el estudio anterior. 10A-7. *a*) No tenemos razón para creer que la proporción de mujeres en nuestra población muestreada es diferente de la de la población seleccionada de Johnsonville.

Conjunto de problemas 10B

10B-2. *a*) 2.074; $p > .05$; no se rechaza *b*) −2.896; $.01 > p > .001$; se rechaza
10B-4. Aquellos mayores de 55 años de edad en esta ciudad no visitan a los médicos un promedio de 5.2 visitas por año. Para aquellas personas en la ciudad mayores de 55 años de edad, estimamos el número promedio de visitas al médico por año en 5.8, .6 visitas más que el promedio nacional. 10B-6. *a*) 5.64 visitas al médico 10B-8. *a*) 56.85 a 61.15

Conjunto de problemas 10C

10C-1. *a*) −.6 años *b*) −.06 10C-3. La proporción de adultos que favorecen el control de revólveres parece ya no ser .53 (o 53%). En la actualidad, aproximadamente 63% de los adultos favorecen el control de los revólveres. 10C-5. Los sujetos en la muestra de tensión alta parecen reportar más síntomas físicos que los sujetos en el estudio previo.
10C-7. *b*) No hay razón para creer que la proporción de adultos con educación superior en nuestra población muestreada es diferente de la de la población seleccionada de Smithville.

Conjunto de problemas 10D

10D-2. *a*) 4.318; $p < .001$; se rechaza *c*) −1.753; $p > .05$; no se rechaza
10D-4. Los paquetes de papel de 200 piezas no parecen promediar 200 hojas cada uno cuando salen de la fábrica. Estimamos que el volumen promedio de los paquetes de papel de 200 es de 194 hojas. 10D-6. *a*) 196.56 hojas 10D-8. *a*) 1 160.40 a 1 239.60 puntos

CAPÍTULO 11

Conjunto de problemas 11A

11A-1. $t_{\bar{X}_1-\bar{X}_2} = -3.07$ SE; $p < .01$; sí 11A-3. $t_{\bar{X}_1-\bar{X}_2} = -2.73$ SE; $p < .01$; sí 11A-5. *a*) $t_{\bar{X}_1-\bar{X}_2} = -.23$ SE; $p > .05$; fueron asignados en forma aleatoria *b*) $t_{\bar{X}_1-\bar{X}_2} = 4.20$ SE; $p < .001$; sí *c*) $t_{\bar{D}} = 35.68$ SE; $p < .001$; sí

Conjunto de problemas 11B

11B-2. $t_{\bar{X}_1-\bar{X}_2} = -3.53$ SE; $p < .001$; sí 11B-4. $t_{\bar{X}_1-\bar{X}_2} = -3.72$ SE; $p < .001$; sí

Conjunto de problemas 11C

11C-1. $t_{\bar{X}_1-\bar{X}_2} = 6.68$ SE; $p < .001$; sí 11C-3. $t_{\bar{X}_1-\bar{X}_2} = 1.75$ SE; $p < .05$; sí 11C-5. *a*) $t_{\bar{D}} = 16.56$ SE; $p < .001$; sí

Conjunto de problemas 11D

11D-2. $t_{\bar{X}_1-\bar{X}_2} = 1.55$ SE; $p > .05$; no 11D-4. $t_{\bar{X}_1-\bar{X}_2} = 4.38$ SE; $p < .001$; sí

CAPÍTULO 12

Conjunto de problemas 12A

12A-1. Efecto principal para las artes y las ciencias = .11 12A-3. $F = 7.76; p < .01$
12A-5. Efecto principal para los afroamericanos = –3.31; $F = 28.19$; $p < .01$; sí, los oficiales blancos tienen las actitudes más punitivas y los afroamericanos las menores.

Conjunto de problemas 12B

12B-2. $Y_{(Kathy\ Schaefer)} = 24\,829 = 16\,489 + (9\,902) + (-1\,562)$
12B-4. $F = 11.27$; $p < .01$; sí 12B-6. Efecto principal para blancos = .37; $F = 54.18$; $p < .01$; sí, hay una relación entre la etnia y la carga del cuidador.

Conjunto de problemas 12C

12C-1. Efecto principal para ingreso moderado = .84 12C-3. $F = 19.73$; $p < .01$; sí
12C-5. Efecto principal para hispanos = –.18; $F = 7.59$; $p < .01$; sí, hay una relación entre la raza y la implicación religiosa.

Conjunto de problemas 12D

12D-2. $Y_{(Nicole\ Owens)} = 81\,247 = 49\,257 + (29\,235) + (2\,755)$
12D-4. $F = .87$; $p > .05$; no 12D-6. Efecto principal para afroamericanos = .71; $F = 17.96$; $p < .01$; sí, hay una relación entre la raza y las sentencias.

CAPÍTULO 13

Conjunto de problemas 13A

13A-1. % [de médicos municipales rurales empleados en especialidades de alto riesgo] = 70.42%
13A-3. $\chi^2 = 7.77$; $p < .01$; rechazar 13A-5. p [de lanzar cuatro monedas y obtener una cara y tres cruces] = .25 13A-7. $p = .0898$; $p > .05$; el procedimiento no es un poco mejor.

Conjunto de problemas 13B

13B-2. $\chi^2 = 0.072$; $p > .05$; no se rechaza. 13B-4. $\chi^2 = 3.18$; $p > .05$; no se rechaza
13B-6. $p = .172$; $p > .001$; no se rechaza. 13B-8. Tamaño mínimo: $n = 8$

Conjunto de problemas 13C

13C-1. % [de empleados de servicios con título universitario] = 12.77%
13C-3. $\chi^2 = 9.97$; $p < .01$; rechazar 13C-5. $p = .0078$ 13C-7. $p = .3769$; $p > .05$; no se rechaza

Conjunto de problemas 13D

13D-2. $\chi^2 = 4.69$; $p < .05$; rechazar 13D-4. $X^2 = 7.82$; $p < .01$; rechazar
13D-6. $p = .2266$; $p > .05$; no se rechaza 13D-8. Tamaño mínimo: $n = 5$

CAPÍTULO 14

Conjunto de problemas 14A

14A-1. *b)* $\acute{Y} = 27.92 + .69(X)$ *c)* $r = .85$ 14A-3. *b)* $\acute{Y} = -4.64 + 8.24(X)$ *c)* $r = .74$

Conjunto de problemas 14B

14B-2. *d.* $\acute{Y} = 27.78 + (-.41)X$ 14B-4. *b)* $r = .36$ *d)* $r = .93$

Conjunto de problemas 14C

14C-1. *b)* $\acute{Y} = (-\$26\,413.60) + (3\,768.54)\,X$ *c)* $r = .90$
14C-3. *b)* $\acute{Y} = \$49\,190 + (-1\,600)\,X$

Conjunto de problemas 14D

14D-2. *d)* $\acute{Y} = 9.20 + .307(X)$ 14D-4. *b)* $r = -.36$ *d)* $r = -.91$

CAPÍTULO 15

Conjunto de problemas 15A

15A-1. *c)* Relación negativa. Conforme aumenta la edad, disminuye el número de películas vistas en los seis meses anteriores. 15A-3. $t_r = 5.63$ SE; $p < .001$
15A-5. $t_r = -3.61$ SE; $p < .01$

Conjunto de problemas 15B

15B-2. *b)* $\acute{Y} = 57.43 + .136(X)$; *c)* $t_r = 2.15$ SE; $p < .01$
15B-4. *a)* La relación entre el ingreso propio y el ingreso del padre es más fuerte, porque su r^2 es más grande. 15B-6. $t_r = 7.73$ SE; $p < .001$

Conjunto de problemas 15C

15C-1. *b)* Relación positiva. Entre mayor es el número de condenas previas, será más larga la sentencia en prisión. 15C-3. $t_r = 1.89$ SE; $p < .05$
15C-5. $t_r = -6.68$ SE; $p < .001$

Conjunto de problemas 15D

15D-2. $\acute{Y} = 4.09 + (-.162)$; $X_{t_r} = 6.21$; $p < .001$. 15D-4. La relación entre la satisfacción en la vida y los eventos estresantes de la vida es más fuerte, porque su r^2 es más grande.
15D-6. $t_r = 13.14$ SE; $p < .001$

APÉNDICE

D Guía para el *SPSS for Windows*

Esta guía para el *Statistical Package for the Social Sciences*, *SPSS for Windows*, proporciona información básica sobre la configuración, ejecución e interpretación de los resultados del *SPSS*. El *SPSS* viene en una versión completa, la cual está disponible en muchos campus universitarios. *The Statistical Imagination* viene con *SPSS for Windows Student Version*, el cual incluye todos los procedimientos empleados en este texto. La versión para estudiantes tiene limitaciones. Para aquellos familiarizados con el *SPSS*, la mayor limitación es la ausencia de una ventana *Syntax* (Sintaxis) que permite que los comandos se peguen en un archivo y se graben para su uso posterior. Por consiguiente, cada vez que un usuario desea ejecutar un procedimiento en el *SPSS*, el usuario debe completar una serie de comandos de señalar y oprimir con el ratón. Para los procedimientos estadísticos cubiertos en este texto, esta limitación no es grave.

Otra limitación del *SPSS for Windows Student Version* es que tiene un uso por tiempo limitado. Trece meses después de la fecha y hora en que el usuario carga el software, ya no se abrirá en esa computadora. En resumen, el software *SPSS for Windows Student Version* es una herramienta de aprendizaje. Para propósitos de investigación, se aconseja al usuario usar la versión completa del *SPSS*.

CONFIGURACIÓN BÁSICA DE ENTRADA Y SALIDA

En jerga de computadora, hay *entrada* y *salida*. La entrada es lo que el usuario proporciona a la computadora, como la introducción de datos. Además, la entrada de instrucciones se introduce de acuerdo con el diseño de un paquete de software como *SPSS*. La entrada de instrucciones o "comandos" le indica a la computadora qué procedimientos ejecutar o "correr". Para *SPSS for Windows Student Version*, estas instrucciones se hacen sólo con apuntar y oprimir el cursor del ratón sobre los iconos y elementos de menú en las ventanas de la pantalla de la computadora. En este punto, este tipo de entrada se llama *secuencia de comandos*, una serie de instrucciones de apuntar y oprimir. Estas secuencias de comandos se presentan en diagramas de árbol fáciles de seguir. La entrada en *SPSS* se inicia a través de una ventana llamada Data Editor (Editor de datos) de *SPSS*. La salida son los resultados de los cálculos que aparecen en una ventana Output-*SPSS* Viewer (Visor de resultados del *SPSS*) como una serie de cuadros llamados tablas dinámicas.

Este apéndice está estructurado de acuerdo con los procedimientos *SPSS* requeridos de cada capítulo. Información adicional sobre la aplicación de los resultados de los procedimientos *SPSS* para los ejercicios de aplicaciones en computadora de un capítulo se proporciona con los ejercicios en el sitio web de *La imaginación estadística* en www.mhhe.com/ritchey2.

SPSS PARA EL CAPÍTULO 1 DE *THE STATISTICAL IMAGINATION*: "INTRODUCCIÓN"

Instalación de SPSS for Windows Student Version

En la mayor parte de las computadoras, para instalar *SPSS for Windows Student Version*, inserta el disco de *SPSS* en la unidad de CD/DVD. Deberá aparecer en forma automática una ventana

de instalación. En esta ventana, oprime "Install SPSS Student Version x.x" (donde x.x representa la versión más reciente, como 15.0). Sigue las instrucciones de Setup (Instalación).

Abrir SPSS for Windows

Secuencia de comandos para abrir *SPSS for Windows*:

Desde la computadora personal
└Start (Inicio)
  └All Programs (Todos los programas)
    └SPSS for Windows
      └haz doble clic en SPSS (version #) for Windows Integrated Student Version

Aparecerá una ventana *SPSS* (version #) for Windows Integrated Student Version. Usa esta ventana después de que te familiarices con el *SPSS*. Por ahora, oprime cancelar y esto dejará una ventana Untitled [data set0]-SPSS Data Editor, la cual puedes ver como una base principal para operar *SPSS*.

Crear un icono de acceso directo en tu escritorio

Desde la computadora personal
└Start (Inicio)
  └All Programs (Todos los programas)
    └SPSS for Windows
      └Coloca el cursor sobre SPSS (version #) for Windows Integrated Student Version; oprime el botón derecho en el ratón; oprime Create Shortcut (Crear acceso directo)

Aparecerá "SPSS (version #) for Windows (2)". Arrastra hasta el escritorio y aparecerá un icono *SPSS*. Una vez que lo hayas completado, abre *SPSS* haciendo doble clic en el icono del escritorio.

Descargar e imprimir archivos de datos, libros de códigos y ejercicios de capítulo desde el sitio web de La imaginación estadística

Ejercicios de aplicaciones para computadora, conjuntos de datos y libros de códigos que describen las variables en conjuntos de datos están disponibles para su descarga en el sitio web de *La imaginación estadística* en www.mhhe.com/ritchey2. Estos archivos se encuentran bajo Student Resources, Computer Application Exercises (Recursos para estudiantes, Ejercicios para aplicación con computadora). Los conjuntos de datos tienen nombres de archivo como FEAR981. El nombre es una referencia al contenido del conjunto de datos y el número indica el número de casos en el archivo. Por ejemplo, FEAR981 es una muestra de 981 médicos que fueron encuestados sobre sus temores de ser demandados por mala práctica médica.

Descargar desde el sitio web Sigue las instrucciones en el sitio web para descargar conjuntos de datos, libros de códigos y ejercicios de aplicación para la computadora del capítulo. Si trabajas con tu propia computadora, puedes descargar estos archivos a tu disco duro. Si trabajas en una computadora comunitaria (por ejemplo, en un laboratorio en el campus), guarda los archivos en tu unidad de disco personal, disco de CD/DVD o memoria portátil. Almacena los archivos en un directorio identificable con facilidad de modo que recuerdes sus ubicaciones.

Imprimir desde el sitio web Los ejercicios de los capítulos también pueden imprimirse en forma directa desde el sitio web de *La imaginación estadística*. Bajo "DOWNLOAD OF COMPUTER APPLICATIONS EXERCISE FILES" (Descarga de archivos de ejercicios de aplicaciones para computadora), haz doble clic en un número de capítulo para abrirlo. En el navegador web, haz clic en File (Archivo) y luego Print (Imprimir).

Respuestas a ejercicios seleccionados También puede descargarse un archivo titulado "Answers to Selected Exercises" (Respuestas a ejercicios seleccionados). Este archivo suministra respuestas parciales a algunos de los ejercicios en cada capítulo.

Secuencia de comandos para abrir un archivo de datos (*.sav) *SPSS* existente Después que se han descargado los conjuntos de datos desde el sitio web a su computadora, abre *SPSS for Windows Student Version* y sigue esta secuencia:

Ventana Data Editor de *SPSS*
└File (Archivo)
  └Open (Abrir)
    └Data (Datos) → ventana Open File (Abrir archivo)
      ├"Look in:" ("Buscar en:") para localizar el archivo *.sav en la unidad de la computadora
      └Haz doble clic en el nombre del archivo

La ventana Data Editor

Cuando se abre un conjunto de datos, los códigos de datos reales aparecerán en la ventana Data Editor como una matriz de puntuaciones. Los nombres de las variables se sitúan en columnas, y los casos en filas. Apunta a un nombre de variable para observar su etiqueta. En el menú a lo largo de la parte superior de la ventana Data Editor de *SPSS*, haz clic en Utilities (Herramientas) y luego en Variables para obtener una visión general de las variables en el conjunto de datos. Si el significado de una variable no es evidente, remítete a la guía de codificación del conjunto de datos (del sitio web) para obtener una descripción completa de la variable.

Vista de datos y vista de variable En la esquina inferior izquierda de la ventana Data Editor de *SPSS* verás dos pestañas marcadas Data View (Vista de datos) y Variable View (Vista de variable). *SPSS for Windows* está predeterminado para abrir en Data View, la cual enlista las puntuaciones reales para cada variable.

Haz clic en la pestaña Variable View. Esto presenta una matriz en la que los nombres de las variables se enlistan en la columna izquierda y características de las variables se enlistan a lo largo de la parte superior. El significado de cada característica se volverá evidente con los ejercicios del capítulo 2.

SPSS PARA EL CAPÍTULO 2 DE *THE STATISTICAL IMAGINATION*: "ORGANIZAR DATOS"

Los ejercicios de aplicaciones para computadora del capítulo 2 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) crear y guardar archivos de datos (*.sav) usando *SPSS for Windows*; (2) aprender la importancia del control de calidad en la

codificación e introducción de datos; (3) producir distribuciones de frecuencias y calcular cuartiles y percentiles, y (4) gestionar y grabar salidas y archivos de datos.

Crear un archivo de datos (.sav) en SPSS for Windows*

Secuencia de comandos para abrir un archivo de datos nuevo (vacío):

Ventana Data Editor de *SPSS*
└File (Archivo)
  └New (Nuevo)
    └Data (Datos)

Aparecerá una matriz de datos vacía como "Untitled-SPSS Data Editor" ("Sin título-SPSS Data Editor").

La tabla D-1 es una hoja de cálculo con datos ficticios de una muestra de estudiantes universitarios y la tabla D-2 es una guía de codificación que describen las variables y sus códigos. Usa este conjunto de datos pequeño manejable para aprender lo básico de la introducción de datos. Inserta los códigos y define las variables de acuerdo con las instrucciones que siguen. Guarda el conjunto de datos como stu7.sav.

Definir las características de cada variable en la ventana Vista de variable

Antes de introducir los datos, informa a la computadora de las variables que vas a crear y definir. Sigue tu guía de codificación, como la tabla D-2 a continuación. Proporciona un nombre para cada variable, una etiqueta que describa la variable y etiquetas para los nombres de categorías de variables nominales/ordinales. Este proceso se llama "Data Definition" ("Definición de datos"), el cual se logra en la Variable View de la ventana Data Editor del *SPSS*. Las cifras en los cuadros de esta ventana son valores predeterminados, los cuales pueden necesitar cambiarse.

TABLA D-1 | Hoja de cálculo de los datos de estudiantes con los nombres de variable del *SPSS*

CASONUM1	GENERO2	EDAD3	HELADO4
1	0	18	1
2	1	21	3
3	9	20	1
4	1	22	2
5	1	999	1
6	0	19	9
7	1	20	2

TABLA D-2 | Guía de codificación de los datos de los estudiantes

Nombre de la variable	Etiquetas de variables y descripción de los códigos de variables	Valores perdidos
CASONUM1	Número de identificación del caso	NA
GENERO2	Género: 0 = Hombre, 1 = Mujer	9
EDAD3	Edad del estudiante: autorreportada	999
HELADO4	Sabor de helado favorito. 1 = Chocolate, 2 = Vainilla, 3 = Otro o no tiene preferencia	9

Secuencia de comandos para definición de datos:

Ventana Data Editor del *SPSS*
└ Pestaña Variable View (Vista de variable)
  ├ Nombre: inserta el nombre de la variable en la columna Name (Nombre)
  ├ Tipo: haz clic para la variable y luego haz clic en la pestaña → en la ventana Variable Type (Tipo de variable)
  │   ├ elige el tipo de variable
  │   ├ para variables numéricas, inserta el número de lugares decimales
  │   ├ para variables de cadena, incrementa el número de caracteres
  │   └ OK (Aceptar)
  ├ Etiqueta: inserta la etiqueta de la variable en la columna Label (Etiqueta)
  ├ Valores: haz clic para la variable y luego haz clic en la pestaña → en la ventana Value Labels (Etiquetas de valor)
  │   ├ Especifica las etiquetas de valor
  │   └ OK (Aceptar)
  ├ Perdido: haz clic para la variable y luego haz clic en la pestaña → en la ventana Missing Values (Valores perdidos)
  │   ├ especifica los códigos para los valores faltantes
  │   └ OK (Aceptar)
  └ Medida: haz clic para la variable y luego haz clic en la pestaña → un menú
      └ Especifica el nivel de medida de la variable, ya sea nominal, ordinal o de *escala* (el término de *SPSS* para una variable de nivel de intervalo/razón).

Nota: aunque los lugares decimales y el ancho pueden establecerse en la ventana Variable Type (Tipo de variable), estos elementos también pueden cambiarse en forma directa en sus columnas de la Variable View de la ventana Data Editor del *SPSS*.

Nombre de la variable En la Variable View de la ventana Data Editor del *SPSS*, las variables se enlistan bajo la columna Name (Nombre) y sus características bajo las columnas restantes. Usando las tablas D-1 y D-2 como guía, para la variable 1 introduce CASONUM1 bajo la columna de nombre (la cual está limitada a ocho caracteres). Comienza el nombre con una letra. No termines con un punto. No incluyas caracteres especiales como comas, apóstrofos, ?, ! o *. Inserta GENERO2 como el nombre de la segunda variable, etcétera.

Tipo de variable y lugares decimales En la ventana Variable View del Data Editor del *SPSS*, en la columna Type (Tipo), ve al cuadro para una variable particular y haz clic en él. Haz clic en la pestaña que aparece a la derecha. Se abrirá una ventana pequeña llamada Variable Type (Tipo de variable). Especifica si la variable es *numérica* (es decir, compuesta de códigos numéricos, o números, y quizá lugares decimales) o de *cadena* (es decir, compuesta de letras y palabras). Para variables numéricas, el ancho del campo es el número de espacios asignado para los códigos. El predeterminado de 8 espacios por lo general será suficiente. También introduce el número de lugares decimales. Por ejemplo, para la variable CASONUM1, estipula Numeric (Numérico) y cero lugares decimales. Para GENERO2, también haz clic en Numeric con cero lugares decimales, porque usaremos números para registrar hombre y mujer. Para variables de cadena, puede requerirse un ancho de campo mayor y se introduce en el cuadro "Caracteres". Haz clic en Continue (Continuar) y la ventana pequeña se cerrará.

Etiquetas de variables Una Variable Label (Etiqueta de variable) es una etiqueta descriptiva que se imprimirá en la salida de resultados. Por ejemplo, para la variable CASONUM1, usaremos la etiqueta "Case Identification Number" ("Número de identificación de caso") de la guía de codificación en la tabla D-2 anterior. En la ventana Variable View del Data Editor del *SPSS*, mueve el cursor bajo la columna Label (Etiqueta) al cuadro vacío para la variable 1. Teclea la etiqueta de la variable. Cuando regreses a Data View de la ventana Data Editor, si no aparece el nombre de la variable entero en su encabezado de columna, puede usarse el puntero del ratón para ampliar el ancho de la columna.

Etiquetas de valor Para las variables nominales y ordinales, las Value Labels (Etiquetas de valor) le asignan a cada código numérico (número) su nombre respectivo. La mayoría de los investigadores usan códigos numéricos para representar un nombre de categoría para las variables nominales/ordinales (por ejemplo, para GENERO4, "Hombre" = "0" y "Mujer" = "1", como en la guía de codificación en la tabla D-2 anterior). Inserta cada código y su nombre en sus respectivos cuadros y haz clic en Add (Añadir). Resalta una etiqueta para Change (Cambiar) o Remove (Eliminar). (Para la variable CASONUM1, las etiquetas de valor son innecesarias.) Cuando termines de asignar etiquetas para una variable, haz clic en Continue (Continuar). No es necesario asignar etiquetas de valor a las variables de intervalo/razón, aunque etiquetas seleccionadas pueden ser útiles. Por ejemplo, en la codificación de edades, podríamos asignar el código 60 a todos los que tienen 60 años de edad o más. Para recordar esto, para un código de 60, asignamos la etiqueta de valor "60 y mayores".

Valores perdidos Algunas variables pueden tener "datos perdidos" debido a la incapacidad de uno de los que respondieron o su negativa a responder una pregunta de encuesta. Por ejemplo, en la tabla D-1 para la variable edad (EDAD3), el número de caso 5 no reportó su edad. Por tanto, codificamos su respuesta como 999 y debemos definir esta puntuación como un código de "valor perdido". Esto instruye a la computadora para que ignore este valor cuando realice cálculos. De manera tradicional, a los valores perdidos se les asignan combinaciones de 9 (9, 99, 999, etc.). Los valores faltantes para las variables en la tabla D-1 se especifican en la guía de codificación de la tabla D-2.

Para definir los valores perdidos: en Variable View de la ventana Data Editor del *SPSS* bajo la columna Missing (Perdido), ve al cuadro para una variable particular y haz clic en él. Haz clic en la pestaña que aparece a la derecha y se abrirá una ventana Missing Values (Valores perdidos). Asigna valores perdidos Discrete (Discretos) específicos o un Range (Rango) de valores perdidos. Por ejemplo, para la variable EDAD3, especificamos el valor faltante discreto de 999.

Nivel de medición de una variable En la ventana Define Variable (Definir variable), en el cuadro Measure (Medida), escoge Nominal, Ordinal o Scale (Escala) (intervalo/razón).

Hacer cambios posteriores en la definición de datos Para hacer cambios en la definición de los datos en cualquier momento, ve a la ventana Variable View y haz clic en la ubicación del dato por cambiar.

Verificación doble de la definición de datos de las variables Después de definir las características de las variables, para verlas en forma rápida, en la parte superior de la ventana del Data Editor del *SPSS* haz clic en Utilities (Herramientas), Variables.

Reacomodo del orden y adición de variables al conjunto de datos En la ventana Data Editor, las variables pueden moverse dentro de la matriz de datos y pueden insertarse nuevas variables y casos usando los menús Edit (Edición), Data (Datos) y Help (Ayuda).

Introducción de datos

En la ventana Data Editor del *SPSS*, introduce los datos resaltando un cuadrado dentro de la matriz de datos haciendo clic sobre él. Teclea el código (puntuación) para ese caso y variable y oprime la tecla Enter (Introducir). Usa el ratón y las teclas de dirección para moverte por la matriz. Para códigos repetitivos, usa los comandos Copy (Copiar) y Paste (Pegar) bajo Edit (Edición). Un método abreviado de copiar y pegar es bloquear una entrada seleccionada con el ratón. Oprime el botón Control y la letra C en forma simultánea (Ctrl-C) para copiar y Ctrl-V para pegar.

Guardar el archivo de datos

Para practicar, crea un archivo de datos usando las entradas de la tabla D-1 y la guía de codificación en la tabla D-2.

Secuencia de comandos para guardar un archivo de datos:

Ventana Data Editor del *SPSS*
└File (Archivo)
  └Save as... (Guardar como...) → ventana Save Data As (Guardar datos como)
    ├elige la ubicación del archivo en el panel "Save in..." (Guardar en...)
    ├teclea el nombre del archivo elegido en el panel File Name (Nombre de archivo)
    └Save (Guardar)

Los archivos de datos en *SPSS* se guardan con el sufijo ".sav". Por ejemplo, los datos de la tabla D-1 podrían guardarse bajo el nombre de archivo stu7.sav.

Volver a guardar archivos alterados Una vez que haz utilizado un conjunto de datos para un ejercicio, es aconsejable que lo vuelvas a guardar. Esto se debe a que algunos ejercicios requieren la modificación de los datos. Si haces modificaciones ligeras y no deseas guardar el archivo original, tan sólo haz clic en el icono Save File (Guardar archivo) con forma de un disco flexible en la parte superior de la ventana Data Editor del *SPSS*. Si haces cambios considerables y deseas guardarlos en un archivo separado del conjunto de datos original, sigue la secuencia de comandos anterior para guardar un archivo de datos pero proporciona un nombre de archivo nuevo, como stu7b.sav.

Elegir estilos de visualización opcionales

Cuando se abre un conjunto de datos, *SPSS* usará estilos predeterminados para su visualización de variables, listas, vistas de pantalla, salida, gráficos, etc. Puedes desear cambiar algunos elementos predeterminados para adecuarlos a tus gustos y propósitos. Por ejemplo, puedes preferir que las listas de variables aparezcan usando nombres de variable en lugar de etiquetas de variable o que las variables sean enlistadas en su orden en el archivo de datos en lugar de hacerlo en orden alfabético. O puedes desear cambiar el estilo de fuente de las tablas de salida para que concuerden con otros documentos para un informe que estás compilando. Para cambiar estas configuraciones predeterminadas, en la ventana Data Editor del *SPSS* usa la secuencia de comandos Edit (Edición), Options (Opciones). Estos cambios tendrán efecto

la próxima vez que se abra el conjunto de datos. La siguiente secuencia de comandos ilustra cómo navegar por la ventana Options.

Secuencia de comandos para estilos opcionales:

Ventana Data Editor del *SPSS*
└Edit (Edición) (para cambiar las configuraciones predeterminadas)
  └Options (Opciones) → ventana Options (Opciones)
    ├General: elige Display (variable) labels (Mostrar etiquetas [variable]) o Display (variable) names (Mostrar nombres [variable]), y otras opciones
    ├Viewer (Visor): establece el estilo de las fuente y el tamaño de la fuente para la salida de los resultados
    ├Output Labels (Etiquetas de salida): elige cómo aparecerán las etiquetas de valor (por ejemplo, para la variable GENERO2 en la tabla D-2, el código "mujer" puede imprimirse como el valor "1" o la etiqueta "mujer" o ambos)
    ├Pivot Table (Tabla dinámica): elige el estilo de las tablas de salida.
    └OK (Aceptar)

Control de calidad: imprimir salida y verificar con la vista una lista de casos

Para verificar la precisión de las entradas de datos, obtén una lista de las entradas de los datos y verifícalos con la vista usando la secuencia de comandos que sigue. La salida será una matriz de datos lista para imprimir en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*).

Secuencia de comandos para imprimir listas de datos de las entradas de datos:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Reports (Reportes)
    └Case Summaries (Resúmenes de casos) → cuadro Summarize Cases (Resumir casos)
      ├insertar variables en el cuadro Variable(s):
      ├seleccionar el botón Display cases (Mostrar casos)
      ├quizá deseleccionar el botón Limit cases to first 100 (Limitar casos a los primeros 100)
      └OK (Aceptar)

Editar e imprimir archivos de salida en la ventana Output-SPSS Viewer (Visor de salida del SPSS)

Cuando se ejecuta un procedimiento, los resultados aparecen en una ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). Varias partes de la salida aparecen en tablas con bordes llamadas tablas dinámicas. Una tabla dinámica o un archivo de salida entero pueden editarse, imprimirse y/o guardarse con el sufijo "*.spo".

Editar la salida Para evitar desperdiciar papel, elimina el material innecesario de las tablas y archivos de salida antes de imprimir. Para editar una tabla dinámica, en la ventana Output-

SPSS Viewer (Visor de salida del *SPSS*), haz doble clic sobre la tabla. Edita bloqueando cosas con el ratón. Los anchos de columna pueden ajustarse ubicando el cursor sobre el borde de una tabla. Aparecerá una flecha. Arrastra la flecha para ampliar o reducir una columna. Cuando la edición esté completa, haz clic fuera de la tabla para regresar a la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). Experimenta con las herramientas de edición hasta que te familiarices con ellas.

Imprimir salida Para imprimir la salida, haz clic sobre una sola tabla dinámica, o haz clic en el árbol de salida en el lado izquierdo de la pantalla para resaltar las partes seleccionadas de la salida. Haz clic en File (Archivo), Print Preview (Vista preliminar de impresión) (o el icono Print Preview) para observar cómo aparecerá la salida. Para hacer ajustes al tamaño de la página, márgenes o diseño, en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*) haz clic en File (Archivo), Page Setup (Configurar página). Cuando esté listo, haz clic en File (Archivo), Print (Imprimir) (o el icono Print).

Corrección de entradas de datos Si se encuentra un error en la entrada (o un "código atípico"), corrígelo moviéndolo a su ubicación en el archivo de datos en la Data View (Vista de datos) de la ventana Data Editor del *SPSS*. Haz clic en la ubicación, teclea la entrada de datos correcta y oprime Enter. Si son necesarias múltiples correcciones para el mismo código, usa el procedimiento Recode (Recodificar) descrito a continuación bajo la sección "Recoding Variables" ("Recodificación de variables").

Moverse entre las ventanas Data Editor y Output del *SPSS* En una ventana Data Editor del *SPSS* o en una ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*), usa el elemento de menú Window (Ventana) para moverte de una a la otra.

Guardar un archivo de salida e imprimir desde otros paquetes de software:

Ventana Data Editor del *SPSS*
└File (Archivo)
  └Save as (Guardar como) → ventana Save Data As (Guardar datos como)
    ├elige la ubicación del archivo en el panel "Save in..." ("Guardar en..."); teclea el nombre de archivo elegido en el panel File Name (Nombre de archivo) en lugar del "output" predeterminado.
    └Save (Guardar)

El archivo se guardará con el sufijo "*.spo" (para salida del *SPSS*). Por ejemplo, el archivo de salida para el resumen de casos de los datos en la tabla D-1 podría guardarse como A: Stu7list.spo. El archivo puede recuperarse más tarde en el *SPSS for Windows*.

Copiar tablas de salida a software de procesador de palabras

Para copiar una tabla dinámica de un archivo de salida a un paquete de procesador de palabras, coloca el puntero del ratón sobre la tabla y haz clic; haz clic en Edit (Edición), Copy (Copiar) (o Ctrl-C). Minimiza *SPSS* y maximiza el paquete del procesador de palabras y haz clic en Edit (Edición), Paste (Pegar) (o Ctrl-V). Usa el editor de tablas del programa procesador de palabras para modificar la salida. En general, sin embargo, es más eficiente editar la salida en *SPSS* antes de moverla a un paquete de procesador de palabras.

Producir distribuciones de frecuencias, cuartiles y percentiles

Secuencia de comandos para distribuciones de frecuencias, cuartiles y percentiles:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Frequencies (Frecuencias) → ventana Frequencies (Frecuencias)
      ├inserta variables en el cuadro Variable(s):
      ├selecciona Display frequency tables (Mostrar tablas de frecuencias)
      ├botón Statistics (Estadísticos)
      │ ├Quartiles (Cuartiles) (seleccionar si se desea)
      │ ├Percentiles; especifique valores y Add (Agregar) (si se desea)
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida La salida para una ejecución de Frequencies (Frecuencias) producirá una tabla dinámica Statistics (Estadísticas). Para cada variable, aparecerá una tabla de distribución Frequencies (Frecuencias) con cinco columnas. Los encabezados de las columnas de izquierda a derecha son: Valid (Válidos) (valores de puntuaciones de la variable, como una edad de 19 años), Frequency (Frecuencia) (el número o "frecuencia" de casos para cada puntuación), Percent (Porcentaje) (el porcentaje de frecuencias de las puntuaciones), Valid Percent (Porcentaje válido) (el porcentaje de frecuencias de las puntuaciones excluyendo los valores faltantes) y Cumulative Percent (Porcentaje acumulado) (el porcentaje acumulado de las frecuencias de puntuaciones).

Recodificación de variables

Hay situaciones en las que podemos desear cambiar un código en el conjunto de datos. Esto se llama recodificación. Por ejemplo, en la tabla D-1, sólo el caso 2 respondió con "otro sabor" de helado. En vista que hay muchas elecciones de chocolate, podemos elegir poner el caso "otro sabor" con las "vainillas" y comparar chocolate con otros sabores. Para estar en el lado seguro, recodificaremos a un nuevo nombre de variable HELADO42, de modo que el código "otro sabor" será conservado bajo el nombre de variable actual HELADO4. Para recodificar con un nombre de variable nuevo: haz clic en Transform (Transformar), Recode (Recodificar) en Different Variables (Variables diferentes); introduce HELADO4 en el cuadro "Input Variable->Output Variable" ("Variable de entrada->Variable de salida"); teclea el nuevo nombre de la variable, HELADO42, en el cuadro "Name" ("Nombre"); haz clic en Change (Cambiar). Haz clic en los botones Old Values (Valores antiguos) y New Values (Valores nuevos) y usa las opciones para recodificación, todas las cuales están diseñadas para "Add" ("Agregar") recodificaciones en el cuadro "Old->New" ("Antiguo->Nuevo") de esta ventana. Para recodificar HELADO4, bajo Old Values (Valores antiguos), teclea 3; bajo New Values (Valores nuevos), teclea 2; haz clic en Add (Agregar) para introducir esta recodificación en el cuadro "Old->New" ("Antiguo->Nuevo"); haz clic en Continue (Continuar). Nota que en la ventana Old and New Values (Valores antiguos y nuevos) hay formas abreviadas de introducir una serie de recodificaciones para especificar un rango de valores a recodificar. Aquí está la secuencia de comandos y la descripción de la salida para recodificación.

Secuencia de comandos para recodificar variables:

Ventana Data Editor del SPSS
└Transform (Transformar)
 └Recode (Recodificar)
  └Into Different Variables (En variables diferentes) → ventana Recode into Different Variables (Recodificar en variables diferentes)
   ├inserta el nombre de la variable que se va a recodificar en el cuadro "Numeric Variable->Output Variable" ("Variable numérica-> Variable de salida")
   ├teclea el nuevo nombre de variable en el cuadro "Name" ("Nombre") de Output Variable (Variable de salida); haz clic en Change (Cambiar)
   ├Haz clic en Old and New Values (Valores antiguos y nuevos) → Recode into Different Variables (Recodificar en variables diferentes): ventana Old and New Values (Valores antiguos y nuevos)
   │ ├usa opciones para recodificar y oprime "Add" ("Agregar")
   │ └Continue (Continuar)
   └OK (Aceptar)

Salida Después de recodificar para crear una variable nueva, regresa a la ventana Data Editor del *SPSS*. Encontrarás la variable nueva en la columna de la extrema derecha de la matriz de datos. Ahora haz clic en Variable View (Vista de variable). Encontrarás el nuevo nombre de variable como la última variable enlistada en la parte inferior. Edita las etiquetas de valor, define los valores perdidos, y así sucesivamente, conforme lo necesites. Puedes desear mover esta variable junto a su original.

Recodificar en las mismas variables Puedes elegir recodificar una variable y no cambiar su nombre. Para hacer esto, después de hacer clic en Transform-Recode (Transformar-Recodificar), selecciona Into Same Variables (En las mismas variables) en lugar de Into Different Variables (En variables diferentes). Sus recodificaciones se harán conservando el nombre de variable actual. Nota, sin embargo, que una vez que una variable es recodificada con el mismo nombre, no puede recuperar los códigos antiguos y conectarlos con los números de casos. Si piensas que podría haber una razón para identificar más adelante casos con los códigos antiguos, mejor recodifica la variable con un nombre nuevo.

Guardar cambios en el conjunto de datos Para conservar los cambios en el conjunto de datos, como las recodificaciones y variables nuevas, recuerda guardar el archivo. Si se hacen muchos cambios y no estás seguro sobre ellos, guarda el archivo bajo un nombre nuevo para conservar el archivo antiguo en su forma original.

SPSS PARA EL CAPÍTULO 3 DE *THE STATISTICAL IMAGINATION*: "CUADROS Y GRÁFICOS"

Los ejercicios de aplicación para computadora del capítulo 3 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) producir gráficos y cuadros usando *SPSS for Windows*, y (2) aprender a editar gráficos y cuadros en *SPSS*.

Gráficos y cuadros en SPSS

Los gráficos y cuadros se extraen del menú Graphs (Gráficos) del menú Data Editor del *SPSS for Windows*.

Gráficos de pastel

Secuencia de comandos para gráficos de pastel:

Ventana Data Editor del *SPSS*
└Graphs (Gráficos)
 └Pie (Pastel) → ventana Pie Charts (Gráficos de pastel)
  └Check Summaries (Verificar resúmenes) para grupos de casos y haga clic en Define (Definir)
   └En Define Pie (Definir pastel): ventana Summaries of Groups of Cases (Resúmenes de grupos de casos)
    ├haz clic en "% of cases" ("% de casos")
    ├inserta el nombre de variable en el cuadro "Define Slices by" ("Definir rebanadas por")
    ├botón Titles (Títulos) en la parte inferior derecha → ventana Titles (Títulos)
    │ ├teclea el título
    │ └Continue (Continuar)
    ├botón Options (Opciones) en la parte inferior derecha → ventana Options (Opciones)
    │ ├deselecciona "Display groups defined by missing values" ("Mostrar grupos definidos por valores faltantes")
    │ └Continue (Continuar)
    └OK (Aceptar)

Salida El gráfico de pastel aparecerá en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). Examínalo con cuidado porque puede requerir edición. Revisa en la página 633 la sección titulada "Editar salida de gráfico".

Gráficos de barras

Secuencia de comandos para gráficos de barras:

Ventana Data Editor del *SPSS*
└Graphs (Gráficos)
 └Bar (Barras) → ventana Bar Charts (Gráficos de barras)
  ├haz clic en Simple
  ├selecciona Summaries for group of cases (Resúmenes para grupo de casos) y haz clic en Define (Definir)
  └En ventana Define Simple Bar: Summaries for group of cases (Definir barras simple: resúmenes para grupo de casos)
   └haz clic en "% of cases" ("% de casos")
    ├inserta el nombre de variable en el panel "Category Axis" ("Eje de categoría")
    ├botón Titles (Títulos) en la parte inferior derecha → ventana Titles (Títulos)
    │ ├teclee el título
    │ └Continue (Continuar)
    ├botón Options (Opciones) en la parte inferior derecha → ventana Options (Opciones)
    │ ├deselecciona "Display groups defined by missing values" ("Mostrar grupos definidos por valores faltantes")
    │ └Continue (Continuar)
    └OK (Aceptar)

Salida El gráfico de barras aparecerá en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). Examínalo con cuidado porque puede requerir edición. Revisa en la página siguiente la sección titulada "Editar salida de gráfico".

Histogramas

Secuencia de comandos para histogramas:

Ventana Data Editor del *SPSS*
- Graphs (Gráficos)
  - Histogram (Histograma) → ventana Histogram (Histograma)
    - Inserta el nombre de variable en el panel "Variable"
    - si lo deseas, selecciona Display normal curve (Mostrar curva normal)
    - botón Titles (Títulos) en la parte inferior derecha → ventana Titles (Títulos)
      - teclea el título
      - Continue (Continuar)
    - OK (Aceptar)

Salida El histograma aparecerá en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). Examínalo con cuidado porque puede requerir edición. Véase en la siguiente página la sección titulada "Editar salida de gráfico".

Polígonos (gráficos de línea en SPSS)

Secuencia de comandos para gráficos de línea:

Ventana Data Editor del *SPSS*
- Graphs (Gráficos)
  - Line (Línea) → ventana Line Charts (Gráficos de línea)
    - haz clic en Simple
    - selecciona Summaries for group of cases (Resúmenes para grupo de casos) y haz clic en Define (Definir)
      - en la ventana Define Simple Line: Summaries of Groups of Cases (Definir línea simple: resúmenes de grupos de casos)
        - haz clic en "% of cases" ("% de casos")
        - inserta el nombre de variable en el panel "Category Axis" ("Eje de categoría")
        - botón Titles (Títulos) en la parte inferior derecha → ventana Titles (Títulos)
          - teclea el título
          - Continue (Continuar)
        - botón Options (Opciones) en la parte inferior derecha → ventana Options (Opciones)
          - deselecciona "Display groups defined by missing values" ("Mostrar grupos definidos por valores faltantes")
          - Continue (Continuar)
        - OK (Aceptar)

Salida El polígono (gráfico de línea) aparecerá en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). Examínalo con cuidado porque puede requerir edición. Revisa en la siguiente página la sección titulada "Editar salida de gráfico".

Editar salida de gráfico

En la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*), haz doble clic sobre el gráfico para tener acceso al Chart Editor (Editor de gráficos) del *SPSS*. Resalta partes seleccionadas del cuadro o gráfico, como una sola rebanada de un gráfico de pastel, haciendo clic sobre ellas. Haz clic en Edit (Edición) y luego en Properties (Propiedades). Se abrirá una ventana Properties (Propiedades) con opciones variadas. Experimenta con ellas hasta que te familiarices con las opciones que son pertinentes para cada tipo de gráfico. El elemento de menú Chart (Gráfico) en la ventana Chart Editor (Editor de gráficos) también proporciona opciones. Cuando termines, haz clic en el cuadro cerrar (el cuadro con X en la esquina superior derecha) de la ventana Chart Editor para regresar a la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*).

Copiar gráficos para su inserción en documentos de procesadores de palabras

Un gráfico *SPSS* puede copiarse e insertarse en el texto de un documento procesador de palabras, como un informe preparado en *MSWord* o *WordPerfect*. En la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*), haz clic sobre el gráfico y aparecerá un borde rodeándolo. Oprime Ctrl-C para copiar. Muévete al paquete de procesador de palabras y oprime Ctrl-V para pegar el gráfico. Una vez en el paquete de procesador de palabras, el gráfico puede ser editado, pero habrá limitaciones en lo que puede cambiarse. Por consiguiente, como regla general, es mejor completar la edición de un gráfico en *SPSS* antes de su inserción en un paquete de procesador de palabras.

SPSS PARA EL CAPÍTULO 4 DE *THE STATISTICAL IMAGINATION*: "MEDICIÓN DE PROMEDIOS"

Los ejercicios de aplicación para computadora del capítulo 4 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) generar estadísticos de tendencia central usando *SPSS for Windows*, y (2) usar estadísticos de tendencia central para obtener un sentido de proporción sobre las formas de las distribuciones de puntuaciones.

Estadísticos de tendencia central: media, mediana y moda

Los estadísticos de tendencia central pueden calcularse en varias formas mediante el *SPSS for Windows*. Primera, son accesibles con Frequency Distributions (Distribuciones de frecuencias). Segunda, son accesibles como "Descriptives" ("Descriptivos"). Por último, son opciones con otros procedimientos estadísticos.

Secuencia de comandos para estadísticas de tendencia central usando "Frequencies":

Ventana Data Editor del *SPSS*
- Analyze (Analizar)
  - Descriptive Statistics (Estadísticas descriptivas)
    - Frequencies (Frecuencias) → ventana Frequencies (Frecuencias)
      - inserta variables en el cuadro Variable(s):
      - selecciona Display frequency tables (Mostrar tablas de frecuencias) (si se desea)
      - botón Statistics (Estadísticas) → ventana Frequencies: Statistics (Frecuencias: Estadísticas)
        - selecciona Mean, Median and Mode (Media, mediana y moda)
        - selecciona Skewness (Asimetría) (si se desea)
        - Continue (Continuar)
      - OK (Aceptar)

Salida La salida consistirá en una tabla dinámica con los estadísticos solicitados y tablas de frecuencias si se solicitaron.

Secuencia de comandos para estadísticas de tendencia central usando "Descriptives":

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Descriptives (Descriptivas) → ventana Descriptives (Descriptivos)
      ├inserta variables en el cuadro Variable(s):
      ├botón Options (Opciones) → ventana Descriptives: Options (Descriptivos: Opciones)
      │ ├seleccionar Mean (Media)
      │ ├seleccionar otros estadísticos deseados
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida La salida consistirá en una tabla dinámica de estadísticos descriptivos con los estadísticos solicitados.

SPSS PARA EL CAPÍTULO 5 DE *THE STATISTICAL IMAGINATION*: "MEDICIÓN DE DISPERSIÓN"

Los ejercicios de aplicación para computadora del capítulo 5 en el sitio Web de *La imaginación estadística* tienen los siguientes objetivos: (1) usar *SPSS for Windows* para calcular el rango, la desviación estándar y las puntuaciones Z; (2) usar estadísticos de tendencia central y de dispersión para discernir las formas de las distribuciones de puntuaciones, y (3) usar puntuaciones Z para identificar puntuaciones extremas en una distribución.

Como los estadísticos de tendencia central, los estadísticos de dispersión se encuentran en las secciones Frequencies (Frecuencias) y Descriptives (Descriptivos) del elemento de menú Descriptive Statistics (Estadísticos descriptivos). Por lo común, obtenemos juntos los estadísticos de tendencia central y de dispersión.

Secuencia de comandos para estadísticas de dispersión usando "Frequencies":

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Frequencies (Frecuencias) → ventana Frequencies (Frecuencias)
      ├Inserta las variables en el cuadro Variable(s):
      ├selecciona Display frequency tables (Mostrar tablas de frecuencia) (si se desea)
      ├botón Statistics (Estadísticos) → ventana Frequencies: Statistics (Frecuencias: Estadísticos)
      │ ├selecciona Std. deviation (Desviación estándar) y Range (Rango)
      │ ├selecciona otros estadísticos deseados
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida La salida consistirá en una tabla dinámica con los estadísticos solicitados y las tablas de frecuencias si se solicitaron.

Secuencia de comandos para estadísticas de dispersión usando "Descriptives":

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Descriptives (Descriptivas) → ventana Descriptives (Descriptivos)
      ├inserta variables en el cuadro Variable(s):
      ├botón Options (Opciones) → ventana Descriptives: Options (Descriptivos: Opciones)
      │ ├selecciona Std. deviation (Desviación estándar) y Range (Rango)
      │ ├selecciona otros estadísticos deseados
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida La salida consistirá en una tabla dinámica de estadísticos descriptivos con los estadísticos solicitados.

Calcular puntuaciones estandarizadas (puntuaciones Z)

Para variables al nivel de intervalo/razón, las puntuaciones estandarizadas (puntuaciones *Z*) son una expresión de una puntuación en bruto (puntuación *X*) en unidades de medida de desviación estándar.

Secuencia de comandos para puntuaciones estandarizadas (puntuaciones Z):

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Descriptives (Descriptivos) → ventana Descriptives (Descriptivos)
      ├inserta las variables en el cuadro Variable(s):
      ├selecciona "Save standardized values as variables" ("Guardar valores estandarizados como variables")
      ├botón Options (Opciones) → ventana Descriptives: Options (Descriptivos: Opciones)
      │ ├selecciona los estadísticos deseados
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida Se crean variables nuevas que constan de puntuaciones Z y se les asignan nombres con la letra Z precediendo a los nombres de las variables originales (por ejemplo, ZEDAD3). Estas variables nuevas se encuentran en la matriz de datos de la ventana Data Editor-Data View (Editor de datos-Vista de datos) en la columna de la extrema derecha. Haz clic en la pestaña para la ventana Variable View (Vista de variable) y las variables nuevas se encontrarán en la parte inferior. Especifica sus propiedades, como las etiquetas. Puedes desear mover estas variables junto a sus variables originales. Si deseas guardar estas variables nuevas, guarda el archivo de datos antes de salir del *SPSS for Windows*.

Es útil generar un listado de números de identificación de casos, puntuaciones brutas y puntuaciones estandarizadas (puntuaciones Z) para variables de intervalo/razón. La lista de puntuaciones Z ayuda a identificar los valores más alto y más bajo de una variable en función de cuánto difieren estas puntuaciones de la media. Las diferencias grandes (es decir, 3 desviaciones estándar o más en cualquier dirección) pueden indicar un valor extremo. Para obtener un listado de valores, véase "Secuencia de comandos para imprimir listas de datos de entradas de datos" para el capítulo 2 de este apéndice.

SPSS PARA EL CAPÍTULO 6 DE *THE STATISTICAL IMAGINATION*: "PROBABILIDAD"

Los ejercicios de aplicación para computadora del capítulo 6 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) aprender a usar distribuciones de frecuencias de puntuaciones como distribuciones de probabilidad, y (2) a usar puntuaciones Z para calcular probabilidades con variables de intervalo/razón distribuidas en forma normal.

Los ejercicios de aplicación para computadora del capítulo 6 utilizan distribuciones de frecuencias y puntuaciones estandarizadas (puntuaciones Z). Para las distribuciones de frecuencia, repasa en el capítulo 2 anterior en este apéndice: "Secuencia de comandos para distribuciones de frecuencias, cuartiles y percentiles". Para puntuaciones Z, repasa en el capítulo 5 anterior en este apéndice: "Secuencia de comandos para puntuaciones estandarizadas (puntuaciones Z)".

SPSS PARA EL CAPÍTULO 7 DE *THE STATISTICAL IMAGINATION*: "DISTRIBUCIONES MUESTRALES"

Los ejercicios de aplicación para computadora del capítulo 7 en el sitio web *La imaginación estadística* tienen los siguientes objetivos: (1) reforzar la comprensión del concepto de distribuciones muestrales, y (2) obtener un sentido de proporción acerca de la relación entre el tamaño de la muestra y el error muestral.

Los ejercicios de aplicación para computadora del capítulo 7 utilizan histogramas de frecuencias y estadísticos descriptivos. Para los histogramas, repasa en el capítulo 3 anterior en este apéndice: "Secuencia de comandos para histogramas". Para los estadísticos descriptivos, repasa en el capítulo 5 anterior en este apéndice: "Secuencia de comandos para estadísticas de dispersión usando frecuencias" y solicita tanto los estadísticos de tendencia central como los de dispersión.

SPSS PARA EL CAPÍTULO 8 DE *THE STATISTICAL IMAGINATION*: "INTERVALOS DE CONFIANZA"

Los ejercicios de aplicación para computadora del capítulo 8 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) calcular intervalos de confianza usando *SPSS*, y (2) entender la importancia de examinar los efectos del sesgo en el cálculo de un intervalo de confianza.

Calcular intervalos de confianza

Además de la siguiente secuencia de comandos, los intervalos de confianza pueden calcularse como una opción con otros procedimientos en *SPSS for Windows*. Estas opciones se señalarán en capítulos posteriores.

Intervalo de confianza de una media poblacional

Secuencia de comandos para calcular intervalos de confianza de la media:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Explore (Explorar) → ventana Explore (Explorar)
      ├inserta variable en el cuadro Dependent List: (Dependiente:)
      ├en el cuadro Display: (Mostrar:), selecciona Statistics (Estadísticos)
      ├botón Statistics (Estadísticos) → ventana Explore: Statistics (Explorar: Estadísticos)
      │ ├selecciona Descriptives (Descriptivos)
      │ ├estipula el nivel de confianza
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida En una tabla dinámica de la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*), el intervalo de confianza se proporciona con el límite de confianza inferior (LCL [lower confidence limit]), estipulado como Lower Bound (Límite inferior) y el límite de confianza superior (UCL [upper confidence limit]), estipulado como Upper Bound (Límite superior).

Intervalo de confianza de una proporción de la población

Para calcular un intervalo de confianza de una proporción, es necesario hacer "códigos simulados" de la variable recodificando la categoría de éxito de la variable nominal/ordinal a un valor de 1 y todas las categorías restantes a cero. Para algunas variables, como GENERO2 en las tablas D-1 y D-2 anteriores, es innecesario recodificar. También nota en la secuencia de comandos que no se solicitan Descriptives (Descriptivos). Para una variable nominal/ordinal, los estadísticos descriptivos de intervalo/razón llevarían una mala interpretación.

Secuencia de comandos para calcular intervalos de confianza de una proporción: recodificar una variable nominal/ordinal como una variable simulada:

Ventana Data Editor del *SPSS*
└Transform (Transformar)
  └Recode (Recodificar)
    └Into Different Variables (En variables diferentes) → cuadro Recode into Different Variables: (Recodificar en variables diferentes:)
      ├inserta el nombre de variable de la variable que se va a recodificar en el cuadro "Numeric Variable-> Output Variable" ("Variable numérica-> Variable de salida")
      ├teclea un nombre de variable nuevo en el cuadro "Name" ("Nombre") de Output Variable (Variable de salida), haz clic en Change (Cambiar)
      ├Haz clic en Old and New Values (Valores antiguos y nuevos) → ventana Recode Into Different Variables: Old and New Values (Recodificar en variables diferentes: valores antiguos y nuevos)
      │ ├recodificar categoría de éxito como "1" y todos los otros valores (excepto los valores faltantes) como "0" y oprima "Add" ("Agregar")
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida Después de recodificar para crear la nueva variable simulada, regresa a la ventana Data Editor del *SPSS*. Encontrarás la variable nueva en la columna de la extrema derecha de la matriz de datos. Haz clic en la pestaña Variable View (Vista de variable). Encuentra la nueva variable simulada en la parte inferior de la columna de nombres de variables. Edita las etiquetas de variable y de valor, define valores faltantes, etc., según sea necesario. Puedes desear mover esta variable junto a su original.

Calcular el intervalo de confianza de una proporción poblacional:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Explore (Explorar) → ventana Explore (Explorar)
      └inserta la variable simulada en el cuadro Dependent List: (Lista de dependientes:)
      ├en el cuadro Display (Mostrar), selecciona Statistics (Estadísticos)
      └botón Statistics (Estadísticos) → ventana Explore: Statistics (Explorar: Estadísticos)
        ├estipule nivel de confianza
        └Continue (Continuar)
      └OK (Aceptar)

Salida En una tabla dinámica de la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*), el intervalo de confianza se proporciona con el límite de confianza inferior (LCL [lower confidence limit]), estipulado como Lower Bound (Límite inferior) y el límite de confianza superior (UCL [upper confidence limit]), estipulado como Upper Bound (Límite superior). (Si cualquiera de estos valores es mayor que 1.00, olvidaste hacer el código simulado de la variable.) Nota en la salida que la "media" reportada de esta variable nominal/ordinal con código simulado es la proporción de casos en la categoría con éxito.

SPSS PARA EL CAPÍTULO 9 DE *THE STATISTICAL IMAGINATION*: "PRUEBA DE HIPÓTESIS 1: PRUEBA DE MEDIAS DE UNA SOLA MUESTRA GRANDE"

Los ejercicios de aplicación para computadora del capítulo 9 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) correr pruebas de medias de una sola muestra, y (2) enfocarse en localizar e interpretar valores *p* en la salida estadística.

La prueba de medias de una sola muestra grande

Aun cuando la prueba de una sola muestra grande utiliza la curva normal, los paquetes de software para computadora, incluyendo el *SPSS*, se refieren a la prueba como una prueba *t*. Se usa el mismo procedimiento estadístico sin importar el tamaño de la muestra. Las razones para esto se hacen evidentes en el capítulo 10.

Secuencia de comandos para la prueba de medias de una sola muestra grande:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Compare Means (Comparar Medias)
    └One-Sample T Test (Prueba T de una muestra) → ventana One-Sample T Test (Prueba T de una muestra)
      ├ inserte variable de intervalo/razón en el cuadro Test Variable(s): (Variable[s] de prueba:)
      ├introduce el valor del parámetro de la hipótesis seleccionada en el cuadro Test Value: (Valor de prueba:)
      └OK (Aceptar)

Salida La salida proporciona estadísticos descriptivos en un cuadro dinámico titulado "One-Sample Statistics" ("Estadísticos de una muestra"). Estos estadísticos incluyen la media, la desviación estándar y el error estándar. En un segundo cuadro dinámico titulado "One-Sample Test" (Prueba de una muestra), la salida proporciona estadísticos del procedimiento de prueba. Éstos incluyen: (1) el estadístico de prueba, *t* (el cual puede verse como una puntuación *Z* si $n > 121$). (2) El valor *p* asumiendo una prueba de dos colas enlistado como "Sig" (por *significancia*). Si la prueba es de una sola cola, divide el valor *p* de dos colas ("Sig") entre 2. (3) El efecto de la prueba estipulado como "mean difference" ("diferencia media"). (4) Los límites reales superior e inferior de un intervalo de confianza de 95% de la diferencia media. *Nota:* este intervalo de confianza de 95% es de esta *diferencia* media (no debe confundirse con un intervalo de confianza de la media en sí). El cuadro también proporciona los grados de libertad (*gl*), los cuales se explican en el capítulo 10.

SPSS PARA EL CAPÍTULO 10 DE *THE STATISTICAL IMAGINATION*: "PRUEBA DE HIPÓTESIS II: PRUEBA DE MEDIAS DE UNA SOLA MUESTRA PEQUEÑA (PRUEBA *t*) Y PRUEBA DE PROPORCIONES DE UNA SOLA MUESTRA GRANDE"

Los ejercicios de aplicación para computadora del capítulo 10 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) aprender cómo correr pruebas de medias de una sola muestra (prueba *t*) para cualquier tamaño de muestra, y (2) aprender cómo correr la prueba de proporciones de una sola muestra grande (prueba *t*).

La prueba de medias de una sola muestra pequeña (prueba t)

La prueba de medias de una sola muestra pequeña usa la misma secuencia de comandos que la prueba de medias de una sola muestra grande. Cuando $n < 121$, los grados de libertad (*gl*) son significativos debido a que la distribución es sólo aproximadamente normal.

Secuencia de comandos para la prueba de medias de una sola muestra pequeña:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  ├Compare Means (Comparar medias)
  ├One-Sample T Test (Prueba T de una muestra) → ventana One-Sample T Test (Prueba T de una muestra)
    ├inserta la variable de intervalo/razón en el cuadro Test Variable(s): (Variables de prueba:)
    └introduce el valor del parámetro de la hipótesis seleccionada en el cuadro Test Value: (Valor de prueba:)
  └OK (Aceptar)

Salida La salida proporciona estadísticos descriptivos en un cuadro dinámico titulado "One-Sample Statistics" ("Estadísticos de una muestra"). Estos estadísticos incluyen la media, la desviación estándar y el error estándar. En un segundo cuadro dinámico titulado "One-Sample Test" (Prueba de una muestra), la salida proporciona estadísticos del procedimiento de prueba. Éstos incluyen: (1) el estadístico de prueba. (2) El valor *p* asumiendo una prueba de dos colas enlistado como "Sig" (por *significancia*). Si la prueba es de una sola cola, divide el valor p de dos colas ("Sig") entre 2. (3) El efecto de la prueba estipulado como "mean difference" ("diferencia media"). (4) Los límites reales superior e inferior de un intervalo de confianza de 95% de la diferencia media. *Nota:* este intervalo de confianza de 95% es de esta *diferencia* media (no debe confundirse con un intervalo de confianza de la media en sí).

Prueba de proporciones de una sola muestra grande

Para una variable nominal/ordinal, una prueba de proporciones de una sola muestra grande usa la misma secuencia de comandos que la de las variables de intervalo/razón. Sin embargo, la variable nominal/ordinal debes hacer un código simulado de tal manera que la categoría de éxito (la categoría que estamos probando) sea codificada "1" y todas las otras categorías, "0". A menos que una variable ya tenga un código simulado, esto requiere recodificación.

Para recodificar una variable nominal/ordinal en una variable simulada:

Véase el capítulo 8, "Secuencia de comandos para calcular intervalos de confianza de una proporción: recodificar una variable nominal/ordinal como una variable simulada".

Secuencia de comandos para la prueba de proporciones de una sola muestra grande:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Compare Means (Comparar medias)
    └One-Sample T Test (Prueba T de una muestra) → ventana One-Sample T Test (Prueba T de una muestra)
      ├inserte la variable simulada en el cuadro Test Variable(s): (Variable[s] de prueba)
      ├introduce el valor del parámetro de la hipótesis seleccionada en el cuadro Test Value: (Valor de prueba:)
      └OK (Aceptar)

Salida La salida es la misma que para la prueba de medias de una sola muestra. Nota, sin embargo, que la "media" reportada de esta variable nominal/ordinal con código simulado es la proporción de casos en la categoría de éxito.

SPSS PARA EL CAPÍTULO 11 DE *THE STATISTICAL IMAGINATION:* "DIFERENCIA DE DOS GRUPOS DE LA PRUEBA DE MEDIAS"

Los ejercicios de aplicación para computadora del capítulo 11 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) aprender la diferencia de dos grupos de las pruebas de medias, y (2) identificar la salida de computadora apropiada para abordar los aspectos de relaciones para estas pruebas.

La diferencia de dos grupos de la prueba de medias para grupos independientes

La prueba para grupos independientes es para grupos separados de sujetos.

Secuencia de comandos para la diferencia de dos grupos de la prueba de medias:

Ventana Data Editor *SPSS*
└Analyze (Analizar)
  └Compare Means (Comparar medias)
    └Independent-Samples T Test (Prueba T de muestras independientes) → ventana Independent-Samples T Test (Prueba T de muestras independientes)
      ├inserta la variable dependiente de intervalo/razón en el cuadro Test Variable(s): (Variable[s] de prueba:)
      ├inserta la variable nominal dicotómica independiente en el cuadro Grouping Variable: (Variable de agrupamiento:)
      ├botón Define Groups (Definir grupos) → ventana Define Groups (Definir grupos)
      │ ├inserta los códigos para los grupos 1 y 2
      │ └Continue (Continuar)
      └OK (Aceptar)

Después de hacer clic en el botón Define Groups (Definir grupos), si no recuerdas los códigos para la Grouping Variable (Variable de agrupamiento), cierra la ventana Define Groups (Definir grupos). Haz clic en Utilities (Herramientas) en la parte superior de la ventana Data Editor del *SPSS*, luego en Variables. Haz clic en el nombre de la variable de agrupamiento y aparecerán sus códigos; toma nota de los códigos. (Esto puede hacerse mientras la ventana Independent-Samples T Test [Prueba T de muestras independientes] está abierta.) Oprime el botón Define Groups (Definir grupos) y continúa con la secuencia de comandos.

Salida Este procedimiento produce dos tablas dinámicas. La primera es una tabla dinámica Group Statistics (Estadísticos de grupo) que proporciona estadísticos descriptivos para la variable de intervalo/razón para ambos grupos de la variable nominal. La segunda tabla es la Independent Samples Test (Prueba de muestras independientes) y tiene dos partes. Una parte, a la derecha, es la "t-test for Equality of Means". Estos estadísticos abordan la hipótesis de una diferencia de medias. Esta parte de la tabla presenta los siguientes resultados para una prueba de hipótesis de una diferencia entre medias: (1) el estadístico de prueba, t, (2) los grados de libertad bajo df (gl), (3) el valor p para una prueba de dos colas bajo "Sig. (2-tailed)" ("Sig. [2 colas]"), si la prueba es de una sola cola, divida este valor p entre 2, (4) el efecto de la prueba (es decir, la diferencia entre las medias) bajo "Mean Difference" ("Diferencia media"), (5) el error estándar y (6) el límite de confianza superior e inferior de la *diferencia* media.

Nota que se presentan dos conjuntos de salida de prueba t. La línea superior son los resultados para "Equal variances assumed" ("Se asumen varianzas iguales"). Estos resultados usan una estimación de la varianza conjunta para el error estándar. La segunda línea de estadísticas es para "Equal variances not assumed" ("No se asumen varianzas iguales"), donde el error estándar es la estimación de la varianza separada. (Revisa las fórmulas en el capítulo 11 en el texto.) Bajo la parte t-test for Equality of Means (prueba t para Igualdad de Medias) de la tabla, sólo uno de estos conjuntos de estadísticas se usan en una prueba de hipótesis dada, dependiendo de si podemos asumir varianzas iguales.

Determinar cuál salida de prueba t de igualdad de medias reportar La parte izquierda de la tabla dinámica Independent Samples Test (Prueba de muestras independientes), bajo "Levene's

Test for Equality of Variances" ("Prueba de Levene para igualdad de varianzas"), determina cuáles resultados de la prueba *t* reportar. La prueba de Levene prueba la hipótesis de que las *varianzas* son iguales. Usa una distribución muestral llamada prueba *F*, y esta estadística de prueba se reporta bajo "F". Bajo "Sig." (significancia), se reporta el valor *p* para la prueba de Levene. Si este valor *p* es menor que .05, entonces se rechaza la hipótesis de *varianzas* iguales. Cuando esto ocurre, se aplica el estadístico "Equal variances *not* assumed" ("Varianzas iguales *no* asumidas") bajo "t-test for Equality of Means" ("prueba t para igualdad de medias"). Si el valor *p* de Levene (Sig.) es mayor que .05, entonces no se rechaza la hipótesis de varianzas iguales. Cuando ocurre esto, se aplica el estadístico "Equal variances assumed" ("Varianzas iguales asumidas") bajo "t-test for Equality of Means" ("prueba t para igualdad de medias"). Una vez que se ha determinado cuál línea de salida reportar de la parte "*t*-test for Equality of Means" ("prueba *t* para igualdad de medias) de la tabla, haz caso omiso del estadístico en la otra línea de esta tabla.

La prueba de la diferencia de medias de dos grupos para grupos no independientes o de comparación de datos relacionados

La prueba de la diferencia de medias de dos grupos para grupos no independientes o de muestras de datos relacionados hace comparaciones sobre variables para el mismo grupo de sujetos. Esta prueba *t* puede usarse para comparar a los mismos sujetos en dos momentos sobre la misma variable, o comparar a los mismos sujetos en dos variables que están codificadas igual.

Secuencia de comandos para la prueba de la diferencia de medias de dos grupos para grupos no independientes o de comparación de datos relacionados

Ventana Data Editor del *SPSS*
- Analyze (Analizar)
  - Compare Means (Comparar medias)
    - Paired-Samples T Test (Prueba T de muestras relacionadas) → ventana Paired-Samples T Test (Prueba T de muestras relacionadas)
      - Desde la lista de variables, haz clic en variable1 para insertarla en el cuadro Current Selections: (Selecciones actuales:)
      - Desde la lista de variables, haz clic en variable2 para insertarla en el cuadro Current Selections: (Selecciones actuales:)
      - Usa el botón de flecha para mover el par de variables al cuadro Paired Variables: (Variables emparejadas:)
      - Sigue el mismo procedimiento para seleccionar pares adicionales de variables (si se desea)
      - OK (Aceptar)

Salida La salida incluirá: (1) una tabla dinámica "Paired Samples Statistics" ("Estadísticos de muestras relacionadas") con estadísticos descriptivos para ambas variables; (2) una tabla de "Paired Samples Correlations" ("Correlaciones de muestras relacionadas") con coeficiente de correlación de Pearson (revisa los capítulos 14 y 15), y (3) una tabla "Paired Samples Test" ("Prueba de muestras relacionadas") con el estadístico de la prueba *t* para probar la hipótesis nula de ninguna diferencia entre medias. Bajo Paired Differences (Diferencias relacionadas), esta tabla proporciona la media, desviación estándar, error estándar y límites de confianza de las *diferencias* entre las puntuaciones. El resto de la tabla proporciona la estadística de prueba (prueba *t*), grados de libertad (*gl*) y el valor *p* de dos colas bajo "Sig. (2-tailed)" ("Sig. [2 colas]"). Si la prueba es de una cola, divida este valor *p* entre 2.

SPSS PARA EL CAPÍTULO 12 DE *THE STATISTICAL IMAGINATION*: "ANÁLISIS DE VARIANZA (ANOVA)"

Los ejercicios de aplicación para computadora del capítulo 12 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) aprender el análisis de varianza (ANOVA) unidireccional, y (2) aprender los aspectos de la relación para las pruebas ANOVA.

Análisis de varianza (ANOVA) unidireccional en SPSS for Windows

Secuencia de comandos para ANOVA unidireccional:

Ventana Data Editor del SPSS
- Analyze (Analizar)
  - Compare Means (Comparar medias)
    - One-Way ANOVA (ANOVA unidireccional) → ventana One-Way ANOVA (ANOVA unidireccional)
      - Inserta la variable dependiente de intervalo/razón en el cuadro Dependent List: (Lista dependiente:)
      - Inserta la variable independiente nominal/ordinal en el cuadro Factor:
      - Para pruebas de rangos, haz clic en el botón Post Hoc → ventana One-Way ANOVA: Post Hoc Multiple Comparisons (ANOVA unidireccional: comparaciones múltiples post hoc)
        - selecciones Tukey bajo el cuadro Equal Variances Assumed (Se asumen varianzas iguales)
        - selecciona T2 de Tamhane bajo el cuadro Equal Variances Not Assumed (No se asumen varianzas iguales)
        - Continue (Continuar)
      - Options (Opciones) → ventana One-Way ANOVA: Options (ANOVA unidireccional: Opciones)
        - selecciona Descriptives (Descriptivas)
        - selecciona Homogeneity of variance test (Prueba de homogeneidad de la varianza)
        - Continue (Continuar)
      - OK (Aceptar)

Salida Una tabla dinámica Descriptives (Descriptivos) proporciona estadísticos descriptivos para cada grupo al igual que intervalos de confianza. Una segunda tabla dinámica titulada "Test of Homogeneity of Variances" ("Prueba de homogeneidad de las varianzas") establece si hay diferencias significativas en las varianzas o dispersiones de la variable dependiente entre los grupos. Recuerda que una prueba de diferencia de medias asume varianzas iguales en la población de los grupos. De modo similar a las opciones presentadas en el capítulo 11, la suposición de varianzas iguales se prueba con el estadístico de Levene. Los resultados determinan cuál conjunto de pruebas de rango se aplican. Cuando el valor *p* del estadístico de Levene bajo "Sig." es mayor que .05, esto indica que es seguro asumir varianzas iguales; por consiguiente, se reportan las pruebas de rango de la salida Tukey HSD de la tabla dinámica Multiple Comparisons (Comparaciones múltiples) descrita más adelante. Cuando el valor *p* del estadístico de Levene bajo "Sig." es menor o igual que .05, esto indica diferencias significativas entre las varianzas grupales; por consiguiente, se reportan pruebas de rango de la salida Tamhane de la tabla dinámica Multiple Comparisons (Comparaciones múltiples) descrita más adelante.

La tercera tabla dinámica, titulada "ANOVA", es la tabla fuente para la prueba F. Las pruebas de rango se presentan en una tabla dinámica Post Hoc Tests-Multiple Comparisons (Pruebas post hoc-comparaciones múltiples). (Reporta los resultados de los cuadros Tukey o Tamhane pero no de ambos, dependiendo de la prueba de homogeneidad de varianzas descrita antes.) Estas tablas de comparaciones múltiples comparan las diferencias entre medias de cada grupo con los otros y usan asteriscos para resaltar diferencias significativas. Por último, una tabla dinámica "Homogeneous Subsets" ("Subconjuntos homogéneos") proporciona una guía aproximada de a cuáles grupos se agrupan (es decir, son más o menos parecidos) en función de sus medias para la variable dependiente.

SPSS PARA EL CAPÍTULO 13 DE *THE STATISTICAL IMAGINATION*: "PRUEBAS DE CHI CUADRADA Y BINOMIAL"

Los ejercicios de aplicación para computadora del capítulo 13 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) aprender los procedimientos Crosstabs (Tabulaciones cruzadas) y Binomial en *SPSS for Windows*, y (2) aprender cómo interpretar estas pruebas estadísticas y sus aspectos de una relación.

La prueba chi cuadrada

Secuencia de comandos para la prueba chi cuadrada:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Descriptive Statistics (Estadísticos descriptivos)
    └Crosstabs (Tabulaciones cruzadas) → ventana Crosstabs (Tablas de contingencia)
      ├Inserta una variable nominal dependiente en el cuadro Row(s): (Fila[s]:)
      ├Inserta una variable nominal independiente en el cuadro Column(s): (Columna[s]:)
      ├Statistics (Estadísticos) → entana Crosstabs: Statistics (Tablas de contingencia: estadísticos)
      │ ├selecciona Chi-Square (Chi cuadrada)
      │ └Continue (Continuar)
      ├Cells (Casillas) → ventana Crosstabs: Cell Display (Tablas de contingencia: mostrar casillas)
      │ ├selecciona Observed (Observadas), Expected (Esperadas) y Column (Columna)
      │ └Continue (Continuar)
      ├Format (Formato) → ventana Crosstabs: Table Format (Tabla de contingencia: formato de la tabla)
      │ ├selecciona Ascending (Ascendente)
      │ └Continue (Continuar)
      └OK (Aceptar)

Salida La salida Crosstabs (Tablas de contingencia) consistirá en tres tablas dinámicas: (1) una tabla Case Processing Summary (Resumen de procesamiento de casos) que reporta el tamaño de la muestra total, perdidos y casos válidos; (2) una tabla de contingencia, y (3) una tabla de la prueba de chi-cuadrada con varios estadísticos de chi cuadrada opcionales, siendo la pertinente la "Pearson Chi-Square" ("Chi cuadrada de Pearson"). El valor del estadístico chi cuadrada calculada se enlista bajo "Value" ("Valor"), grados de libertad bajo "df" ("gl") y el valor p bajo "Asymp. Sig. (2-sided)" ("saint. Sig. (bilateral"), con bilateral significando no direccional. Si hay una frecuencia esperada de casilla menor que 5, aparecerá en la salida ya sea la prueba exacta de Fisher o la "Continuity Correction" ("Corrección por continuidad") de Yates.

La prueba binomial

Secuencia de comandos para la prueba binomial:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
  └Nonparametric Tests (Pruebas no paramétricas)
    └Binomial → ventana Binomial Test (Prueba binomial)
      ├inserte variable dicotómica en el cuadro Test Variable List: (Lista de variables de prueba:)
      ├estipule el valor hipotético de P_u en el cuadro Test Proportion: (Proporción de prueba:)
      └OK (Aceptar)

Salida La salida aparece en una tabla dinámica Binomial Test (Prueba binomial). La tabla reporta: los nombres de cada categoría y sus frecuencias (N); la proporción muestral observada, P_s, como "Observed Prop." ("Prop. observada"); la proporción de prueba de la hipótesis, P_u, bajo "Test Prop." ("Prop. de prueba"), y un valor p de dos colas bajo "Asymp. Sig. (2-tailed)" ("Asymp. Sig. [de dos colas]"). Para obtener el valor p para una prueba de una cola, divide el valor para dos colas entre 2.

SPSS PARA LOS CAPÍTULOS 14 Y 15 DE *THE STATISTICAL IMAGINATION*: "CORRELACIÓN SIMPLE Y ANÁLISIS DE REGRESIÓN"

Los ejercicios de aplicación para computadora de los capítulos 14 y 15 en el sitio web de *La imaginación estadística* tienen los siguientes objetivos: (1) aprender a producir diagramas de dispersión y estadísticos de correlación-regresión; (2) aprender la relación entre la forma de un diagrama de dispersión y el tamaño del coeficiente de correlación r de Pearson, y (3) obtener estadísticos de correlación y regresión para probar hipótesis y abordar los cuatro aspectos de una relación entre dos variables de intervalo/razón.

Diagramas de dispersión

Un diagrama de dispersión se usa para verificar que existe una relación lineal entre las dos variables de intervalo/razón. El diagrama de dispersión también revela las coordenadas de valores extremos en caso de que existan.

Secuencia de comandos para producir diagramas de dispersión:

Ventana Data Editor del *SPSS*
└Graphs (Gráficos)
  └Scatter/Dot (Dispersión/Puntos) → ventana Scatter/Dot (Dispersión/Puntos)
    └Simple Scatter (Dispersión simple)
      └Define (Definir) → ventana Simple Scatterplot (Diagrama de dispersión simple)
        ├inserta la variable dependiente en el cuadro Y Axis: (Eje Y:)
        ├inserta la variable independiente en el cuadro X Axis: (Eje X:)
        ├botón Titles (Títulos) en la parte inferior derecha → ventana Titles (Títulos)
        │ ├teclea el título
        │ └Continue (Continuar)
        └OK (Aceptar)

Salida El diagrama de dispersión aparecerá en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*). El diagrama de dispersión puede editarse haciendo doble clic sobre él para abrir la ventana Chart Editor (Editor de gráficos) del *SPSS*. Prueba varios botones y opciones para familiarizarse con las capacidades de edición. Las siguientes secuencias de comandos utilizan algunas de las opciones.

Graficar la línea de regresión y la línea de la media en el diagrama de dispersión

La regresión y la media de la línea de referencia *Y* pueden trazarse en el diagrama de dispersión. Esto se hace en el Chart Editor (Editor de gráficos) después que el diagrama de dispersión aparece en la salida.

Secuencia de comandos para graficar la línea de regresión y la línea de la media:

En la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*)
└Haz doble clic sobre el diagrama de dispersión → ventana Chart Editor (Editor de gráficos)
 ├Para la línea de regresión, selecciona las coordenadas de datos haciendo clic con el botón derecho sobre una sola coordenada de datos → aparece la ventana del menú Properties (Propiedades)
 │ └haz clic en Add Fit Line at Total (Agregar línea de ajuste al total) → ventana Properties (Propiedades)
 ├Cierra (a menos que desees alterar el estilo del gráfico)
 ├La línea de regresión aparece con el valor de r^2. Si deseas eliminar el valor de r^2, haz clic sobre él para revelar su cuadro de texto y eliminar el cuadro.
 │ └Para la media de la línea de referencia *Y*, selecciona (resalta) las coordenadas de datos haciendo clic con el botón derecho sobre una sola coordenada de datos.
 │ → aparece la ventana de menú Properties (Propiedades)
 │   └haz clic en Add Y Axis Reference Line (Agregar línea de referencia en el eje Y) → ventana Properties (Propiedades)
 │   └Cierra (a menos que desees alterar el estilo del gráfico)
 └Haz clic en el cuadro cerrar (el cuadro con X en la esquina superior derecha) de la ventana Chart Editor (Editor de gráficos) para regresar a la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*).

El diagrama reaparecerá con las modificaciones estipuladas.

Agregar títulos y ejes truncados al diagrama de dispersión

En el diagrama de dispersión, puedes agregar títulos, cambiar de forma o truncar los ejes, modificar el título de un eje, agregar o eliminar las líneas de cuadrícula, etc. Mientras te encuentres en la ventana Chart Editor (Editor de gráficos) del *SPSS*, haz clic en Options (Opciones) o Elements (Elementos) para obtener menús. También puedes hacer doble clic en partes del gráfico, como en la etiqueta de un eje, para abrir en forma automática una sección de edición. Experimenta con las opciones disponibles. Cuando la edición esté completa, haz clic en el cuadro cerrar (el cuadro con la X en la esquina superior derecha) de la ventana Chart Editor (Editor de gráficos) para regresar a la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*).

Copiar un diagrama de dispersión a un paquete de procesador de palabras

Para copiar un diagrama de dispersión a un programa de procesador de palabras, en la ventana Output-*SPSS* Viewer (Visor de salida del *SPSS*), desplázate para dejar a la vista el diagrama de dispersión. Haz clic una vez sobre el diagrama para encerrarlo en un cuadro, luego oprime Ctrl-C. En el paquete del procesador de palabras, oprime Ctrl-V para insertar el gráfico en el texto. En algunos paquetes de procesador de palabras, puede realizarse una edición adicional al gráfico, a menos que requiera cálculos o graficación. En general, es mejor completar la edición mientras te encuentras en el *SPSS* antes de copiar el gráfico a otro software.

Coeficiente de correlación de Pearson

Los estadísticos de correlación se encuentran en una variedad de menús en el *SPSS for Windows*. Para obtenerlos con rapidez usa la siguiente secuencia de comandos. Sin embargo, la *r* de Pearson también se reporta con los estadísticos de regresión. (Revisa la sección "Coeficientes de regresión" más adelante.)

Secuencia de comandos para el coeficiente de correlación *r* de Pearson:

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
 └Correlate (Correlación)
  └Bivariate (Bivariado) → ventana Bivariate Correlations (Correlaciones bivariadas)
   ├inserta dos o más variables de intervalo/razón deseadas en el cuadro Variables:
   ├selecciona Pearson en el cuadro Correlations Coefficients (Coeficientes de correlaciones)
   ├selecciona Two-Tailed (Dos colas) o One-Tailed (Una cola) en el cuadro Test of Significance (Prueba de significancia)
   ├Options (Opciones) → ventana Bivariate Correlations: Options (Correlaciones divariadas: Opciones)
   │ ├selecciona los estadísticos deseados
   │ └Continue (Continuar)
   └OK (Aceptar)

Salida Si se solicitan estadísticos opcionales, su salida aparece en una tabla dinámica Descriptive Statistics (Estadísticos descriptivos). Una tabla Correlations (Correlaciones) proporciona los coeficientes de correlación, el valor *p* en la línea "Sig.", una indicación de si el valor *p* es de una cola o de dos colas, y el número de casos usados (N).

Coeficientes de regresión

En *SPSS for Windows* usa la siguiente secuencia de comandos para obtener coeficientes de regresión en la ecuación lineal, $\hat{Y} = a + bX$. Esta salida también proporciona el coeficiente de correlación *r* de Pearson.

Secuencia de comandos para coeficientes de regresión, *a* y *b* (al igual que la *r* de Pearson):

Ventana Data Editor del *SPSS*
└Analyze (Analizar)
 └Regression (Regresión)
  └Linear (Lineal) → ventana Linear Regression (Regresión lineal)
   ├inserta la variable dependiente (*Y*) en el cuadro Dependent: (Dependiente:)
   ├inserta la o las variables independientes (*X*) en el cuadro Independent(s): (Independiente[s]:)
   ├Statistics (Estadísticos) → ventana Linear Regression: Statistics (Regresión lineal: estadísticos)
   │ ├selecciona Estimates (Estimados), Model Fit (Ajuste del modelo), Descriptives (Descriptivos)
   │ └Continue (Continuar)
   └OK (Aceptar)

Salida La salida para Linear Regression (Regresión lineal) suministra seis tablas dinámicas. Esta salida extensa está diseñada para correlación y regresión *múltiple*, la situación de diversas variables independientes. Para correlación y regresión bivariada simple, la situación de sólo una variable independiente, enfócate en las tablas dinámicas 1, 2 y 6 que se describen a continuación: (1) una tabla dinámica Descriptive Statistics (Estadísticos descriptivos) con medias, desviaciones estándar y tamaño de la muestra (N); (2) una tabla Correlations (Correlaciones) (la cual es valiosa en especial para regresión múltiple donde hay más de una variable independiente); (3) una tabla Variables Entered/Removed (Variables introducidas/eliminadas) la cual también se aplica para regresión múltiple; (4) una tabla Model Summary (Resumen del modelo) que proporciona r y r cuadrada de Pearson (pero reporta estos estadísticos como R y R cuadrada, las cuales son aplicables a la regresión múltiple); (5) una tabla de ANOVA, la cual se aplica a la regresión múltiple, y (6) una tabla Coefficients (Coeficientes) que contiene los coeficientes de regresión. En la tabla dinámica 6, la interección de Y, a, está junto a "(Constant)" ("[Constante]") bajo la columna "B" situada bajo "Unstandardized coefficients" ("Coeficientes no estandarizados"). La pendiente, b, está en la misma columna junto a la etiqueta (nombre) de variable de la variable independiente. El nombre de la variable dependiente se señala en una nota al pie de este cuadro. El valor p para la pendiente se enlista bajo "Sig.". Compara este valor p con el obtenido para la r de Pearson usando la secuencia Analyze-Correlate-Bivariate (Analizar-Correlacionar-Bivariada) bajo "Secuencia de comandos para el coeficiente de correlación r de Pearson", y encontrarás que es el mismo. Recuerda del capítulo 14 del texto que para la correlación y regresión bivariadas, la r de Pearson es una forma estandarizada de la pendiente no estandarizada, b.

E

APÉNDICE

Potencia estadística

En la etapa de análisis de resultados de una investigación, el investigador se vale de las herramientas de la estadística inferencial para probar su hipótesis. Con este fin, debe seleccionar y realizar una prueba estadística que le permita concluir si acepta o no la hipótesis de investigación.

Cuando el investigador realiza una prueba de hipótesis y decide, con base en los resultados del análisis estadístico que realizó, aceptarla o rechazarla está en riesgo de cometer un error.

Sabemos que en la toma de decisiones siempre existe el riesgo de cometer un error. Sin embargo, resulta prácticamente imposible para el investigador saber al momento de tomar una decisión, si ha cometido un error. Con el tiempo, y principalmente mediante la réplica de su estudio por parte de otros investigadores, surgirán evidencias para conocer si su decisión fue correcta o incorrecta. Por eso se dice que la realidad surge tiempo después, conforme su conclusión se confirma o se desconfirma.

En la investigación se pueden cometer dos tipos de errores:

1. que se rechace una hipótesis nula que en realidad sea verdadera, o
2. que se acepte una hipótesis nula que en realidad sea falsa.

Ambas situaciones implican un error por parte del investigador y son indeseables en la investigación, pero en ocasiones un tipo de error conlleva consecuencias más graves que el otro. No se puede asegurar categóricamente que un tipo de error sea más grave que el otro: dependiendo del contexto de la investigación y de sus consecuencias, será la gravedad de cada uno. El investigador debe analizar cuidadosamente los posibles riesgos de su decisión y asumir el riesgo que la misma conlleve.

Hay quienes afirman que en la investigación básica (encaminada a la generación de conocimiento) es peor cometer un tipo de error, y que en la investigación aplicada (encaminada a la aplicación del conocimiento y a la solución de problemas) es peor cometer el otro tipo de error, sin embargo resulta conveniente insistir en que las consecuencias de cada error constituyen el mejor criterio para determinar la gravedad de cada uno.

Para que el lector juzgue la gravedad de cada tipo de error basta presentar dos ejemplos, uno en el terreno de la investigación básica y otro en el de la investigación aplicada.

INVESTIGACIÓN BÁSICA

Supongamos que un investigador realiza una investigación con la finalidad de probar la teoría que está desarrollando. Plantearemos dos posibles situaciones:

Situación 1. En realidad su teoría es acertada, el investigador tiene un gran descubrimiento en las manos.

Si el investigador decide, con base en la evidencia, que su teoría carece de valor y la abandona, estaría cometiendo un error (de omisión) pues dejaría pasar de largo un gran descubrimiento que podría explicar un fenómeno determinado. Las consecuencias de este error girarían en torno a la labor principal de un investigador que busca generar conocimiento. Lograr un descubrimiento y dejarlo ir, probablemente sería lo peor que le podría ocurrir a un investigador de este tipo.

Situación 2. En realidad su teoría no es acertada, la teoría del investigador no explica cabalmente la ocurrencia de un fenómeno.

En caso de que la evidencia sugiera que la teoría del investigador sí explica el fenómeno de estudio (caso contrario a la realidad), probablemente el investigador la compartirá con la comunidad científica en Congresos o mediante su publicación, cometiendo un error (de acción). La reacción natural e inmediata de la comunidad científica será ponerla a prueba para ver si logran replicar sus resultados y validar sus conclusiones. Sin embargo, conforme se describe en este caso, los demás investigadores descartarán la teoría al no poder comprobarla y se hará evidente que el investigador cometió un error en su investigación. Las consecuencias de su error serían el desviar por cierto tiempo la atención de algunos investigadores y la desilusión de pensar que había descubierto algo de valor y darse cuenta que no era así.

INVESTIGACIÓN APLICADA

Imaginemos que en una empresa se analiza si llevar a cabo un cambio importante en cuanto al proceso de producción. Supongamos dos posibles situaciones:

Situación 1. En realidad ese cambio SÍ supondría una mejoría importante en la producción de la empresa (recordemos que al momento de decidir no se tiene conocimiento de que esa era la realidad).

En caso de que las personas responsables de tomar esa decisión de cambio supusieran que la evidencia apunta a que no es conveniente llevar a cabo ese cambio, y por ende no lo realizaran, estarían cometiendo un error, podríamos clasificarlo como un error de omisión ya que dejarán de implementar algo que hubiera tenido un efecto favorable en la producción.

Las principales consecuencias del error en este caso serían que en la empresa habrían *dejado pasar* (error de omisión) la oportunidad de mejorar sus procesos y permanecerían con el proceso actual, o probablemente perderían competitividad si otra empresa realizara ese cambio.

Situación 2. En realidad ese cambio NO supondría una mejora importante en la producción de la empresa.

Si los responsables decidieran, con base en las evidencias que tienen en ese momento, que el cambio sí sería conveniente y lo llevaran a cabo estarían cometiendo otro tipo de error, uno más bien de acción ya que implementarían algo que en realidad no tendría el efecto esperado. Las consecuencias de este error estarían relacionadas a la *realización* (error de acción) de una inversión y cambio ineficaces que podrían significar la disminución de la producción en la empresa hasta llegar, en caso extremo, a su quiebra).

Resulta esencial aclarar que estos errores no se refieren a un descuido intencional o a una conducta encaminada a engañar a otros. Durante la toma de decisiones el investigador desconoce la realidad, y en ese estado debe decidir si aceptar o rechazar la hipótesis. Su decisión va

acompañada del riesgo de cometer alguno de estos tipos de errores y vale decir que conforme se protege de cometer un tipo de error, incrementa el riesgo de cometer el otro tipo de error.

Ambos errores encierran gran complejidad teórica y moral, sin embargo no deja de sorprender la simplicidad y falta de originalidad de sus nombres. Estos errores se denominan error Tipo I y error Tipo II.

ERROR TIPO I

Si en realidad una hipótesis nula es verdadera y el investigador decide rechazarla, se dice que comete un error Tipo I o "alfa". El error consiste en afirmar que un resultado es estadísticamente significativo cuando en realidad no lo es. Se puede decir que es un error de acción al asegurar que un resultado sí es significativo cuando en realidad no lo es.

El riesgo de cometer un error de este tipo se puede disminuir estableciendo un nivel de significancia más estricto, por ejemplo .01 o .001, así se reduciría la probabilidad de cometer un error Tipo I. Sin embargo al tomar esta medida incrementa el riesgo de cometer otro tipo de error; volvernos tan estrictos o exigentes, que pocos resultados serán significativos cuando en realidad tal vez lo sean.

ERROR TIPO II

Cuando una hipótesis nula es falsa y con base en los resultados obtenidos decide aceptarla, comete un error Tipo II o "beta". Este error consiste en afirmar que un resultado no es significativo cuando en realidad sí lo es. Se puede calificar como un error de omisión al asegurar que un resultado no es significativo cuando en realidad sí lo es.

El riesgo de cometer un error tipo II se puede disminuir estableciendo un nivel de significancia menos estricto, por ejemplo .10 o .15, así se reduciría la probabilidad de cometer un error de este tipo. Sin embargo al tomar esta medida incrementa el riesgo de cometer un error Tipo I.

Condición real de la hipótesis nula (el investigador la desconoce al momento de tomar la decisión)

	La hipótesis nula es verdadera (los resultados no son significativos)	La hipótesis nula es falsa (los resultados son significativos)
Rechazar la hipótesis nula	Error Tipo I p = alfa	Decisión correcta p = 1 – beta potencia
Aceptar la hipótesis nula	Decisión correcta p = 1 – alfa	Error Tipo II p = beta

¿QUÉ ES LA POTENCIA ESTADÍSTICA DE UNA PRUEBA?

Al analizar los reportes de investigación se hace evidente que hasta ahora, se le ha dado mucha importancia a la significancia estadística (alfa) en los estudios, pero en pocos se determina su potencia.

En el desarrollo de una investigación, el investigador no puede asumir de manera inmediata que si la hipótesis de investigación es verdadera, automáticamente se obtendrá un resultado estadísticamente significativo. Existen diferentes razones por las cuales esto puede no ocurrir.

Al hablar de la potencia de una prueba, nos referimos a la probabilidad de que ésta rechace la hipótesis nula cuando en realidad es falsa, es decir, se refiere a la probabilidad de no cometer un error Tipo II. Se refiere a la habilidad de la prueba para alcanzar las metas de una investigación.

Cuando el estadístico de prueba de un estudio tiene una probabilidad menor del 50% de resultar significativo, aun cuando la hipótesis del investigador fuera verdadera, se concluye que el estudio tiene baja potencia.

La potencia se relaciona de manera directa con el tamaño de la muestra empleada; conforme mayor sea el tamaño de la muestra, mayor será la potencia de la prueba para correctamente rechazar una hipótesis nula falsa. Por ello resulta importante para el investigador calcular la potencia de la prueba estadística que empleará, con el fin de establecer cuál es el tamaño de muestra recomendable en un experimento. La potencia es una función de la prueba estadística elegida y del tamaño de la muestra empleada.

Calcular la potencia es un proceso laborioso pero para los fines de este libro resultará valioso hacer algunas recomendaciones para incrementar la potencia de un estudio. Incluso aunque la potencia de un estudio fuera pequeña, se podrían realizar algunas acciones para maximizarla. A continuación se mencionan algunas medidas para incrementar la potencia de un estudio (Dunn, 2001).

1. Incrementar el tamaño de la muestra

   Generalmente, al incrementar el tamaño de una muestra, se incrementa la potencia de la prueba, y la probabilidad de cometer un error Tipo II disminuye.

   Cuando se cuenta con un número reducido de sujetos (menos de 30), resulta difícil detectar diferencias significativas. Si el número de sujetos incrementa alrededor de 100, la potencia mejora de manera sustancial y se tiene por lo tanto una mayor probabilidad de rechazar la hipótesis nula.

2. Usar instrumentos precisos

   Cuando se emplean instrumentos que han probado ser confiables, se obtienen mediciones precisas con un reducido grado de error. Esto es importante ya que cuando la desviación estándar dentro de los grupos está inflada debido a la falta de precisión en las medidas, resulta difícil detectar diferencias significativas entre los grupos.

3. Seleccionar variables independientes relevantes

   Un principio importante en investigación tiene que ver con la selección de las variables independientes. Se deben escoger para su estudio aquellas variables independientes que provocan una diferencia importante en la conducta de los grupos que se comparan. Esto se puede hacer mediante el análisis de otros estudios realizados sobre el tema y mediante una prueba piloto una vez identificadas esas posibles variables de estudio.

4. Controlar los factores aleatorios

   Esta medida es fundamental en la realización de cualquier experimento y se refiere a controlar el efecto de todas aquellas variables que pudieran interferir en el mismo, excepto la variación de los niveles de la variable independiente.

5. Realizar un contraste unilateral

   Resulta más probable rechazar una hipótesis nula al realizar un contraste unilateral que uno bilateral. Dado que el valor z para un contraste unilateral es 1.65 y el valor para uno bilateral es 1.96, naturalmente resulta más probable que un resultado alcance entrar a la zona crítica del primer valor z que del segundo.

6. Evitar la reducción de los niveles de significancia

   Al establecer un nivel alfa de .05 se admite una mayor probabilidad de cometer un error Tipo I que al fijar el alfa al nivel de .01. Sin embargo al reducir el nivel de significancia de una prueba estadística, también se reduce su potencia porque es menos probable que aceptemos una hipótesis nula verdadera pero también es menos probable que rechacemos una hipótesis nula falsa.

7. La magnitud del tamaño del efecto

   El efecto se refiere a la magnitud de la asociación entre las variables del estudio o al efecto experimental, es decir a la separación entre los grupos del estudio. El tamaño del efecto y la potencia se relacionan de manera directa, ya que conforme mayor sea el tamaño del efecto, mayor será la potencia de un estudio. Resulta conveniente estimar el tamaño del efecto en la etapa de diseño de un estudio.

TAMAÑO DEL EFECTO

El coeficiente del tamaño del efecto se reporta mediante un índice que varía entre 0 y 1.00. Conforme mayor es el tamaño del efecto denota que la variable independiente tuvo una mayor influencia en una medida dependiente. El investigador puede anticipar el tamaño del efecto con base en la diferencia observada entre las dos muestras, con base en lo que sugiere la teoría, por investigaciones anteriores sobre el tema o por la desviación estándar en la población

Cuando un estudio realizado con una muestra pequeña resulta significativo, muy probablemente se deba a un tamaño del efecto grande. Pero un estudio con una muestra grande que resulta significativo puede tener o no un tamaño del efecto grande.

Cohen (1992) brinda los índices o valores de referencia para distintos tamaños del efecto para un ANOVA de un factor:

Tamaño del efecto	Valor
Pequeño	.10
Mediano	.25
Grande	.40

CÁLCULO DEL TAMAÑO DEL EFECTO Y POTENCIA PARA EL ANOVA

Según Cohen, el tamaño del efecto de un estudio que emplea el análisis de varianza se calcula dividiendo la desviación estándar de la distribución de medias entre la desviación estándar de las observaciones individuales.

A continuación se desarrolla un ejemplo para calcular el tamaño del efecto para un análisis de varianza con tres grupos (datos ficticios). En la siguiente tabla se muestran las puntuaciones de tres grupos con 30 participantes cada uno.

Grupo 1			Grupo 2			Grupo 3		
9	9	9	7	7	7	10	10	10
7	7	7	5	5	5	7	7	7
7	7	7	7	7	7	8	8	8
8	8	8	7	7	7	7	7	7
9	9	9	6	6	6	8	8	8
9	9	9	7	7	7	10	10	10
7	7	7	5	5	5	7	7	7
7	7	7	7	7	7	8	8	8
8	8	8	7	7	7	7	7	7
9	9	9	6	6	6	8	8	8
N = 30			N = 30			N = 30		
Media = 8			Media = 6.4			Media = 8		
Varianza = 0.828			Varianza = 0.662			Varianza = 1.241		

Realizamos un ANOVA de un factor para probar si hay diferencias significativas entre los tres grupos de puntuaciones y la tabla del ANOVA de un factor (Oneway) que genera la computadora se muestra aquí:

Tabla de análisis de varianza de un factor para la comparación de los 3 grupos

Fuente de variación	Suma de cuadrados	gl	Cuadrado medio	F	Sig.
Entre grupos	51.200	2	25.600	28.121	.000
Dentro de grupos	79.200	87	.910		
Total	130.400	89			

A continuación se enlistan los pasos para calcular el tamaño del efecto para la prueba de ANOVA que se realizó.

Medias de los grupos	(media – media total)	(media – media total)^2
8.0	0.53	.28
6.4	– 1.07	1.14
8.0	0.53	.28
$\Sigma =$ 22.4		$\Sigma =$ 1.70
Media total = 7.47		

1. Calcula la media de cada muestra (columna 1): 8, 6.4 y 8.
2. Calcula la media total de los tres grupos (columna 1): 8 + 6.4 + 8 = 22.4 / 3 = 7.47.
3. Calcula la diferencia de cada media respecto a la media total de 7.47 (columna 2).
4. Eleva al cuadrado cada diferencia del paso anterior y calcula la sumatoria (columna 3): 0.28 + 1.14 + .28 = 1.70
5. Divide lo anterior entre el número de grupos (k) menos 1 (grados de libertad $_{entre}$): 1.70 / 2 = 0.85

6. Saca la raíz cuadrada del resultado anterior: $\sqrt{.85} = .922$
7. Calcula la varianza total, sumando las varianzas de cada grupo y dividiendo entre 3 = 0.828 + 0.662 + 1.241 = 2.731 / 3 = 0.910 (Nota que el resultado es el mismo que se observa en la tabla del ANOVA para el cuadrado medio $_{dentro}$).
8. Saca la raíz de la varianza total, o la raíz del cuadrado medio $_{dentro}$ de la razón F obtenida: $\sqrt{.920} = .954$
9. Divide el resultado del paso 6 entre el resultado del paso 8 para obtener el tamaño del efecto: $\dfrac{0.922}{.954} = 0.97$

CÁLCULO RÁPIDO DEL TAMAÑO DEL EFECTO A PARTIR DEL VALOR DE F

El tamaño del efecto también se puede calcular rápidamente si se conoce el valor de F y el número de sujetos en cada grupo:

$$\text{Tamaño del efecto} = \frac{\sqrt{F}}{\sqrt{n}} = \frac{\sqrt{28.121}}{\sqrt{30}} = \frac{5.30}{5.48} = .97$$

ÍNDICES DEL TAMAÑO DEL EFECTO

Como mencionamos anteriormente, para el ANOVA de un factor un tamaño del efecto es pequeño cuando es cercano a .10, mediano alrededor de .25 y alto cuando su valor es cercano a .40 (Cohen 1992). Con base en estos valores, podemos ver que el resultado obtenido de .97 en el ejemplo sugiere un tamaño del efecto muy grande. Es decir, la diferencia entre las medias es muy pronunciada.

POTENCIA

Recuerda que la potencia de una prueba se refiere a la probabilidad de que ésta rechace la hipótesis nula cuando en realidad es falsa. Cuando un estudio tiene una probabilidad igual o menor del 50% de resultar significativo, aun cuando la hipótesis del investigador fuera verdadera, se concluye que el estudio tiene baja potencia, porque ese resultado es similar al esperado al lanzar una moneda al aire. Se considera que una potencia aceptable es alrededor de 80-85%.

Existen tablas y programas de cómputo sencillos para calcular la potencia de diferentes pruebas estadísticas con varios niveles de significancia. A continuación se presenta una tabla que contiene los valores aproximados de la potencia para el ANOVA con una significancia al .05 (Aron y Aron 2001).

En el caso de nuestro ejemplo, encontramos un tamaño del efecto muy grande (.97) para la prueba de ANOVA que realizamos, por lo que nos ubicamos en la última columna de esta tabla para un tamaño del efecto grande. Al consultar los valores podemos ver que para un ANOVA con un nivel de significancia al .05, tres grupos, 30 participantes en cada uno, y un tamaño del efecto grande, la potencia es aproximadamente .93 lo que representa una potencia muy elevada. Esto significa que existía una probabilidad de 93% de que el estudio resultara significativo.

Valores aproximados de la potencia para el ANOVA con una significancia de .05

	Tamaño del efecto		
Participantes por grupo	Pequeño (0.10)	Mediano (0.25)	Grande (0.40)
Tres grupos (2 gl)			
10	0.07	0.20	0.45
20	0.09	0.38	0.78
30	0.12	0.55	0.93
40	0.15	0.68	0.98
50	0.18	0.79	0.99
100	0.32	0.98	*
Cuatro grupos (3 gl)			
10	0.07	0.21	0.51
20	0.10	0.43	0.85
30	0.13	0.61	0.96
40	0.16	0.76	0.99
50	0.19	0.85	*
100	0.36	0.99	*
Cinco grupos (4 gl)			
10	0.07	0.23	0.56
20	0.10	0.47	0.90
30	0.13	0.67	0.98
40	01.7	0.81	*
50	0.21	0.90	*
100	0.40	*	*

* La potencia es cercana a 1.

En el caso de obtener una baja potencia, ésta sugiere que a pesar de que exista un efecto leve, mediano o grande en la población, resultará poco probable que el estudio fuera significativo.

PLANEACIÓN DEL TAMAÑO DE LA MUESTRA

Una buena práctica metodológica exige que, previamente a la realización de un estudio, se realice una estimación del tamaño muestral que garantice una adecuada potencia del mismo. La utilidad de calcular la potencia durante la etapa de planeación de un estudio radica en determinar cuántos participantes deben incluirse en la muestra. Como explicamos anteriormente, la potencia se relaciona de manera directa con el tamaño de la muestra empleada; conforme mayor sea el tamaño de la muestra, mayor será la potencia de la prueba para correctamente rechazar una hipótesis nula falsa.

En su artículo clásico sobre potencia (1992) Cohen presenta una tabla con el número aproximado de participantes para diferentes pruebas. En esa tabla se puede observar que para alcanzar una potencia de 80%, conforme mayor es el tamaño del efecto, menor será el número de participantes necesarios para el estudio.

A continuación se presentan los tamaños muestrales aproximados para alcanzar una potencia de 80% en un ANOVA, con un nivel alfa de .05.

$\alpha = .05$

	Tamaño del efecto		
Grupos	Pequeño (0.10)	Mediano (0.25)	Grande (0.40)
2	393	64	26
3	322	52	21
4	274	45	18
5	240	39	16
6	215	35	14
7	195	32	13

Cada vez con mayor frecuencia se reportan los resultados del análisis de potencia en las investigaciones y esto resulta fundamental en el análisis crítico de los resultados de un estudio. Como el resto de las herramientas analíticas que se te presentan en este libro, el análisis de la potencia permite una comprensión más precisa de los fenómenos que estudiamos.

REFERENCIAS

Aron, A. y Aron, E (2001). *Estadística para psicología*. México, Prentice Hall.

Cohen, J. (1992). *A Power Primer. Psychological Bulletin*, vol.112-1, pp. 155-159.

Dunn, D. (2001). *Statistics and data analysis for the behavioral sciences*. Nueva York, McGraw-Hill.

REFERENCIAS

Abbotts, Joanne E., Rory G.A. Williams, Helen N. Sweeting, and Patrick B. West. 2004. "Is Going to Church Good or Bad for You? Denomination, Attendance, and Mental Health of Children in West Scotland." *Social Science and Medicine* 58:645–56.

Aiken, L.H., S.P. Clarke, D.M. Sloane, J. Sochalski, and J.H. Silber. 2002. "Hospital Nurse Staffing and Patient Mortality, Nurse Burnout, and Job Dissatisfaction." *Journal of the American Medical Association* 288:1987–93.

Alabama Center for Health Statistics, 2004. http://ph.state.al.us/csc/vs/Query/Mortality/MortalityQrySLT.htm.

Alba, Richard D., John R. Logan, and Kyle Crowder. 1997. "White Ethnic Neighborhoods and Assimilation: The Greater New York Region, 1980–1990." *Social Forces* 75:883–909.

American Medical Association. 1997. *Physician Characteristics and Distribution in the U.S., 1996–97*. Chicago: The American Medical Association.

Arthur, Winfred, Jr., and William G. Graziano. 1996. "The Five-Factor Model, Conscientiousness, and Driving Accident Involvement." *Journal of Personality* 64(3):593–615.

Babbie, Earl. 1992. *The Practice of Social Research*, 6th ed. New York: Wadsworth.

Bailey, Kenneth D. 1978. *Methods of Social Research*. New York: Free Press.

Bastiaens, Leo. 2004. "Response to Antidepressant Treatment in a Community Mental Health Center." *Community Mental Health Journal* 40:561–67.

Betts, Julian R., and Darlene Morell. 1999. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects." *The Journal of Human Resources* 34:268–93.

Blalock, Hubert M. 1979. *Social Statistics*, 3rd ed. New York: McGraw-Hill.

Boardman, Jason D. 2004. "Stress and Physical Health: The Role of Neighborhoods as Mediating and Moderating Mechanisms." *Social Science and Medicine* 58:2473–83.

Browning, Christopher R., Tama Leventhal, and Jeanne Brooks-Gunn. 2004. "Neighborhood Context and Racial Differences in Early Adolescent Sexual Activity." *Demography* 41: 697–720.

Bryant, Richard A., Michelle L. Moulds, and Rachel M. Guthrie. 2000. "Acute Stress Disorder Scale: A Self-Report Measure of Acute Stress Disorder." *Psychological Assessment* 12:61–68.

Bryson, Bethany. 1996. "'Anything But Heavy Metal': Symbolic Exclusion and Musical Dislikes." *American Sociological Review* 61:884–99.

Cardano, Mario, Giuseppe Costa, and Moreno Demaria. 2004. "Social Mobility and Health in the Turin Longitudinal Study." *Social Science and Medicine* 58:1563–74.

Central Bank of the Russian Federation. 2000. *Bulletin of Banking Statistics*. Moscow, Russia: Prime-TASS.

Central Bank of the Russian Federation. 2002. *Bulletin of Banking Statistics*. Moscow, Russia: Prime-TASS.

_____ 2003. *Bulletin of Banking Statistics*. Moscow, Russia: Prime-TASS.

Clair, Jeffrey M., Ferris J. Ritchey, and Richard M. Allman. 1993. "Satisfaction with Medical Encounters Among Caregivers of Geriatric Outpatients." *Sociological Practice* 11:139–57.

Cockerham, William C., M. Christine Snead, and Derek F. DeWaal. 2002. "Health Lifestyles in Russia and the Socialist Heritage." *Journal of Health and Social Behavior* 43:42–55.

Crawford, Charles. 2000. "Gender, Race, and Habitual Offender Sentencing in Florida." *Criminology* 38:263–80.

Crawford, Charles, Ted Chiricos, and Gary Kleck. 1998. "Race, Racial Threat, and Sentencing of Habitual Offenders." *Criminology* 36:481–510.

David, F.N. 1962. *Games, Gods and Gambling*. London: Charles Griffen and Company.

Deschenes, M.R. 2004. "Effects of Aging on Muscle Fibre Type and Size." *Sports Medicine* 34:809–24.

DuBois, David L., and Naida Silverthorn. 2005. "Natural Mentoring-Relationships and Adolescent Health: Evidence from a National Study." *American Journal of Public Health*-95:518–25.

Durkheim, Emile. 1951 [1897]. *Suicide: A Study in Sociology*. Translated by John A. Spaulding, and George Simpson. Glencoe, Ill.: The Free Press.

Ebrahim, Shah, Olia Papacosta, Goya Wannamethee, and Joy Adamson. 2004. "Social Inequalities and Disability in Older Men: Prospective Findings from the British Regional Heart Study." *Social Science and Medicine* 59:2109–20.

Edin, Kathryn, and Laura Lein. 1997. "Work, Welfare, and Single Mothers' Economic Survival Strategies." *American Sociological Review* 61:253–66.

Egan, Marcia, and Goldie Kadushin. 2004. "Job Satisfaction of Home Health Social Workers in the Environment of Cost Containment." *Health and Social Work* 29:287–96.

Elder, Randy W., Ruth A. Shults, Monica H. Swahn, Brian J. Strife, and George W. Ryan. 2004. "Alcohol-Related Emergency Department Visits Among Young People Ages 13 to 25 Years." *Journal of Studies on Alcohol* 65:297–300.

Ellickson, Phyllis L., Chloe E. Bird, Maria Orlando, David J. Klein, and Daniel F. McCaffrey. 2003. "Social Context and Adolescent Health Behavior: Does School-Level Smoking Prevalence Affect Students' Subsequent Smoking Behavior?" *Journal of Health and Social Behavior* 44:525–35.

Ensminger, Margaret E. 1995. "Welfare and Psychological Distress: A Longitudinal Study of African American Urban Mothers." *Journal of Health and Social Behavior* 36:346–59.

Farrelly, Matthew C., Kevin C. Davis, M. Lyndon Haviland, Peter Messeri, and Cheryl G. Healton. 2005. "Evidence of Dose-Response Relationship Between 'Truth' Antismoking Ads and Youth Smoking Prevalence." *American Journal of Public Health* 95:425–31.

Federal Bureau of Investigation. 2002a. *Crime in the United States, 2002: Uniform Crime Reports*. Washington, D.C.: U.S. Government Printing Office.

_____ 2002b. "Hate Crime Statistics Press Release." November 25, 2002. Washington, D.C.: FBI National Press Office.

Ferraro, Kenneth F., and Yan Yu. 1995. "Body Weight and Self-Ratings of Health." *Journal of Health and Social Behavior* 36:274–84.

Fischer, Hans. 2000. *The Central Limit Theorem from Laplace to Cauchy: Changes in Stochastic Objectives and in Analytical Methods*. Aachen, Netherlands: Shaker Publishing.

Fisher, Ronald A., and Frank Yates. 1963. *Statistical Tables for Biological, Agricultural and Medical Research*. New York: Hafner Publishing Company. (Previously published by Longman Group Ltd., London, and Oliver and Boyd, Edinburgh.)

Franks, Peter, Marthe R. Gold, and Kevin Fiscella. 2003. "Sociodemographics, Self-Rated Health, and Mortality in the U.S." *Social Science and Medicine* 56:2505–14.

Freedle, Roy O. 2003. "Correcting the SAT's Ethnic and Social-Class Bias: A Method for Reestimating SAT Scores." *Harvard Educational Review* 73:1–43.

Freund, John E., and Gary A. Simon. 1991. *Statistics: A First Course*, 5th ed. Englewood Cliffs, N.J.: Prentice-Hall.

Friedman, Michael A., and Kelly D. Brownell. 1995. "Psychological Correlates of Obesity: Moving to the Next Research Generation." *Psychological Bulletin* 117:3–20.

Funk, Jeanne B., Debra D. Buchman, Jennifer Jenks, and Heidi Bechtoldt. 2003. "Playing Violent Video Games, Densensitization, and Moral Evaluation in Children." *Applied Developmental Psychology* 24:413–36.

Gardner, Donald G., Linn Van Dyne, and Jon L. Pierce. 2004. "The Effects of Pay Level on Organization-Based Self-Esteem and Performance: A Field Study." *Journal of Occupational and Organizational Psychology* 77:307–22.

Garroutte, Eva M., Robert M. Kunovich, Clemma Jacobsen, and Jack Goldberg. 2004. "Patient Satisfaction and Ethnic Identity Among American Indian Older Adults." *Social Science and Medicine* 59:2233–44.

Gaughan, Monica. 2006. "The Gender Structure of Adolescent Peer Influence on Drinking." *Journal of Health and Social Behavior* 47:47–61.

Gillings, Richard J. 1972. *Mathematics in the Time of the Pharaohs*. Cambridge, Mass.: MIT Press.

Goesling, Brian. 2001. "Changing Income Inequalities Within and Between Nations: New Evidence." *American Sociological Review* 66:745–61.

Goldberg, Amie, and Joseph Pedulla. 2002. "Performance Differences According to Test Mode and Computer Familiarity on a Practice Graduate Record Exam." *Educational and Psychological Measurement* 62:1053–67.

Green, Donald P., Laurence H. McFalls, and Jennifer-K. Smith. 2001. "Hate Crime: An Emergent Research Agenda." *Annual Review of Sociology* 27:479–504.

Greiner, Birgit A., Niklas Krause, David Ragland, and June M. Fisher. 2004. "Occupational Stressors and Hypertension: A Multi-Method Study Using Observer-Based Job Analysis and Self-Reports in Urban Transit Operators." *Social Science and Medicine* 59:1081–94.

Groome, David, and Anastasia Soureti. 2004. "Post-Traumatic Stress Disorder and Anxiety Symptoms in Children Exposed to the 1999 Greek Earthquake." *British Journal of Psychology* 95:387–97.

Grove, Wayne A., and Tim Wasserman. 2004. "The Life-Cycle Pattern-of Collegiate GPA: Longitudinal Cohort Analysis and-Grade Inflation." *Journal of Economic Education* 35:162–74.

Guilamo-Ramos, Vincent, James Jaccard, Robert Turrisi, and Margaret Johansson. 2005. "Parental and School Correlates of Binge Drinking Among Middle School Students." *American Journal of Public Health* 95:894–99.

Guo, Chiquan. 2004. "Marketing Research: Cui Bono?" *Business Horizons* 47:33–38.

Guth, James L., John C. Green, Lyman A. Kellstedt, and Corwin E. Smidt. 1995. "Faith and the Environment: Religious Beliefs and Attitudes on Environmental Policy." *American Journal of Political Science* 39:364–82.

Hajjar, Ihab, and Theodore Kotchen. 2003. "Regional Variations of Blood Pressure in the United States Are Associated with Regional Variations in Dietary Intakes: The NHANES-III Data." *The Journal of Nutrition* 133:211–14.

Harmelink, Philip J., and William M. VanDenburgh. 2003. "On the CPA's Role in Guarding Clients' Investments." *The CPA Journal* 73:6–10.

He, Meizi, and Judy Sutton. 2004. "Using Routine Growth Monitoring Data in Tracking Overweight Prevalence in Young Children." *Canadian Journal of Public Health* 95:419–23.

Henning, Kris, and Lynette Feder. 2004. "A Comparison of Men and Women Arrested for Domestic Violence: Who Presents the Greater Threat?" *Journal of Family Violence* 19:69–80.

Hersch, Joni, and Leslie S. Stratton. 2002. "Housework and Wages." *The Journal of Human Resources* 37:217–29.

Hoff, Timothy J. 2003. "How Physician-Employees Experience Their Work Lives in a Changing HMO." *Journal of Health and Social Behavior* 44:75–96.

Hollinger, Constance, and Carla Baldwin. 1990. "The Stanford-Binet, 4th ed.: A Small Study of Concurrent Validity." *Psychological Reports* 66:1331–36.

Hughes, Mary Elizabeth, and Linda J. Waite. 2002. "Health in Household Context: Living Arrangements and Health in Late Middle Age." *Journal of Health and Social Behavior* 43:1–21.

Hunt, Larry L., and Matthew O. Hunt. 2001. "Race, Region, and Religious Involvement: A Comparative Study of Whites and African Americans." *Social Forces* 80:605–31.

Jackson, Jerome E., and Sue Ammen. 1996. "Race and Correctional Officers' Punitive Attitudes Toward Treatment Programs for Inmates." *Journal of Criminal Justice* 24:153–66.

Jewkes, Rachel, Jonathan Levin, and Loveday Penn-Kekana. 2002. "Risk Factors for Domestic Violence: Findings from a South African Cross-Sectional Study." *Social Science and Medicine* 55:1603–17.

Johnson, Elmer H. 1973. *Social Problems of Urban Man*. Homewood, Ill.: Dorsey Press.

Kahn, Joan R., and Leonard I. Pearlin. 2006. "Financial Strain Over the Life Course and Health Among Older Adults." *Journal of Health and Social Behavior* 47:17–31.

Kim, Sung Soo, Stan Kaplowitz, and Mark V. Johnston. 2004. "The Effects of Physician Empathy on Patient Satisfaction and Compliance." *Evaluation & Health Professions* 27:237–51.

Klem, Adena M., and James P. Connell. 2004. "Relationships Matter: Linking Teacher Support to Student Engagement and Achievement." *The Journal of School Health* 74:262–73.

Laplace, Pierre S. 1951 (origin. 1820). *A Philosophical Essay on Probabilities*. Translated from the sixth French edition by F.W. Truscott and F.L. Emory. New York: Dover.

Lee, Ivy, and Minako Maykovich. 1995. *Statistics: A Tool for Understanding Society*. Needham Heights, Mass.: Allyn and Bacon.

Lewis, LaVonna B., David C. Sloane, Lori M. Nascimento, Allison L. Diament, et al. 2005. "African Americans' Access to Healthy Food Options in South Los Angeles Restaurants." *American Journal of Public Health* 95:668–74.

Likert, Rensis. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 21 (140).

Loureiro, Maria L., and Rodolfo M. Nayga, Jr. 2006. "Obesity, Weight Loss, and Physician's Advice." *Social Science and Medicine* 62:2458–68.

Lueschen, Guenther, William Cockerham, Jouke van der Zee, Fred Stevens, Jos Diederiks, Manuel Garcia Ferrando, Alphonse d'Houtaud, Ruud Peeters, Thomas Abel, and Steffen-Niemann. 1995. *Health Systems in the European Union: Diversity, Convergence, and Integration*. Munich: Oldenbourg.

Lynch, Wendy L., Paul K. Maciejewski, and Mark N. Potenza. 2004. "Psychiatric Correlates of Gambling in Adolescents and Young Adults Grouped by Age at Gambling Onset." *Archives of General Psychiatry* 61:1116–22.

Macassa, Gloria, Gebrenegus Ghilagaber, Eva Bernhardt, Finn Diderichsen, and Bo Burström. 2003. "Inequalities in Child Mortality in Mozambique: Differentials by Parental Socio-Economic Position." *Social Science and Medicine* 57:2255–64.

Madriz, Esther. 1996. "The Perception of Risk in the Workplace: A Test of Routine Activity Theory." *Journal of Criminal Justice* 24:407–18.

Marshall, Grant N., M. Audrey Burnam, Paul Koegel, Greer Sullivan, and Bernadette Benjamin. 1996. "Objective Life Circumstances and Life Satisfaction: Results from the Course of Homelessness Study." *Journal of Health and Social Behavior* 37:44–58.

Martin, Paul. 2003. "Voting Rewards: Voter Turnout, Attentive Publics, and Congressional Allocation of Federal Money." *American Journal of Political Science* 47:110–27.

Matthews, Ruth J., Carol Jagger, and Ruth M. Hancock. 2006. "Does Socioeconomic Advantage Lead to a Longer, Healthier Old Age?" *Social Science and Medicine* 62:2489–99.

McCarthy, John D., and Mark Wolfson. 1996. "Resource Mobilization by Local Social Movement Organizations: Agency, Strategy, and Organization in the Movement Against Drinking and Driving." *American Sociological Review* 61:1070–88.

McCrae, Robert R. 1996. "Social Consequences of Experiential Openness." *Psychological Bulletin* 120:323–37.

Meshel, David S., and Richard P. McGlynn. 2004. "Intergenerational Contact, Attitudes, and Stereotypes of Adolescents and Older People." *Educational Gerontology* 30:457–79.

Mills, C. Wright. 1959. *The Sociological Imagination*. New York: Oxford University Press.

Molyneux, C.S., N. Peshu, and K. Marsh. 2004. "Understanding of Informed Consent in a Low-Income Setting: Three Case Studies from the Kenyan Coast." *Social Science and Medicine* 59:2547–59.

National Advisory Commission on Civil Disorders. 1968. *Report of the National Advisory Commission on Civil Disorders*. Washington, D.C.: U.S. Government Printing Office.

National Oceanic and Atmospheric Administration. 2003. National Weather Service. http://www.nws.noaa.gov/om/severe_weather/63yrstat.pdf.

Neugebauer, O. 1962. *The Exact Sciences in Antiquity*. New York: Harper.

Nishi, Nobuo, Kae Makino, Hideki Fukuda, and Kozo Tatara. 2004. "Effects of Socioeconomic Indicators on Coronary Risk Factors, Self-Rated Health and Psychological Well-Being Among Urban Japanese Civil Servants." *Social Science and Medicine* 58:1159–70.

Orbuch, Terri L., and Sandra L. Eyster. 1997. "Division of Household Labor among Black Couples and White Couples." *Social Forces* 76:301–32.

Pearson, E.S., and H.O. Hartley. 1976. *Biometrika Tables for Statisticians, Volume I*. London: Biometrika Trust.

Pearson, Jane L., Andrea G. Hunter, Margaret E. Ensminger, and Sheppard G. Kellam. 1990. "Black Grandmothers in Multigenerational Households: Diversity in Family Structure and Parenting Involvement in the Woodlawn Community." *Child Development* 61:434–42.

Peterson, Richard A., and Roger M. Kern. 1996. "Changing Highbrow Taste: From Snob to Omnivore." *American Sociological Review* 61:900–7.

Pikhart, Hynek, Martin Bobak, Andrzej Pajak, Sofia Malyutina, Ruzena Kubinova, Roman Topor, Helena Sebakova, Yuri Nikitin, and Michael Marmot. 2004. "Psychosocial Factors at Work and Depression in Three Countries of Central and Eastern Europe." *Social Science and Medicine* 58:1475–82.

Pinquart, Martin, and Silvia Sorenson. 2005. "Ethnic Differences in Stressors, Resources, and Psychological Outcomes of Family Caregiving: A Meta-analysis." *The Gerontologist* 45:90–106.

Prosser, Helen, and Tom Walley. 2005. "A Qualitative Study of GP's and PCO Stakeholders' Views on the Importance and Influence of Cost on Prescribing." *Social Science and Medicine* 60:1335–46.

Ram, Rati. 2004. "School Expenditures and Student Achievement: Evidence from the United States."*Education Economics* 12:169–76.

Ramstedt, Mats. 2004. "Alcohol Consumption and Alcohol-Related Mortality in Canada, 1950–2000." *Canadian Journal of Public Health* 95:121–26.

Reynolds, John R. 1997. "The Effects of Industrial Employment Conditions on Job-Related Distress." *Journal of Health and Social Behavior* 38:105–16.

Riebschleger, Joanne. 2004. "Good Days and Bad Days: The Experiences of Children of a Parent with a Psychiatric Disability." *Psychiatric Rehabilitation Journal* 28:25–31.

Rogge, Mary E. 1996. "Social Vulnerability to Toxic Risk." *Journal of Social Service Research* 22:109–29.

Roose, Steven P., Harold A. Sackeim, K. Ranga Rama Krishnan, Bruce G. Pollock, George Alexopoulos, Helen Lavretsky, Ira R. Katz, and Heikki Hakkarainen. 2004. "Antidepressant Pharmacotherapy in the Treatment of Depression in the Very Old: A Randomized, Placebo-Controlled Trial." *The American Journal of Psychiatry* 161:2050–59.

Sagan, Carl. 1995a. "Crop Circles and Aliens: What's the Evidence?" *Parade Magazine*, December 3:10–13.

_____. 1995b. *The Demon-Haunted World*. New York: Random House.

Sheynin, O.B. 1970. "On the Early History of the Law of Large Numbers." In E.S. Pearson amd M.G. Kendall, *Studies in the History of Statistics and Probability*, pp. 231–39. London: Griffin.

Sibicky, Mark E., David A. Schroeder, and John F. Dovidio. 1995. "Empathy and Helping: Considering the Consequences of Intervention." *Basic and Applied Psychology* 16:435–53.

Siebert, Darcy C. 2004. "Depression in North Carolina Social Workers: Implications for Practice and Research." *Social Work Research* 28:30–40.

Slater, Michael D., Kimberly L. Henry, Randall C. Swaim, and Lori L. Anderson. 2003. "Violent Media Content and Aggressiveness in Adolescents." *Communication Research* 30:713–36.

Snyder, Douglas K., Robert M. Willis, and Averta Grady-Fletcher. 1991. "Long-term Effectiveness of Behavioral Versus Insight-Oriented Marital Therapy: A 4-Year Follow-up Study." *Journal of Consulting and Clinical Psychology* 59:138–41.

Spoge, Liga, and Janet Trewin. 2003. "The Basics of E-File." *Strategic Finance* 85:13–14.

Steelman, Lala Carr, Brian Powell, and Robert M. Carini. 2000. "Do Teacher Unions Hinder Educational Performance? Lessons Learned from State SAT and ACT Scores." *Harvard Educational Review* 70:437–66.

Struik, Dirk J. 1948. *A Concise History of Mathematics*. New York: Dover.

Takao, Soshi, Norito Kawakami, Tadahiro Ohtsu, and the Japan Work Stress and Health Cohort Study Group. 2003. "Occupational Class and Physical Activity Among Japanese Employees." *Social Science and Medicine* 57:2281–89.

Tohill, Beth C., Jennifer Seymour, Mary Serdula, Laura Kettel-Khan, and Barbara J. Rolls. 2004. "What Epidemiologic Studies Tell Us About the Relationship Between Fruit and Vegetable Consumption and Body Weight." *Nutrition Reviews* 62:365–74.

Tompkins, Peter. 1971. *Secrets of the Great Pyramid*. New York: Harper and Row.

Tukey, J.W. 1953. *The Problem of Multiple Comparisons*. Princeton, N.J.: Princeton University, mimeographed monograph.

Turner, Jonathon B. 1995. "Economic Context and the Effects of Unemployment." *Journal of Health and Social Behavior* 36:213–29.

U.S. Bureau of the Census. 2000. *Census of the Population: General Population Characteristics*. Washington, D.C.: U.S. Government Printing Office. http://eire.census.gov/popest/ national_dataset.csv.

_____. 2003. Population Division. *Children's Living Arrangements and Characteristics: March 2002*. http://www.census.gov/population/socdemo/hh-fam/cps2002/tabA1-all.pdf.

U.S. Department of Energy. 2004. *U.S. Environmental Protection Agency Fuel Economy Guide: Model Year 2004*. Pub. No. DOE/EE-0283. Washington, D.C.: U.S. Government Printing Office. http://www.fueleconomy.gov/feg/ FEG2004_GasolineVehicles.pdf.

U.S. Department of Health and Human Services. 1999. *Mental Health: A Report of the Surgeon General—Executive Summary*. Rockville, Md.: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health. 1999.

U.S. Department of Homeland Security. 2003. *Yearbook of Immigration Statistics, 2002*. Washington, D.C.: U.S. Government Printing Office.

U.S. Federal Bureau of Prisons. "Federal Bureau of Prisons Quick Facts," August 2003. www.bop.gov/fact0598.html.

Varano, Sean P., John D. McCluskey, Justin W. Patchin, and Timothy S. Bynum. 2004. "Exploring the Drugs–Homicide Connection." *Journal of Contemporary Criminal Justice* 20:369–92.

Wee, Christina C., Russell S. Phillips, Anna T.R. Legedza, Roger B. Davis, Jane R. Soukup, Graham A. Colditz, and Mary Beth Hamel. 2005. "Health Care Expenditures Associated with Overweight and Obesity Among U.S. Adults: Importance of Age and Race." *American Journal of Public Health* 95:159–65.

Wiesner, Margit. 2003. "A Longitudinal Latent Variable Analysis of Reciprocal Relations Between Depressive Symptoms and Delinquency During Adolescence." *Journal of Abnormal Psychology* 112:633–45.

Williams, Robin M., Jr. 1970. *American Society*, 3rd ed. New York: Knopf.

Wong, Yin-Ling Irene, and Irving Piliavin. 2001. "Stressors, Resources, and Distress Among Homeless Persons: A Longitudinal Analysis." *Social Science & Medicine* 52:1029–42.

Xiaoxing, Z. He, and David W. Baker. 2004. "Body Mass Index, Physical Activity, and the Risk of Decline in Overall Health and Physical Functioning in Late Middle Age." *American Journal of Public Health* 94:1567–73.

ÍNDICE