CENGAGE Learning

Estadística elemental:

Lo esencial

10a. edición

JOHNSON | KUBY

Estadística elemental:

Lo esencial

DÉCIMA EDICIÓN

Robert Johnson
Patricia Kuby

Monroe Community College

Traducción:
Jorge Humberto Romo Muñoz
Traductor profesional

Revisión técnica:
Ofelia Vizcaíno Díaz
ITESM CCM

CENGAGE Learning™

Australia • Brasil • Corea • España • Estados Unidos • Japón • México • Reino Unido • Singapur

CENGAGE Learning™

Estadística elemental: Lo esencial
Décima edición
Robert Johnson,
Patricia Kuby

Presidente de Cengage Learning Latinoamérica:
Javier Arellano Gutiérrez

Director General México y Centroamérica:
Héctor Enrique Galindo Iturribarría

Director Editorial Latinoamérica:
José Tomás Pérez Bonilla

Director Editorial:
Lilia Moreno Olvera

Editor:
Felipe de Jesús Castro Pérez

Coordinador de preprensa:
Alejandro Gómez Ruiz

Editor de producción:
Timoteo Eliosa García

Director de producción:
Raúl D. Zendejas Espejel

Supervisor de manufactura:
Israel Robles Martínez

Composición tipográfica:
Ediciones OVA

Imagen de la portada:
Getty Images

© D.R. 2008 por Cengage Learning Editores, S.A. de C.V., una Compañía de Cengage Learning, Inc.
Corporativo Santa Fe
Av. Santa Fe, núm. 505, piso 12
Col. Cruz Manca, Santa Fe
C.P. 05349, México, D.F.
Cengage Learning™ es una marca registrada usada bajo permiso.

DERECHOS RESERVADOS. Ninguna parte de este trabajo amparado por la Ley Federal del Derecho de Autor, podrá ser reproducida, transmitida, almacenada o utilizada en cualquier forma o por cualquier medio, ya sea gráfico, electrónico o mecánico, incluyendo, pero sin limitarse a lo siguiente: fotocopiado, reproducción, escaneo, digitalización, grabación en audio, distribución en Internet, distribución en redes de información o almacenamiento y recopilación en sistemas de información a excepción de lo permitido en el Capítulo III, Artículo 27 de la Ley Federal del Derecho de Autor, sin el consentimiento por escrito de la Editorial.

Traducido del libro *Just the Essentials of Elementary Statistics* , tenth edition
Publicado en inglés por
Cengage/Brooks/Cole
ISBN-10: 0-495-31487-0
ISBN-13: 978-0495-31487-5
Datos para catalogación bibliográfica
Johnson, Kuby
Estadística elemental: Lo esencial, Décima edición
ISBN-13: 978-607-481-199-5
ISBN-10: 607-481-199-7

Visite nuestro sitio en:
http://latinoamerica.cengage.com

Impreso en México
1 2 3 4 5 6 7 11 10 09 08

Contenido breve

CAPÍTULO 1 **Estadística** 1

CAPÍTULO 2 **Análisis descriptivo y presentación de datos de una sola variable (univariados)** 38

CAPÍTULO 3 **Análisis descriptivo y presentación de datos bivariados** 144

CAPÍTULO 4 **Probabilidad** 204

CAPÍTULO 5 **Distribuciones de probabilidad (variables discretas)** 268

CAPÍTULO 6 **Distribuciones de probabilidad normal** 312

CAPÍTULO 7 **Variabilidad de la muestra** 360

CAPÍTULO 8 **Introducción a la inferencia estadística** 394

CAPÍTULO 9 **Inferencias que involucran a una población** 472

CAPÍTULO 10 **Inferencias que involucran a dos poblaciones** 544

CAPÍTULO 11 **Aplicaciones de *Ji* cuadrada** 618

Contenido

PARTE 1 Estadística descriptiva

Capítulo 1 **Estadística 1**

1.1 Los norteamericanos, una mirada a sí mismos 1
1.2 ¿Qué es la estadística? 3
1.3 Medibilidad y variabilidad 17
1.4 Recolección (obtención) de datos 18
1.5 Comparación entre probabilidad y estadística 27
1.6 Estadística y la tecnología 28

Capítulo 2 **Análisis descriptivo y presentación de datos de una sola variable (univariados) 38**

2.1 Usted y la Internet 39
2.2 Gráficas, diagramas de Pareto y diagramas de tallo y hoja 40
2.3 Distribuciones de frecuencias e histogramas 55
2.4 Medidas de tendencia central 73
2.5 Medidas de dispersión 84
2.6 Medidas de posición 92
2.7 Interpretación y comprensión de la desviación estándar 106
2.8 El arte de la mentira estadística 114
2.9 Media y desviación estándar de una distribución de frecuencias (opcional) 117

Capítulo 3 **Análisis descriptivo y presentación de datos bivariados 144**

3.1 El chico ha crecido 145
3.2 Datos bivariados 146
3.3 Correlación lineal 162
3.4 Regresión lineal 173

PARTE 2 Probabilidad

Capítulo 4 **Probabilidad 204**

4.1 Estadística y los dulces 205
4.2 Probabilidad de eventos 207

4.3 Probabilidad condicional de eventos 223
4.4 Reglas de probabilidad 228
4.5 Eventos mutuamente excluyentes 236
4.6 Eventos independientes 243
4.7 ¿Existe relación entre los eventos mutuamente excluyentes y la independencia? 249

Capítulo 5 Distribuciones de probabilidad (variables discretas) 268
5.1 Bebidas con cafeína 269
5.2 Variables aleatorias 270
5.3 Distribuciones de probabilidad de una variable aleatoria discreta 273
5.4 Media y varianza de una distribución de probabilidad discreta 278
5.5 Distribución de probabilidad binomial 284
5.6 Media y desviación estándar de la distribución binomial 300

Capítulo 6 Distribuciones de probabilidad normal 312
6.1 Medición de la inteligencia 313
6.2 Distribuciones de probabilidad normal 315
6.3 La distribución normal estándar 316
6.4 Aplicaciones de la distribución normal 323
6.5 Notación 338
6.6 Aproximación normal de la binomial 343

Capítulo 7 Variabilidad de la muestra 360
7.1 275 millones de norteamericanos 361
7.2 Distribuciones muestrales 363
7.3 Distribución de medias muestrales 369
7.4 Aplicación de la distribución de medias muestrales 377

PARTE 3 Estadística inferencial

Capítulo 8 Introducción a la inferencia estadística 394
8.1 ¿La gente era menos alta en otros tiempos? 395
8.2 La naturaleza de la estimación 397
8.3 Estimación de la media μ (σ conocida) 402
8.4 La naturaleza de la prueba de hipótesis 416
8.5 Prueba de hipótesis para la media μ (σ conocida): un acercamiento al valor probabilístico 426
8.6 Prueba de hipótesis para la media μ (σ conocida): un enfoque clásico 444

Capítulo 9 **Inferencias que involucran a una población 472**

9.1 ¿Se ejercita lo suficiente todos los días? 473

9.2 Inferencias relacionadas con la media μ (σ desconocida) 474

9.3 Inferencias acerca de los éxitos de la distribución binomial 496

9.4 Inferencias relacionadas con la varianza y la desviación estándar 516

Capítulo 10 **Inferencias que involucran a dos poblaciones 544**

10.1 Estudiantes, tarjetas de crédito y débito 545

10.2 Muestras dependientes e independientes 547

10.3 Inferencias relacionadas con la diferencia de medias usando dos muestras dependientes 550

10.4 Inferencias relacionadas con las diferencias de medias usando dos muestras independientes 564

10.5 Inferencias relacionadas con las diferencias entre proporciones usando dos muestras independientes 581

10.6 Inferencias relacionadas con la razón de varianzas usando dos muestras independientes 592

PARTE 4 **Más acerca de la inferencia estadística**

Capítulo 11 **Aplicaciones de *Ji* cuadrada 618**

11.1 Algo dulce para contrarrestar el sabor picante 619

11.2 Estadístico *Ji* cuadrada 620

11.3 Inferencias relacionadas con experimentos multinomiales 622

11.4 Inferencias relacionadas con las tablas de contingencia 633

Apéndice A: Principios básicos de conteo 656

Apéndice B: Tablas 657

Respuestas a ejercicios seleccionados 681

Respuesta a exámenes de práctica de los capítulos 716

Índice 721

Prefacio

Nuestro enfoque

A través de los años, el libro de texto *Estadística elemental* se ha transformado en un libro de texto introductorio muy accesible que promueve el aprendizaje, la comprensión y la motivación al presentar la estadística a los estudiantes en un contexto real, sin sacrificar el rigor matemático. Además, esta obra ha respondido a la aceptación gradual en muchas disciplinas donde la estadística se ha vuelto una herramienta importante. Como resultado de lo anterior, las aplicaciones, los ejemplos, proyectos y ejercicios contienen datos que abarcan una amplia variedad de áreas de interés, incluyendo la física y las ciencias sociales, la opinión pública y la ciencia política, los negocios, la economía y la medicina.

En la actualidad, toda vez que han transcurrido 30 años desde la primera publicación de *Estadística elemental*, se recomienda que los estudiantes de todas las disciplinas se inscriban en por lo menos un curso de estadística, ya que la estadística actual está llegando a múltiples áreas de la vida cotidiana. A pesar de este cambio en la percepción, nuestra filosofía no ha cambiado, continuamos esforzándonos para darle a la estadística un tono de accesibilidad y sentido común que motive a los estudiantes que están más interesados en las aplicaciones que en la teoría.

Cambios en esta edición

***NOVEDAD Capítulo 1.* Estadística: este capítulo se ha reescrito** para dar mayor énfasis a la interpretación de la información estadística cuando aprendemos términos y procedimientos que son clave para la estadística.

***Capítulo 3.* Análisis descriptivo y presentación de datos bivariados: los temas de regresión descriptiva y correlación se introducen tempranamente** para todos aquellos que prefieran este enfoque. Luego se continúa con relaciones entre dos variables; esta secuencia de ideas genera una presentación lógica en el material, misma que satisface la curiosidad natural de los alumnos respecto a dos variables; esto sucede después de iniciar el estudio de la estadística descriptiva de una variable. Además, esta introducción temprana permite que los profesores tengan un acercamiento a todos los procesos de pensamiento que se realizan en la prueba de hipótesis, sin tener que utilizar nombres o procedimientos técnicos. Después, en el capítulo 8, cuando llega el momento de introducir el procedimiento de la prueba de hipótesis, mediante el uso nuevamente de la decisión de correlación como un ejemplo de introducción, los estudiantes se sentirán a gusto con el "nuevo" proceso de prueba.

***NOVEDAD Capítulo 4.* Probabilidad: este capítulo se revisó completamente,** se aplicó un enfoque creciente en el análisis, en oposición a las fórmulas, para aumentar el interés y la comprensión (por parte del alumno) de este tema que siempre resulta ser desafiante.

Los temas valor *p* y enfoque clásico a la prueba de hipótesis se presentan en forma individual, pero a partir de este punto se mostrarán "en forma intercalada" para ofrecer flexibilidad pedagógica y enfatizar su comparabilidad.

Viaje por esta nueva edición

Los objetivos del capítulo aparecen al inicio de cada capítulo para dar una descripción breve de los temas que se presentan.

CAPÍTULO 8

Introducción a la inferencia estadística

8.1 ¿La gente era menos alta en otros tiempos?

8.2 La naturaleza de la estimación

8.3 Estimación de la media μ (σ conocida)

8.4 La naturaleza de la prueba de hipótesis

8.5 Prueba de hipótesis para la media μ (σ conocida): un acercamiento al valor probabilístico

8.6 Prueba de hipótesis para la media μ (σ conocida): un enfoque clásico

NOVEDAD y Parte actualizada
Las secciones de inicio del capítulo se utilizan como un "ejemplo de introducción", que muestra a la estadística en acción respecto al material específico que se presenta en cada capítulo. Cada ejemplo presenta una situación familiar donde se aplica la estadística en forma relevante para el alumno.

© Christa Renee/Getty Images

8.1 ¿La gente era menos alta en otros tiempos?

¿LA GENTE ERA MENOS ALTA EN OTROS TIEMPOS?

La estatura promedio para un inglés de principios del siglo XVII era aproximadamente 5'6"; para una inglesa del siglo XVII era de 5'½". Si bien la estatura promedio en Inglaterra permaneció prácticamente sin cambio en los siglos XVII y XVIII, los colonizadores norteamericanos eran más altos. Los promedios para norteamericanos de la época actual son un poco más de 5'9" para hombres y 5'3¾" para mujeres. Las razones principales para esta diferencia son una mejor nutrición, consumo notablemente mayor de carne y leche, y uso de antibióticos.

Fuente: http://www.plimoth.org/Library/l-short.htm

El National Center for Health Statistics (NCHS) da información estadística que guía acciones y políticas para mejorar la salud del pueblo norteamericano. Datos recientes del NCHS dan la estatura promedio de mujeres en Estados Unidos de 63.7 pulgadas, con una desviación estándar de 2.75 pulgadas.

NOVEDAD y Parte actualizada
Los proyectos de capítulo que se presentan al final de cada capítulo cierran el círculo que inició con las secciones de inicio del capítulo, ya que giran en torno al material que se presenta en cada sección. También incluyen un análisis en miniatura que deberá desarrollarse en forma individual o en pequeños grupos de investigación.

Proyecto del capítulo

¿La gente era menos alta en otros tiempos?

Los datos del Centro Nacional para Estadísticas de la Salud indican que la estatura promedio de una mujer en Estados Unidos es 63.7 pulgadas, con una desviación estándar de 2.75 pulgadas. Use los datos de estaturas de mujeres de la profesión de la salud de la sección 8.1, "la gente era menos alta en otros tiempos?" (p. 395), para contestar las siguientes preguntas.

65.0	66.0	64.0	67.0	59.0	69.0	66.0	69.0	64.0	61.5
63.0	62.0	63.0	64.0	72.0	66.0	65.0	64.0	67.0	68.0
70.0	63.0	63.0	68.0	58.0	60.0	63.5	66.0	64.0	62.0
64.5	69.0	63.5	69.0	62.0	58.0	66.0	68.0	59.0	56.0
64.0	66.0	65.0	69.0	67.0	66.5	67.5	62.0	70.0	62.0

Trabajando en el contenido del capítulo 8

8.199 a. ¿Se satisfacen las suposiciones del intervalo de confianza y métodos de prueba de hipótesis de este capítulo? Explique.

b. Usando los datos muestrales y un nivel de confianza de 95%, estime la estatura media de mujeres de la profesión de la salud. Use la desviación estándar poblacional dada de 2.75 pulgadas.

c. Pruebe lo dicho de que la estatura media de mujeres de la profesión de la salud es diferente de 63.7 pulgadas, que es la estatura media para todas las mujeres de Estados Unidos. Use un nivel de significación de 0.05.

d. En el mismo histograma empleado en la parte b del ejercicio 8.1 de la página 396:

(i) Trace una recta vertical en el valor medio poblacional hipotético, 63.7.

(ii) Trace un segmento de recta horizontal que muestre el intervalo de confianza de 95% de la parte b.

e. ¿La media $\mu = 63.7$ cae en el intervalo? Explique lo que esto significa.

f. Describa la relación entre las dos rectas trazadas en su gráfica para la parte c, del ejercicio 8.2 de la página 396, y las dos rectas trazadas para la parte d de este ejercicio.

g. Con base en los resultados obtenidos antes, ¿parece que las mujeres de este estudio, en promedio, tienen la misma estatura que todas las mujeres de Estados Unidos como lo reporta el NCHS? Explique.

Parte actualizada
En todo el texto se incluye una cantidad abundante de **ejemplos**, mismos que presentan el proceso de resolución paso a paso para los conceptos estadísticos clave y los métodos estadísticos.

318 **CAPÍTULO 6** *Distribuciones de probabilidad normal*

EJEMPLO 6.2 **Para hallar el área en la cola derecha de una curva normal**

Encuentre el área bajo la curva normal a la derecha de $z = 1.52$: $P(z > 1.52)$.

SOLUCIÓN El área a la derecha de la media (toda el área sombreada de la figura) es exactamente 0.5000. El problema pide el área sombreada que no está incluida en 0.4357. Por tanto, restamos 0.4357 de 0.5000:

$$P(z > 1.52) = 0.5000 - 0.4357 = \mathbf{0.0643}$$

Notas: 1. Como lo hemos hecho aquí, siempre trace y aplique leyendas a un dibujo; es muy útil. 2. Fórmese el hábito de escribir z con dos lugares decimales y áreas y probabilidades con cuatro lugares decimales, como en la tabla 3.

Parte actualizada
El libro contiene una amplia cantidad de **casos prácticos** que incorporan conceptos estadísticos para demostrar cómo trabaja la estadística en el mundo real.

CASO PRÁCTICO 1.1 **Explicación de nuestra conducta temprana**

¿Recuerda cuando asistía al jardín de niños? ¡Puede que sí, o puede que no! Si lo recuerda, es muy posible que su preocupación fuera la de hacer amigos y divertirse. ¿Cuál sería la preocupación de sus maestros?

Considere la información que se incluye en la gráfica "Las relaciones fallan hasta en el jardín de niños." La gráfica describe las habilidades que los maestros de jardín de niños consideran esenciales o muy importantes. Ochocientos maestros (sólo una parte de todos ellos) fueron encuestados e informaron de las habilidades y porcentajes indicados. A la cabeza de la lista están "Pone atención" y "No interviene." De los 800 maestros encuestados, 86% consideraron estas habilidades como esenciales o muy importantes. Al ver los porcentajes, se observa que suman más de 100%. Al parecer, a los maestros se les permitió dar más de una habilidad como respuesta.

NOVEDAD
¿Sabía usted...?
Se incluyen historias breves y hechos curiosos que proporcionan información y una mirada divertida a los conceptos que están relacionados o los métodos que se presentan en la sección correspondiente.

¿SABÍA USTED...?

DISPUTAS EN MÉTODO

Las estadísticas no son sólo matemáticas. Hay diferentes modos de abordar inferencias estadísticas y diferentes modos de interpretar lo que nos dicen los datos. Cuanto más significativas sean las diferencias, es más probable que haya acalorados desacuerdos entre quienes tengan puntos de vista opuestos. Una de estas disputas surgió en 1935 en una discusión en la Royal Statistical Society cuando R. A. Fisher desafió a Jerzy Neyman para que se pusiera al corriente en el tema que estaban tratando. La disputa se centró en el uso de intervalos de confianza y el método para probar hipótesis de Pearson y Neyman, contra los intervalos y concepto de valores *p* de Fischer en pruebas de significación. La enemistad duró hasta la muerte de Fisher en 1962.

La variable z normal estándar es nuestra estadística de prueba para esta prueba de hipótesis.

Región crítica: es el conjunto de valores para la estadística de prueba que nos llevará a rechazar la hipótesis nula. El conjunto de valores que no están en la región crítica se denomina **región no crítica** (a veces llamada *región de aceptación*.)

Recuerde que estamos trabajando bajo la suposición de que la hipótesis nula es verdadera. Así, estamos suponiendo que la resistencia media al corte de todos los remaches de la población muestreada es 925. Si éste es el caso, entonces cuando seleccionamos una muestra aleatoria de 50 remaches, podemos esperar que esta media muestral, $\bar{x}$, sea parte de una distribución normal que se centra en 925 y tener un error estándar de $\sigma/\sqrt{n} = 18/\sqrt{50}$, o sea aproximadamente 2.55. Alrededor de 95% de los valores de media muestral serán mayores a 920.8 (un valor de 1.65 errores estándar debajo de la media: $925 - (1.65)(2.55) - 920.8$). Así, si H_o es verdadera y $\mu = 925$, entonces esperamos que $\bar{x}$ sea mayor a 920.8 aproximadamente 95% del tiempo y menos de 920.8 sólo 5% del tiempo.

Si, no obstante, el valor de $\bar{x}$ que obtenemos de nuestra muestra es menor a 920.9, por ejemplo 919.5, tendremos que escoger. Podría ser que: (A)ese valor de $\bar{x}$ (919.5) sea miembro de la distribución con media de 925 aun cuando tiene muy baja probabilidad de que se presente (menor a 0.05), o bien, (B) $\bar{x} = 919.5$ es miembro de una distribución muestral cuya media es menor a 925, que la haría un valor que es más probable que ocurra.

NOVEDAD y Parte actualizada
Incluye cerca de 550 nuevos ejercicios y casi 100 ejercicios actualizados,
esta nueva edición de *Estadística elemental* proporciona tareas actualizadas y relevantes que pueden usar los profesores. Estas tareas tienen la finalidad de fomentar el interés de los estudiantes.

SECCIÓN 8.3 EJERCICIOS

8.19 Discuta las condiciones que deben existir antes que podamos estimar la media poblacional usando las técnicas de intervalo de la fórmula (8.1).

8.20 Determine el valor del coeficiente de confianza $z(\alpha/2)$ para cada una de las situaciones descritas:
a. $1 - \alpha = 0.90$ b. $1 - \alpha = 0.95$

8.21 Determine el valor del coeficiente de confianza $z(\alpha/2)$ para cada una de las situaciones descritas:
a. 98% de confianza b. 99% de confianza

8.22 Determine el nivel de confianza dado el coeficiente de confianza $z(\alpha/2)$ para cada situación:
a. $z(\alpha/2) = 1.645$ b. $z(\alpha/2) = 1.96$
c. $z(\alpha/2) = 2.575$ d. $z(\alpha/2) = 2.05$

8.23 Dada la información, la población muestreada

8.24 Dada la información, la población muestreada está normalmente distribuida, $n = 55$, $\bar{x} = 78.2$, y $\sigma = 12$:
a. Encuentre 0.98 de intervalo de confianza para μ.
b. ¿Se satisfacen las suposiciones? Explique.

8.25 Dada la información, $n = 86$, $\bar{x} = 128.5$, y $\sigma = 16.4$:
a. Encuentre 0.90 de intervalo de confianza para μ.
b. ¿Se satisfacen las suposiciones? Explique.

8.26 Dada la información, $n = 22$, $\bar{x} = 72.3$, y $\sigma = 6.4$:
a. Encuentre 0.99 de intervalo de confianza para μ.
b. ¿Se satisfacen las suposiciones? Explique.

Ejercicios del capítulo

6.101 Según el teorema de Chebyshev, ¿al menos cuánta área hay bajo la distribución normal estándar entre $z = -2$ y $z = +2$? ¿Cuál es el área real bajo la distribución normal estándar entre $z = -2$ y $z = +2$?

6.102 ¿El 60% central de la población normalmente distribuida está entre cuáles dos puntajes estándar?

6.103 Encuentre el puntaje estándar (z) tal que el área arriba de la media y debajo de z bajo la curva normal es:
a. 0.3962 b. 0.4846 c. 0.3712

6.104 Encuentre el puntaje estándar (z) tal que el área bajo la media y arriba de z bajo la curva normal es:
a. 0.3212 b. 0.4788 c. 0.2700

NOVEDAD y Parte actualizada La sección de **Repaso del capítulo** fue adaptada a las necesidades de los estudiantes, funciona como una guía de estudio que se incluye al final del capítulo. Cada sección incluye:

- **En retrospectiva:** es un resumen de los conceptos que se aprendieron en cada capítulo, aquí se indican las relaciones que tiene el material que se analizó con todo lo que se cubrió previamente.

En retrospectiva

Hemos estudiado la distribución de probabilidad normal estándar, la familia más importante de variables aleatorias continuas. Hemos aprendido a aplicarla a todas las otras distribuciones de probabilidad normal y cómo usarla para estimar probabilidades de distribuciones binomiales. Hemos visto una amplia diversidad de variables que tienen esta distribución normal o que son aproximados razonablemente bien por ella.

En el siguiente capítulo examinaremos distribuciones muestrales y aprenderemos a usar la probabilidad normal estándar para resolver aplicaciones adicionales.

- **Vocabulario y conceptos clave:** proporciona una idea a los estudiantes de la cantidad de material que realmente han comprendido.

Vocabulario y conceptos clave

aproximación normal de la binomial (p. 343)
continuidad (p. 344)
curva en forma de campana (p. 315)
curva normal (p. 316)
distribución binomial (p. 343)
distribución normal estándar (pp. 316, 323, 338)
factor de corrección de
porcentaje (p. 316)
probabilidad (p. 316)
probabilidad (p. 316)
proporción (p. 316)
puntaje z (pp. 316, 323)
representación de área para
variable aleatoria (p. 315)
variable aleatoria continua (pp. 315, 344)
variable aleatoria discreta (pp. 315, 344)

- **Objetivos de aprendizaje:** presenta una lista de conceptos clave que debieron haberse aprendido durante el análisis del capítulo; estos conceptos van acompañados de los ejercicios de repaso y las referencias a las secciones para asegurar la comprensión del material del capítulo.

Objetivos de aprendizaje

✓ Entender la diferencia entre una variable discreta y una continua.	p. 315
✓ Entender la relación entre la regla empírica y la curva normal.	p. 313-314, Ejer. 6.1
✓ Entender que una curva normal es una curva en forma de campana, con área total bajo la curva igual a 1.	pp. 315-316, EJ. 6.1, Ejer. 6.40
✓ Entender que la curva normal es simétrica alrededor de la media, con un área de 0.5000 en cada lado de la media.	pp. 315-317, EJ. 6-2
✓ Ser capaz de trazar una curva normal, aplicando leyenda a la media y diversos puntajes z.	p. 314
✓ Entender y ser capaz de usar la tabla 3, áreas de la distribución normal estándar,	EJ.6.1-6.7

- **Ejercicios del capítulo:** ofrece la práctica de todos los conceptos que se presentan en el capítulo, pero también muestra su relación con el material que se aprendió en los capítulos anteriores.

Ejercicios del capítulo

6.101 Según el teorema de Chebyshev, ¿al menos cuánta área hay bajo la distribución normal estándar entre $z = -2$ y $z = +2$? ¿Cuál es el área real bajo la distribución normal estándar entre $z = -2$ y $z = +2$?

6.102 ¿El 60% central de la población normalmente distribuida está entre cuáles dos puntajes estándar?

6.103 Encuentre el puntaje estándar (z) tal que el área arriba de la media y debajo de z bajo la curva normal es:

a. 0.3962 b. 0.4846 c. 0.3712

6.104 Encuentre el puntaje estándar (z) tal que el área bajo la media y arriba de z bajo la curva normal es:

a. 0.3212 b. 0.4788 c. 0.2700

- **Proyecto del capítulo:** ofrece a los estudiantes la oportunidad de visitar nuevamente las secciones de inicio del capítulo para responder las preguntas que se presentaron al inicio del mismo, aplicando los conocimientos que se obtuvieron a partir del estudio del material presentado.

Proyecto del capítulo

Medición de la inteligencia

Todas las distribuciones de probabilidad normales tienen la misma forma y distribución respecto a la media y desviación estándar. En este capítulo aprendimos a usar la distribución de probabilidad normal estándar para contestar preguntas acerca de todas las distribuciones normales. Regresemos a la distribución de puntuaciones de IQ que estudiamos en la sección 6.1, "Medición de la inteligencia" (p. 313), y pongamos a prueba nuestro nuevo conocimiento.

j. ¿Qué proporción de las calificaciones de IQ rebasan 125?

k. ¿Qué porcentaje de las calificaciones del SAT está debajo de 450?

l. ¿Qué porcentaje de las calificaciones del SAT está arriba de 575?

m. ¿Qué calificación del SAT está en el 95avo percentil? Explique lo que esto significa.

- **Examen de práctica del capítulo:** proporciona una autoevaluación formal del dominio del material antes de ser evaluado por el profesor. Las respuestas correctas están al final del libro de texto.

c. ¿Qué porcentaje de la población adulta tiene inteligencia "superior"?

d. ¿Cuál es la probabilidad de seleccionar al azar una persona de esta población que se clasifique abajo del "promedio"?

e. ¿Qué puntuación del IQ está en el 95avo percentil? Explique lo que significa.

Examen de práctica del capítulo

PRIMERA PARTE: Conocer las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

6.1 La distribución de probabilidad normal es simétrica alrededor de **cero.**

6.2 El área total bajo la curva de cualquier distribución normal es **1.0.**

6.3 La probabilidad teórica de que ocurra un valor particular de una variable aleatoria **continua** es exactamente cero.

6.4 La unidad de medida para la calificación estándar es la **misma que la unidad de medida de los datos**.

representan probabilidades de eventos **independientes**.

6.10 La distribución más común de una variable aleatoria continua es la probabilidad **binomial.**

SEGUNDA PARTE: Aplicación de conceptos

6.11 Encuentre las siguientes probabilidades para z, la calificación normal estándar:

a. $P(0 < z < 2.42)$ b. $P(z < 1.38)$

c. $P(z < -1.27)$ d. $P(-1.35 < z2.72)$

6.12 Encuentre el valor de cada puntaje z:

a. $P(z > ?) = 0.2643$ b. $P(z < ?) = 0.17$

c. $z(0.04)$

6.13 Use la notación simbólica $z()$ para dar el nombre simbólico para cada puntaje z que se muestra en la figura de esta página.

6.14 La vida útil de baterías para linternas eléctricas está normalmente distribuida alrededor de una media de 35.6 horas, con una desviación estándar de 5.4 horas. Kevin seleccionó al azar una de estas baterías y la probó. ¿Cuál es la probabilidad de que esta batería dure menos de 40.0 horas?

6.15 Se cree que los tiempos, x, que estudiantes pierden

NOVEDAD y Parte actualizada **El texto incluye instrucciones para el uso de Minitab, Excel y la calculadora TI-83/84** mismas que se presentan a lo largo del texto. Este enfoque didáctico permite que el profesor seleccione la tecnología estadística de su preferencia para que pueda incorporarla en su curso.

NOVEDAD y Parte actualizada **Contiene más de 400 conjuntos de datos,** ordenados desde el más pequeño hasta el más grande; su uso permite que los estudiantes practiquen usando su calculadora estadística o la computadora personal.

Trabajando con sus propios datos

Estos apartados aparecen al final de cada una de las cuatro partes principales del libro. Están diseñados para fomentar la exploración, el aprendizaje independiente de los estudiantes y el pensamiento crítico. Estos apartados se pueden utilizar como un proyecto individual del curso o se pueden trabajar en equipos pequeños.

Trabajando con sus propios datos

Poniendo la probabilidad a trabajar

La distribución de medias muestrales y el teorema de límite central son muy importantes para el desarrollo del resto de este curso. La prueba, que requiere el uso de cálculo, no está incluida en este libro pero la verdad de la SDSM y el CLT se puede demostrar teóricamente y por experimentación. Las siguientes actividades pueden ayudar a verificar ambos enunciados.

A La población

Considere la población teórica que contiene los números 0, 3 y 6 en iguales proporciones.

1 a. Construya la distribución de probabilidad teórica para la toma de un solo número, con restitución, de esta población.

b. Trace un histograma de esta distribución de probabilidad.

c. Calcule la media, μ, y la desviación estándar, σ, para esta población.

B La distribución muestral, teóricamente

Estudiemos la distribución muestral teórica formada por las medias de todas las posibles muestras de tamaño 3 que puedan sacarse de una población dada.

2. Construya una lista que muestra todas las posibles muestras de tamaño 3 que puedan ser sacadas de esta población. (Hay 27 posibilidades.)
3. Encuentre la media de cada una de las 27 posibles muestras de la lista de la respuesta a la pregunta 2.
4. Construya la distribución de probabilidad (la distribución muestral teórica de medias muestrales) para estas 27 medias muestrales.
5. Construya un histograma para esta distribución muestral de medias muestrales.
6. Calcule la media $\mu_{\bar{x}}$ y el error estándar de la media $\sigma_{\bar{x}}$ usando la distribución de probabilidad hallada en la pregunta 4.
7. Demuestre que los resultados hallados en las preguntas 1c, 5 y 6 apoyan las tres afirmaciones hechas por la distribución muestral de medias muestrales y el teorema de límite central. Cite valores específicos para apoyar sus conclusiones.

C La distribución muestral, empíricamente

Veamos ahora si la distribución muestral de medias muestrales y el teorema de límite central se pueden verificar empíricamente, es decir, ¿se cumple cuando la distribución muestral está formada por las medias muestrales que resultan de varias muestras aleatorias?

8. Saque una muestra aleatoria de tamaño 3 de la población dada. Haga una lista de su muestra de tres números y calcule la media para esta muestra.

Puede usar computadora para generar sus muestras. Puede tomar tres "etiquetas" idénticas numeradas 0, 3 y 6, ponerlas en un "sombrero," y sacar su muestra usando restitución entre cada toma. También puede usar dados; sea un 0 representado por 1 y 2; 3 por 3 y 4; y 6, por 5 y 6. También es posible usar números aleatorios para simular la toma de sus muestras, o bien, puede sacar su muestra de la lista de muestras aleatorias que aparecen al final de esta sección. Describa el método que decida usar. (Pida ayuda a su profesor.)

9. Repita la pregunta ocho 49 veces más, de modo que tenga un total de 50 medias muestrales que han resultado de muestras de tamaño 3.

10. Construya una distribución de frecuencia de las 50 medias muestrales halladas en las preguntas 8 y 9.
11. Construya un histograma de la distribución de frecuencia de medias muestrales observadas.
12. Calcule la media $\bar{\bar{x}}$ y desviación estándar $s_{\bar{x}}$, de la distribución de frecuencia formada por las 50 medias muestrales.
13. Compare los valores observados de $\bar{\bar{x}}$ y $s_{\bar{x}}$ con los valores de $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$. ¿Están de acuerdo? ¿La distribución empírica de $\bar{x}$ se parece a la teórica?

A continuación aparecen 100 muestras aleatorias de tamaño 3 que fueron generadas por computadora:

6 3 0	0 3 0	6 6 0	3 3 6	6 6 3	6 3 3
0 0 3	3 0 6	3 3 0	3 6 6	0 3 0	6 6 3
6 6 6	0 3 0	6 3 6	0 6 3	6 0 3	6 3 3
6 0 0	3 0 6	6 3 3	3 3 0	3 3 0	3 3 3
3 3 3	3 0 0	6 6 6	3 3 6	0 0 6	0 6 3
6 6 6	0 0 6	3 3 0	0 6 6	0 0 3	6 6 3
0 0 6	0 0 6	6 6 6	6 3 6	6 6 0	3 0 0
3 6 6	6 3 0	3 6 3	3 0 0	3 3 6	0 6 0
3 0 0	0 3 6	6 3 3	6 0 6	3 3 6	6 0 3
0 3 6	3 6 3	6 6 3	6 6 0	3 3 3	3 0 0
6 3 0	6 6 0	0 3 0	6 6 0	3 6 6	0 3 6
6 3 3	0 3 0	6 6 0	6 6 3	6 6 0	3 0 3
3 6 3	3 6 0	0 0 6	0 3 3	3 6 6	0 3 6
0 6 0	6 0 0	0 6 0	0 6 6	0 3 3	0 3 6
3 3 6	3 3 3	3 3 6	6 3 6	3 3 3	3 6 6
6 3 3	3 0 0	3 0 6	6 0 3	3 6 6	6 0 3
0 3 3	6 3 0	0 3 6	0 3 6		

Material de apoyo para el profesor

Este libro cuenta con una serie de recursos para el profesor, los cuales están disponibles en inglés y sólo se proporcionan a los docentes que lo adopten como texto en sus cursos. Para mayor información, póngase en contacto con el área de servicio a clientes en las siguientes direcciones de correo electrónico:

Cengage Learning México y Centroamérica	clientes.mexicoca@cengage.com
Cengage Learning Caribe	clientes.caribe@cengage.com
Cengage Learning Cono Sur	clientes.conosur@cengage.com
Cengage Learning Paraninfo	clientes.paraninfo@cengage.com
Cengage Learning Pacto Andino	clientes.pactoandino@cengage.com

Los recursos disponibles se encuentran disponibles en el sitio web del libro:

http://latinoamerica.cengage.com/johnson

Las direcciones de los sitios web referidas en el texto no son administradas por Cengage Learning Latinoamérica, por lo que ésta no es responsable de los cambios o actualizaciones de las mismas.

NOTA: En diversos capítulos del libro hay problemas cuyos conjuntos de datos se encuentran disponibles en la página web de este libro. La dirección es http://latinoamerica.cengage.com/johnson **aquí podrá consultar y bajar la información relacionada con estos problemas.**

Agradecimientos

Es un placer reconocer la ayuda y estímulo que hemos recibido durante el desarrollo de este libro de parte de los estudiantes y nuestros colegas del Monroe Community College. Además, deseamos enviar un agradecimiento especial a todos los revisores que leyeron y ofrecieron sugerencias a ésta y todas las ediciones anteriores:

Nancy Adcox, *Mt. San Antonio College*
Paul Alper, *College of St. Thomas*
William D. Bandes, *San Diego Mesa College*
Matrese Benkofske, *Missouri Western State College*
Tim Biehler, *Fingerlakes Community College*
Barbara Jean Blass, *Oakland Community College*
Austin Bonis, *Rochester Institute of Technology*
Nancy C. Bowers, *Pennsylvania College of Technology*
Shane Brewer, *College of Eastern Utah, San Juan Campus*
Robert Buck, *Slippery Rock University*
Louis F. Bush, *San Diego City College*
Ronnie Catipon, *Franklin University*
Rodney E. Chase, *Oakland Community College*
Pinyuen Chen, *Syracuse University*
Wayne Clark, *Parkland College*
David M. Crystal, *Rochester Institute of Technology*
Joyce Curry and Frank C. Denny, *Chabot College*
Larry Dorn, *Fresno Community College*
Shirley Dowdy, *West Virginia University*
Thomas English, *Pennsylvania State University, Erie*
Kenneth Fairbanks, *Murray State University*
Dr. William P. Fox, *Francis Marion University*
Joan Garfield, *University of Minnesota General College*
Monica Geist, *Front Range Community College*
David Gurney, *Southeastern Louisiana University*
Edwin Hackleman
Carol Hall, *New Mexico State University*
Silas Halperin, *Syracuse University*
Noal Harbertson, *California State University, Fresno*
Hank Harmeling, *North Shore Community College*
Bryan A. Haworth, *California State College at Bakersfield*
Harold Hayford, *Pennsylvania State University, Altoona*
Jim Helms, *Waycross College*
Marty Hodges, *Colorado Technical University*
John C. Holahan, *Xerox Corporation*
James E. Holstein, *University of Missouri*
Soon B. Hong, *Grand Valley State University*
Robert Hoyt, *Southwestern Montana University*
Peter Intarapanach, *Southern Connecticut State University*
T. Henry Jablonski, Jr., *East Tennessee State University*
Brian Jean, *Bakersfield University*
Jann-Huei Jinn, *Grand Valley State University*
Sherry Johnson
Meyer M. Kaplan, *The William Patterson College of New Jersey*
Michael Karelius, *American River College*
Anand S. Katiyar, *McNeese State University*
Jane Keller, *Metropolitan Community College*
Gayle S. Kent, *Florida Southern College*
Andrew Kim, *Westfield State College*
Amy Kimchuk, *University of the Sciences in Philadelphia*
Raymond Knodel, *Bemidji State University*
Larry Lesser, *University of Northern Colorado*
Natalie Lochner, *Rollins College*
Robert O. Maier, *El Camino College*
Linda McCarley, *Bevill State Community College*
Mark Anthony McComb, *Mississippi College*
Carolyn Meitler, *Concordia University Wisconsin*
John Meyer, *Muhlenberg College*
Jeffrey Mock, *Diablo Valley College*
David Naccarato, *University of New Haven*
Harold Nemer, *Riverside Community College*
John Noonan, *Mount Vernon Nazarene University*
Dennis O'Brien, *University of Wisconsin, LaCrosse*
Chandler Pike, *University of Georgia*
Daniel Powers, *University of Texas, Austin*
Janet M. Rich, *Miami-Dade Junior College*
Larry J. Ringer, *Texas A & M University*
John T. Ritschdorff, *Marist College*
John Rogers, *California Polytechnic Institute at San Luis Obispo*
Neil Rogness, *Grand Valley State University*
Thomas Rotolo, *University of Arizona*
Barbara F. Ryan and Thomas A. Ryan, *Pennsylvania State University*
Robert J. Salhany, *Rhode Island College*
Melody Smith, *Dyersburg State Community College*
Dr. Sherman Sowby, *California State University, Fresno*
Roger Spalding, *Monroe County Community College*
Timothy Stebbins, *Kalamazoo Valley Community College*
Howard Stratton, *State University of New York at Albany*
Larry Stephens, *University of Nebraska-Omaha*
Paul Stephenson, *Grand Valley State University*
Richard Stockbridge, *University of Wisconsin, Milwaukee*
Thomas Sturm, *College of St. Thomas*
Edward A. Sylvestre, *Eastman Kodak Co.*
Gwen Terwilliger
William K. Tomhave, *Concordia College, Moorhead, MN*
Bruce Trumbo, *California State University, Hayward*
Richard Uschold, *Canisius College*
John C. Van Druff, *Fort Steilacoom Community College*
Philip A. Van Veidhuizen, *University of Alaska*
John Vincenzi, *Saddleback College*
Kenneth D. Wantling, *Montgomery College*
Joan Weiss, *Fairfield University*
Mary Wheeler, *Monroe Community College*
Barbara Whitney, *Big Bend Community College*
Sharon Whitton, *Hofstra University*
Don Williams, *Austin College*
Rebecca Wong, *West Valley College*
Pablo Zafra, *Kean University*
Yvonne Zubovic, *Indiana University Purdue University, Fort Wayne*

Robert Johnson
Patricia Kuby

CAPÍTULO

1 Estadística

1.1 Los norteamericanos, una mirada a sí mismos

1.2 ¿Qué es la estadística?

1.3 Medibilidad y variabilidad

1.4 Recolección (obtención) de datos

1.5 Comparación entre probabilidad y estadística

1.6 Estadística y la tecnología

© Rudi Von Briel/PhotoEdit

1.1 Los norteamericanos, una mirada a sí mismos

La Oficina Norteamericana del Censo (U.S. Census Bureau) publica anualmente el *Statistical Abstract of the United States* (Resumen estadístico de Estados Unidos), libro de más de 1000 páginas que nos da una idea de muchas de las más oscuras y poco comunes facetas de nuestras vidas. Ésta es sólo una de miles de fuentes de toda clase de datos que siempre hemos deseado saber y que nunca preguntamos. ¿Le interesa saber cuántas horas trabajamos y jugamos? ¿Cuánto gastamos en bocadillos? ¿Cuánto ha subido de precio la manzana roja? Todo esto y más, mucho más, se puede hallar en el *Statistical Abstract* (http://www.census.gov/statab/www).

Los extractos estadísticos que veremos a continuación provienen de diversas fuentes y representan sólo una pequeña muestra de lo que puede conocerse estadísticamente respecto a los norteamericanos. ¡Veamos!

MÉTODO DE COMUNICACIÓN PREFERIDO POR LOS TRABAJADORES

Datos de Anne R. Carey and Ron Coddington, © 2004 USA Today.

¿DEBE ELIMINARSE EL CENTAVO?

Datos de Shannon Reilly and Chad Palmer, © 2004 USA Today.

¿LE GUSTARÍA CUMPLIR CIEN AÑOS?

Datos de USA Today, 10/13/2003.

VIAJAR EN AUTO ES UN RIESGO IMPORTANTE PARA LOS ADOLESCENTES

Cerca de 3 500 adolescentes perdieron la vida en accidentes de tránsito en Estados Unidos. Los vehículos eran conducidos por jóvenes. Esta causa de muerte es superior a cualquier otra enfermedad o lesiones que se presenten en los adolescentes

Los conductores de 16 años tienen el mayor porcentaje de sufrir accidentes fatales.

Edad del conductor	Participación en accidentes mortales por 100 millones de millas recorridas.
16	9.3
17	8.3
18	6.5
19	7.2
20–24	4.3
25–29	2.3
30–59	1.6
60–69	1.6
70+	4.1

Datos de USA Today, © 2003.

Los ejemplos precedentes y una gran cantidad de mediciones adicionales se emplean para describir la vida en Estados Unidos.

Considere la gráfica "¿Le gustaría cumplir cien años?" Si alguien le pregunta "¿Le gustaría vivir hasta los 100 años?" ¿Cuál hubiera sido su respuesta? ¿Considera que la gráfica representa correctamente su respuesta? ¿Le hace detenerse y preguntarse cómo se obtuvo la información y de dónde proviene? ¿Cree usted en el material "impreso"? Cuando estudie el capítulo 1, empezará por aprender a leer y analizar medidas estadísticas para obtener las conclusiones adecuadas. A continuación podrá investigar más a fondo sobre "Los norteamericanos, una mirada a sí mismos" en la sección de Proyecto del capítulo en los ejercicios 1.88 y 1.89 (p. 35).

SECCIÓN 1.1 EJERCICIOS

1.1 a. ¿Cada una de las gráficas estadísticas que se presentan en esta sección parecen sugerir que la información está relacionada con algún tipo de población? ¿Es ése el caso? Justifique su respuesta.

b. Describa la información que se obtuvo y se utilizó para determinar las estadísticas que se reportaron en el "Método de comunicación preferido por trabajadores."

c. "63%:sí" fue una de las estadísticas específicas reportadas en la gráfica "¿Le gustaría cumplir cien años?" Describa lo que le dice esa estadística.

d. Considere la gráfica ¿Debe eliminarse el centavo? Si a usted se le hubiera hecho esa pregunta, ¿cuál habría sido su respuesta? ¿Piensa que su respuesta está representada con precisión en el diagrama? ¿Qué significa realmente el porcentaje asociado con su respuesta? Explique.

e. ¿Cómo interpreta el 7.2 que aparece para la edad de 19 años del conductor en la gráfica "¿Viajar en auto es un riesgo importante para los adolescentes?"

1.2 a. Escriba un párrafo de 50 palabras que describa lo que significa para usted la palabra *estadística* en este momento.

b. Escriba un párrafo de 50 palabras que describa lo que significa para usted la palabra *aleatorio*.

c. Escriba un párrafo de 50 palabras que describa lo que significa para usted la palabra *muestra*.

1.3 ¿Trabaja duro para ganar dinero? Los profesionales de Java piensan que sí, ya que reportan una gran cantidad de horas trabajadas en sus sitios de trabajo. Se preguntó a varios desarrolladores de Java en todo el mundo cuántas horas trabajaban por semana. A continuación aparece la cantidad promedio de horas trabajadas por semana en varias regiones de Estados Unidos y en el mundo.

Región	Horas trabajadas	Región	Horas trabajadas
U.S.	48	California	50
Noreste	47	NO del Pacífico	47
Atlántico	49	Canada	43
Sur	47	Europa	48
Medio Oeste	47	Asia	47
Central	51	América del Sur y África	49

a. ¿Cuántas horas trabaja usted por semana (o espera trabajar una vez que se haya titulado)?

b. ¿Qué sucedió con la semana de trabajo de 40 horas? ¿Parece existir para el profesional de Java?

c. La información de esta tabla, ¿hace aparecer atractiva la carrera de ser un desarrollador profesional de Java?

1.4 "Lo que hagas depende de dónde trabajes." Cuando se agrupan de acuerdo al tipo de organización para la que trabajan, quienes se arriesgan (trabajadores por honorarios) suben de nuevo hasta la cima. Para los creadores de Java, los trabajadores por honorarios ganan más dinero, seguidos por quienes se emplean en empresas públicas; ambos grupos ganan casi el doble que quienes trabajan para instituciones educativas.

Fuente: Jupitermedia Corporation

a. Examine la gráfica y describa cuidadosamente la "imagen" que la gráfica le ha transmitido.

b. ¿La información de esta tabla le hace pensar que la profesión de desarrollador Java es atractiva?

c. ¿Puede usted concluir algo acerca de la disponibilidad de empleos en estos seis grupos de lugares de trabajo?

d. ¿Puede usted concluir algo acerca del número de horas que trabaja por semana un profesional de Java para obtener estos ingresos?

1.2 ¿Qué es la estadística?

Cuando iniciamos nuestro viaje en el estudio de la estadística, debemos empezar por la definición de *estadística* y extendernos en los detalles necesarios.

La estadística es el lenguaje universal de las ciencias. Como usuarios potenciales de la estadística, es necesario dominar la "ciencia" y el "arte" de utilizar correctamente su metodología. El empleo cuidadoso de los métodos estadísticos permite obtener información precisa de los datos. Estos métodos incluyen: (1) definir cuidadosamente la situación, (2) obtener los datos, (3) resumir con precisión los datos y (4) obtener y comunicar las conclusiones importantes.

La estadística implica información, números y gráficas visuales para resumir esta información, y su interpretación. El término *estadística* posee varios significados para personas de diversos entornos e intereses. Para algunos, es un medio para hacer "trucos" en los que la persona trata de confundir a otros con información y conclusiones incorrectas. Para otros, es una forma de obtener y presentar información. Aún más, para otro grupo de personas es una forma de "tomar decisiones de

frente a la incertidumbre." En la perspectiva idónea, cada uno de estos puntos de vista es correcto.

El terreno de la estadística puede dividirse a grandes rasgos en dos campos de acción: estadística descriptiva y estadística inferencial. La *estadística descriptiva* es lo que piensa la mayoría de las personas al escuchar la palabra *estadística.* Incluye la obtención, presentación y descripción de los datos muestrales. El término *estadística inferencial* se refiere a la técnica de interpretación de los valores resultantes de las técnicas descriptivas y la toma de decisiones, así como a la obtención de conclusiones relativas a la población.

La estadística es más que sólo números: son los datos, lo que se hace con ellos, lo que se aprende de los datos y las conclusiones resultantes. Se utilizará la siguiente definición:

Estadística: es la ciencia que se encarga de obtener, describir e interpretar los datos.

Antes de comenzar el estudio detallado de la estadística, veamos algunos ejemplos de cómo y cuándo es posible aplicar la estadística.

CASO PRÁCTICO 1.1 Explicación de nuestra conducta temprana

¿Recuerda cuando asistía al jardín de niños? ¡Puede que sí, o puede que no! Si lo recuerda, es muy posible que su preocupación fuera la de hacer amigos y divertirse. ¿Cuál sería la preocupación de sus maestros?

Considere la información que se incluye en la gráfica "Las relaciones fallan hasta en el jardín de niños." La gráfica describe las habilidades que los maestros de jardín de niños consideran esenciales o muy importantes. Ochocientos maestros (sólo una parte de todos ellos) fueron encuestados e informaron de las habilidades y porcentajes indicados. A la cabeza de la lista están "Pone atención" y "No interviene." De los 800 maestros encuestados, 86% consideraron estas habilidades como esenciales o muy importantes. Al ver los porcentajes, se observa que suman más de 100%. Al parecer, a los maestros se les permitió dar más de una habilidad como respuesta.

LAS RELACIONES FALLAN HASTA EN EL JARDÍN DE NIÑOS

Datos de Julia Neyman y Alejandro Gonzalez, © 2004 USA Today.

CASO PRÁCTICO 1.2 Descripción de nuestro lado más amable

La industria del turismo (SPA) está en auge. La International SPA Association reporta estadísticas que demuestran que atender solícitamente a las personas puede rendir utilidades. Los ingresos de los sitios de atracción turística y los salones de SPA

han aumentado en 409% entre los años 1997 y 2003. De hecho, la industria del turismo es la cuarta industria más grande de tiempo libre en Estados Unidos, más que los parques de diversiones y cines.

Fuente: Rochester, *Democrat and Chronicle*, 12/5/2004. Reimpreso con permiso.

Las gráficas anteriores brindan una gran cantidad de información con relación a la industria del turismo. Considere la información que tendría que recolectarse para construir las tablas y gráficas, no sólo el número de centros de turismo sino el tipo o categoría de los mismos, y el género de los visitantes. Pero, ¿de dónde provienen estas cifras? Siempre observe la fuente de las estadísticas publicadas. En este caso la fuente es la International SPA Association. La asociación es reconocida mundialmente como una organización profesional y es la voz de la industria del turismo.

CASO PRÁCTICO 1.3 Información acerca de lo que piensan las empresas

Los periódicos publican gráficas y tablas que indican lo que piensan en conjunto diversas organizaciones o personas. ¿Alguna vez se ha preguntado cuánto de lo que pensamos está influenciado directamente por la información que leemos en estos artículos?

¿SE PREOCUPAN LAS EMPRESAS POR EL PESO DE SUS TRABAJADORES?

Datos de Darryl Haralson y Alejandro González, © 2004, USA Today.
Margen de error ±5 puntos porcentuales.

La siguiente gráfica reporta que 65% de las empresas no se preocupan del siguiente hecho: el aumento en la obesidad de su personal tiene un impacto directo en los ingresos o la productividad. ¿De dónde llegó esta información? Observe la fuente, Duffey Communications. ¿Cómo se obtuvo la información? Esta empresa realizó un estudio en 450 empresas y personajes de la política. Se da un margen de error de ±5 puntos porcentuales. (Recuerde leer las letras pequeñas, por lo general en la parte inferior de una gráfica o tabla de estadísticas.) Con base en esta información, entre 60% y 70% de las empresas no se preocupan porque un personal cada vez más obeso tenga impacto sobre los ingresos o la productividad. Esto parece sorprendente, dada la cantidad de información que aparece

en los noticieros y medios impresos con relación a la obesidad y sus efectos en la salud, así como la cantidad de dinero y la atención que se presta a las dietas y los métodos para perder peso.

CASO PRÁCTICO 1.4 La estadística es una cuestión engañosa

"Una onza de técnica de estadística exige una onza de sentido común para su correcta aplicación."

Considere la **International Shark Attack File** (ISAF, por sus siglas en inglés) (Archivo Internacional de Ataques de Tiburones). El ISAF es administrado por la American Elasmobranch Society y el Florida Museum of Natural History (Museo de Historia Natural de Florida) es una compilación de todos los ataques conocidos de tiburones, misma que se ilustra en la siguiente gráfica.

© lodrakon /Shutterstock

Territorio	Total de ataques	Ataques mortales	Último ataque mortal
Estados Unidos (sin Hawai)	761	39	2004
(Australia	294	134	2004
Africa	264	69	2004
Asia	116	55	2000
Pacífico/Oceanía	114	47	2003
Islas Hawai (sin Hawai)			
Hawai	100	15	2004
Sudamérica	96	22	2004

Territorio	Total de ataques	Ataques mortales	Último ataque mortal
Antillas y Bahamas	59	19	1972
Centroamérica	58	31	1997
Nueva Zelanda	45	9	1968
Europe	38	18	1984
Bermuda	4	0	
General	20	6	1965
Mundo	1969	464	2004

Fuente: http://gerber.iwarp.com/Attack/GAttack/World.htm

¿Sentido común? Si se usa el sentido común y se revisa el párrafo anterior, de seguro que cualquiera se alejaría de Estados Unidos si disfrutan del mar. Casi dos quintos de los ataques mundiales de tiburones ocurrieron en Estados Unidos. ¡Las aguas de ese país deben estar llenas de tiburones, y los tiburones deben estar locos!¿Recuerda lo que le dice el sentido común? ¿Es un poco confusa la gráfica?

¿Qué más podría influir en las estadísticas que se presentan? En primer lugar se debe considerar la porción de la frontera del país o del continente que está en contacto con el océano. En segundo lugar, ¿quién da seguimiento a estos ataques? En este caso, se indica en la parte superior de la gráfica, el Museo de Historia Natural de Florida, de Estados Unidos. Aparentemente, este país está tratando de dar seguimiento a los ataques de tiburones sin provocación. ¿Qué otras diferencias hay en Estados Unidos en comparación con las otras regiones? ¿Es el océano una zona de recreación en los otros lugares? ¿Cuál es la economía de estas otras zonas, y/o quién da seguimiento a los ataques de tiburones?

Recuerde considerar la fuente siempre que analice un reporte estadístico. Asegúrese de observar un panorama completo.

Los usos de la estadística son ilimitados. Es mucho más difícil citar un campo de acción donde no se use estadística, en comparación con el hecho de mencionar uno en el que la estadística desempeñe un papel integral. Los siguientes son algunos ejemplos de cómo y dónde se emplean estadísticas:

- En educación, la estadística descriptiva se emplea para describir los resultados de exámenes.
- En ciencias, los datos resultantes de experimentos deben obtenerse y analizarse.
- En el gobierno se obtiene una gran cantidad de diferentes tipos de datos estadísticos en todo momento. De hecho, el gobierno de Estados Unidos es probablemente el mayor recolector de datos estadísticos del mundo.

Una parte muy importante del proceso estadístico es estudiar los resultados estadísticos y formular las conclusiones apropiadas. Estas conclusiones deben comunicarse con precisión, porque nada se gana en una investigación a menos que los descubrimientos se compartan con otros. En todas partes se informa de estadísticas: periódicos, revistas, radio y televisión. Leemos y escuchamos acerca de toda clase de nuevos resultados de investigaciones, en especial en el campo relacionado con la salud.

Para continuar nuestro estudio de estadística, necesitamos "hablar claro". La estadística tiene su propia terminología, es decir, términos fuera de la *estadística descriptiva* y de la *estadística inferencial*, que debe definirse e ilustrarse. El concepto de una población es la idea más importante en estadística.

Población: es la colección, o conjunto, de individuos, objetos o eventos cuyas propiedades serán analizadas.

La población es la colección completa de individuos u objetos de interés para la persona que obtiene los datos de la muestra. La población de interés debe definirse cuidadosamente y se considera que está definida por completo sólo cuando se especifica la lista de elementos que pertenecen a ella. El conjunto de "todos los estudiantes que han asistido alguna vez a una universidad estadounidense" es un ejemplo de una población bien definida.

Por lo general se piensa que una población es una colección de personas, pero en estadística la población puede ser una colección de animales, de objetos manufacturados o de cualquier cosa. Por ejemplo, el conjunto de todos los árboles de secuoya en California puede ser una población.

Hay dos tipos de poblaciones: finitas e infinitas. Cuando se puede enumerar físicamente a todos los elementos que componen a una población se dice que la

¿SABÍA USTED...?

Sólo un momento

Un *momento* es una unidad real de tiempo muy corto que se usa en ingeniería de computación. Si usted toma su desayuno en un momento, entonces tendrá que hacerlo en 10 milisegundos ¡exactamente 0.01 segundo!

población es **finita.** Cuando los elementos son ilimitados, se dice que la población es **infinita.** Los libros de una biblioteca universitaria constituyen una población finita; sin embargo, el OPAC (Online Public Access Catalog, que es el catálogo computarizado de tarjetas para bibliotecas) enumera exactamente los elementos que le pertenecen. Todos los electores registrados en Estados Unidos constituyen una población finita muy grande; en caso necesario, se puede compilar una composición de todos los padrones electorales. Por otra parte, la población de todas las personas que podrían tomar aspirina y la población de todos los focos de 40 w que se producirán en la planta de Sylvania son infinitas. El estudio de grandes poblaciones se dificulta grandemente, en consecuencia, se acostumbra seleccionar una *muestra* y estudiar los datos que la integran.

Muestra: es el subconjunto de una población.

Una muestra está integrada por los individuos, objetos o medidas seleccionados de la población por la persona que obtiene los elementos de la muestra.

Variable (o variable de respuesta): es una característica de interés relacionada con cada elemento individual de una población o muestra.

La edad de un estudiante que ingresa a una universidad, el color de su cabello, la estatura y su peso son cuatro variables.

Dato: es el valor de la variable asociada a un elemento de una población o muestra. Este valor puede ser un número, una palabra o un símbolo.

Por ejemplo, Juan Pérez ingresó a la universidad a la edad de 23 años, su cabello es café, mide 1.80 m y su peso es de 83 kg. Estas cuatro piezas de datos son los valores de las cuatro variables aplicadas a Juan Pérez.

Datos: son el conjunto de valores que se obtienen de la variable a partir de cada uno de los elementos que pertenecen a la muestra.

El conjunto de los 25 valores de estatura que se obtuvieron de los 25 estudiantes es un ejemplo de un conjunto de datos.

Experimento: es una actividad planeada cuyos resultados producen un conjunto de datos.

Un experimento incluye las actividades tanto para seleccionar los elementos como para obtener los valores de los datos.

Parámetro: es un valor numérico que resume todos los datos de una población completa.

La edad "promedio" al momento de inscribirse para todos los estudiantes que han asistido alguna vez a una universidad, y la "proporción" de estudiantes que tenían más de 21 años de edad cuando ingresaron a la universidad, son ejemplos de dos parámetros poblacionales. Un parámetro es un valor que describe a toda la población. A menudo se utiliza una letra griega para simbolizar la denominación de un

parámetro. Estos símbolos serán asignados a medida que se estudien los parámetros específicos.

Para todo parámetro existe un *estadístico muestral correspondiente*. La estadística describe a la muestra en la misma forma que el parámetro describe a la población.

PARA SU INFORMACIÓN
Los **p**arámetros describen la población. Cabe observar que las dos palabras empiezan con la letra **p**. Un estadístico describe a la muestra. En el idioma inglés tanto la palabra estadístico como muestra inician con la letra ***s*** (*statistic* y *sample*, respectivamente).

Estadístico: es un valor numérico que resume los datos de la muestra.

La estatura "promedio" encontrada al utilizar el conjunto de 25 estaturas es un ejemplo de un estadístico muestral. Un estadístico es un valor que describe una muestra. Casi todos los estadísticos muestrales se determinan con ayuda de fórmulas y suele asignárseles denominaciones simbólicas con el uso de letras del alfabeto español (por ejemplo $\bar{x}$, s y r).

EJEMPLO 1.5 Aplicación de términos básicos

Un estudiante de estadística está interesado en determinar algo sobre el valor promedio en dólares de los automóviles que pertenecen al cuerpo docente de nuestra universidad. Cada uno de los ocho términos que acabamos de describir puede identificarse en esta situación.

1. La *población* es la colección de todos los automóviles que pertenecen a todos los miembros del cuerpo docente de nuestra universidad.
2. Una *muestra* es cualquier subconjunto de esa población. Por ejemplo, los automóviles que pertenecen a los profesores del departamento de matemáticas integran a la muestra.
3. La *variable* es el "valor en dólares" de cada automóvil individual.
4. Un *dato* es el valor en dólares de un automóvil en particular. El automóvil del Sr. Sánchez, por ejemplo, está valuado en 9400 dólares.
5. Los *datos* serían el conjunto de valores que corresponden a la muestra obtenida (9 400, 8 700, 15 950…).
6. El *experimento* serían los métodos que se aplican para seleccionar los automóviles que integran a la muestra y determinar el valor de cada automóvil de la muestra. El experimento podría realizarse preguntando a cada miembro del departamento de matemáticas, o de otras formas.
7. El *parámetro* sobre el que se está buscando información es el valor "promedio" de todos los automóviles de la población.
8. El *estadístico* que encontrará es el valor "promedio" de todos los automóviles de la muestra.

PARA SU INFORMACIÓN
Los parámetros tienen valor fijo, mientras que los estadísticos varían su valor.

Nota: Si se toma una segunda muestra, quizá el conjunto de personas seleccionadas sería diferente, por ejemplo el departamento de inglés, y en consecuencia, el estadístico promedio se anticiparía para un valor diferente. No obstante, el valor promedio de "todos los automóviles del profesorado" no cambiaría.

Básicamente, hay dos clases de variables: 1) variables que resultan en información *cualitativa* y (2) variables que resultan en información *cuantitativa.*

Variable cualitativa, de atributos, o categórica: es una variable que clasifica o describe a un elemento de una población.

Variable cuantitativa o numérica: es aquella que cuantifica un elemento de una población.

Una muestra de cuatro clientes de una peluquería fue cuestionada en cuanto al "color de su cabello", "la ciudad donde vive" y el "nivel de satisfacción" respecto a los resultados en la peluquería. Las tres variables son ejemplos de variables cualitativas (de atributos), ya que describen alguna característica de la persona, y todas las personas con el mismo atributo pertenecen a la misma categoría. Los datos recolectados fueron {rubio, café, negro, café}, {Brighton, Columbus, Albany, Jacksonville}, y {muy satisfecho, satisfecho, algo satisfecho}.

El "costo total" de los libros de texto adquiridos por cada estudiante para las clases de este semestre es un ejemplo de variable cuantitativa (numérica). Se obtuvo una muestra con los datos siguientes: $238.87, $94.57, $139.24. [Para determinar el "costo promedio", simplemente se suman los tres números y el resultado se divide entre tres: (238.87 + 94.57 + 139.24)/3 = $157.56.]

Nota: Algunas operaciones aritméticas, como sumar y promediar, tienen sentido para los datos que resultan de una variable cuantitativa.

Cada uno de estos tipos de variables (cualitativas y cuantitativas) pueden subdividirse aún más, como se ilustra en el diagrama siguiente.

Las variables cualitativas pueden caracterizarse como nominales u ordinales.

Variable nominal: es una variable cualitativa que caracteriza (describe o identifica) a un elemento de una población. Para los datos resultantes de una variable nominal, las operaciones aritméticas no sólo carecen de sentido sino que tampoco se puede asignar un orden a las categorías.

En la encuesta anterior que se aplicó a los cuatro clientes de una peluquería, dos de las variables, "color de su cabello" y "ciudad donde vive", son ejemplos de variables nominales, ya que ambas identifican alguna característica de la persona y carecerían de sentido para encontrar el promedio muestral al sumar y dividir entre cuatro. Por ejemplo, (rubio + café + negro + café)/4 no está definido. Además, el color del cabello y la ciudad donde vive no tienen un orden en sus categorías.

Variable ordinal: es una variable cualitativa que presenta una posición, o clasificación, ordenada.

En la encuesta anterior de cuatro clientes de una peluquería, la variable "nivel de satisfacción" es un ejemplo de variable ordinal, ya que presenta una clasificación ordenada: "muy satisfecho" está antes que "satisfecho", que se encuentra antes que "algo satisfecho". Otro ejemplo de una variable ordinal sería la clasificación de cin-

co fotografías de paisaje según la preferencia de alguien: primera elección, segunda elección, etcétera.

Las variables cuantitativas o numéricas también pueden subdividirse en dos clasificaciones: variables *discretas* y variables *continuas.*

Variable discreta: es una variable cuantitativa que puede asumir un número contable (o finito) de valores. Intuitivamente, la variable discreta puede asumir los valores correspondientes a puntos aislados a lo largo de un intervalo de recta. Es decir, entre dos valores cualesquiera siempre hay un hueco.

Variable continua: es una variable cuantitativa que puede asumir una cantidad incontable de valores. Intuitivamente, la variable continua puede asumir cualquier valor a lo largo de un intervalo de recta, incluyendo cualquier valor posible entre dos variables determinadas.

En muchos casos, es posible distinguir los dos tipos de variables decidiendo si las variables están relacionadas con un conteo o una medición. La variable "número de cursos en los que usted está inscrito actualmente" es un ejemplo de una variable discreta; sus valores se determinan al contar el número de cursos. (Al contar, no es posible que ocurran valores fraccionarios; en consecuencia, entre los valores que puedan ocurrir hay huecos (números fraccionarios). La variable "peso de los libros y material que carga al asistir hoy a clases" es un ejemplo de variable aleatoria continua; los valores de la variable se encuentran midiendo el peso. (Al medir, puede ocurrir cualquier valor fraccionario; así, a lo largo de la recta es posible obtener cualquier valor.)

Cuando intente determinar si una variable es continua o discreta, recuerde analizar la variable y piense en los valores que podrían ocurrir. No considere los valores de datos que se han registrado porque pueden ser engañosos.

Considere la variable "calificación asignada por un juez" en una competencia de patinaje de figuras. Si se consideran algunas calificaciones que ya se han asignado, 9.9, 9.5, 8.8, 10.0, y se observa la presencia de cifras decimales, podría pensarse que todas las fracciones son posibles y concluir que la variable es continua. Sin embargo, esto no es cierto; de hecho, entre los valores posibles hay huecos y la variable es discreta.

Nota: No permita que la apariencia de los datos lo engañe al momento de determinar el tipo de la variable. Las variables cualitativas no siempre son fáciles de reconocer, algunas veces se presentan como números. La muestra anterior de colores de cabello pueden codificarse como: 1 = negro, 2 = rubio, 3 = café. Los datos de la muestra se verían así: {2, 3, 1, 3} aunque siguen siendo datos de atributos. Luego al calcular el promedio aritmético de la variable "color de cabello" [(2 + 3 + 1 + 3)]/4 = 9/4 = 2.25] se obtiene un dato que sigue careciendo de sentido. Las ciudades de residencia pueden clasificarse usando números de códigos postales, pero también en este caso el código postal promedio tampoco tiene sentido; en consecuencia, los códigos postales también serían variables nominales cualitativas.

Se considerará otro ejemplo. Suponga que después de inspeccionar un estacionamiento, los datos muestrales se resumen al reportar 5 automóviles rojos, 8 azules, 6 verdes y 2 amarillos. Es necesario considerar cada fuente individual para determinar el tipo de información que se está recolectando. Un automóvil específico era rojo; "rojo" es el dato de ese automóvil, y "rojo" es un atributo. Por tanto, esta colección (5 rojos, 8 azules, etc.) es un resumen de los datos nominales.

Otro ejemplo de información engañosa es un número de identificación. Por ejemplo el vuelo 249 y la habitación 168 parecen ser datos numéricos. El número 249 no

describe ninguna propiedad del vuelo: no indica si está retrasado o llegó a tiempo, la calidad del refrigerio servido, el número de pasajeros o cualquier detalle adicional relacionado con el vuelo. El número de vuelo sólo identifica un vuelo específico. Los números de las licencias para conducir, de afiliación a la Seguridad Social y de las cuentas bancarias son, en todos los casos, números de identificación en el sentido nominal, no en el sentido cuantitativo. En consecuencia, por sí mismos no funcionan como variables.

Recuerde revisar la variable individual y un dato individual, de esta forma tendrá poca dificultad para distinguir entre los diferentes tipos de variables, tanto cualitativas como cuantitativas.

CASO PRÁCTICO 1.6 Datos del censo

Es frecuente que la información de un censo sea noticia, independientemente de si se trata de un censo local o nacional. Los resultados del censo tienen diversas aplicaciones: desde ayudar a determinar los diferentes puestos legislativos y asignar impuestos hasta el otorgamiento de información a los visitantes de una ciudad (como aquí se muestra). Todos somos parte del censo de población y hemos visto informes semejantes a los que se presentan a continuación.

Quiénes somos: condado Lee, Florida

La Nueva Prensa-Visitantes en 2001

Fuente: The News-Press

Ahora veamos cómo podemos aplicar nuestra nueva terminología al informe anterior. Considerando el título general para las gráficas presentadas, la población general de interés serían los residentes del condado de Lee, Florida. Para ser más

específicos en cada gráfica, la población serían todas las familias del condado de Lee para la gráfica "Ingreso familiar", todos los residentes del condado de Lee para la gráfica "Edad", y todos los adultos registrados para la gráfica "Partido político". Las variables necesarias para completar estas gráficas son ingreso, edad y afiliación a un partido político. El ingreso y la edad son variables continuas, mientras que la afiliación a un partido es una variable nominal. Es muy posible que los datos de las gráficas en círculo fueran recolectados por categorías y luego se hayan calculado los porcentajes. Se emplearon conteos para generar la gráfica de barras de partidos políticos.

SECCIÓN 1.2 EJERCICIOS

1.5 La *estadística* se define en la página 4 como "la ciencia de recolectar, describir e interpretar datos". Con sus propias palabras, escriba una oración que describa cada una de las tres actividades estadísticas. Conserve su trabajo para el ejercicio 1.87.

1.6 Determine cuál de las siguientes frases es descriptiva en su naturaleza y cuál es inferencial. Consulte el caso "Las relaciones fallan hasta en el jardín de niños" en el caso práctico 1.1 (pp. 4-5).

a. De todos los maestros de jardín de niños en Estados Unidos, 32% dicen que "Conocer el alfabeto" es una habilidad esencial.

b. De los 800 maestros que fueron entrevistados en Estados Unidos, 32% indican que "Conocer el alfabeto" es una habilidad esencial.

1.7 Determine cuál de las siguientes frases es descriptiva en su naturaleza y cuál es inferencial. Consulte "Atender bien a las personas produce utilidades" en el caso práctico 1.2 (pág. 5).

a. De los centros vacacionales estudiados en 2004, 72% se clasificaron como centros vacacionales de un día.

b. De todas las visitas a centros vacacionales en 2003, 23% fueron visitas por hombres.

1.8 Consulte la gráfica "Estudiantes de E.U. de acuerdo al grado de estudio".

El informe del último censo sobre escuelas encontró que la población total es de 70 millones de estudiantes (27.8% de la población) que cursan desde maternal hasta la universidad.

a. ¿Cuál es la población?

b. ¿Qué información se obtuvo de cada persona?

c. Con base en la información dada, estime el número de estudiantes que están inscritos en las universidades.

d. Con la información dada, estime el tamaño de la población total de Estados Unidos.

1.9 La International Communications Research (ICR) realizó el Estudio Nacional de Limpieza de Primavera para la Asociación de Jabones y Detergentes (Soap and Detergent Association). La ICR entrevistó a 1000 hombres y mujeres jefes de familia respecto a las actitudes de limpieza que toman en sus casas. El estudio tiene un margen de error de más o menos 5%.

a. ¿Cuál es la población?

b. ¿Cuántas personas fueron interrogadas?

c. ¿Qué información se obtuvo de cada persona?

ESOS LUGARES DIFÍCILES DE LIMPIAR

Datos de Anne R. Carey y Gia Kereselidze, USA TODAY; Fuente: Swiffer

d. Con la información dada, estime el número de adultos entrevistados que piensan que limpiar bajo el sillón es el trabajo de limpieza más difícil.

e. ¿Qué piensa usted que significa el "margen de error de más o menos 5%?"

f. ¿Cómo usaría usted el "margen de error" para calcular el porcentaje de todos los adultos que piensan que las persianas venecianas son los elementos cuya limpieza es más difícil?

1.10 Consulte la gráfica "Distracciones de los automovilistas debido al uso de los teléfonos celulares."

DISTRACCIONES DE LOS AUTOMOVILISTAS DEBIDO AL USO DE LOS TELÉFONOS CELULARES

Datos de Lori Joseph y Sam Ward, © 2001 USA Today

a. ¿Qué grupo de personas fue entrevistado?

b. ¿Cuántas personas fueron entrevistadas?

c. ¿Qué información se obtuvo de cada persona?

d. Explique el significado de "41% comentan que rebasaron inadvertidamente los límites de velocidad".

e. ¿Cuántas personas contestaron "Afirman que rebasaron inadvertidamente los límites de velocidad"?

1.11

¿CÓMO GASTARÁ LA DEVOLUCIÓN DE SUS IMPUESTOS?

Datos de Darryl Haraison y Jerry Mosemak, USA TODAY, Fuente: turbotax.com

a. ¿Cuál es la población?

b. Describa la muestra empleada para elaborar este informe.

c. Identifique las variables empleadas para recolectar esta información.

d. ¿Qué es lo que la mayoría de personas va a hacer con la devolución de sus impuestos? ¿Cómo se muestra esta mayoría en la gráfica?

1.12 Durante un programa de radio que se transmitió el 16 de agosto de 1998, David Essel informó las siguientes estadísticas: (1) el porcentaje de divorcios en Estados Unidos es 55%; y cuando se preguntó a adultos casados si seguirían con su esposa, (2) 75% de las mujeres dijeron que sí y (3) 65% de los hombres dijeron que sí.

a. ¿Cuál es el porcentaje de personas que indican que prefieren "seguir casados"?

b. Parece haber una contradicción en esta información. ¿Cómo es posible que estas tres frases sean correctas? Explique.

1.13 El conocimiento práctico de las estadísticas es muy útil cuando se desean comprender los estadísticos que aparecen en los noticieros informativos. Es frecuente que los medios de información y nuestro gobierno hagan comentarios como "El porcentaje de delincuencia aumentó 50% en esta ciudad."

a. El aumento en el porcentaje de delincuencia de 4% a 6%, ¿representa un aumento de 50%? Explique.

b. ¿Por qué alguien informaría que un aumento de 4% a 6% es "un salto de 50% en la delincuencia"?

1.14 Encuentre un artículo en un periódico reciente que ilustre un tipo de informe del tipo "las manzanas son malas".

1.15 De la población de adultos en Estados Unidos, 36% tiene una alergia. Una muestra de 1200 adultos seleccionados al azar resultó que 33.2% tenían algún tipo de alergia.

a. Describa a la población.

b. ¿Cuál es la muestra?

c. Describa la variable.

d. Identifique el estadístico y dé su valor.

e. Identifique el parámetro y dé su valor.

1.16 En sus propias palabras, explique por qué el parámetro es fijo y la estadística varía.

1.17 El número de la camiseta de un equipo de futbol, ¿es una variable cuantitativa o es categórica? Apoye su respuesta con una explicación detallada.

1.18 a. Mencione dos variables de atributo relacionadas con los clientes de una tienda departamental que se abrió recientemente. Cabe mencionar que el estudio de estas variables debe resultar informativo para la tienda.

b. Mencione dos variables numéricas relacionadas con los clientes de una tienda departamental de apertura reciente. El estudio de estas variables debe ser informativo para la tienda.

1.19 a. Mencione dos variables nominales relacionadas con los clientes de una tienda departamental que se abrió recientemente. Cabe mencionar que el estudio de estas variables debe resultar informativo para la tienda.

b. Mencione dos variables ordinales relacionadas con los clientes de una tienda departamental que se abrió recientemente. Cabe mencionar que el estudio de estas variables debe resultar informativo para la tienda.

1.20 Cierto ejercicio simula **el hecho de tomar una muestra** de tamaño 10 de una población de 100 estudiantes universitarios. Tome una muestra y observe el resultado.

a. Mencione la variable de atributos que está involucrada en este experimento. ¿Es nominal u ordinal?

b. Mencione la variable numérica involucrada en este experimento. ¿Es discreta o continua?

1.21 a. Explique por qué la variable "anotación" para el equipo local en un juego de baloncesto es discreta.

b. Explique por qué la variable "cantidad de minutos para ir al trabajo" es continua.

1.22 En la actualidad se estudia la severidad de los efectos colaterales que experimentan ciertos pacientes cuando reciben tratamiento con un medicamento particular. Esa severidad se mide en la escala: ninguna, benigna, moderada, grave, muy grave.

a. Mencione la variable de interés.

b. Identifique el tipo de variable.

1.23 Al cuerpo docente de la universidad estatal de Boise se le hizo la siguiente pregunta "¿Qué tan satisfecho estuvo usted con el programa de verano 2002?" Sus respuestas fueron clasificadas como "muy satisfecho", "un poco satisfecho", "ni satisfecho ni insatisfecho", "un poco insatisfecho", o "muy insatisfecho".

a. Mencione la variable de interés.

b. Identifique el tipo de variable.

1.24 Se pregunta a varios estudiantes el peso de los libros y demás material que llevan a clases.

a. Identifique la variable de interés.

b. Identifique el tipo de variable.

c. Elabore una lista de algunos valores que podrían presentarse en una muestra.

1.25 Un fabricante de medicamentos está interesado en la proporción de personas que tienen hipertensión (presión sanguínea elevada) considerando que esta condición puede controlarse con un medicamento nuevo que la compañía ha perfeccionado. Se efectúa un estudio que abarca 5000 personas que padecen hipertensión, y se encuentra que 80% de ellas pueden controlar su hipertensión con el medicamento. Suponiendo que las 5000 personas sean representativas del grupo que tiene hipertensión, conteste las siguientes preguntas:

a. ¿Cuál es la población?

b. ¿Cuál es la muestra?

c. Identifique el parámetro de interés.

d. Identifique el estadístico y dé su valor.

e. ¿Conocemos el valor del parámetro?

1.26 La dirección desea calcular el costo de los libros de texto para los estudiantes de cierta universidad. Sea x la variable del costo total de todos los libros de texto comprados por un estudiante este semestre. El plan es identificar al azar a 100 estudiantes y obtener los costos totales de sus libros de texto. El costo promedio para los 100 estudiantes se empleará para calcular el costo promedio para todos los estudiantes.

a. Describa el parámetro que la dirección desea calcular.

b. Describa la población.

c. Describa la variable involucrada.

d. Describa la muestra.

e. Describa el estadístico y la forma en que usted usaría los 100 datos recolectados para calcular el estadístico.

1.27 Un técnico de control de calidad selecciona ciertas piezas ensambladas de una línea de montaje y registra la siguiente información sobre cada pieza:

A: defectuosa o no defectuosa

B: el número de identificación del trabajador que ensambló la pieza

C: el peso de la pieza

a. ¿Cuál es la población?

b. La población, ¿es finita o infinita?

c. ¿Cuál es la muestra?

d. Clasifique las tres variables como datos de atributo o numéricos.

1.28 Seleccione 10 estudiantes que estén inscritos en su ciclo escolar y recolecte datos para las tres variables siguientes:

X: número de cursos en los que está inscrito

Y: costo total de libros de texto y el material para los cursos

Z: método de pago para cubrir el importe de los libros de texto y el material de los cursos

a. ¿Cuál es la población?

b. La población, ¿es finita o infinita?

c. ¿Cuál es la muestra?

d. Clasifique las tres variables como nominal, ordinal, discreta o continua.

1.29 La empresa Aventis Pharmaceuticas realizó un estudio para medir los efectos colaterales de Allegra, un medicamento que se utiliza en el tratamiento de alergias estacionales. A una muestra de 679 personas que sufren de alergias en Estados Unidos se les administraron 60 mg del medicamento dos veces al día. Los pacientes informaron si durante ese periodo experimentaron alivio de sus alergias, así como cualesquier otro efecto colateral adverso (infección viral, náusea, somnolencia, etcétera).

Fuente: Good Housekeeping, febrero de 2005, p. 120.

a. ¿Cuál es la población bajo estudio?

b. ¿Cuál es la muestra?

c. ¿Cuáles son las características de interés acerca de cada uno de los elementos de la población?

d. Los datos recolectados, ¿son cualitativos o cuantitativos?

1.30 Simule mediante un ejercicio el hecho de tomar una muestra de tamaño 10 de una población de 100 estudiantes universitarios. Tome una muestra de tamaño 10.

a. ¿Cuál es la población?

b. La población, ¿es finita o infinita?

c. Mencione dos parámetros y dé sus valores.

POBLACIÓN: Media = 66.9 pulgadas
Porcentaje de mujeres = 64.0%

MUESTRA: Media = 67.3 pulgadas
Porcentaje de mujeres = 60.0%

d. ¿Cuál es la muestra?

e. Mencione los dos estadísticos correspondientes e indique sus valores.

f. Tome otra muestra de tamaño 10. De los elementos precedentes, ¿cuáles permanecen fijos y cuáles cambiaron?

1.31 Identifique las siguientes expresiones como ejemplos de (1) variables de atributo (cualitativas) o (2) variables numéricas (cuantitativas):

a. Resistencia a la ruptura de un tipo de cuerda determinado

b. El color del pelo de los niños que dan una audición para la revista musical *Annie*

c. El número de señalamientos de "alto" en poblaciones de menos de 500 habitantes

d. Si una válvula de agua es defectuosa o no lo es

e. El número de reactivos contestados correctamente en un examen estandarizado

f. El tiempo necesario para contestar una llamada telefónica en cierta oficina de bienes raíces

1.32 Identifique las siguientes expresiones como ejemplos de variables (1) nominales, (2) ordinales, (3) discretas, o (4) continuas:

a. Una encuesta de electores registrados donde se preguntaba a qué candidato daban su apoyo

b. El tiempo necesario para que sane una herida cuando se aplica un nuevo medicamento

c. El número de receptores de televisión en un hogar

d. La distancia que viaja un balón de futbol cuando es pateado por las alumnas universitarias de primer año

e. El número de páginas impresas que se procesan en la impresora de una computadora

f. El tipo de árbol que se utiliza como árbol de Navidad

1.33 Suponga que un niño de 12 años le pide que le explique la diferencia entre una muestra y una población.

a. ¿Qué información debe incluir en su respuesta?

b. ¿Qué razones proporcionaría al niño sobre por qué debe tomarse una muestra en lugar de encuestar a todos los elementos de la población?

1.34 Suponga que un niño de 12 años le pide que le explique la diferencia entre una estadística y un parámetro.

a. ¿Qué información debe incluir en su respuesta?

b. ¿Qué razones proporcionaría al niño sobre por qué debe reportarse el valor de una estadística en vez de un parámetro?

1.3 Medibilidad y variabilidad

En un conjunto de datos siempre se espera variación. Si se encuentra poca variación o no se encuentra variación, podría suponerse que el instrumento de medición no está calibrado con una unidad que sea lo suficientemente pequeña. Por ejemplo, se toma un paquete de 24 barras de algún dulce favorito y cada barra se pesa en forma individual. Se observa que cada una de las 24 barras pesa 24.8 gramos, aproximado a la décima de gramo más próxima. ¿Significa esto que todas las barras pesan exactamente lo mismo? No necesariamente. Suponga que las barras se pesan en una balanza analítica que registra la centésima de gramo más próxima. En este caso los pesos de las 24 barras mostrarían **variabilidad.**

No importa cuál sea la variable de respuesta: si la herramienta de medición es suficientemente exacta, habrá variabilidad en los datos. Uno de los objetivos primordiales del análisis estadístico es la medición de la variabilidad. Por ejemplo, en el estudio del control de calidad, la medición de la variabilidad es absolutamente indispensable. Controlar (o reducir) la variabilidad en un proceso de manufactura es todo un campo por sí mismo: el control estadístico de los procesos.

SECCIÓN 1.3 EJERCICIOS

1.35 Suponga que se miden los pesos (en kilogramos) de las personas que pertenecen a cada uno de los siguientes grupos:

Grupo 1: porristas (de los equipos de la Liga Nacional de Futbol (NFL)

Grupo 2: jugadores de los equipos de la NFL

¿Para qué grupo esperaría que los datos tengan más variabilidad? Explique su respuesta

1.36 Suponga que trata de decidir la compra de una máquina entre dos opciones. Además, suponga que es importante la longitud a la que las máquinas cortan una pieza de un producto particular. Si ambas máquinas producen piezas de la misma longitud en promedio, ¿qué otra consideración sobre las longitudes sería importante?, ¿Por qué?

1.37 Durante años, grupos activistas de consumidores han pugnado por que los comerciantes al menudeo utilicen precios unitarios en sus productos. Argumentan que los precios de los alimentos, por ejemplo, siempre deben etiquetarse como $/onza, $/libra, $/gramo, $/litro, etc., además de estarlo como $/paquete, $/lata, $/caja, $/botella, etcétera. Explique por qué.

1.38 Una máquina vendedora de café proporciona, en promedio, 170 mililitros de café por taza. Esta afirmación, ¿puede ser verdadera para una máquina vendedora que a veces proporciona sólo el café suficiente para que la taza quede llena apenas a la mitad (por ejemplo, 114 mililitros)? Explique su respuesta.

1.39 Los profesores aplican exámenes para medir el grado de conocimiento de sus estudiantes acerca de su materia. Explique cómo es que "una falta de variabilidad en las calificaciones de estudiantes podría indicar que el examen no fue una herramienta de medición muy eficaz". Ideas a considerar: ¿Qué significaría si todos los estudiantes obtienen una calificación de 100% en un examen? ¿Qué significaría si todos los estudiantes alcanzaran un 0%? ¿Qué significaría si las calificaciones varían de 40% a 95%?

1.40 Realice una simulación donde se pretende tomar la muestra de una población de estudiantes universitarios.

a. Tome 10 muestras de tamaño 4 y realice un seguimiento de los promedios de las muestras de las horas por semana que estudian los alumnos. Encuentre el intervalo de estos promedios al restar el promedio más bajo del promedio más alto.

b. Tome 10 muestras de tamaño 10 y efectúe un seguimiento de los promedios de las muestras de las horas por semana que estudian los alumnos. Encuentre el intervalo (o rango) de estos promedios al restar el promedio más bajo del promedio más alto.

c. ¿Cuál tamaño de muestra demostró más variabilidad?

d. Si el promedio de población es de alrededor de 15 horas por semana, ¿qué tamaño de muestra demostró esto con más precisión? ¿Por qué?

1.4 Recolección (obtención) de datos

Por lo general es imposible estudiar toda una población (cada uno de los individuos de un país, todos los estudiantes universitarios, cada paciente de un médico, etc.), los investigadores suelen apoyarse en *muestras* para adquirir la información, o los *datos*, necesarios. Es importante obtener "buenos datos" porque en última instancia las inferencias efectuadas se basan en los estadísticos obtenidos a partir de los datos. Estas inferencias sólo pueden ser tan buenas como lo sean los datos.

Aunque es relativamente fácil definir "buenos datos" como datos que representen con exactitud a la población de la cual se tomaron, no es fácil garantizar que un método particular de muestreo produzca "buenos datos". Se desea usar métodos de muestreo (recolección de datos) que producirán datos que sean representativos de la población y que no sean *sesgados.*

Método de muestreo sesgado: produce valores que difieren sistemáticamente de la población que está siendo muestreada. Un *método de muestreo no-sesgado* es aquel que no está sesgado.

Dos métodos de muestreo de uso común que a menudo producen muestras sesgadas son las *muestras por conveniencia* y *de voluntarios.*

Una muestra de conveniencia, a veces conocida también como *al azar*, ocurre cuando los elementos son seleccionados arbitrariamente y en forma no estructurada de una población, mientras que una muestra de voluntarios está formada por resultados recolectados a partir de los elementos de la población que por su propia iniciativa eligen contribuir con la información necesaria.

¿Alguna vez ha comprado una cesta de frutas en el mercado, basando la decisión de compra en la apariencia apetitosa de la fruta de arriba, para descubrir después que la fruta de abajo no era tan fresca? No consideró conveniente inspeccionar la fruta de abajo, de modo que usted confió en una muestra por conveniencia. ¿Alguna vez el maestro ha utilizado a los alumnos de un salón como muestra de dónde recabar datos? Los alumnos forman un grupo conveniente, pero, ¿el grupo es realmente representativo de la población de la escuela? (Considere las diferencias entre estudiantes del turno matutino, nocturno, y/o de fin de semana; tipo de curso; etcétera).

¿Alguna vez ha enviado el cuestionario de la encuesta de una revista? ¿Bajo qué condiciones se tomó el tiempo para llenar un cuestionario? La actitud inmediata de casi todas las personas es no hacer caso de la encuesta. Quienes tengan ideas firmes harán el esfuerzo por responder el cuestionario; por tanto, no deben esperarse muestras representativas cuando se recolectan muestras de voluntarios.

Proceso de recolección de datos

La obtención de datos para el análisis estadístico es un proceso complicado que incluye los siguientes pasos:

1. Definir los objetivos de la investigación o del experimento.

   Ejemplos: comparar la eficacia de un nuevo medicamento con la eficacia del medicamento estándar; estimar el ingreso familiar promedio en Estados Unidos.

2. Definir la variable y la población de interés.

   Ejemplos: duración del tiempo de recuperación de los pacientes que sufren alguna enfermedad particular; estimar el ingreso total en las familias de Norteamérica.

3. Definir los esquemas para recolectar y medir los datos.

   Esto incluye el marco muestral, el procedimiento de muestreo, el tamaño de la muestra y el instrumento de medición (cuestionario, por teléfono, etc.) de los datos.

4. Recolectar su muestra. Seleccionar los sujetos a ser muestreados y recolectar los datos.

5. Revisar el proceso de muestreo al terminar la recolección

   A menudo ocurre que un analista se encuentra con datos ya recolectados, tal vez, incluso, recolectados para otros fines, lo cual imposibilita determinar si los datos son "buenos" o no. Es mucho mejor que usted recolecte los datos utilizando técnicas aprobadas. Aun cuando el interés primordial de este libro está relacionado con diversas técnicas de análisis de datos, usted debe estar al tanto de los problemas de la recolección de datos.

El siguiente ejemplo describe la población y la variable de interés para la investigación específica:

Los dos métodos que por lo general se utilizan para recolectar datos son *experimentos* y *estudios observacionales.* En un experimento, el investigador controla o

modifica el entorno y observa el efecto sobre la variable bajo estudio. A menudo leemos sobre resultados de laboratorio obtenidos usando ratas blancas para probar dosis diferentes de un nuevo medicamento y analizar su efecto en la presión arterial. Los tratamientos experimentales fueron diseñados específicamente para obtener los datos necesarios para estudiar el efecto sobre la variable. En un estudio observacional, el investigador no modifica el entorno y no controla el proceso en observación. Los datos se obtienen al muestrear a alguien de la población de interés. Las encuestas son estudios observacionales de personas.

CASO PRÁCTICO 1.7

¿Experimento o estudio observacional?

UNA INFECCIÓN QUIRÚRGICA ES CUESTIÓN DE TIEMPO

En Archivos de Cirugía, unos investigadores escriben acerca de numerosos pacientes de quirófanos que no obtienen dosis oportunas de los medicamentos apropiados, lo cual eleva el riesgo de infecciones. Dicho informe dice que de 30 millones de operaciones que se realizan cada año en Estados Unidos, alrededor del 2% se complican por una infección en el sitio. El estudio de 34 000 pacientes de cirugías en casi 3 000 hospitales en 2001 encontraron que sólo 56% obtuvo medicamentos profilácticos hasta una hora antes de la cirugía, cuando pueden ser eficaces.

Fuente: USA Today, 22 de febrero, 2005

Este estudio es un ejemplo de un estudio observacional. Los investigadores no modificaron ni trataron de controlar el entorno. Observaron lo que estaba ocurriendo y escribieron lo que vieron.

Si todos los elementos de la población se pueden poner en lista, o enumerarse, y observarse, entonces se compila un censo. No obstante, los censos raras veces se utilizan porque con frecuencia son difíciles de compilar y consumen mucho tiempo, razón por la que son muy costosos. Imagine el trabajo de compilar un censo de cada persona que sea un cliente potencial de una empresa de corretaje. En situaciones semejantes a ésta, suele efectuarse una *encuesta muestral.*

Cuando se seleccione una muestra para una encuesta, es necesario construir un *marco muestral.*

Marco muestral: es una lista, o conjunto, de los elementos que pertenecen a la población de la cual se toma la muestra.

En el ideal, el marco muestral debe ser idéntico a la población con cada uno de los elementos de la población incluido una vez y sólo una. En este caso, un censo sería el marco muestral. En otras situaciones, un censo podría no ser tan fácil de obtener, porque no se dispone de una lista completa. Las listas de electores registrados o del directorio telefónico se utilizan a veces como marcos de muestreo del público en general. Dependiendo de la naturaleza de la información que se busque, la lista de electores registrados o el directorio telefónico pueden o no pueden servir como marco no-sesgado de muestreo. Debido a que sólo los elementos del marco tienen oportunidad de ser seleccionados como parte de la muestra, es importante que el marco muestral sea representativo de la población.

¿SABÍA USTED...?

Una parte es mejor que todo

En la década de 1930, Prasanta Chandra Mahalanobis dio alta prioridad a producir una muestra representativa apropiada. Él deseaba determinar las características de grandes poblaciones cuando fuera casi imposible obtener todas las mediciones de una población estadística. Las muestras de juicio parecían ser una buena opción, pero tienen fallas importantes: si se sabe lo suficiente acerca de la población para recolectar una buena muestra de juicio, es probable que no haya necesidad de una muestra; si la muestra es errónea, no hay forma de saber qué tan mala es. La respuesta a esta pregunta fue una muestra aleatoria.

Una vez establecido el marco muestral representativo, a partir de éste se procede a la selección de los elementos de muestra. Este proceso de selección se denomina diseño muestral. Hay numerosos tipos diferentes de diseños muestrales; no obstante, todos pueden clasificarse en dos categorías: *muestras de juicio* y *muestras probabilísticas.*

Muestras de juicio: se refiere a las muestras que son seleccionadas con base en el hecho de ser "típicas".

Cuando se obtiene una muestra de juicio, la persona que elabora la muestra elige unidades que considera representativas de la población. La validez de los resultados de una muestra de juicio refleja la solidez del juicio del recolector de datos. Éste no es un procedimiento estadístico aceptable.

Muestras probabilísticas: son aquellas en las que los elementos a seleccionar se obtienen con base en la probabilidad. Cada elemento de una población tiene cierta probabilidad de ser elegido como parte de la muestra.

Las inferencias que estudiaremos más adelante en este libro están basadas en la suposición de que nuestros datos muestrales se obtienen con el uso de una muestra probabilística. Hay muchas formas de diseñar muestras probabilísticas. Veremos dos de ellas, métodos de una etapa y métodos de etapas múltiples, y estudiaremos unos pocos de los muchos diseños específicos que son posibles.

Métodos de una etapa

Muestreo de una etapa: es un diseño muestral en el que los elementos del marco muestral se tratan en la misma forma y no hay subdivisión o partición del marco.

Uno de los métodos más comúnmente usados para recolectar datos es el *muestreo aleatorio simple.*

Muestreo aleatorio simple: surge cuando una muestra se selecciona de modo que todos los elementos de la población o marco muestral tengan la misma probabilidad de ser elegidos. Del mismo modo, todas las muestras de tamaño *n* tienen igual probabilidad de ser elegidas.

Nota: Las muestras aleatorias se obtienen por muestreo con reemplazo de una población finita o por muestreo sin reemplazo de una población infinita.

El concepto de aleatoriedad conlleva implícita la idea de que el siguiente resultado (u ocurrencia) no es predecible. Cuando se extrae una muestra aleatoria simple, es necesario realizar todos los esfuerzos para asegurar que cada elemento tenga la misma probabilidad de ser elegido y que el siguiente resultado no se vuelva predecible. El procedimiento idóneo para obtener una muestra aleatoria simple es utilizar números aleatorios. A menudo se cometen errores porque el término *aleatorio* (mismas probabilidades) se confunde con el *fortuito* (sin patrón).

Para elegir una muestra aleatoria simple, primero se asigna un número de identificación a cada elemento del marco muestral. Esto suele hacerse de manera secuencial usando el mismo número de dígitos para cada elemento. A continuación, usando números aleatorios con el mismo número de dígitos, se selecciona tantos números con esa cantidad de dígitos como sea necesario para obtener el tamaño de muestra deseado. Cada elemento numerado del marco muestral que corresponda a un número aleatorio seleccionado se elige para la muestra.

EJEMPLO 1.8 Uso de números aleatorios

La dirección de nuestra universidad desea calcular el costo "promedio" actual de libros de texto por semestre, por estudiante. La población de interés es el "cuerpo de estudiantes actualmente inscritos", y la variable es la "cantidad total gastada para libros de texto" por cada estudiante este semestre. Como se desea una muestra aleatoria, el Sr. Clark, que trabaja en la dirección, ha obtenido una lista por computadora de la matrícula de tiempo completo para este semestre. Hubo 4265 nombres de estudiantes en la lista. El Sr. Clark numeró los estudiantes 0001, 0002, 0003, y así sucesivamente, hasta 4265; a continuación, usando números aleatorios de cuatro dígitos, identificó una muestra: 1288, 2188, 1952, 2463, 1644, 1004, etcétera. (Vea en el *Manual de Soluciones para el Estudiante* un análisis sobre el empleo de números aleatorios.)

Una muestra aleatoria simple es nuestro primer paso hacia una muestra no-sesgada. Las muestras aleatorias se requieren para la mayor parte de los procedimientos estadísticos presentados en este libro. Sin un diseño aleatorio, las conclusiones que obtengamos de los procedimientos estadísticos pueden no ser confiables.

CASO PRÁCTICO 1.9 Proceso para recolección de datos

Considere la gráfica "Las relaciones fallan hasta en el jardín de niños" de la página 4 y los cinco pasos del proceso de recolección de datos.

1. *Definir los objetivos del estudio o experimento.* Determine la opinión de maestros de jardines de niños en Estados Unidos respecto a lo que consideran habilidades importantes para un niño.
2. *Definir la variable y la población de interés.* La variable es la opinión o respuesta a una pregunta respecto a las habilidades en un jardín de niños. La población de interés es todos los maestros de jardín de niños de Estados Unidos.
3. *Definir la recolección de datos y los esquemas de medición de datos.* Con base en la gráfica misma, se puede ver que la fuente para los porcentajes presentados fue Mason-Dixon Polling. Después de una investigación a conciencia, Mason-Dixon Polling & Research, Inc., de Washington, D.C., realizó el estudio titulado "El Delito de Pelear: Invierta en el Kids National Kindergarten Teacher Survey." Fue una encuesta por teléfono de 800 maestros de jardines de niños de Estados Unidos, escogidos al azar, realizado del 9 al 14 de julio de 2004. El marco muestral se compiló de una lista de registros estatales de licencias de manejo.
4. *Recolectar la muestra.* La información recolectada de cada maestro encuestado fueron las opiniones de los maestros en varias habilidades que consideran esenciales o muy importantes para sus estudiantes.
5. *Revisar el proceso de muestreo al completar la recolección.* Como el proceso de muestreo fue una encuesta por teléfono, ¿cuál fue la proporción de quienes no respondieron? Si la proporción fue alta, la validez de los porcentajes recibidos sería cuestionable. ¿Se disponía de registros de todos los estados, de modo que cada maestro tuviera la misma probabilidad de ser elegido?

En concepto, la muestra aleatoria simple es la más sencilla de las técnicas de muestreo de probabilidades, pero raras veces se emplea en la práctica porque con frecuencia es una técnica que resulta ser poco eficiente. Uno de los métodos más fáciles de usar para aproximar una muestra aleatoria simple es el *método de muestreo sistemático.*

Muestra sistemática: es aquella en la que se selecciona todo *k*-ésimo elemento del marco muestral, empezando con un primer elemento que se elige de manera aleatoria de los primeros *k* elementos.

Para seleccionar una muestra sistemática al x%, primero se selecciona aleatoriamente un elemento de los primeros $\frac{100}{x}$ elementos. Después de que aleatoriamente se selecciona el primer elemento dentro de los primeros $\frac{100}{x}$ elementos, de ahí se continúa para seleccionar todo $\frac{100}{x}$avo elemento hasta que se cuenta con el número de datos que se desea tener para nuestra muestra.

Por ejemplo, si se desea tener una muestra sistemática al 3%, el primer elemento se encontraría seleccionando un entero entre 1 y33 ($\frac{100}{x} = \frac{100}{3} = 33.33$, que al redondear se convierte en 33). Supóngase que se eligió arbitrariamente 23, esto significa que el primer dato se obtiene a partir del sujeto ubicado en la posición 23 del marco muestral. El segundo dato proviene del sujeto ubicado en la posición 56 (23 + 33 = 56); el tercero proviene del sujeto ubicado en la posición 89 (56 + 33); y así sucesivamente hasta que la muestra esté completa.

La técnica sistemática es fácil de describir y ejecutar; no obstante, conlleva algunos riesgos cuando el marco muestral es repetitivo o de naturaleza cíclica.

Por ejemplo, una muestra sistemática de toda k-ésima casa a lo largo de una calle prolongada podría resultar en una muestra desproporcionada con respecto a casas de lotes situados en esquina. Es probable que la información resultante sea sesgada si la finalidad del muestreo sea la de aprender acerca del apoyo para un impuesto de una banqueta propuesta. En estas situaciones los resultados pueden no aproximar una muestra aleatoria simple.

Métodos de etapas múltiples

Cuando se muestrean poblaciones muy grandes, a veces es necesario usar un diseño de *muestreo de etapas múltiples* para aproximar el muestreo aleatorio.

Muestreo aleatorio de etapas múltiples: es un diseño de muestra en el que los elementos del marco muestral se subdividen y la muestra se elige en más de una etapa.

A veces los diseños de muestreo de etapas múltiples empiezan por dividir una población muy grande en subpoblaciones con base en alguna característica. Estas subpoblaciones se denominan *estratos.* Estos estratos más pequeños y fáciles de trabajar se muestren por separado. Uno de los diseños muestrales es el *método de muestreo aleatorio estratificado.*

Muestra aleatoria estratificada: se obtiene al estratificar la población, o marco muestral, y luego seleccionar un número de elementos para cada uno de los estratos por medio de una técnica sencilla de muestreo aleatorio.

Una muestra aleatoria estratificada resulta cuando la población, o marco muestral, se subdivide en varios estratos que en general es alguna subdivisión que ya ocurrió de alguna manera natural, y luego de cada uno de estos estratos se extrae una submuestra. Estas submuestras pueden extraerse de los diversos estratos usando métodos aleatorios o sistemáticos. Las submuestras se resumen primero por separado y luego se combinan para obtener conclusiones acerca de toda la población.

Cuando se muestrea una población con varios estratos, con frecuencia se requiere que el número de elementos recolectados de cada estrato sea proporcional al tamaño de los estratos; este método recibe el nombre de *muestreo estratificado proporcional.*

Muestra estratificada proporcional: se obtiene al estratificar la población, o marco muestral, y luego seleccionar un número de elementos en proporción al tamaño de los estratos de cada uno de los estratos por medio de una técnica de muestreo aleatorio sencilla.

Una forma cómoda de expresar la idea de muestreo proporcional es establecer una proporción. Por ejemplo, la proporción, "1 por cada 150" nos induce a seleccionar 1 dato por cada 150 elementos de cada estrato. De esa forma, el tamaño de los estratos determina el tamaño de la submuestra de ese estrato. Las submuestras se resumen por separado y luego se combinan para extraer conclusiones acerca de toda la población.

Otro método de muestreo que empieza por estratificar la población, o marco muestral, es una *muestra por conglomerados.*

Muestra por conglomerados: se obtiene al estratificar la población, o marco muestral, y luego seleccionar algunos o todos los elementos de alguno, pero no de todos, los estratos.

La muestra por conglomerados es un diseño de etapas múltiples. Utiliza métodos aleatorios o sistemáticos para elegir los estratos (conglomerados) que serán muestreados (primera etapa) y luego emplea métodos aleatorios o sistemáticos para seleccionar elementos de cada uno de los conglomerados identificados (segunda etapa). El método de muestreo por conglomerados también permite la posibilidad de seleccionar todos los elementos de cada conglomerado identificado. En cualquier caso, las submuestras se resumen por separado y luego se combina la información.

Para ilustrar un proceso de muestreo aleatorio de posibles etapas múltiples, considere que es necesaria una muestra de un país grande. En la primera etapa, el país se divide en regiones más pequeñas, por ejemplo estados, y se elige una muestra aleatoria de estos estados. En la segunda etapa, se selecciona una muestra aleatoria de regiones más pequeñas dentro de los estados seleccionados (condados). En la tercera etapa, se toma una muestra aleatoria de zonas incluso más pequeñas (poblados) dentro de cada condado. Por último, en la cuarta etapa, si los poblados son suficientemente pequeños para los fines del estudio, el investigador podría continuar seleccionando muestras aleatorias simples de cada uno de los poblados identificados. Esto significaría que toda la muestra estuvo formada por varias submuestras "locales" identificadas como resultado de las diversas etapas.

El diseño muestral no es una cuestión fácil; muchas universidades e instituciones de educación superior ofrecen cursos sobre encuestas muestrales y diseño experimental. El tema de encuestas muestrales es todo un libro por sí mismo. Se pretende que la información que acaba de proporcionarse constituya una revisión sobre lo que es el muestreo y ponga el papel de éste en perspectiva.

SECCIÓN 1.4 EJERCICIOS

1.41 *USA Today* hace regularmente a sus lectores la siguiente pregunta: "Tiene alguna queja acerca del equipaje, devoluciones, publicidad o servicio a clientes de una línea aérea? Escriba: . . ." ¿Qué clase de método de muestreo es éste? ¿Es probable que los resultados sean sesgados? Explique.

1.42 *USA Today* realizó una encuesta preguntando a lectores "¿Qué es lo más divertido que le ha ocurrido a usted en camino o durante un viaje de trabajo?"

a. ¿Qué clase de método de muestreo es éste?

b. ¿Es probable que los resultados sean sesgados? Explique.

1.43 En una encuesta acerca de familias, Ann Landers preguntó a padres de familia si tendrían más hijos: 70% respondió "No". Una encuesta aleatoria independiente que hizo la misma pregunta encontró 90% de respuestas "Sí". Dé al menos una explicación de por qué el porcentaje resultante de la encuesta de Landers es tan diferente del porcentaje de la muestra aleatoria.

1.44 Considere la pregunta siguiente, que fue tomada de Quick Vote de CNN en Internet el 16 de febrero de 2005: "¿Qué debe hacerse con el programa *Star Trek: Enterprise*?" La respuesta fue como sigue: 45%, déjenlo seguir; 55%, por su bien, sáquenlo.

a. ¿Qué clase de encuesta se utilizó?

b. ¿Piensa usted que estos resultados podrían estar sesgados? ¿Por qué?

1.45 Todos sabemos que el ejercicio es bueno para nosotros. Pero, ¿puede el ejercicio evitar o retardar los síntomas de la enfermedad de Parkinson? Un estudio reciente efectuado por la Harvard School of Public Health examinó a 48 000 hombres y 77 000 mujeres que estaban relativamente sanos y eran de mediana edad o mayores. Durante el curso del estudio, la enfermedad se desarrolló en 387 personas. El estudio encontró que los hombres que habían practicado alguna actividad vigorosa al menos dos veces por semana en preparatoria, universidad y hasta una edad de 40 años, tenían 60% menos riesgo de que se les formara la enfermedad de Parkinson. El estudio no encontró esa reducción para mujeres. ¿Qué tipo de muestreo representa esto?

Fuente: El ejercicio puede evitar el mal de Parkinson, *USA Today,* 22 de febrero, 2005. p. 7D.

1.46 Un distribuidor de alimentos al mayoreo en una gran zona metropolitana quisiera probar la demanda de un nuevo producto alimenticio. Él distribuye alimentos a través de cinco grandes cadenas de supermercados. El distribuidor de alimentos selecciona una muestra de tiendas ubicadas en zonas donde considera que los compradores están dispuestos a probar nuevos productos. ¿Qué tipo de muestreo representa lo anterior?

1.47 Considere una población simple que sólo consta de los números 1, 2 y 3 (un número ilimitado de cada uno). De esta población podrían extraerse nueve muestras diferentes de tamaño 2: (1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3).

a. Si la población consta de los números 1, 2, 3 y 4, haga una lista de todas las muestras de tamaño 2 que puedan elegirse.

b. Si la población consta de los números 1, 2 y 3, haga una lista de todas las muestras de tamaño 3 que puedan elegirse.

1.48 a. ¿Qué es un marco muestral?

b. ¿Qué empleó el Sr. Clark para un marco muestral en el ejemplo 1.8 (p. 22)?

c. ¿De dónde proviene el número 1 288, y cómo se utilizó?

1.49 Un artículo titulado "Surface Sampling in Gravel Streams" (*Journal of Hydraulic Engineering,* abril de 1993) analiza el muestreo de criba y muestreo de zonas. El muestreo de criba comprende la remoción a mano de piedras que se encuentren en puntos específicos. Estos puntos se establecen en la superficie de grava usando ya sea malla de alambre o distancias predeterminadas en una cinta de estudio. En general, el material recolectado por muestreo de criba se analiza como distribución de frecuencia. Una muestra de zona se recolecta removiendo todas las partículas halladas en una zona predeterminada del lecho de un canal. Es frecuente que el material recuperado sea analizado como distribución de frecuencia por peso. ¿Clasificaría usted estos diseños muestrales como muestras de juicio o muestras de probabilidad?

1.50 Una muestra aleatoria puede ser difícil de obtener. ¿Por qué?

1.51 ¿Por qué la muestra aleatoria es tan importante en estadística?

1.52 Sheila Jones trabaja para una reconocida empresa de investigación de mercados ubicada en Cincinnati, Ohio. Su supervisor le proporcionó una lista de 500 números aleatorios de 4 dígitos cada uno, extraídos de una tabla estadística de números aleatorios. El supervisor solicitó a Sheila que hiciese una encuesta telefónica a 500 residentes de Cincinnati, en el supuesto de que los 4 últimos dígitos del número telefónico de éstos coinciden con uno de los números de la lista. Si Sheila sigue las instrucciones de su supervisor, ¿tiene él la certeza de obtener una muestra aleatoria de los encuestados? Explique su respuesta.

1.53 Describa en detalle cómo se seleccionaría una muestra sistemática al 4% de los adultos en una gran ciudad vecina, a fin de determinar una encuesta sobre un asunto político.

1.54 a. ¿Qué organismo del gobierno federal ilustra un muestreo estratificado de la población? (No se emplea un proceso aleatorio de selección.)

b. ¿Qué organismo del gobierno federal ilustra un muestreo proporcional de la población? (No se emplea un proceso aleatorio de selección.)

1.55 Suponga que usted ha sido contratado por un grupo de estaciones radiofónicas deportivas para determinar la distribución de edad de su auditorio. Describa en detalle cómo seleccionaría una muestra

aleatoria de 2500 de las 35 zonas de radioescuchas involucradas.

1.56 Explique por qué las encuestas que se citan con tanta frecuencia durante las primeras devoluciones, en reportajes por TV el día de elecciones, son un ejemplo de muestreo por conglomerados.

1.57 La guía telefónica podría no ser un marco muestral representativo. Explique por qué.

1.58 El padrón de electores del consejo electoral no es un censo de la población de adultos. Explique por qué.

1.5 Comparación entre probabilidad y estadística

La **probabilidad** y la **estadística** son dos campos ajenos pero relacionados de las matemáticas. Se ha dicho que "la probabilidad es el vehículo de la estadística". Es decir, que si no fuera por las leyes de la probabilidad, la teoría de la estadística no sería posible.

A continuación se ilustrará la relación y la diferencia entre estas dos ramas de las matemáticas mediante la observación de dos cajas. Se sabe que la caja de probabilidad contiene fichas de póquer: cinco azules, cinco rojas y cinco blancas. La probabilidad intenta responder preguntas como "si se extrae una ficha de la caja, ¿cuál es la posibilidad de que sea azul?" Por otra parte, en la caja de estadística se ignora cuál es la combinación de fichas. Se extrae una muestra y, con base en los resultados obtenidos en ésta, se hacen conjeturas sobre lo que se cree que hay en la caja. Observe la diferencia: la probabilidad pregunta sobre la posibilidad de que ocurra algo específico, por ejemplo sacar una ficha azul, cuando se conocen las posibilidades (es decir, que se conoce la población). Por otra parte, la estadística pide extraer una muestra, describirla (estadística descriptiva) y luego hacer inferencias sobre la población con base en la información que se obtuvo en la muestra (estadística inferencial).

SECCIÓN 1.5 EJERCICIOS

1.59 De lo siguiente, ¿cuál ilustra la probabilidad? ¿y cuál representa a la estadística?

a. Determinar cuál es la probabilidad de que se obtenga un "6" cuando se lanza un dado.

b. Estudiar los pesos de 35 bebés para calcular el aumento de peso en el primer mes después de su nacimiento.

1.60 De lo siguiente, ¿cuál ilustra la probabilidad? ¿y la estadística?

a. Recolectar el número de horas de crédito de 100 estudiantes, para calcular el número promedio de horas de crédito por estudiante en una universidad particular de la comunidad.

b. Determinar cuál es la probabilidad de ganar en la Lotería de Nueva York.

1.61 Clasifique cada una de las siguientes afirmaciones como problema de probabilidad o de estadística.

a. Determinar si un nuevo medicamento reduce el tiempo de recuperación de cierta enfermedad.

b. Determinar la posibilidad de obtener "cara" cuando se lanza una moneda.

c. Determinar el tiempo de espera necesario para pagar y salir de una tienda.

d. Determinar la posibilidad de obtener una mano de "veintiuno".

1.62 Clasifique cada una de las siguientes afirmaciones como problema de probabilidad o de estadística.

a. Determinar cuánto tiempo es necesario para manejar un sondeo telefónico típico en una oficina de bienes raíces.

b. Determinar la duración de la vida de los focos de 100 w producidos por una empresa.

c. Determinar la posibilidad de extraer una bola azul de un tazón que contiene 15 bolas, de las cuales 5 son azules.

d. Determinar la resistencia al corte de los remaches recientemente adquiridos por una compañía constructora de aviones.

e. Determinar la posibilidad de obtener "dobles" al lanzar un par de dados.

1.6 Estadística y la tecnología

En años recientes, la tecnología electrónica ha afectado enormemente casi todos los aspectos de la vida diaria. El campo de la estadística no es la excepción. Como se verá mas adelante, en este campo se utilizan muchas técnicas de naturaleza repetitiva: cálculos de estadísticas numéricas, procedimientos para elaborar representaciones gráficas de datos y de procedimientos previos a la formulación de inferencias estadísticas. Las computadoras y calculadoras son muy buenas para ejecutar estas operaciones, que a veces son largas y tediosas. Si su computadora cuenta con uno de los paquetes estadísticos estándares en línea o usted tiene una calculadora estadística, será fácil llevar a cabo el análisis.

En todo este libro, a medida que se estudian procedimientos estadísticos, encontrará la información necesaria para que una computadora ejecute los mismos procedimientos utilizando el software de MINITAB (versión 14) y Excel. Los procedimientos con calculadora también se ilustrarán usando la calculadora TI-83/84 Plus.

A continuación se proporciona una explicación de los convencionalismos tipográficos más comunes que se utilizarán en este texto. En caso de que se requieran explicaciones o selecciones adicionales, éstas se proporcionarán según sea necesario.

INSTRUCCIONES DE TECNOLOGÍA: CONVENCIONES BÁSICAS

MINITAB (versión 14)

Choose:	tells you to make a menu selection by a mouse "point and click" entry. For example: **Choose: Stat** > **Quality Tools** > **Pareto Chart** instructs you to, in sequence, "point and click on" **Stat** on the menu bar, "followed by" **Quality Tools** on the pull-down, and then "followed by" **Pareto Chart** on the second pull-down.
Select:	indicates that you should click on the small box or circle to the left of a speci?ed item.
Enter:	instructs you to type or select information needed for a speci?c item.

PARA SU INFORMACIÓN

Para obtener más información acerca de cómo obtener el MINITAB consulte la siguiente dirección Internet http://www.minitab.com.

Excel

PARA SU INFORMACIÓN
Excel es parte de Microsoft Office y se puede hallar en numerosas computadoras personales.

Choose:	tells you to make a menu or tab selection by a mouse "point and click" entry. For example: **Choose: Chart Wizard** > **XY(Scatter)** > **1st graph picture** > **Next** instructs you to, in sequence, "point and click on" the **Chart Wizard** icon, followed by **XY(Scatter)** under Chart type, followed by **1st graph picture** on the Chart subtype, and then followed by **Next** on the dialog window.
Select:	indicates that you should click on the small box or circle to the left of a speci?ed item. It is often followed by a "point and click on" **Next** or **Finish** on the dialog window.
Enter:	instructs you to type or select information needed for a speci?c item.

TI-83/84 Plus

PARA SU INFORMACIÓN
Para obtener información acerca de cómo adquirir la calculadora TI-83/84 Plus consulte la siguiente página de Internet http://www.ti.com/calc.

Choose:	tells you which keys to press or menu selections to make. For example: **Choose: Zoom** > **9:ZoomStat** > **Trace** > > > instructs you to press the **Zoom** key, followed by selecting **9:ZoomStat** from the menu, followed by pressing the **Trace** key; > > > indicates to press arrow keys repeatedly to move along a graph to obtain important points.
Enter:	instructs you to type or select information needed for a speci?c item.
Screen Capture:	gives pictures of what your calculator screen should look like with chosen speci?cations highlighted.

Los detalles adicionales sobre el uso de MINITAB y Excel están a su disposición mediante el empleo del sistema Help en el software de MINITAB y de Excel. Los detalles adicionales para la TI-83/84 se encuentran en la *TI-83/84 Plus Graphing Calculator Guidebook.* Los detalles específicos sobre el uso de computadoras disponibles para las necesidades de usted debe obtenerlos de su profesor, o con el responsable del laboratorio de computación local.

Su centro local de computadoras puede proporcionarle una lista de los programas que hay a su disposición. Algunos de los paquetes estadísticos más fáciles de obtener son MINITAB, JMP-IN, y SPSS (Paquete Estadístico para Ciencias Sociales).

Nota: *Hay una gran tentación en cuanto a usar la computadora o calculadora para analizar cualquier conjunto de datos y luego considerar los resultados como si las estadísticas fuesen correctas. Recuerde el viejo dicho "basura entra, basura sale". Es muy importante el uso responsable de la metodología estadística. Corresponde al usuario asegurar la correcta aplicación de los métodos idóneos, la extracción de conclusiones exactas y la comunicación de éstas a los demás.*

SECCIÓN 1.6 EJERCICIOS

1.63 ¿De qué forma han incrementado las computadoras la utilidad de la estadística para profesionales como investigadores, analistas de datos que trabajan para el gobierno, asesores en estadística, etcétera?

1.64 ¿Cómo podrían ayudarle las computadoras en estadística?

1.65 ¿Alguna vez escuchó usted a alguien decir "esto debe estar bien, porque así lo dice mi computadora"? Explique por qué la calculadora puede o no puede haber dado la respuesta correcta.

1.66 ¿Qué es lo que significa decir "basura entra, basura sale" y cómo es que las computadoras han aumentado la probabilidad de que los estudios puedan ser víctimas de este adagio?

REPASO DEL CAPÍTULO

En retrospectiva

Ahora ya debe tener una idea general de lo que trata la estadística; imagen que crecerá y cambiará a medida que avance en el estudio de este libro. Sabe lo que son una muestra y una población, y conoce la distinción entre variables cualitativas (atributos) y cuantitativas (numéricas). Incluso conoce la diferencia entre la estadística y la probabilidad (aun cuando la probabilidad no se estudiará en detalle sino hasta el capítulo 4). También debe tener una idea y un entendimiento parcial sobre la importancia de las muestras aleatorias en estadística.

A lo largo del capítulo ha leído varios artículos que presentan diversos aspectos de la estadística. Las gráficas estadísticas presentan una variedad de información acerca de nosotros mismos, como las describimos, y otros aspectos del mundo que nos rodea. La estadística puede ser hasta divertida. Los ejemplos son innumerables. Mire a su alrededor y encuentre algunos ejemplos de la estadística en su vida cotidiana (vea los ejercicios 1.85 y 1.86, página 35).

Vocabulario y conceptos clave

censo (pp. 12, 20)
datos (p. 8)
datos numéricos (p. 10)
diseño muestral (p. 21)
encuesta (p. 20)
estadística (pp. 3, 4, 27)
estadística descriptiva (p. 4)
estadística inferencial (p. 4)
estadístico (p. 9)
estratos (p. 24)
estudio observacional (p. 19)
experimento (pp. 8, 19)
fortuito (p. 22)
marco muestral (p. 20)
marco muestral representativo (p. 21)
método muestral no-sesgado (p. 18)
método sesgado de muestreo (p. 18)
muestra (p. 8)
muestra aleatoria (p. 21)
muestra aleatoria estratificada (p. 24)
muestra aleatoria simple (p. 22)
muestra de conveniencia (p. 19)
muestra de juicio (p. 21)
muestra de voluntarios (p. 19)
muestra por conglomerados (p. 25)
muestra probabilística (p. 21)
muestra proporcional (p. 24)
muestra sistemática (p. 23)
muestreo de etapas múltiples (p. 24)
muestreo de una etapa (p. 21)
parámetro (p. 8)
población (p. 7)
población finita (p. 7)
población infinita (p. 7)
probabilidad (p. 27)
recolección de datos (pp. 18, 22)
variabilidad (p. 17)
variable (p. 8)
variable categórica (p. 11)
variable continua (p. 11)
variable cualitativa (p. 9)
variable cuantitativa (p. 10)
variable de atributos (p. 9)
variable discreta (p. 11)
variable nominal (p. 10)
variable ordinal (p. 10)

Objetivos de aprendizaje

✓ Comprender y describir la diferencia entre estadística descriptiva e inferencial.	pp.3-4, Ejer. 1.6, 1.7, 1.69
✓ Entender y ser capaz de identificar e interpretar las relaciones entre muestra y población, entre estadístico y parámetro.	pp. 7-9, EJ. 1.5
✓ Conocer y ser capaz de identificar y describir los diferentes tipos de variables.	pp. 9-12, Ejer. 1.31, 1.32

- ✓ Comprender e identificar la forma en que las muestras de conveniencia y de voluntarios resultan en muestras sesgadas. pp. 18-19, Ejer. 1.43
- ✓ Entender las diferencias entre experimentos, estudios observacionales y muestras de juicio, así como identificarlos. pp. 19-21
- ✓ Comprender y ser capaz de describir los métodos de muestreo de una etapa de "muestra aleatoria simple" y "muestreo sistemático". pp. 21-24
- ✓ Entender y ser capaz de describir los métodos de muestreo de varias etapas de "muestreo estratificado" y "muestreo por conglomerados". pp. 24-25
- ✓ Comprender y explicar la diferencia entre probabilidad y estadística. p. 27. Ejer. 16.1
- ✓ Entender que la variabilidad es inherente en todo y en el proceso de muestreo. p.17, Ejer. 1.36

Ejercicios del capítulo

1.67 Se desea describir al estudiante típico en su universidad. Describa una variable que mida algunas características de un estudiante y sus resultados en:

a. Datos de atributos

b. Datos numéricos

1.68 Un candidato para un cargo político dice que ganará las elecciones. Se lleva a cabo una encuesta, y 35 de 150 electores indican que votarán por el candidato, 100 electores indican que votarán por su oponente, y 15 no se han decidido.

a. ¿Cuál es el parámetro poblacional de interés?

b. ¿Cuál es el valor del estadístico de la muestra que podría usarse para estimar el parámetro de la población?

c. ¿Tendería usted a creerle al candidato con base en los resultados de la encuesta?

1.69 Una investigadora que estudia los hábitos de compras de consumidores pregunta, a una de cada 20 personas que entran al supermercado Publix, cuántas veces por semana va de compras a esa tienda. Hecho esto, registra la respuesta como T.

a. ¿Es $T = 3$ un ejemplo de una muestra, una variable, un estadístico, un parámetro, o un valor de datos?

Suponga que la investigadora interroga a 427 compradores durante la encuesta.

b. Indique un ejemplo de una pregunta que pueda ser contestada usando las herramientas de estadística descriptiva.

c. Proporcione un ejemplo de una pregunta que pueda ser contestada usando las herramientas de estadística inferencial.

1.70 Un investigador que estudia las actitudes de padres de niños de preescolar entrevista a una muestra aleatoria de 50 madres, cada una de las cuales tiene un niño de preescolar. Él pregunta a cada una de las madres "¿cuántas veces felicitó ayer a su hijo?" Él registra la respuesta como C.

a. ¿Es $C = 4$ un ejemplo de un valor de datos, un estadístico, un parámetro, una variable, o una muestra?

b. Dé un ejemplo de una pregunta que pueda ser contestada usando las herramientas de estadística descriptiva.

c. Dé un ejemplo de una pregunta que pueda ser contestada usando las herramientas de estadística inferencial.

1.71 La empresa Harris Interactive realizó una encuesta en línea de adultos estadounidenses, en diciembre de 2004, para el *Wall Street Journal Online's Health Industry Edition.*

Éstos son algunos de los resultados de una encuesta de Harris Interactive®, misma que fue aplicada a 2 013 adultos estadounidenses, efectuada el 14 y 16 de diciembre de 2004, para el Wall Street Journal Online's Health Industry Edition.

De todos los adultos que este año han recibido vacuna antigripal, 43% la recibió en el consultorio del médico. Otras instituciones donde se ha aplicado la vacuna este año incluyen clínicas de vacunación (18%), lugares de trabajo o escuelas (12%), y farmacias (10%). Es notable que nadie reportó haber comprado una vacuna antigripal vía Internet o en Canadá.

De 83% de adultos que no se han aplicado la vacuna este año, la mayoría (77%) dice que están tomando precauciones para reducir sus riesgos de infectarse esta temporada. La precaución más común reportada es lavarse las manos, usan guantes o productos de limpieza para las manos con más frecuencia (63%). Otras precauciones son tomar vitaminas o suplementos (49%), tratar de seguir una dieta sana (42%), descansar más (34%) y evitar lugares atestados (24%).

Fuente: http://www.harrisinteractive.com/news/

a. ¿Cuál es la población?

b. Mencione al menos cuatro variables que deben haberse usado.

c. Clasifique todas las variables del estudio como de atributos o numéricas.

1.72 Una instantánea de *USA Today* del 4 de junio de 2002 describe cómo se sienten los ejecutivos cuando buscan un nuevo trabajo cuando todavía están empleados. Según la instantánea, una encuesta de 150 ejecutivos de las 1000 empresas más grandes del país dio como resultado las respuestas siguientes: 36% se sentía muy cómodo, 33% se sentía un poco cómodo, 26% se sentía un poco incómodo, y 5% se sentía muy incómodo. ¿Clasificaría usted los datos recolectados, y los emplearía para determinar estos porcentajes como cualitativos (nominales u ordinales) o cuantitativos (discretos o continuos)?

1.73 Los resultados de un estudio titulado *Academic Atrophy: The Condition of the Liberal Arts in America's Public Schools* se publicaron el 8 de marzo de 2004. Fue el primer estudio sobre cómo es que la ley de No Dejar Niños Atrás podría influir en el tiempo de instrucción en los estudios sociales, es decir, las artes, geografía, historia y lenguas extranjeras. El estudio hizo una encuesta a más de 1 000 directores de escuelas en cuatro estados, y encontró que 47% de los directores de escuelas de altas minorías reportaron disminuciones en estudios sociales elementales.

Fuente: http://music-for-all.org/CBESurvey.html

a. ¿Cuál es la población?

b. ¿Cuál es la muestra?

c. ¿Es ésta una muestra de juicio o una muestra probabilística?

d. Si este estudio es una muestra de probabilidad, ¿qué tipo de método de muestreo piensa usted que se utilizó?

1.74 Con base en una encuesta de más de 125 000 personas, el National Center for Health Statistics informó que las personas casadas tienden a ser más sanas que otros grupos. Entre otras cosas, el estudio observó el número de quienes se reportaban como en estado de salud regular o malo. El estudio informó de los siguientes descubrimientos para todos los adultos de 18 años de edad o más: 11.9% informó que estaban con salud regular o mala, incluyendo 10.5% de casados, 19.6% de viudas, 16.7% de divorciados o separados, 12.5% de los que nunca se habían casado, y 14% de quienes vivían en pareja sin casarse.

Fuente: Finger Lakes Times, 19 de diciembre de 2004.

a. ¿Cuál es la población?

b. ¿Cuál es la muestra?

c. Con base en el tamaño de la muestra, ¿qué clase de muestra sospecha usted que se tomó?

1.75 La siguiente gráfica muestra la relación entre tres variables: número de conductores con licencia de manejo, número de vehículos registrados, y el tamaño de la población residente para Estados Unidos de 1961 a 2003.

Fuente: United States Department of Transportation - Federal Highway Administration

Observe la gráfica y responda las siguientes preguntas:

a. ¿Parece razonable que la línea de Población y la línea de Conductores corran básicamente paralelas entre sí, y que la línea de Población está arriba de la de Conductores? Explique qué significa que las líneas sean paralelas. ¿Qué sería si no fueran paralelas?

b. Las líneas de Conductores y Automotores se cruzan. ¿Qué significa lo anterior? ¿Cuándo se cruzaron las líneas, y qué representa la intersección en un punto?

1.76 El club de 700: Barry Bonds, de los Gigantes de San Francisco, está en camino de convertirse en el rey de cuadrangulares (*home-runs*) esta temporada o la siguiente. El año pasado, se unió a Hank Aaron y Babe Ruth como los únicos jugadores de las ligas mayores en haber conectado más de 700 cuadrangulares en sus carreras. A continuación veamos cómo es que conformaron sus totales.

a. Describa y compare la apariencia general de las tres gráficas. Incluya ideas acerca de aspectos como la duración de sus carreras, cuándo batearon más cuadrangulares por año y sus relaciones con el proceso de envejecimiento, y cualquier otra cosa que usted considere.

b. ¿Le parece a usted que uno de ellos fue más consistente con la producción anual de cuadrangulares?

c. A partir de la evidencia aquí presentada, ¿quién piensa usted que debe ser llamado el "Rey del Cuadrangular"?

d. Los 73 cuadrangulares de Barry Bonds en una temporada, ¿fueron de "suerte"?

e. Si usted fuera el propietario de un equipo y estuviera interesado en la producción de cuadrangulares en los siguientes varios años, ¿a quién le gustaría tener en su equipo? Digamos que lo estuviera usted firmando a la edad de 21 años. Y a los 35.

1.77 En el otoño de 2003, la National Safe Kids Campaign llevó a cabo un estudio sobre el uso del casco entre niños de 5 a 14 años que participan en deportes sobre ruedas. Se recolectaron datos de varios lugares por todo Estados Unidos que fueron designados como lugares donde es frecuente que participen niños en deportes sobre ruedas. La actividad, género aparente, y edad estimada se registraron por cada participante, así como información sobre el uso de casco. Se encontró que, en general, 41% de los niños llevaban casco cuando participaron en una competencia de deportes sobre ruedas.

a. Este estudio, ¿fue un experimento o un estudio observacional?

b. Identifique el parámetro de interés.

c. Identifique el estadístico y dé su valor.

d. Clasifique las cuatro variables como numéricas o de atributos.

1.78 *USA Today*, en un artículo de diciembre de 2004 titulado "No hay lugar como el trabajo para las vacaciones", presentó los resultados de un estudio de 600 trabajadores de tiempo completo en Estados Unidos hecho por Penn Schoen & Berland Associates. Los resultados revelaron que 33% de quienes respondieron a la encuesta no se tomaron días de descanso durante las vacaciones. De quienes fueron entrevistados, 28% reportó haberse tomado 1 o 2 días durante las vacaciones.

a. ¿Cuál es la población?

b. ¿Cuál es la muestra?

c. ¿Es ésta una muestra de juicio o una muestra de probabilidad?

Ejercicio 1.76

Fuente: The Washington Post

1.79 Dos tercios de los estadounidenses más viejos toman parte en actividades físicas en su tiempo libre pero la mala nutrición sigue siendo un problema, en especial cuando se trata de frutas y verduras, según un estudio hecho por los Centros for Disease Control and Prevention y el Merck Institute of Aging & Health. Entre lo que descubrió el estudio, la tercera parte de estadounidenses de más de 65 años no tienen ninguna actividad física en su tiempo libre. Entre todos los estadounidenses, este valor es 24.5%. Dos terceras partes de los estadounidenses más viejos no comen cinco porciones de frutas y verduras al día, pero lo hacen mejor que los estadounidenses en su conjunto, de quienes 73% no satisfacen las recomendaciones diarias.

a. ¿Cuál es la población?

b. ¿Cuáles son las características de interés?

c. Clasifique todas las variables del estudio ya sea como de atributos o numéricas.

1.80 El 2001 National Aging Research Survey (Estudio Nacional de 2001 para Investigación del Envejecimiento) reveló que los estadounidenses tienen grandes expectativas de tener una vida larga, sana e independiente. Las generaciones X e Y, de 18 a 36 años, son quienes más desean llegar a 100 años. De este grupo de edades, 69% reportó que les gustaría vivir hasta los 100 años. El método de muestreo utilizado para recolectar esta información fue el marcar dígitos aleatorios (RDD). Este método da a toda familia que cuente con teléfono en Estados Unidos la misma probabilidad de ser llamado. El RDD es una popular herramienta de encuestas. Utilice el Internet para buscar información y escriba una explicación de 100 palabras de cómo funciona.

1.81 ¿Quién toma más medicamentos? La Asociación Nacional de Cadenas de Farmacias proporciona algunas respuestas a esa pregunta.

a. ¿Cuál variable se emplea en la elaboración de esta información?

b. ¿Qué género toma más medicamentos? ¿En cuánto más?

c. ¿Qué grupo de edades toma más medicamentos? ¿Cuánto toman?

d. ¿Tiene importancia la geografía? ¿Qué estado tiene el uso más alto? ¿Y el más bajo?

Fuente: National Association of Chain Drug Stores

1.82 "Medicamentos selectos" gráfica de un artículo de *Democrat & Chronicle* de marzo de 2005, titulado "La píldora desengancha las garras de la heroína, pero pocos la consiguen", presenta los números de personas que abusan de diversos medicamentos. El número estimado de personas que abusaron de la mariguana en Estados Unidos durante 2002 fue de 4.3 millones. La fuente para esta gráfica fue la Substance Abuse and Mental Health Services administration. ¿Qué método de muestreo piensa usted que esa Administración utilizó para obtener esta estadística? Explique.

1.83 Describa en sus propias palabras y dé un ejemplo de cada uno de los siguientes términos. Sus ejemplos no deben ser los dados en clase ni en este libro de texto.

a. Variable b. Datos c. Muestra

d. Población e. Estadístico f. Parámetro

1.84 Describa en sus propias palabras y dé un ejemplo de los términos siguientes. Sus ejemplos no deben ser los dados en clase ni en este libro de texto.

a. Muestra aleatoria

b. Muestra probabilística

c. Muestra de juicio

1.85 Encuentre un artículo o un anuncio en un periódico o revista que ejemplifique el uso de la estadística.

a. Identifique y describa un estadístico reportado en el artículo.

b. Identifique y describa la variable relacionada con el estadístico del inciso a.

c. Identifique y describa la muestra relacionada con el estadístico del inciso a.

d. Identifique y describa la población de la que se tomó la muestra del inciso c.

1.86 a. Encuentre un artículo en un periódico o revista que ejemplifique el uso de la estadística en una forma que pudiera ser considerada "entretenimiento" o "de recreación". Describa por qué piensa usted que este artículo se ajusta a una de estas categorías.

b. Encuentre un artículo en un periódico o revista que ejemplifique el uso de la estadística el uso de la estadística, y presente un descubrimiento poco común como resultado de un estudio. Describa por qué estos resultados son (o no son) "de interés periodístico".

1.87 En el ejercicio 1.5, se le pidió a usted que escribiera una oración por cada una de las tres actividades estadísticas dadas en la definición de *estadística.* Ahora que ha terminado el capítulo, repase su trabajo. De nueva cuenta, con sus propias palabras, cambie y/o mejore su trabajo para completar un párrafo sobre la definición de *estadística.*

Proyecto del capítulo

¡Ciudadanos, los están observando!

El proyecto del capítulo nos lleva de nuevo a la sección 1.1, ¡Ciudadanos, les están observando!, como una forma de evaluar lo que ha aprendido usted en este capítulo. Estudie la información estadística presentada por las gráficas y tablas, y hágase la pregunta de cómo es que los términos (población, muestra, variable, estadístico, tipo de variable) estudiados en este capítulo se aplican a cada uno, y cómo los compara con el caso estadístico que se expone.

Trabajando en el contenido del capítulo 1

1.88 Con respecto a las cuatro gráficas de la sección 1.1 de las páginas 1 y 2, complete lo siguiente:

a. ¿Qué población estadística es de interés para todas estas gráficas?

b. Identifique una gráfica específica. ¿Qué variables se emplearon para recolectar la información necesaria para determinar las estadísticas reportadas?

c. Mencione un estadístico que esté siendo reportado en su gráfica.

d. Para obtener los datos para su gráfica, ¿qué métodos piensa usted que se emplearon: muestra de conveniencia, muestra de voluntarios, muestra aleatoria, encuesta, estudio observacional, experimento, o muestra de juicio?

e. Considerando el método, ¿cuánta fe tiene usted en las estadísticas impresas? Describa posibles sesgos.

Para su estudio

1.89 Seleccione una de las gráficas de ¡Ciudadanos, les están observando! (p.1); a continuación, usando los estudiantes de su escuela o universidad como la población de interés, recolecte datos muestrales de 30 estudiantes y produzca su propia versión de la gráfica. Escriba un párrafo que describa cómo se comparan los resultados de usted con los reportados en la gráfica seleccionada.

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si no siempre es verdadero, cambie las palabras impresas en negritas con las que hagan siempre verdadero el enunciado.

1.1 La estadística **inferencial** es el estudio y descripción de datos que resultan de un experimento.

1.2 La **estadística descriptiva** es el estudio de una muestra que hace posible que hagamos proyecciones o estimaciones acerca de la población de la que se extrae la muestra.

1.3 Una **población** es típicamente un conjunto muy grande de individuos u objetos acerca de los cuales deseamos información.

1.4 Un estadístico es la medida calculada de alguna característica de una **población.**

1.5 Un parámetro es la medida de alguna característica de una **muestra.**

1.6 Como resultado de encuestar a 50 estudiantes de primer año, se encontró que 16 habían participado en deportes intramuros, 23 habían trabajado como oficiales de grupos y clubes, y 18 habían estado en obras escolares durante sus años de preparatoria. Éste es un ejemplo de **datos numéricos.**

1.7 El "número de manzanas podridas por caja de embarque" es un ejemplo de una variable **cualitativa.**

1.8 El "grueso de una lámina metálica" empleada en un proceso de manufactura es un ejemplo de una variable **cuantitativa.**

1.9 Una muestra **representativa** es una muestra obtenida en forma tal que todos los individuos tenían igual probabilidad de ser elegidos.

1.10 Los objetivos básicos de la **estadística** son obtener una muestra, inspeccionar esta muestra, y luego hacer inferencias acerca de las características desconocidas de la población de la cual se extrajo la muestra.

SEGUNDA PARTE: Aplicación de los conceptos

Los propietarios de Corner Convenience Store están preocupados por la calidad del servicio que sus clientes reciben. Para estudiar el servicio, recolectaron muestras de cada una de las diversas variables.

1.11 Clasifique cada una de las variables siguientes como nominales, ordinales, discretas o continuas:

a. Método de pago por compras (de contado, tarjeta de crédito, cheque)

b. Satisfacción del cliente (muy satisfecho, satisfecho, nada satisfecho)

c. Cantidad de impuesto de ventas sobre compras

d. Número de artículos comprados

e. Número de licencia de conductor del cliente

1.12 El tiempo medio de salida para todos los clientes de la Corner Convenience Store se ha de calcular usando el tiempo medio de salida para 75 clientes seleccionados al azar. Compare los renglones de la columna 2 con los términos estadísticos de la columna 1.

1	2
_____dato	(a) los 75 clientes
_____datos	(b) el tiempo medio para todos los clientes
_____experimento	
_____parámetro	(c) 2 minutos, tiempo de salida de un cliente
_____población	
_____muestra	(d) el tiempo medio para los 75 clientes
_____estadística	
_____variable	(e) todos los clientes de Corner Convenience Store
	(f) el tiempo de salida para un cliente
	(g) los 75 tiempos
	(h) el proceso usado para seleccionar 75 clientes y medir sus tiempos

TERCERA PARTE: Para entender los conceptos

Escriba un breve párrafo en respuesta a cada una de estas preguntas.

1.13 La población y la muestra son conjuntos de objetos. Describa la relación entre ellas y dé un ejemplo.

1.14 La variable y los datos para una situación específica están estrechamente relacionados. Explique esta relación y dé un ejemplo.

1.15 Los datos, el estadístico, y el parámetro son todos ellos valores empleados para describir una situación estadística. ¿Cómo distinguimos entre estos tres términos? Dé un ejemplo.

1.16 ¿Qué condiciones se requieren para que una muestra sea aleatoria? Explique e incluya un ejemplo de una muestra que sea aleatoria y uno donde no sea aleatoria.

CAPÍTULO

2

Análisis descriptivo y presentación de datos de una sola variable (univariados)

2.1 Usted y la Internet

2.2 Gráficas, diagramas de Pareto y diagramas de tallo y hoja

2.3 Distribuciones de frecuencias e histogramas

2.4 Medidas de tendencia central

2.5 Medidas de dispersión

2.6 Medidas de posición

2.7 Interpretación y comprensión de la desviación estándar

2.8 El arte de la mentira estadística

2.9 Media y desviación estándar de una distribución de frecuencias (opcional)

© Photodisc Red/Getty Images

2.1 Usted y la Internet

¿Alguna vez se ha preguntado lo que hacen otras personas cuando están en la Internet? Bien, no es usted el único. El Stanford Institute for the Quantitative Study of Society (SIQSS) apoyó un estudio para analizar cómo es que las personas utilizan la Internet. A cuatrocientos encuestados se les pidió seleccionaran cuál de las 17 actividades comunes realizaron (o no) en Internet. El e-mail fue identificado por 90% de las personas que respondieron la encuesta como una de las aplicaciones de mayor uso en la Internet. Otros usos comunes son la búsqueda y obtención de información, actividades de entretenimiento, el uso de salas de chat y la realización de transacciones comerciales.

LO QUE HACEN USUARIOS DE LA INTERNET

Fuente: **Stanford Institute for the Quantitative Study of Society [http://www.stanford.edu/group/siqss/Press_Release/Chart9gif].**
Internet and Society–**study por Mie y Erbring, patrocinado por el SIQSS (vea ejemplar). Reimpreso con permiso.**

El párrafo precedente resume toda la información obtenida del estudio de 4000 usuarios de Internet. ¿Puede acaso imaginar que toda esa información se escribiera en oraciones? Las gráficas (imágenes) pueden verdaderamente valer mil palabras.

No sólo es más clara la información y su formato es más conciso, sino que el formato también nos permite hacer algunas conclusiones al mismo tiempo. De inmediato sabemos en cuáles actividades participan más usuarios y cuáles son las más populares.

Si le hubieran preguntado, "Qué actividad realiza usted con más frecuencia en la Internet?" ¿Qué hubiera contestado? ¿Piensa que su respuesta está representada con claridad en el diagrama? Cuando trabaje el capítulo 2, empezará a aprender a organizar y resumir datos en gráficas y estadísticos numéricos para describir a los datos con claridad y propiedad. Después podrá realizar una investigación a fondo en el apartado "Usted y la Internet" que se incluye en la sección de proyecto, mediante los ejercicios 2.224 y 2.225 (p.140).

SECCIÓN 2.1 EJERCICIOS

2.1 A unos estudiantes de un curso de estadística en línea se les preguntó en cuántas actividades diferentes de Internet estuvieron durante una semana típica. Los siguientes datos muestran el número de actividades:

6	7	3	6	9	10	8	9	9	6	4	9	4	9
4	2	3	5	13	12	4	6	4	9	5	6	9	
11	5	6	5	3	7	9	6	5	12	2	6	9	

a. Si se le pidiera a usted que presentara estos datos, ¿cómo los organizaría y resumiría?

b. Este capítulo estudiará varios métodos para presentar y describir datos. ¿Qué tipo de información o conclusiones le gustaría conocer acerca de estos datos, si una de las partes de los datos se refirieran a usted?

2.2 a. ¿En cuántas actividades diferentes en Internet participó usted la semana pasada?

b. ¿Cómo piensa que se compara usted con los 40 usuarios de Internet en la muestra del ejercicio 2.1?

c. ¿Cómo piensa que se compara usted con todos los usuarios de Internet?

2.2 Gráficas, diagramas de Pareto y diagramas de tallo y hoja

Una vez que hayan sido recolectados los datos muestrales, debemos "conocerlos". Una de las formas más útiles de conocer los datos es usar una técnica inicial de exploración de análisis de datos que resultará en una representación gráfica de los datos. La gráfica revelará visualmente patrones de comportamiento de la variable bajo estudio. Hay diversas formas gráficas de describir datos. El tipo de datos y la idea que se va representar determina el método que se va a utilizar.

Nota: no existe una respuesta correcta única cuando se construye una gráfica. El juicio del analista y las circunstancias que rodean al problema desempeñan papeles importantes en el desarrollo de la gráfica.

Datos cualitativos

Gráficas de círculos y gráficas de barras: son gráficas que se usan para resumir datos cualitativos, o por atributos, o datos categóricos. Las gráficas de círculos (diagramas de pastel) muestran la cantidad de datos que pertenecen a cada una de las categorías como parte proporcional de un círculo. Las gráficas de barras muestran la cantidad de datos que pertenecen a cada una de las categorías como un área rectangular de tamaño proporcional.

EJEMPLO 2.1 Gráficas de datos cualitativos

La tabla 2.1 ilustra el número de casos de cada tipo de operación efectuada en el Hospital General el año pasado.

TABLA 2.1

Operaciones efectuadas en el Hospital General el año pasado

Tipo de operación	Número de casos
Torácica	20
Huesos y articulaciones	45
Ojos, oídos, nariz y garganta	58
General	98
Abdominal	115
Urológico	74
Proctológico	65
Neurocirugía	23
Total	498

Los datos de la tabla 2.1 se muestran en una gráfica circular en la figura 2.1, con cada tipo de operación representado por una proporción relativa del círculo, que se encuentra al dividir el número de casos entre el tamaño total de la muestra, es decir, 498. Las proporciones se presentan entonces como porcentajes (por ejemplo, 25% es ¼ del círculo). La figura 2.2 muestra los mismos datos de "tipo de opera-

FIGURA 2.1 Gráfica circular

FIGURA 2.2 Gráfica de barras

PARA SU INFORMACIÓN

Todas las representaciones gráficas necesitan explicarse completamente por sí mismas. Esto incluye un título descriptivo, que tenga sentido y la identificación apropiada de las cantidades y variables que se incluyen.

ción", pero en forma de una gráfica de barras. Las gráficas de barras de datos por atributos deben dibujarse con un espacio entre las barras de la misma anchura.

INSTRUCCIONES DE TECNOLOGÍA: GRÁFICA CIRCULAR

MINITAB (Versión 14) Introduzca las categorías en C1 y las frecuencias correspondientes en C2; luego continúe con:

```
Choose:     Graph > Pie Chart . . .
Select:     Chart values from a table
Enter:      Categorical variable:  C1  Summary variables: C2
Select:     Labels > Title/Footnotes  Enter: Title: your title
Select:     Slice Labels > Select desired labels > OK > OK
```

Excel Introduzca las categorías en la columna A y las frecuencias correspondientes en la columna B; luego continúe con:

```
Choose:     Chart Wizard > Pie > 1st picture (usually) > Next
Enter:      Data range:  (A1:B5 or select cells)
Check:      Series in:  columns > Next
Choose:     Titles
Enter:      Chart title:  Your title
Choose:     Data Labels
Select:     Category name and Percentage > Next > Finish
```

Para editar la gráfica de pastel:

```
Click On:   Anywhere clear on the chart
            –use handles to size
            Any cell in the category or frequency column
            and type in different name or amount > ENTER
```

TI-83/84 Plus Introduzca en L1 las frecuencias para las diversas categorías, luego continúe con:

```
Choose:     PRGM > EXEC > CIRCLE*
Enter:      LIST:  L1 > ENTER
            DATA DISPLAYED?: 1:PERCENTAGES
                       OR
                             2:DATA
```

*El *Programa 'CIRCLE' de la TI-83/84 Plus, y otros programas, se pueden descargar desde la siguiente dirección de Internet: http://statistics.duxbury.com/jkes10e. Los programas y archivos de datos de la TI-83/84 Plus son jkprogs.zip y jklists.zip. Guarde los archivos para su computadora y descomprímalos usando una función zip. Descargue los programas a su calculadora usando software TI-Graph Link.

Cuando la gráfica de barras se presenta en la forma de un *diagrama de Pareto,* presenta información adicional y muy útil.

Diagrama de Pareto: es una gráfica especial de barras. En ésta, las barras se presentan de la categoría más numerosa a la menos numerosa. Incluye una gráfica de líneas que muestra los porcentajes acumulativos y las cantidades para las barras.

EJEMPLO 2.2 Diagrama de Pareto de delitos con violencia

El FBI informó del número de delitos con violencia, por categoría, para el año 2003 http://www.fbigov/ucr/ucr.htm#hate). El diagrama de Pareto de la figura 2.3 muestra 8706 delitos con violencia clasificados, sus porcentajes y los porcentajes acumulativos.

FIGURA 2.3 Diagrama de Pareto

Delito	Raza	Orientación Sexual	Religión	Etnia	Otros
Cantidad	4574	1430	1426	1236	49
Porcentaje	52.5	16.4	16.4	14.2	0.6
Cum %	52.5	68.9	85.3	99.4	100.0

El diagrama de Pareto es el preferido en aplicaciones de control de calidad. Un diagrama de Pareto para los tipos de defectos mostrará aquellos que tienen el mayor efecto en la valoración de las partes defectuosas respecto al orden de efecto. En consecuencia, es fácil ver qué defectos deben ser identificados, para disminuir con mayor eficiencia el porcentaje de partes defectuosas.

INSTRUCCIONES DE TECNOLOGÍA: DIAGRAMA DE PARETO

MINITAB (Versión 14) Introduzca las categorías en C1 y las frecuencias correspondientes en C2; luego continúe con:

```
Choose:     Stat > Quality Tools > Pareto Chart
Select:     Chart defects table
Enter:      Labels in:  C1   Frequencies in:  C2
Select:     Options
Enter:      Title: your title > OK > OK
```

Excel Introduzca las categorías en la columna A y las frecuencias correspondientes en la columna B (los encabezados de las columnas son opcionales), luego continúe con:

Primero, ordenando la tabla:

```
Activate both columns of the distribution
Choose:     Data > Sort > Sort by: Column B (freq or rel freq col.)
Select:     Descending
            My list has: Header row or No Header row > OK
Choose:     Chart Wizard > Column > 1st picture (usually) > Next
```

```
Choose:      Data Range
Enter:       Data Range:  (A1:B5 or select cells)
Select:      Series in:   Columns > Next
Choose:      Titles
Enter:       Chart title: your title
             Category (x) axis: title for x-axis
             Value (y) axis: title for y-axis > Next > Finish
```

Para editar el diagrama de Pareto:

```
Click on:    Anywhere clear on the chart
             —use handles to size
             Any title name to change
             Any cell in the category column and type in a name > Enter
```

Excel no incluye la gráfica de líneas.

TI-83/84 Plus Introduzca en L1 las categorías numeradas y en L2 las frecuencias correspondientes, luego continúe con:

```
Choose:      PRGM > EXEC > PARETO*
Enter:       LIST: L2 > ENTER
 Ymax:       at least the sum of the frequencies > ENTER
 Yscl:       increment for y-axis > ENTER
```

*El programa 'PARETO' es uno de los diferentes programas que se pueden descargar desde una página web. Consulte las instrucciones específicas en la página 42.

Datos cuantitativos

Una razón importante para construir una gráfica de datos cuantitativos es el hecho de presentar la *distribución* de los mismos.

Distribución: es el patrón de variabilidad que presentan los datos de una variable. La distribución exhibe la frecuencia de cada valor de la variable.

Una de las gráficas más sencillas empleadas para exhibir una distribución es la *gráfica de puntos*.

Gráfica de puntos: presenta los datos de una muestra al representar cada dato con un punto ubicado a lo largo de una escala que puede ser horizontal o vertical. La frecuencia de los valores se representa a lo largo de la otra escala.

EJEMPLO 2.3 Gráfica de datos de las calificaciones de un examen

La tabla 2.2 es una muestra de 19 calificaciones de un examen, mismas que fueron seleccionadas aleatoriamente de un grupo de alumnos.

TABLA 2.2

Muestra de 19 calificaciones de examen

76	74	82	96	66	76	78	72	52	68
86	84	62	76	78	92	82	74	88	

La figura 2.4 es una gráfica de puntos de las 19 calificaciones de examen.

FIGURA 2.4
Gráfica de puntos

Observe la forma en la que se "agrupan" los datos de la figura 2.4: están concentrados cerca del centro y más dispersados cerca de los extremos.

La presentación de una gráfica de puntos es una técnica que conviene usar cuando se empiezan a analizar los datos. Produce una imagen de los datos que los clasifica en orden numérico. (*Ordenar* datos es ponerlos en una lista en orden de jerarquía según el valor numérico.)

INSTRUCCIONES DE TECNOLOGÍA: GRÁFICA DE PUNTOS

Minitab (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:      Graph > Dotplot . . . > One Y. Simple > OK
Enter:       Graph Variables: C1 > OK
```

Excel No se dispone de un diagrama de gráficas de puntos, pero se puede realizar el paso inicial que implica clasificar los datos. Introduzca los datos en la columna A y active la columna de datos; luego continúe con:

```
Choose:      Data > Sort
Enter:       Sort by: Column A
Select:      Ascending > My list has: Header row or No Header row
```

Utilice los datos ordenados para terminar de construir la gráfica de puntos.

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:      PRGM > EXEC > DOTPLOT*
Enter:       LIST: L1 > ENTER
             Xmin: at most the lowest x value
             Xmax: at least the highest x value
             Xscl: 0 or increment
             Ymax: at least the highest frequency
```

*El programa 'DOTPLOT' es uno de numerosos programas de que se dispone para descargarlos de un sitio web. En la página 42 vea instrucciones específicas.

En años recientes se ha popularizado ampliamente el uso de una técnica que se conoce con *diagrama de tallo y hoja*, ya que permite resumir datos numéricos. Es una combinación de la técnica de gráficas y la de ordenar. Estos diagramas son fáciles de crear y usar, además de ser muy propios para aplicaciones en computadoras.

Diagrama de tallo y hoja: presenta los datos de una muestra usando los dígitos reales que conforman el dato. Cada valor numérico se divide en dos partes: El dígito o dígitos principales son el tallo, y los dígitos que le siguen son las hojas. Los tallos están situados a lo largo del eje principal, y una hoja por cada dato está dispuesta de modo que muestra la distribución de los datos.

EJEMPLO 2.4 Diagrama de tallo y hoja sin terminar

Construyamos ahora un diagrama de tallo y hoja para las 19 calificaciones de examen que se presentan en la tabla 2.2 de la página 45.

Con un vistazo rápido vemos que hay calificaciones de los 50, 60, 70, 80 y 90. Usemos el primer dígito de cada calificación como el tallo y el segundo dígito como la hoja. Por lo general, el diagrama se construye verticalmente. Trazamos una línea vertical y ponemos los tallos, en orden, a la izquierda de la línea.

```
5 |
6 |
7 |
8 |
9 |
```

A continuación ponemos cada hoja en su tallo. Esto se hace al poner el dígito siguiente en el lado derecho de la línea vertical opuesto a su dígito principal correspondiente. Nuestro primer valor es 76; 7 es el tallo y 6 es la hoja. Así, ponemos un 6 en el lado opuesto al tallo 7:

```
7|6
```

El siguiente valor es 74, de modo que la hoja de 4 se pone en el tallo 7, a continuación del 6.

```
7|6 4
```

El siguiente dato es 82, por lo que la hoja de 2 se pone en el tallo 8.

```
7 | 6 4
8 | 2
```

Continuamos hasta que cada una de las otras 16 hojas aparezca en el diagrama. La figura 2.5A muestra el diagrama de tallo y hoja resultante, la figura 2.5B muestra el diagrama de tallo y hoja completo después que se han ordenado las hojas.

De la figura 2.5B, vemos que las calificaciones están centradas alrededor de 70. En este caso, todas las calificaciones con el mismo dígito de decenas se colocan en la misma rama, pero no siempre puede ser deseable. Supóngase que reconstruimos el diagrama; esta vez, en lugar de agrupar 10 posibles valores en cada tallo, agrupamos los valores de modo que sólo 5 valores posibles puedan caer en cada

FIGURA 2.5A
Diagrama de tallo y hoja no terminado

Calificaciones de 19 exámenes

5	2
6	6 8 2
7	6 4 6 8 2 6 8 4
8	2 6 4 2 8
9	6 2

FIGURA 2.5B
Diagrama de tallo y hoja no terminado

Calificaciones de 19 exámenes

5	2
6	2 6 8
7	2 4 4 6 6 6 8 8
8	2 2 4 6 8
9	2 6

FIGURA 2.5C
Diagrama de tallo y hoja no terminado

Calificaciones de 19 exámenes

(50–54) 5	2
(55–59) 5	
(60–64) 6	2
(65–69) 6	6 8
(70–74) 7	2 4 4
(75–79) 7	6 6 6 8 8
(80–84) 8	2 2 4
(85–89) 8	6 8
(90–94) 9	2
(95–99) 9	6

tallo, como se observa en la figura 2.6. ¿Observa una diferencia en el aspecto de la figura 2.6? La forma general es aproximadamente simétrica alrededor de los 70 altos. Nuestra información está un poco más refinada, pero básicamente vemos la misma distribución.

INSTRUCCIONES DE TECNOLOGÍA: DIAGRAMA DE TALLO Y HOJA

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:      Graph > Stem-and-Leaf . . .
Enter:       Graph variables: C1
             Increment:  stem width (optional) > OK
```

Excel Introduzca los datos en la columna A; luego continúe con:

```
Choose:      Tools > Data Analysis Plus* > Stem and Leaf Display > OK
Enter:       Input Range: (A2:A6 or select cells)
             Increment: Stem Increment
```

*Data Analysis Plus es un conjunto de macros estadísticos para Excel. Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

TI-83/84 Plus Teclee los datos en L1; luego continúe con:

```
Choose:      STAT > EDIT > 2:SortA(
Enter:       L1
```

Use datos ordenados para terminar de construir en forma manual el diagrama de tallo y hoja.

Es más común que numerosas variables exhiban una distribución que esté concentrada (agrupada) alrededor de un valor central, y después de algún modo se disperse en una o ambas direcciones. Es frecuente que una gráfica revele algo que el analista pueda haber anticipado. El ejemplo 2.5 demuestra lo que generalmente ocurre cuando dos poblaciones se muestrean juntas.

EJEMPLO 2.5 Traslape de distribuciones

Se selecciona una muestra aleatoria de 50 estudiantes universitarios. El peso de los alumnos se obtuvo a partir de sus registros médicos. En la tabla 2.3 aparecen los datos resultantes.

Nótese que los pesos van de 98 a 215 libras. Agrupemos los pesos en tallos de 10 unidades usando los dígitos de centenas y de decenas como tallos y el dígito de unidades como la hoja (vea la figura 2.7). Las hojas se han dispuesto en orden numérico.

Una inspección minuciosa de la figura 2.7 sugiere que pueden aparecer dos distribuciones que se traslapan. Esto es exactamente lo que tenemos: una distribución de pesos de jovencitas y una distribución de pesos de jóvenes. La figura 2.8 muestra un diagrama de tallo y hoja "consecutivo" de este conjunto de datos, y resulta obvio que se trata de dos distribuciones distintas.

La figura 2.9, que es una gráfica de puntos "consecutiva" (misma escala) de los mismos datos de 50 pesos, muestra la misma distinción entre los dos subconjuntos.

TABLA 2.3

Pesos de 50 estudiantes universitarios

Estudiante	1	2	3	4	5	6	7	8	9	10
Hombre/mujer	M	H	M	H	H	M	M	H	H	M
Peso	98	150	108	158	162	112	118	167	170	120
Estudiante	11	12	13	14	15	16	17	18	19	20
Hombre/mujer	H	H	H	M	M	H	M	H	H	M
Peso	177	186	191	128	135	195	137	205	190	120
Estudiante	21	22	23	24	25	26	27	28	29	30
Hombre/mujer	H	H	M	H	M	M	H	H	H	H
Peso	188	176	118	168	115	115	162	157	154	148
Estudiante	31	32	33	34	35	36	37	38	39	40
Hombre/mujer	M	H	H	M	H	M	H	M	H	H
Peso	101	143	145	108	155	110	154	116	161	165
Estudiante	41	42	43	44	45	46	47	48	49	50
Hombre/mujer	M	H	M	H	H	M	M	H	H	H
Peso	142	184	120	170	195	132	129	215	176	183

FIGURA 2.7 Diagrama de tallo y hoja

Pesos de 50 estudiantes universitarios (lb) Tallo y hoja de PESO
***N* = 50 Unidad de hoja = 1.0**

9	8
10	1 8 8
11	0 2 5 5 6 8 8
12	0 0 0 8 9
13	2 5 7
14	2 3 5 8
15	0 4 4 5 7 8
16	1 2 2 5 7 8
17	0 0 6 6 7
18	3 4 6 8
19	0 1 5 5
20	5
21	5

FIGURA 2.8 Diagrama de tallo y hoja "consecutivo"

Pesos de 50 estudiantes universitarios (lb)

Mujer		Hombre
8	09	
1 8 8	10	
0 2 5 5 6 8 8	11	
0 0 0 8 9	12	
2 5 7	13	
2	14	3 5 8
	15	0 4 4 5 7 8
	16	1 2 2 5 7 8
	17	0 0 6 6 7
	18	3 4 6 8
	19	0 1 5 5
	20	5
	21	5

Con base en la información mostrada en las figuras 2.8 y 2.9, y partiendo de lo que sabemos acerca del peso de las personas, parece razonable concluir que las estudiantes universitarias pesan menos que los estudiantes universitarios. Las situaciones donde aparece más de un conjunto de datos se estudian en el capítulo 3.

FIGURA 2.9
Gráficas de puntos con escala común

INSTRUCCIONES DE TECNOLOGÍA: GRÁFICAS DE PUNTOS MÚLTIPLES

MINITAB (Versión 14) Introduzca los datos en C1 y en C2 las categorías numéricas correspondientes; luego continúe con:

```
Choose:     Graph > Dotplot . . .
Select:     One Y, With Groups > OK
Enter:      Graph variable: C1
            Categorical variables for grouping: C2 > OK
```

Si las diversas categorías están en columnas separadas, seleccione *Multiple Y Simple* e introduzca las demás columnas bajo las variables Graph.

Excel No se dispone de un diagrama de gráficas de puntos, pero es posible hacer el paso inicial de clasificar los datos. Utilice los comandos como se muestra, con el diagrama de gráfica de puntos de la página 45, luego termine construyendo manualmente las gráficas de puntos.

TI-83/84 Introduzca en L1 los datos para la primera gráfica de puntos y en L3 los datos para la segunda gráfica de puntos; luego continúe con:

```
Choose:     STAT > EDIT > 2:SortA(
Enter:      L1 > ENTER
            In L2, enter counting numbers for each category.
            Ex.      L1      L2
                     15      1
                     16      1
                     16      2
                     17      1
Choose:     STAT > EDIT > 2:SortA(
Enter:      L3 > ENTER
            In L4, enter counting numbers (a higher set*) for each category;
            *for example: use 10,10,11,10,10,11,12, . . . (offsets the two dot-
            plots).
Choose:     2nd > FORMAT > AxesOff (Optional—
            must return to AxesOn)
Choose:     2nd > STAT PLOT > 1:PLOT1
```

Choose: **2nd** > **STAT PLOT** > **2:PLOT2**

Choose: **Window**
Enter: **at most lowest value for both, at least highest value for both, 0 or increment, −2, at least highest counting number,1,1**
Choose: **Graph** > **Trace** > > > > (gives data values)

SECCIÓN 2.2 EJERCICIOS

2.3 Como gráfica estadística, la gráfica circular tiene limitaciones. Examine la gráfica circular de la figura 2.1 y la gráfica de barras de la figura 2.2.

a. ¿Qué información demuestran ambas?

b. ¿Qué información se muestra en la gráfica circular que no se puede mostrar en la gráfica de barras?

c. "Hablando generalmente, la gráfica de barras resulta ser una opción más recomendable que la gráfica circular." Justifique este enunciado.

2.4 La forma en que los estadounidenses prefieren comer una manzana se informó en *USA Today,* 11 de noviembre de 2004: morderla 47%, cortarla en rebanadas 39%, pelarla 11%, no saben 3%.

a. Construya una gráfica circular que muestre la forma en la que los estadounidenses prefieren comer una manzana.

b. Elabore una gráfica de barras que muestre la forma en la que los estadounidenses prefieren comer una manzana.

c. En su opinión, respecto a la gráfica circular del inciso a) o la gráfica de barras del inciso b), ¿resultan en una mejor representación de la información? Explique.

2.5 La Asociación Norteamericana de Nóminas obtuvo una gran respuesta a esta pregunta acerca del código de vestido de la compañía: "El actual código de vestido en mi compañía es…" Resultados finales:

a. Un poco relajado 27%

b. Un poco formal 15%

c. Está bien 58%

La mayoría de personas mencionaron la importancia del "confort" en sus explicaciones. La inmensa mayoría de las personas que respondieron a la encuesta estaban muy contentos con el código o política de vestido de su compañía.

a. Construya una gráfica circular que describa esta información. Aplique todas las leyendas.

b. Construya una gráfica de barras que describa esta misma información. Aplique todas las etiquetas.

c. Compare las dos gráficas previas, describiendo lo que usted ve en cada una ahora que las gráficas han sido trazadas y se les han aplicado todas sus leyendas. ¿Tiene usted la misma impresión acerca de las ideas de estas personas a partir de ambas gráficas? ¿Una de ellas destaca algo que no haga la otra?

2.6 La American Community Survey está limitada a la población familiar y excluye la población que viva en instituciones, dormitorios de universidades, y otros alojamientos en grupo. La composición de familias de Montana en 2003 es como sigue:

	Población familiar
Jefe de familia	374 879
Esposa	197 379
Hijo	243 609
Otros familiares	27 583
No familiares	49 047
Total	892 497

Fuente: U.S. Census Bureau

a. Construya una gráfica circular de este desglose.

b. Construya una gráfica de barras de este desglose.

c. Compare las dos gráficas que usted construyó en las partes a y b. ¿Cuál parece ser más informativa? Explique por qué.

2.7 A continuación aparece el número de puntos anotados por los equipos ganadores el 2 de noviembre de 2004, la noche de inauguración de la temporada 2005-2005 de la NBA.

Equipo	Detroit	Dallas	LA Lakers
Marcador	87	107	89

Fuente: http://www.nba.com/schedules/2004_2005_game_schedule/November.html#scheds

a. Trace una gráfica de barras de estos marcadores usando una escala vertical que vaya de 80 a 110.

b. Elabore una gráfica de barras de los marcadores usando una escala vertical que vaya de 50 a 110.

c. ¿En cuál gráfica de barras aparece que los marcadores de la NBA varían más? ¿Por qué?

d. ¿Cómo podría usted crear una representación precisa del tamaño relativo y variación entre estos marcadores?

2.8 Se identificó una muestra de vehículos propiedad de estudiantes, producidos por General Motors, con la marca de cada uno. A continuación veamos la muestra resultante (Ch = Chevrolet, P = Pontiac, B = Buick, O = Oldsmobile, Ca = Cadillac, G = GMC):

Ch	Ch	Ca	P	P	Ca	P	Ch	O	B
P	Ch	Ch	B	P	Ch	Ch	P	Ch	B
B	P	B	G	Ch	Ch	Ch	B	O	Ch
B	G	Ch	P	Ca	G	B	B	Ch	Ch
O	Ch	Ch	Ch	Ch	O	O	Ca	B	G

a. Encuentre el número de automóviles de cada marca de la muestra.

b. ¿Qué porcentaje de estos automóviles fueron Chevrolet? ¿Pontiac? ¿Oldsmobile? ¿Buick? ¿Cadillac? ¿GMC?

c. Trace una gráfica de barras que muestre los porcentajes encontrados en el inciso b).

2.9 La cantidad de personas, por grupo de edades, que en 2004 vivían en los 50 estados de Estados Unidos y el Distrito de Columbia, es la siguiente.

Grupo de edades	Número (millones)	Grupo de edades	Número (millones
0-17	73.45	35-45	66.62
18-24	28.86	50+	84.12
25-34	39.89		

Fuente: Sales & Marketing Management Survey of Buying Power, septiembre de 2004 para los 50 estados de E.U. y el Distrito de Columbia.

Elabore una gráfica de barras que muestre el número de personas por grupos de edad.

2.10 Las opciones limpieza de mostradores, desinfección de superficies e higiene personal están entre los usos favoritos de consumidores para productos de limpieza, según la última encuesta nacional de limpieza de la Soap and Detergent Association (SDA). La International Communications Research (ICR) completó el estudio independiente de investigación de mercado en diciembre de 2004. La pregunta inicial del estudio fue hecha a 1021 adultos estadounidenses (509 hombres y 512 mujeres).

Pregunta formulada: ¿alguna vez ha usado usted algún tipo de material para limpieza, desinfectante o bactericida?

Resultados: Sí: 66%

No: 34%

La cantidad de mujeres que han usado alguna vez un artículo de limpieza (72%) es mayor a la de los hombres (60%).

Consulte en http://www.cleaning101.com/whatsnew/01-17-05.html.

a. Construya y ponga todas las etiquetas necesarias a una gráfica de barras que muestre los resultados de todos los adultos encuestados.

b. Construya y coloque todas las etiquetas necesarias a una gráfica de barras que muestre los resultados que comparen a mujeres y hombres por separado.

c. Analice las gráficas de los incisos a y b, asegurándose de comentar sobre la precisión, o falta de ésta, en que las gráficas presentan la información.

2.11 Un inspector de camisas en una fábrica de prendas de vestir clasifica los últimos 500 defectos como sigue: 67 con botón faltante, 153 presentan una costura mal hecha, 258 no tienen la medida adecuada, 22 presentan defectos en el tejido. Construya un diagrama de Pareto para esta información.

2.12 ¡Qué es lo que NO se debe comprar el día de San Valentín (fiesta del amor y la amistad)!

REGALOS NO DESEADOS

Cuando se trate de regalos del día de San Valentín, los estadounidenses adultos dicen que prefieren NO recibir ositos de peluche.

Datos de Anne R. Carey y Juan Thomassie, USA Today

a. Elabore una gráfica de barras que describa los porcentajes de "Regalos no Deseados".

b. Trace un diagrama de Pareto que descrita los "Regalos no Deseados".

c. Si usted desea estar 80% seguro de no recibir algo no deseado el día de San Valentín, ¿qué debe evitar comprar? ¿Cómo es que el diagrama de Pareto representa lo anterior?

d. Si se van a aplicar una encuesta a 300 adultos, ¿qué frecuencias esperaría que se presentan para cada artículo no deseado de la gráfica?

2.13 Un estudio elaborado por Bruskin-Goldring para la Whirlpool Corp. presenta una lista de las principales tareas domésticas que las amas de casa están deseosas de recibir ayuda. La respuesta más popular fue limpieza (53%), seguida de lavandería (18%), cocinar (9%), lavar loza (8%), y otras (12%).

Fuente: http://pqasb.pqarchiver.com/USAToday/

a. Construya un diagrama de Pareto que muestre esta información.

b. Debido al tamaño de la categoría de "otras", el diagrama de Pareto puede no ser la mejor gráfica a usar. Explique por qué, y describa qué información adicional es necesaria para hacer más apropiado el diagrama de Pareto.

2.14 La Office of Aviation Enforcement and Proceedings, U.S. Department of Transportation, publica la tabla siguiente que presenta el número de quejas de consumidores contra las principales líneas aéreas de Estados Unidos, por categoría de queja.

Categoría de queja	Número	Categoría de queja	Número
Publicidad	68	Problemas en vuelos	2031
Equipaje	1421	Sobreventa	454
Servicio a clientes	1715	Devoluciones	1106
Incapacidad	477	Reservaciones/boletos/abordar	1159
Tarifas	523	Otras	322

Fuente: Office of Aviation Enforcement and Proceedings, U.S. Department of Transportation, Air Travel Consumer Report, http://www.infoplease.com/ipa/A0198353.html

a. Construya un diagrama de Pareto que describa esta información.

b. ¿En cuáles quejas recomendaría usted a aerolíneas poner más atención para corregir, si desean tener el mejor efecto en el número total de quejas? Explique cómo es que el diagrama de Pareto de la parte a demuestra la validez de su respuesta.

2.15 El reporte de defectos de inspección final para la línea de producción A12 se presenta en un diagrama de Pareto.

a. ¿Cuál es la cantidad total de defectos del informe?

b. Verifique el 30% indicado por "arañazo."

Defecto	Defecto	Arañazo	Rebaba	Doblez	Abollado	Otros
Cantidad	56	45	23	12	8	6
Porcentaje	37.3	30.0	15.3	8.0	5.3	4.0
% acum.	37.3	67.3	82.7	90.7	96.0	100.0

c. Explique cómo se obtuvo el valor de "90.7% acum. por doblez" e indique lo que significa.

d. La administración ha notificado a la línea de producción la meta de reducir en 50% los defectos. ¿Cuáles son los dos defectos que según usted deben recibir especial atención para trabajar hacia esta meta? Explique.

2.16 Se pronosticó que la producción mundial de cacao en 2004-2005 iría a la baja después del récord de 3 396 toneladas métricas obtenidas en 2003-2004. Se espera que casi toda la producción se presente en África Occidental, con la producción en Asia y América Latina relativamente sin cambio. África Occidental seguirá siendo el líder en la producción mundial de cacao en 2005.

Producción mundial de cacao (1000 toneladas)

País	2003-2004	2004-2005 (pronóstico)
Brasil	163	164
Camerún	150	150
Ecuador	100	93
Ghana	605	530
Indonesia	420	415
Costa de Marfil	1500	1275
Malasia	25	26
Nigeria	165	170
Otros en África	38	39
Otros en América	165	170
Otros en Asia	65	67

Fuente: World Cocoa Foundation, http://www.chocolateandcocoa.org/stats/supply/default.asp

a. ¿En cuánto se espera que disminuya la producción total? ¿Qué porcentaje de reducción es éste?

b. Elabore un diagrama de Pareto que describa la producción 2003-2004.

c. Construya un diagrama de Pareto que describa el pronóstico de la producción 2004-2005.

d. Se espera que los países africanos suministren ¿qué porcentaje del cacao del mundo para 2004-2005? ¿Los de América? ¿Los de Asia?

2.17 El número de puntos anotados durante cada juego por un equipo de baloncesto de preparatoria la temporada pasada fue como sigue: 56, 54, 61, 71, 46, 61, 55, 68, 60, 66, 54, 61, 52, 36, 64, 51. Construya una gráfica de puntos de estos datos.

2.18 La tabla siguiente enumera la mediana de precios de venta de casas (en miles de dólares) para 20 ciudades de Estados Unidos, como los presenta Realtor.com.

120	120	120	119	117	117	116	116	115	115
114	114	112	112	112	111	109	107	105	105

Fuente: http://www.realtor.org/Research.nsf/pages/EconHousingData

a. Construya una gráfica de puntos de estos datos.

b. Describa la distribución exhibida por la gráfica de puntos encontrada en la parte a.

2.19 HoopsHype.com regularmente publica lo más reciente de la NBA. A continuación están las estaturas (en pulgadas) de los jugadores de baloncesto seleccionados en la primera ronda por los equipos profesionales el 24 de junio, 2004:

82	82	74	79	75	79	80	83	78	79
83	85	71	81	81	78	80	78	79	72
89	81	80	74	76	79	78	75	84	

Fuente: http://www.hoopshype.com/draft.htm

a. Construya una gráfica de puntos de las estaturas de estos jugadores.

b. Utilice la gráfica de puntos para descubrir los jugadores de menos y de más estatura.

c. ¿Cuál es la estatura más común, y cuántos jugadores comparten esa estatura?

d. ¿Qué característica de la gráfica de puntos ilustra la estatura más común?

2.20 Como jugadores de béisbol, Babe Ruth y Hank Aaron fueron bien conocidos por su habilidad para conectar *home-runs* (cuadrangulares). Mark McGwire y Sammy Sosa fueron bien conocidos por su habilidad para conectar *home-runs* durante la "gran persecución del *home-run*" de 1998. Bobby Bonds ganó su fama en 2001. A continuación aparece el número de cuadrangulares que conectó cada jugador en cada temporada de ligas mayores en que jugó.

Ejercicio 2.20

Ruth	4	3	2	11	29	54	59	35	41	46	25	47	60	54	46	49	46	41	34	22	6		
Aar	13	27	26	44	30	39	40	34	45	44	24	32	44	39	29	44	38	47	34	40	20	12	10
McGwire	3	49	32	33	39	22	42	9	9	39	52	58	70	65	32	29							
Sosa	4	15	10	8	33	25	36	40	36	66	63	50	64	49	40	35							
Bonds	16	25	24	19	33	25	34	46	37	33	42	40	37	34	49	73	46	45	45				

a. Construya una gráfica de puntos para los datos de Ruth y Aaron, usando el mismo eje.

b. Con el uso de las gráficas de puntos encontradas en la parte a, plantee un caso para cada uno de los enunciados siguientes respecto a jugadores del pasado: "¡Aaron es el rey del *home-run*!" "¡Ruth es el rey del *home-run*!"

c. Construya una gráfica de puntos de los datos para McGwire, Sosa y Bonds usando el mismo eje.

d. Usando las gráficas de puntos encontradas en la parte c, realice un caso para los enunciados "¡McGwire es el rey del *home-run*!" "¡Bonds es el rey del *home-run*!" y "¡Sosa no es ahora el rey del *home-run*!" respecto a los jugadores actuales. ¿En qué forma apoyan las gráficas de puntos a cada enunciado?

PARA SU INFORMACION Si usted usa su computadora como calculadora, use los comandos de la página 49.

2.21 Delco Products, división de la General Motors, produce conmutadores diseñados para medir 18.810 mm de longitud total. (Un conmutador es un dispositivo que se utiliza en el sistema eléctrico de un automóvil.) La siguiente muestra de 35 longitudes de conmutadores se tomó al hacer vigilancia del proceso de manufactura:

18.802	18.810	18.780	18.757	18.824	18.827	18.825
18.809	18.794	18.787	18.844	18.824	18.829	18.817
18.785	18.747	18.802	18.826	18.810	18.802	18.780
18.830	18.874	18.836	18.758	18.813	18.844	18.861
18.824	18.835	18.794	18.853	18.823	18.863	18.808

Fuente: Con permiso de Delco Products División, GMC

Use una computadora para construir una gráfica de puntos de estos datos.

2.22 Se empleó una computadora para construir la gráfica de puntos de la parte inferior de esta página.

a. ¿Cuántos datos se muestran?

b. Escriba los valores de los cinco datos más pequeños.

c. ¿Cuál es el valor del mayor renglón de los datos?

d. ¿Qué valor se presenta el mayor número de veces? ¿Cuántas veces se presentó?

2.23 Construya un diagrama de tallo y hoja del número de puntos anotados durante cada juego de baloncesto la última temporada:

56	54	61	71	46	61	55	68
60	66	54	61	52	36	64	51

2.24 Forbes.com publicó las devoluciones totales de un periodo de 5 años (2000-2004), en porcentajes, para 17 compañías de la industria bancaria.

Nombre	% de devolución	Nombre	% de devolución
Astoria Financial	23.9	Popular	15.3
Banknorth Group	18.6	State Street	5.0
Bank of America	13.2	Synovus Finl	8.5
BB&T	8.1	UnionBanCal	10.4
Compass Bancshares	16.7	Wachovia	10.5
Golden West Finl	29.1	Wells Fargo	8.8
M&T Bank	19.1	Westcorp	25.9
National City	12.6	Zions Bancorp	2.9
North Fork Bancorp	19.8		

Fuente: http://www.forbes.com/lists/results.jhtml

a. Construya un diagrama de tallo y hoja de los datos.

b. Con base en el diagrama de tallo y hoja, describa la distribución de porcentajes de rentabilidad.

2.25 Las cantidades que se muestran a continuación son las tarifas cobradas por Quik Delivery para los 40 paquetes pequeños que entregó el jueves pasado por la tarde:

4.03	3.56	3.10	6.04	5.62	3.16	2.93	3.82	4.30	3.86
4.57	3.59	4.57	6.16	2.88	5.03	5.46	3.87	6.81	4.91
3.62	3.62	3.80	3.70	4.15	2.07	3.77	5.77	7.86	4.63
4.81	2.86	5.02	5.24	4.02	5.44	4.65	3.89	4.00	2.99

a. Construya un diagrama de tallo y hoja.

b. Con base en el diagrama de tallo y hoja, describa la distribución de los datos.

Figura para ejercicio 2.22

2.26 Una de las muchas cosas de las que la U.S. Census Bureau informa al público es el aumento en población para diversas zonas geográficas dentro del país. Los porcentajes de aumento en población para los 24 condados de más rápido crecimiento en Estados Unidos, del 1 de abril de 2000 al 1 de julio de 2003, aparecen en una tabla que se publica en la siguiente página de Internet: http://www.census.gov/popest/countis/CD-EST2003-09.html.

Condado, Estado	%
Condado de Collin, TX	21.4

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.census.gov/popest/countis/CO-EST2003-09.html

a. Construya un diagrama de tallo y hoja.

b. Con base en el diagrama de tallo y hoja, describa la distribución de los datos.

2.27 Dado el siguiente diagrama de tallo y hoja:

```
Stem-and-Leaf of C1  N = 16
Leaf Unit = 0.010
 1       59   7
 4       60   148
(5)      61   02669
 7       62   0247
 3       63   58
 1       64   3
```

a. ¿Qué significa "Leaf Unit = 0.010"?

b. ¿Cuántos datos se muestran en este diagrama de tallo y hoja?

c. Haga una lista de los primeros cuatro datos.

d. ¿Cuál es la columna de números hacia abajo, a la izquierda de la figura?

2.28 Un término que con frecuencia se usa en investigación de energía solar es *calentamiento-grado-días.* Este concepto está relacionado con la diferencia entre la temperatura de 65°F de interiores y el promedio de temperatura a la intemperie para un día determinado. Un promedio de temperatura a la intemperie de 5°F da 60 calentamiento-grado-días. Los valores normales de calentamiento-grado-días, para varios lugares de Nebraska, se muestran en el siguiente diagrama de tallo y hoja construido usando el MINITAB.

a. ¿Cuál es el significa do de "Leaf Unit = 10"?

b. Haga una lista de los primeros cuatro datos.

c. Haga una lista de todos los datos que se presentaron más de una vez.

```
Stem-and-leaf of C1  N=25
Leaf Unit=10
     2      60  78
     7      61  03699
     9      62  69
    11      63  26
   (3)      64  233
    11      65  48
     9      66  8
     8      67  249
     5      68  18
     3      69  145
```

2.3 Distribuciones de frecuencias e histogramas

Las listas de grandes conjuntos de datos no presentan una imagen útil de los datos. A veces deseamos condensar los datos en una forma más manejable; esto puede lograrse con ayuda de una *distribución de frecuencia.*

Distribución de frecuencia: es una lista que se expresa con cierta frecuencia en forma de gráfica, que enlaza pares de valores de una variable con su frecuencia.

Para demostrar el concepto de una distribución de frecuencia, usemos este conjunto de datos:

3	2	2	3	2	4	4	1	2	2
4	3	2	0	2	2	1	3	3	1

TABLA 2.4

Distribución de frecuencia no agrupada

x	f
0	1
1	3
2	8
3	5
4	3

Si con x representamos la variable, entonces podemos usar una distribución de frecuencia para representar este conjunto de datos con una lista de los valores x con sus frecuencias. Por ejemplo, el valor 1 se presenta tres veces en la muestra; por tanto, la **frecuencia** para $x = 1$ es 3. El conjunto completo de datos se muestra en la distribución de frecuencia de la tabla 2.4.

La frecuencia, f, es el número de veces que el valor x se presenta en la muestra. La tabla 2.4 muestra una *distribución de frecuencia no agrupada* —"no agrupada" porque cada valor de x de la distribución se encuentra solo—. Cuando un gran conjunto de datos tiene muchos valores x diferentes en lugar de unos pocos valores repetidos, como en el ejemplo previo, podemos agrupar los valores en un conjunto de clases y construir una *distribución de frecuencia agrupada.* El diagrama de tallo y hoja de la figura 2.5B (p. 46) presenta, en forma gráfica, una distribución de frecuencia agrupada. Cada tallo representa una clase. El número de hojas de cada tallo es el mismo que la frecuencia para la misma **clase** (a veces llamada *depósito*). Los datos representados en la figura 2.5B aparecen como distribución de frecuencia agrupada en la tabla 2.5.

TABLA 2.5

Distribución de frecuencia agrupada

		Clase	Frecuencia
50 o más a menos de 60	⟶	$50 \leq x < 60$	1
60 o más a menos de 70	⟶	$60 \leq x < 70$	3
70 o más a menos de 80	⟶	$70 \leq x < 80$	8
80 o más a menos de 90	⟶	$80 \leq x < 90$	5
90 o más a menos de 100	⟶	$90 \leq x < 100$	2
			19

El proceso de tallo y hoja se puede usar para construir una distribución de frecuencia; no obstante, la representación del tallo no es compatible con todos los **anchos de clase.** Por ejemplo, los anchos de clase de 3, 4 y 7 son difíciles de usar. Así, a veces es ventajoso tener un procedimiento separado para construir una distribución de frecuencia agrupada.

EJEMPLO 2.6

Agrupación de datos para formar una distribución de frecuencia

Para ilustrar este procedimiento de agrupación o clasificación, usemos una muestra de 50 calificaciones de un examen final tomadas del grupo (o clase) de estadística elemental del último semestre. La tabla 2.6 muestra las 50 calificaciones.

Procedimiento para construir una distribución de frecuencia agrupada

1. Identifique la calificación alta ($H = 98$) y la calificación baja ($L = 39$), y encuentre el rango:

$$\text{rango} = H - L = 98 - 39 = 59$$

2. Seleccione un número de grupo ($m = 7$) y un ancho de grupos ($c = 10$) de modo que el producto ($mc = 70$) sea un poco mayor que el rango (rango = 59).

TABLA 2.6

Calificaciones de exámenes de estadística

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

3. Seleccione un punto inicial; éste debe ser un poco menor que la calificación más baja, L. Supóngase que empezamos en 35; continuando ahí por decenas (el ancho del grupo) obtenemos 35, 45, 55, 65,...95, 105. Éstos reciben el nombre de **fronteras de grupo.** Los grupos para los datos de la tabla 2.6 son:

35 o más a menos de 45	$\longrightarrow$	$35 \leq x < 45$
45 o más a menos de 55	$\longrightarrow$	$45 \leq x < 55$
55 o más a menos de 65	$\longrightarrow$	$55 \leq x < 65$
65 o más a menos de 75	$\longrightarrow$	$65 \leq x < 75$
	$\vdots$	$75 \leq x < 85$
		$85 \leq x < 95$
95 o más hasta e incluyendo 105	$\longrightarrow$	$95 \leq x \leq 105$

Notas:

1. De una mirada se puede comprobar el patrón numérico para determinar si la aritmética empleada para formar los grupos fue correcta (35, 45, 55, ..., 105).
2. Para el intervalo $35 \leq x < 45$, el 35 es la frontera inferior del grupo y 45 es la frontera superior del grupo. Las observaciones que caen en la frontera más baja del grupo permanecen en ese intervalo; las que caen en la frontera superior del grupo entran en el siguiente intervalo más alto, excepto para el último grupo.
3. El ancho del grupo es la diferencia entre las fronteras superior e inferior del grupo.
4. Numerosas combinaciones de anchos de grupo, números de grupos y puntos iniciales son posibles cuando se clasifican datos. No hay una opción mejor que otra. Pruebe unas cuantas combinaciones diferentes y con sentido común decida cuál de ellas usar.

Por lo tanto, se emplean los siguientes **principios básicos** para construir una distribución de frecuencia agrupada:

1. Cada grupo debe tener el mismo ancho.
2. Los grupos (a veces llamados *depósitos*) deben establecerse de modo que no se traslapen y que cada dato pertenezca a exactamente un grupo.
3. Para los ejercicios dados en este libro, de 5 a 12 grupos son más deseables porque todas las muestras contienen menos de 125 datos. (La raíz cuadrada de n es un principio razonable para el número de grupos con muestras de menos de 125 datos.)
4. Utilice un sistema que aplique las ventajas de un patrón numérico para garantizar precisión.
5. Cuando sea conveniente, con frecuencia es ventajoso un ancho de grupo constante.

Una vez establecidos los grupos, es necesario ordenar los datos en esos grupos. El método que se usa para ordenar dependerá del formato actual de los datos: si los

datos están ordenados por rangos, las frecuencias pueden contarse; si no lo están, **registraremos** los datos para hallar los números de frecuencia. Cuando se clasifiquen datos, es útil usar una tabla estándar (ver la tabla 2.7).

TABLA 2.7

Tabla estándar para distribución de frecuencia

Número de grupo	Sumas de grupo	Fronteras	Frecuencia
1	||	$35 \leq x < 45$	2
2	||	$45 \leq x < 55$	2
3	||||| ||	$55 \leq x < 65$	7
4	||||| ||||| |||	$65 \leq x < 75$	13
5	||||| ||||| |	$75 \leq x < 85$	11
6	||||| ||||| |	$85 \leq x < 95$	11
7	||||	$95 \leq x \leq 105$	4
			50

Notas:

1. Si los datos han sido ordenados por rangos (en forma de lista, gráfica de puntos o tallo y hoja), no es necesario registrar; simplemente se cuentan los datos que pertenezcan a cada grupo.
2. Si los datos no han sido ordenados por rangos, al registrar proceda con cuidado.
3. La frecuencia, f, para cada grupo es el número de datos que pertenecen a ese grupo.
4. La suma de las frecuencias debe ser igual al número de partes de datos, n ($n = \Sigma f$). Esta sumatoria sirve como una buena prueba.

Nota: vea en el *Manual de Soluciones para el Estudiante* información acerca de la **notación** Σ (léase "**notación de sumatoria**")

TABLA 2.8

Distribución de frecuencia con puntos medios de grupo

Número de grupo	Fronteras de grupo	Frecuencia, f	Puntos medios de grupo, x
1	$35 \leq x < 45$	2	40
2	$45 \leq x < 55$	2	50
3	$55 \leq x < 65$	7	60
4	$65 \leq x < 75$	13	70
5	$75 \leq x < 85$	11	80
6	$85 \leq x < 95$	11	90
7	$95 \leq x \leq 105$	4	100
		50	

Nota: ahora puede usted ver por qué es útil tener un ancho de clase constante. Un ancho impar de grupo hubiera resultado en un punto medio de grupo con un dígito extra. (Por ejemplo, el grupo 45-54 es de ancho 9 y el punto medio de grupo es 49.5.)

Cada grupo necesita un solo valor numérico para representar todos los valores de datos que caigan en ese grupo. El **punto medio de grupo** (a veces llamado

marca de grupo) es el valor numérico que está exactamente en la mitad de cada grupo. Se encuentra al sumar las fronteras de grupo y dividiendo entre 2. La tabla 2.8 muestra una columna adicional para el punto medio de grupo, x. Como prueba de su aritmética, los puntos medios sucesivos de grupo deben ser un ancho de grupo aparte, que es 10 en esta ilustración (40, 50, 60,…,100 es un patrón reconocible).

CASO PRÁCTICO 2.7

Limpieza de casas

La gráfica de "Horas por semana dedicadas a limpiar la casa" presenta una versión de gráfica circular de una distribución de frecuencia. Cada sector del círculo representa el tiempo empleado en limpieza semanal por cada persona, y el "tamaño relativo" del sector representa el porcentaje o frecuencia relativa.

HORAS A LA SEMANA DEDICADAS A LIMPIAR LA CASA

Los estadounidenses pasan en promedio 3.4 horas por semana limpiando sus casas. ¿Cuánto tiempo se pasa en la limpieza semanal?

Datos de Cindy Hall y Sam Ward, USA TODAY; Fuente: Yankelovich Partners for GCI/ZEP Chemicals.

Ahora, usando terminología de estadísticas, podemos decir que la *variable* "tiempo empleado en limpieza" está representada en la gráfica por sectores del círculo. La *frecuencia relativa* está representada por el tamaño del ángulo que forma el sector. Para mostrar esta información en una distribución de frecuencia agrupada "relativa", cada intervalo de la variable se expresará en la forma $a \leq x < b$. Por ejemplo, la categoría de 2 a 4 horas se expresaría $2 \leq x < 4$. (De este modo, la frontera inferior es parte del intervalo, pero la frontera superior es parte del siguiente intervalo más grande.) La tabla de distribución para esta gráfica circular aparecería entonces como se muestra enseguida.

Fronteras de grupo	Frecuencia relativa
$0 \leq x < 1$	0.05
$1 \leq x < 2$	0.20
$2 \leq x < 4$	0.33
$4 \leq x$	0.39
No sabe	0.03

Cuando clasificamos datos en grupos, perdemos parte de la información. Sólo cuando tenemos todos los datos sin procesar conocemos los valores exactos que en realidad fueron observados para cada grupo. Por ejemplo, ponemos un 47 y un 50 en el grupo 2, con fronteras de grupo de 45 y 55. Una vez colocados en el grupo, sus valores se pierden para nosotros y empleados el punto medio del grupo, 50, como su valor representativo.

Histograma: es una gráfica de barras que representa una distribución de frecuencia de una variable cuantitativa. Un histograma consta de los siguientes componentes:

1. Un título, que identifica la población o muestra de interés.
2. Una escala vertical, que identifica las frecuencias de los diversos grupos.
3. Una escala horizontal, que identifica la variable x. A lo largo del eje x pueden marcarse valores para las fronteras del grupo o puntos medios de grupo. Utilice el método de marcar el eje que mejor presente la variable.

La distribución de frecuencias de la tabla 2.8 aparece en forma de histograma en la figura 2.10.

En ocasiones es importante la **frecuencia relativa** de un valor. La frecuencia relativa es una medida proporcional de la frecuencia para que ocurra un evento. Se encuentra al dividir la frecuencia del grupo entre el número total de observaciones. La frecuencia relativa puede expresarse como una fracción común, en forma decimal, o como porcentaje. Como se puede observar en el ejemplo 2.6, la frecuencia asociada con el tercer grupo (55-65) es 7. La frecuencia relativa para el tercer grupo es $\frac{7}{50}$, o 0.14, o 14%. A veces son útiles las frecuencias relativas en una presentación porque casi todas las personas entienden partes fraccionarias cuando se expresan como porcentajes. Las frecuencias relativas son particularmente útiles cuando se comparan distribuciones de frecuencia de dos conjuntos de datos de diferentes tamaños. La figura 2.11 muestra un **histograma de frecuencia relativa** de la muestra de las 50 calificaciones de examen final de la tabla 2.8.

PARA SU INFORMACIÓN
Observe que el histograma de frecuencia y el histograma de frecuencias relativas tienen la misma forma (suponiendo que para ambos se usen los mismos grupos); sólo cambian las leyendas del eje vertical.

PARA SU INFORMACIÓN
Asegúrese de identificar ambas escalas de tal forma que el histograma presente la historia completa.

FIGURA 2.10 Histograma de frecuencia

FIGURA 2.11 Histograma de frecuencia relativa

Un diagrama de tallo y hoja contiene toda la información necesaria para crear un histograma. La figura 2.5 B (p. 46) muestra el diagrama de tallo y hoja que se generó en el ejemplo 2.4. En la figura 2.12 A el diagrama de tallo y hoja ha girado 90° y se han agregado leyendas para mostrar su relación con un histograma. La figura 2.12B ilustra el mismo conjunto de datos como un histograma completo.

FIGURA 2.12A Diagrama de tallo y hoja modificado

FIGURA 2.12B Histograma

INSTRUCCIONES DE TECNOLOGÍA: HISTOGRAMA

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:      Graph > Histogram > Simple > OK
Enter:       Graph variables:C1
Choose:      Labels > Titles/Footnote
Enter:       Your title and/or footnote > OK
Choose:      Scale > Y-Scale Type
Select:      Y scale Type: Frequency or Percent or Density > OK > OK
To adjust histogram: Double click anywhere on bars of histogram.
Select:      Binning
Select:      Interval Type: Midpoint or Cutpoint
             Interval De?nitions:
               Automatic or. Number of intervals; Enter: N or, Midpt/cutpt
               positions; Enter: A:B/C > OK
```

Notas:

1. Los puntos medios son los puntos medios del grupo, y los puntos de corte son las fronteras del grupo.
2. El porcentaje es frecuencia relativa.
3. Automático significa que MINITAB hará todas las selecciones; N = número de intervalos, es decir, el número de grupos que usted desea se usen.
4. A = punto medio más pequeño de grupo o frontera, B = punto medio más grande de grupo o frontera, C = ancho del grupo que usted desea especificar.

Los comandos siguientes dibujarán el histograma de una distribución de frecuencia. Los grupos de extremo se pueden hacer de ancho completo al agregar un grupo extra con frecuencia cero a cada extremo de la distribución de frecuencia. Introduzca los puntos medios del grupo en C1 y las frecuencias correspondientes en C2.

```
Choose:      Graph > Scatterplot > With Connect Line > OK
Enter:       Y variables: C2   X variables: C1
Select:      Data View: Data Display: Symbols Connect > OK > OK
Double click on a connect line.
Select:      Options
             Connection Function: Step > OK
```

Excel Introduzca los datos en la columna A y los límites* superiores de grupo en la columna B (opcional) y (encabezados de columna son opcionales); luego continúe con:

```
Choose:      Tools > Data Analysis† > Histogram > OK
Enter:       Input Range: Data (A1:A6 or select cells)
             Bin Range: upper class limits (B1:B6 or select cells)
             [leave blank if Excel determines the intervals]
Select:      Labels (if column headings are used)
             Output Range
Enter:       area for freq. distr. & graph (C1 or select cell)
Select:      Chart Output
```

Para eliminar brechas entre barras:

```
Click on:    Any bar on graph
Click on:    Right mouse button
Choose:      Format Data Series > Options
Enter:       Gap Width: 0
```

Para editar el histograma:

```
Click on:    Anywhere clear on the chart
              —use handles to size
             Any title or axis name to change
             Any upper class limit§ or frequency in the frequency distribu-
             tion to change value > Enter
```

*Si frontera = 50, entonces límite = 49.9 (dependiendo del número de lugares decimales de los datos).

§ Si Data Analysis (análisis de datos) no aparece en el menú Tools (Herramientas):

Elija:	Tools > Add-Ins
Seleccione:	Analysis ToolPak
	Analysis ToolPak-VBA

§ Observe que los límites superiores de grupo aparecen en el centro de las barras. Sustituya con puntos medios de grupo. La celda "More" (Más) de la distribución de frecuencia también se puede borrar.

Para datos en tablas, introduzca los grupos en la columna A (ej. 30-40) y las frecuencias en la columna B; luego continúe con:

```
Choose:      Chart Wizard > Column > 1st picture (usually) > Next
Enter:       Data Range: (A1:B4 or select cells)
Select:      Series in: Columns > Next
Choose:      Titles
Enter:       Chart title: your title
             Category (x) axis: title for x-axis
             Value (y) axis: title for y-axis > Next > Finish
```

Realice este trabajo exactamente como se describe para eliminar brechas y ajustar.

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:      2nd > STAT PLOT > 1:Plot1
```

La calculadora selecciona grupos:

```
Choose: Zoom > 9:ZoomStat > Trace > > >
```

El usuario selecciona grupos:

```
Choose:      Window
Enter:       at most lowest value, at least highest value, class width, —1,
             at least highest frequency, 1 (depends on frequency numbers), 1
Choose:      Graph > Trace (use values to construct frequency distribution)
```

Para datos en tablas, introduzca los puntos medios de grupo en L1 y las frecuencias en L2; luego continúe con:

```
Choose:      2nd > STAT PLOT > 1:Plot1
Choose:      Window
Enter:       smallest lower class boundary,
             largest upper class boundary, class
             width, —ymax/4, highest frequency,
             0 (for no tick marks), 1
Choose:      Graph > Trace > > >
```

En cambio, para obtener un histograma de frecuencia relativa de datos en tabla:

```
Choose:      STAT > EDIT > 1:EDIT...
Highlight:   L3
Enter:       L3 = L2/SUM(L2) [SUM - 2nd LIST > MATH > 5:sum]
Choose:      2nd > STAT PLOT > 1:Plot1
Choose:      Window
Enter:       smallest lower class boundary,
             largest upper class boundary,
             class width, −ymax/4, highest rel.
             frequency, 0 (for no tick marks), 1
Choose:      Graph > Trace > > >
```

Los histogramas son herramientas valiosas. Por ejemplo, el histograma de una muestra debe tener una forma de distribución muy semejante a la de la población de la cual se extrajo la muestra. Si el lector de un histograma está familiarizado con la variable de que se trate, él o ella por lo general podrán interpretar diversos datos importantes. La figura 2.13 presenta histogramas con formas específicas que sugieren leyendas descriptivas. Las posibles leyendas descriptivas se indican bajo cada histograma.

Brevemente, los términos empleados para describir histogramas son como sigue:

Simétrico: ambos lados de esta distribución son idénticos (las mitades son imágenes de espejo).

Normal: una distribución simétrica se agrupa alrededor de la media y se hace escasa en los extremos. (Más adelante veremos propiedades adicionales.)

Uniforme (rectangular): todo valor aparece con la misma frecuencia.

FIGURA 2.13 Formas de histogramas

Sesgado: una cola está estirada más que la otra. La dirección de sesgo está en el lado de la cola más larga.

En forma de J: no existe cola en el lado del grupo con la frecuencia más alta.

Bimodal: los dos grupos de mayor población están separados por uno o más grupos. Esta situación con frecuencia implica que dos poblaciones se estén muestreando. (Vea la figura 2.7, p. 48.)

Notas:

1. La **moda** es el valor de los datos que se presenta con mayor frecuencia. (El modo se estudiará en la sección 2.4, p. 76.)
2. La **clase modal** es la clase con la frecuencia más alta.
3. Una **distribución bimodal** tiene dos grupos de mayor frecuencia separados por grupos con frecuencias más bajas. No es necesario que las dos frecuencias más elevadas sean iguales.

Otra forma de expresar una distribución de frecuencia es usar una *distribución acumulativa de frecuencia.*

Distribución acumulativa de frecuencia: distribución de frecuencia que parea las frecuencias acumulativas con valores de la variable.

La **frecuencia acumulativa** para cualquier grupo determinado es la suma de la frecuencia para ese grupo y las frecuencias de todos los grupos de valores menores. La tabla 2.9 muestra la distribución de frecuencia acumulativa de la tabla 2.8 (p. 58).

TABLA 2.9

Usando distribución de frecuencia para formar una distribución acumulativa de frecuencia

Número de grupo	Fronteras de grupo	Frecuencia,	Frecuencia acumulativa
1	$35 \leq x < 45$	2	2 (2)
2	$45 \leq x < 55$	2	4 (2 + 2)
3	$55 \leq x < 65$	7	11 (7 + 4)
4	$65 \leq x < 75$	13	24 (13 + 11)
5	$75 \leq x < 85$	11	35 (11 + 24)
6	$85 \leq x < 95$	11	46 (11 + 35)
7	$95 \leq x \leq 105$	4	50 (4 + 46)
		50	

Puede presentarse la misma información mediante el uso de una *distribución acumulativa de frecuencia* (ver la tabla 2.10). Esto combina las ideas de frecuencia acumulativa y de frecuencia relativa.

TABLA 2.10

Distribución de frecuencia relativa acumulativa

Número de grupo	Fronteras de grupo	Frecuencia acumulativa relativa	*Las frecuencias acumulativas son para el intervalo de 35 hasta la frontera superior de ese grupo.*
1	$35 \leq x < 45$	2/50, o 0.04	← *más de 35 a menos de 45*
2	$45 \leq x < 55$	4/50, o 0.08	← *más de 35 a menos de 55*
3	$55 \leq x < 65$	11/50, o 0.22	← *más de 35 a menos de 65*
4	$65 \leq x < 75$	24/50, o 0.48	.
5	$75 \leq x < 85$	35/50, o 0.70	:
6	$85 \leq x < 95$	46/50, o 0.92	
7	$95 \leq x \leq 105$	50/50, o 1.00	← *más de 35 hasta 105 inclusive*

Las distribuciones acumulativas se pueden presentar en forma gráfica.

Ojiva: se trata de una gráfica de línea de una frecuencia acumulativa o distribución de frecuencia relativa acumulativa. Una ojiva tiene los componentes siguientes:

1. Un título, que identifica la población o muestra.
2. Una escala vertical, que identifica ya sea las frecuencias acumulativas o las frecuencias acumulativas relativas. (La figura 2.14 muestra una ojiva con frecuencias acumulativas relativas.)
3. Una escala horizontal, que identifica las fronteras superiores de grupo. (Hasta que se haya alcanzado la frontera superior de un grupo, no se puede estar seguro de haber acumulado todos los datos en ese grupo. Por tanto, la escala horizontal para una ojiva está siempre basada en las fronteras superiores de grupo.)

FIGURA 2.14
Ojiva

50 calificaciones de examen final en estadística elemental

Nota: cualquier ojiva se inicia a la izquierda con una frecuencia relativa de cero en la frontera más baja de grupo del primer grupo, y termina a la derecha con una frecuencia acumulativa relativa de 1.00 (o 100%) en la frontera superior de grupo del último grupo.

INSTRUCCIONES DE TECNOLOGÍA: OJIVA

MINITAB (Versión 14) Teclee las fronteras de grupo en C1 y los porcentajes acumulativos en C2 (introduzca 0 [cero] para el porcentaje pareado con la frontera inferior del primer grupo e iguale cada porcentaje acumulativo con la frontera superior de grupo).

Use porcentajes; es decir, use 25% en lugar de 0.25.

```
Choose:     Graph > Scatterplot > With Connect Line > OK
Enter:      Y variables: C2    X variables: C1
Select:     Data View: Data Display: Symbols Connect > OK
Select:     Labels > Titles/Footnotes
Enter:      your title or footnotes > OK > OK
```

Excel Capture los datos en la columna A y los límites* superiores de grupo en la columna B (incluya un grupo adicional al principio).

```
Choose:     Tools > Data Analysis > Histogram > OK
Enter:      Input Range: data (A1:A6 or select cells)
            Bin Range: upper class limits (B1:B6 or select cells)
Select:     Labels (if column headings were used)
            Output Range
            Enter: area for freq. distr. & graph: (C1 or se-
            lect cell)
            Cumulative Percentage
            Chart Output
```

Para cerrar espacios y editar, consulte los comandos del histograma que se presentan en las páginas 61-62.

Para construir la tabla, capture las fronteras superiores de grupo en la columna A y las frecuencias acumulativas relativas en la columna B (incluya una frontera adicional de grupo al principio con una frecuencia acumulativa relativa igual a 0 [cero]); luego continúe con:

```
Choose:     Tools > Data Analysis > Histogram > OK
Enter:      Input Range: data (A1:A6 or select cells)
            Bin Range: upper class limits (B1:B6 or select cells)
Select:     Labels (if column headings were used)
            Output Range
            Enter: area for freq. distr. & graph: (C1 or se-
            lect cell)
            Cumulative Percentage
            Chart Output
```

Para editar, vea los comandos de histograma de la página 62.

*Si la frontera = 50, entonces el límite = 49.9 (dependiendo del número de lugares decimales de los datos).

TI-83/84 Plus Teclee las fronteras del grupo en L1 y las frecuencias en L2 (incluya una frontera extra de grupo al principio con una frecuencia de cero); luego continúe con:

```
Choose:     STAT > EDIT > 1:EDIT...
Highlight:  L3
Enter:      L3 = 2nd > LIST > OPS > 6:cum sum(L2)
Highlight:  L4
Enter:      L4 = L3 / 2nd > LIST > Math > 5:sum (L2)
```

Choose: **2nd > STAT PLOT > 1:Plot**

Choose: **Zoom > 9:ZoomStat > Trace > > >**

Para leerla mejor, ajuste la pantalla si es necesario.

SECCIÓN 2.3 EJERCICIOS

2.29 a. Forme una distribución de frecuencia no agrupada de los datos siguientes:

1, 2, 1, 0, 4, 2, 1, 1, 0, 1, 2, 4

Con referencia a la distribución precedente:

b. Explique lo que representa $f = 5$.

c. ¿Cuál es la suma de la columna de frecuencia?

d. ¿Qué representa esta suma?

2.30 Las gráficas de barras e histogramas no son la misma cosa. Explique sus similitudes y diferencias.

2.31 Los jugadores del equipo de futbol soccer profesional Rinos Furiosos de Rochester anotaron 38 goles durante la temporada de 2002.

Jugador	1	2	3	4	5	6	7	8	9	10	11	12	13
Goles	2	8	1	2	2	6	2	1	5	2	3	2	2

Fuente: http://www.rhinossoccer.com/team.asp

a. Si usted desea presentar el número de goles anotados por cada jugador, ¿sería más apropiado exhibir esta información en una gráfica de barras o en un histograma? Explique.

b. Construya la gráfica apropiada para la parte a.

c. Si usted desea mostrar (destacar) la distribución de anotaciones del equipo, ¿sería más apropiado exhibir esta información en una gráfica de barras o en un histograma? Explique.

d. Construya la gráfica apropiada para la parte c.

2.32 El Departamento de Educación de California da un informe anual de los resultados de la prueba de Colocación Avanzada (AP) para cada año. En el año escolar 2003-2004, el Condado de Modoc tuvo estudiantes con las calificaciones siguientes:

2	3	2	1	2	3	3	2	2	3	5	3	2	2
2	4	5	1	2	2	4	3	2	2	5	4	1	4
2	4	5	2	3	2	3	2	3	3	5	1	5	5

Fuente: http://data1.cde.ca.gov/dataquest/

a. Construya una distribución de frecuencia no agrupada para las calificaciones de la prueba.

b. Construya un histograma de frecuencia de esta distribución.

c. Elabore una distribución relativa de frecuencia para estos mismos datos.

d. Si para la transferencia en universidades se requieren a veces calificaciones de al menos 3 de la AP, ¿qué porcentaje de calificaciones Modoc AP recibirá crédito universitario?

(Retenga estas soluciones para usarlas en el Ejercicio 2.52 de la p. 71.)

2.33 El equipo de futbol soccer olímpico femenil de Estados Unidos tuvo un gran año en 2004. Una forma de describir las jugadoras de ese equipo es por sus estaturas individuales.

Estatura (pulgadas)

68	67	65	66	65	67	64	69	69
65	64	71	66	67	68	66	65	71

Fuente: http://www.SoccerTimes.com

a. Construya una distribución de frecuencia no agrupada para las estaturas.

b. Construya un histograma de frecuencia de esta distribución.

c. Elabore una distribución relativa de frecuencia para estos mismos datos.

d. ¿Qué porcentaje del equipo mide al menos 5 pies 6 pulgadas de estatura?

2.34 La U.S. Census Bureau publicó el siguiente informe de 2003 sobre las Familias y Arreglos de Vivienda de Estados Unidos, para todas las razas.

Núm. por familia	Porcentaje	Núm. por familia	Porcentaje
1	26.4%	5	6.3%
2	33.3%	6	2.3%
3	16.1%	7+	1.2%
4	14.3%		

Fuente: http://www.census.gov/population/www/socdemo/hh-fam/cps2003.html

a. Elabore un histograma de frecuencia relativa para el número de personas por familia.

b. ¿Qué forma de distribución sugiere el histograma?

c. Con base en la gráfica, ¿qué sabe usted acerca de las familias en Estados Unidos?

2.35 El universo de la Encuesta de 2003 de la Comunidad de Estados Unidos está limitado a la población de familias, excluyendo la población que vive en instituciones, dormitorios universitarios y otros alojamientos. La tabla siguiente es una lista del número de cuartos en cada una de las 8 658 290 unidades de habitación en Texas.

Cuartos	Unidades de habitación	Cuartos	Unidades de habitación
1 cuarto	124 486	6 cuartos	1 649 479
2 cuartos	349 496	7 cuartos	913 138
3 cuartos	1 007 873	8 cuartos	520 248
4 cuartos	1 548 984	9 + cuartos	485 506
5 cuartos	2 059 080		

Fuente: U.S. Census Bureau, American Community Survey Office

a. Trace un histograma de frecuencia relativa para el número de cuartos por familia.

b. ¿Qué forma de distribución sugiere el histograma?

c. Con base en la gráfica, ¿qué sabe usted acerca del número de cuartos por familia en Texas?

2.36 A continuación aparecen las edades de 50 bailarines que respondieron a un llamado para realizar una audición para una comedia musical:

21	19	22	19	18	20	23	19	19	20
19	20	21	22	21	20	22	20	21	20
21	19	21	21	19	19	20	19	19	19
20	20	19	21	21	22	19	19	21	19
18	21	19	18	22	21	24	20	24	17

a. Elabore una distribución de frecuencia no agrupada de estas edades.

b. Elabore una distribución de frecuencia relativa no agrupada de los mismos datos.

c. Elabore un histograma de frecuencia relativa de estos datos.

d. Elabore una distribución de frecuencia relativa acumulativa de los mismos datos.

e. Elabore una ojiva de estos datos.

2.37 Los marcadores de la ronda inicial para el torneo de la Asociación de Golf Profesional para Damas, en el club campestre del Cerro del Chapulín, se publicaron como sigue:

69	73	72	74	77	80	75	74	72	83	68	73	75	78
76	74	73	68	71	72	75	79	74	75	74	74	68	79
75	76	75	77	74	74	75	75	72	73	73	72	72	71
71	70	82	77	76	73	72	72	72	75	75	74	74	74
76	76	74	73	74	73	72	72	74	71	72	73	72	72
74	74	67	69	71	70	72	74	76	75	75	74	73	74
74	78	77	81	73	73	74	68	71	74	78	70	68	71
72	72	75	74	76	77	74	74	73	73	70	68	69	71
77	78	68	72	73	78	77	79	79	77	75	75	74	73
73	72	71	68	70	71	78	78	76	74	75	72	72	72
75	74	76	77	78	78								

a. Forme una distribución de frecuencia no agrupada de estos marcadores.

b. Trace un histograma de los marcadores de golf de la primera ronda. Use la distribución de frecuencia de la parte a.

2.38 Pensar *dónde* caerá un rayo es un trabajo casi imposible. *Cuándo* caerá ese rayo, no obstante, se ha hecho más predecible con base en investigaciones. Para una pequeña zona en Colorado, se recolectaron datos y los resultados se presentan en el histograma que sigue.

Con base en el histograma:

a. ¿Para qué variable se recolectaron los datos?

b. ¿Qué representa cada barra (intervalo)?

c. ¿A qué conclusión se puede llegar acerca de "cuándo" caerá un rayo en esta pequeña zona de Colorado?

d. ¿Qué características de la gráfica apoyan la conclusión?

2.39 Una encuesta de 100 gerentes de clubes de balnearios acerca de sus salarios anuales resultó en la siguiente distribución de frecuencia:

Salario anual (miles de $)	15-25	25-35	35-45	45-55	55-65
Núm. de gerentes	12	37	26	19	6

a. El dato "35" pertenece a ¿qué grupo?

b. Explique el significado de "35-45"

c. Explique qué es el "ancho de grupo", dé su valor, y describa tres formas en que se pueda determinar.

d. Trace un histograma de frecuencia de los salarios anuales para gerentes de clubes de balnearios. Aplique leyendas a fronteras de grupo.

(Retenga estas soluciones para usarlas en el ejercicio 2.51, p. 71.)

2.40 El siguiente ejercicio demuestra el procedimiento de transformar un diagrama de tallo y hoja en un histograma. Escriba las hojas para el número de casos en el diagrama de tallo y hoja. Luego elabore usted el histograma correspondiente. Comente sobre las similitudes y diferencias.

2.41 El examen de aptitud en ciencias de la computación KSW se dio a 50 estudiantes. De sus calificaciones, resultó la siguiente distribución de frecuencia:

Calificación examen KSW	0-4	4-8	8-12	12-16	16-20	20-24	24-28
Frecuencia	4	8	8	20	6	3	1

a. ¿Cuáles son las fronteras para el grupo que tiene la frecuencia más grande?

b. Dé todos los puntos medios de grupo asociados con esta distribución de frecuencia.

c. ¿Cuál es el ancho del grupo?

d. Dé las frecuencias relativas para los grupos.

e. Trace un histograma de frecuencia relativa de las calificaciones del examen.

2.42 La USA Snapshot titulada "Monjas, un orden que envejece" informa que la edad media de las 94,022 monjas católicas romanas de Estados Unidos es 65 años, y los porcentajes de monjas estadounidenses por grupo de edad son como sigue:

Menos de 50	51 70	Más de 70	Negaron su edad
16%	42%	37%	5%

Esta información está basada en una encuesta de 1049 monjas católicas romanas. Suponga que la encuesta hubiera resultado en la siguiente distribución de frecuencia (52 edades desconocidas).

Edad	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frecuencia	34	58	76	187	254	241	147

a. Trace y aplique leyendas completas a un histograma de frecuencia.

b. Elabore y aplique leyendas completas a un histograma de frecuencia relativa de la misma distribución.

c. Con todo cuidado examine los dos histogramas de las partes a y b y explique por qué uno de ellos podría ser más fácil de entender. (Retenga estas soluciones para usarlas en el ejercicio 2.166, p. 123.)

PARA SU INFORMACIÓN Utilice los comandos de computadora o calculadora de las páginas 61-63 para construir un histograma de una distribución de frecuencia.

2.43 Las velocidades de 55 automóviles fueron medidas por un aparato de radar en una calle citadina:

27	23	22	38	43	24	35	26	28	18	20
25	23	22	52	31	30	41	45	29	27	43
29	28	27	25	29	28	24	37	28	29	18
26	33	25	27	25	34	32	36	22	32	33
21	23	24	18	48	23	16	38	26	21	23

a. Clasifique estos datos en una distribución de frecuencia agrupada, usando las fronteras de grupo 12-18, 18-24,… 48-54.

b. Encuentre el ancho de grupo.

c. Para el grupo 24-30, encuentre el punto medio de grupo, la frontera inferior de grupo, y la frontera superior de grupo.

d. Construya un histograma de frecuencia de estos datos

PARA SU INFORMACIÓN Utilice los comandos de computadora o calculadora de las páginas 61-63 para construir un histograma para un conjunto determinado de datos.

2.44 Una prueba de hemoglobina A_{1c}, que es una prueba sanguínea que se toma a pacientes diabéticos durante sus exámenes periódicos, indica el nivel de control de azúcar en la sangre durante los últimos 2 o 3 meses. Los siguientes datos se obtuvieron para 40 pacientes diabéticos diferentes en una clínica universitaria:

6.5	5.0	5.6	7.6	4.8	8.0	7.5	7.9	8.0	9.2
6.4	6.0	5.6	6.0	5.7	9.2	8.1	8.0	6.5	6.6
5.0	8.0	6.5	6.1	6.4	6.6	7.2	5.9	4.0	5.7
7.9	6.0	5.6	6.0	6.2	7.7	6.7	7.7	8.2	9.0

a. Clasifique estos valores A_{1c} en una distribución de frecuencia agrupada usando los grupos 3.7-4.7, 4.7-5.7, etcétera.

b. ¿Cuáles son los puntos medios de grupo para estos grupos?

c. Construya un histograma de frecuencia de estos datos.

2.45 A todos los estudiantes de tercer grado de la Escuela Elemental Roth se les aplicó una prueba de resistencia y condiciones físicas. Resultaron los siguientes datos:

12	22	6	9	2	9	5	9	3	5	16	1	22
18	6	12	21	23	9	10	24	21	17	11	18	19
17	5	14	16	19	19	18	3	4	21	16	20	15
14	17	4	5	22	12	15	18	20	8	10	13	20
6	9	2	17	15	9	4	15	14	19	3	24	

a. Construya una gráfica de puntos.

b. Elabore una distribución de frecuencia agrupada usando grupos 1-4, 4-7, etc., y trace un histograma de la distribución. (Retenga la solución para usarla en la respuesta al ejercicio 2.75, p. 81)

c. Elabore una distribución de frecuencia agrupada usando grupos 0-3, 3-6, 6-9, etc., y trace un histograma de la distribución.

d. Elabore una distribución de frecuencia agrupada usando grupos –2.5, 2.5, 7.5, 12.5, etc., y trace un histograma de la distribución.

e. Elabore una distribución de frecuencia agrupada usando los grupos que usted elija y trace un histograma de la distribución.

f. Describa la forma del histograma encontrado en las partes b-e separadamente. Relacione la distribución vista en el histograma con la distribución vista en la gráfica de puntos.

g. Analice cómo es que el número de grupos empleados y la selección de fronteras de grupo empleadas afectan el aspecto del histograma resultante.

2.46 Las personas se han maravillado durante años por las continuas erupciones del géiser Old Faithful (Viejo Fiel) del Parque Nacional de Yellowstone . Los tiempos de duración, en minutos, para muestra de 50 erupciones del Old Faithful aparecen enseguida.

4.00	3.75	2.25	1.67	4.25	3.92
4.53	1.85	4.63	2.00	1.80	4.00
4.33	3.77	3.67	3.68	1.88	1.97
4.00	4.50	4.43	3.87	3.43	4.13
4.13	2.33	4.08	4.35	2.03	4.57
4.62	4.25	1.82	4.65	4.50	4.10
4.28	4.25	1.68	3.43	4.63	2.50
4.58	4.00	4.60	4.05	4.70	3.20
4.60	4.73				

Fuente: http://www.stat.sc.edu/~west/javahtml/Histogram.html

a. Trace una gráfica de puntos que exhiba los datos de duración de erupciones.

b. Trace un histograma de los datos de duración de erupciones, usando fronteras de grupo 1.6–2.0,–2.4–⋯–4.8.

c. Trace otro histograma de los datos usando diferentes fronteras y anchos de grupo.

d. ¿Cuál gráfica, en su opinión, hace el mejor trabajo de mostrar la distribución? ¿Por qué?

e. Escriba un breve párrafo que describa la distribución.

2.47 La Oficina de Carbón, Nuclear, Eléctrico y Combustibles Alternos informó de los siguientes datos como los costos (en centavos) del ingreso promedio por kilowatt-hora para sectores en Arkansas:

6.61	7.61	6.99	7.48	5.10	7.56	6.65	5.93	7.92
5.52	7.47	6.79	8.27	7.50	7.44	6.36	5.20	5.48
7.69	8.74	5.75	6.94	7.70	6.67	4.59	5.96	7.26
5.38	8.88	7.49	6.89	7.25	6.89	6.41	5.86	8.04

a. Elabore una distribución de frecuencia agrupada para el ingreso promedio por kilowatt-hora usando fronteras de grupo 4, 5, 6, 7, 8, 9.

b. Encuentre el ancho de grupo.

c. Haga una lista de los puntos medios de grupo.

d. Construya un histograma de frecuencia relativa de estos datos.

2.48 Durante mucho tiempo, la educación ha sido considerada como el boleto para ascender en Estados Unidos. En la era de la información de hoy día, una educación universitaria se ha convertido en el nivel mínimo de educación para entrar a un cada vez más competitivo mercado de trabajos que pagan salarios más que de subsistencia. Un informe del Centro Médico SUNY Downstate incluyó un estudio de zonas suburbanas que rodean a ciudades de Estados Unidos. Una variable que se reportó era el porcentaje de residentes suburbanos de 25 años de edad o mayores que cursaron al menos parte de universidad:

49.3	75.2	64.7	66.1	51.8

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: SUNY Downstate Medical Center, 2004

a. Elabore una distribución de frecuencia agrupada para el porcentaje de población suburbana de 25 años de edad o mayores con algo de educación universitaria, usando los puntos medios de la clase 25, 30, 35,…, 75.

b. Haga una lista de las fronteras del grupo.

c. Construya un histograma de frecuencia relativa de estos datos.

2.49 ¿Puede usted pensar en variables cuya distribución pudiera dar las siguientes formas diferentes? (Si es necesario, vea la figura 2.13, p. 63.)

a. Una forma simétrica, o normal

b. Una forma uniforme

c. Una forma sesgada a la derecha

d. Una forma sesgada a la izquierda

e. Una forma bimodal

2.50 El siguiente ejercicio demuestra el efecto que el número de grupos o depósitos tiene en la forma de un histograma.

a. ¿Qué forma de distribución se generaría al usar un grupo o clase?

b. ¿Qué forma de distribución se produce al usar dos grupos o clases?

c. ¿Qué forma de distribución se produce al usar 10 o 20 grupos o clases?

2.51 Una encuesta se realizó entre 100 gerentes de balnearios o centros vacacionales respecto a sus salarios anuales resultó en la siguiente distribución de frecuencia. (Vea ejercicio 2.39, p. 69.)

Salario anual (miles de $)	15-25	25-35	35-45	45-55	55-65
Núm. de gerentes	12	37	26	19	6

a. Elabore una distribución de frecuencia acumulativa para los salarios anuales.

b. Elabore una distribución de frecuencia relativa acumulativa para los salarios anuales.

c. Construya una ojiva para la distribución de frecuencia relativa acumulativa encontrada en la parte b.

PARA SU INFORMACIÓN Utilice los comandos de computadora o calculadora de la página 66 para construir una ojiva para un conjunto determinado de datos.

2.52 a. Elabore una distribución de frecuencia relativa acumulativa para la variable "AP score" (calificación de Colocación Avanzada) del ejercicio 2.32

b. Construya una ojiva de la distribución.

2.53 a. Elabore una distribución de frecuencia relativa acumulativa para la variable "KSW test score" del ejercicio 2.41.

b. Construya una ojiva de la distribución.

2.54 Los pasantes que usan préstamos para pagar sus estudios promedian $16 500 en deuda cuando se gradúan. A continuación se muestra la distribución de frecuencia relativa de su deuda mensual:

Deuda mensual, $	Menos de 100	100-149	150-199	200-249	250-299	300 o más
Frecuencia relativa	0.17	0.17	0.17	0.19	0.1	0.2

Fuente: USA Today Snapshot, 23 de diciembre, 2004

a. Elabore una distribución de frecuencia relativa acumulativa para la deuda mensual.

b. Construya una ojiva para la distribución de frecuencia relativa acumulativa encontrada en la parte a.

2.55 El siguiente estudio *La calidad de vida en las 100 ciudades más grandes de Norte América y sus suburbios: nuevos y continuos desafíos para mejorar la salud y el bienestar*, junio de 2004, informa sobre el porcentaje de personas de bajos recursos que viven en vecindarios de alta pobreza en 82 ciudades de Estados Unidos:

29.8 21.4 32.0 5.9 27.8

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: SUNY Downstate Medical Center

a. Elabore una distribución de frecuencia agrupada de los datos de porcentaje usando puntos medios de grupo de 0, 5, 10, . . . 45.

b. Elabore una distribución de frecuencia relativa agrupada de estos datos.

c. Trace un histograma de frecuencia relativa de estos datos.

d. Elabore una distribución de frecuencia relativa acumulativa de los mismos datos.

e. Trace una ojiva de estos datos.

2.56 Los niveles de diversos compuestos resultaron en las gráficas de distribución que siguen. Todos parecen ser bastante simétricos alrededor de sus centros, pero difieren en sus dispersiones.

a. ¿Para cuál histograma, A, B, C o D, anticiparía usted que la medida numérica de dispersión sea la más grande?, ¿y la más pequeña?

b. ¿Cuáles dos de los cuatro histogramas anticiparía usted que tienen alrededor de la misma diferencia entre sus valores más pequeños y sus valores más grandes?

Figuras para ejercicio 2.56

2.4 Medidas de tendencia central

Las **medidas de tendencia central** son valores numéricos que localizan, en algún sentido, el centro de un conjunto de datos. Es frecuente que el término *promedio* se asocie con todas las medidas de tendencia central.

Media (media aritmética): es el promedio con el que probablemente el lector esté más familiarizado. La media muestral se representa por $\bar{x}$ (léase "x barra" o "media muestral"). La media se encuentra al sumar todos los valores de la variable x (esta suma de x valores se simboliza como Σx) y dividir la suma entre el número de estos valores, n (el "tamaño muestral"). Expresamos esto mediante fórmula como

$$\textit{media muestral:} \qquad \textit{x barra} = \frac{\textit{suma de todas las x}}{\textit{número de las x}}$$

$$\bar{x} = \frac{\sum x}{n} \tag{2.1}$$

Nota:

1. La media poblacional, μ, es la media de todos los valores de x para toda la población.

EJEMPLO 2.8 **Para hallar la media**

Un conjunto de datos consta de los cinco valores 6, 3, 8, 6 y 4. Encuentre la media.

SOLUCIÓN Con la fórmula (2.1), encontramos

$$\bar{x} = \frac{\sum x}{n} = \frac{6 + 3 + 8 + 6 + 4}{5} = \frac{27}{5} = 5.4$$

Por lo tanto, la media de esta muestra es **5.4.**

PARA SU INFORMACIÓN
La media es el punto medio por peso.

Se puede construir una representación física de la media si consideramos una recta numérica equilibrada en prisma que tiene forma triangular como se muestra en la figura 2.15. Todo el conjunto forma una balanza. Después se coloca un peso sobre la recta en el número correspondiente a cada dato de la muestra del ejemplo 2.8. En la figura 2.15 hay un peso por cada uno de los números 3, 8 y 4 y dos pesos en el 6, puesto que hay dos números 6 en la muestra. La media es el valor que equilibra los pesos sobre la recta numérica, en este caso, 5.4.

FIGURA 2.15
Representación física de la media

$\bar{x}$ = **5.4** (el centro de gravedad, o punto de equilibrio)

INSTRUCCIONES DE TECNOLOGÍA: MEDIA

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:     Calc > Column Statistics
Select:     Mean
Enter:      Input variable: C1 > OK
```

Excel Teclee los datos en la columna A y active una celda para la respuesta; luego continúe con:

```
Choose:     Insert Function, fx > Statistical > AVERAGE > OK
Enter:      Number 1: (A2:A6 or select cells)
            [Start at A1 if no header row (column title) is used.]
```

TI-83/84 Plus Capture los datos en L1; luego continúe con:

```
Choose:     2nd > LIST > Math > 3:mean(
Enter:      L1
```

Mediana: es el valor de los datos que ocupa la posición media cuando los datos están clasificados en orden de acuerdo con su tamaño. La mediana muestral se representa por $\tilde{x}$ (léase "x tilde" o "mediana muestral").

¿SABÍA USTED...?

Las aportaciones de Sir Francis Galton a la estadística son casi ilimitadas. En 1875 estaba experimentando con semillas de chícharo (guisantes); usando 100 semillas de cada uno de siete diferentes diámetros, construyó un esquema de dos direcciones que igualaba semillas con semillas descendientes. Observó que el diámetro medio de los descendientes de las más grandes era menor que los de sus padres, mientras que el diámetro medio de los descendientes de las menores era mayor que los de sus padres. A este fenómeno de resultados que caen hacia el centro de una distribución estadística, él lo llamó *regresión a la media.*

Nota: la mediana poblacional, M (mayúscula de la letra mu del alfabeto griego), es el dato situado en la posición media de toda la población clasificada.

Procedimiento para determinar la mediana

PASO 1: **Clasifique los datos.**

PASO 2: **Determine la profundidad de la mediana.** La **profundidad,** o posición (número de posiciones desde cualquiera de los extremos), de la mediana se determina con la fórmula

Profundidad de mediana: $$\textit{profundidad de mediana} = \frac{\textit{número} + 1}{2}$$

$$d(\tilde{x}) = \frac{n+1}{2} \qquad (2.2)$$

La profundidad (o posición) de la mediana se encuentra al sumar los números de posición de los datos más pequeños (1) y los datos más grandes (n), y dividiendo entre 2 la suma (n es el número de piezas de datos).

PASO 3: **Determine el valor de la mediana.** Cuente los datos clasificados, localizando los datos en la posición $d(\tilde{x})$. La mediana será la misma cualquiera que sea el extremo de los datos clasificados (alto o bajo)

desde el que se cuente. De hecho, contar desde ambos extremos servirá como excelente prueba.

Los siguientes dos ejemplos demuestran este procedimiento cuando se aplica a conjuntos de datos de número impar o par.

EJEMPLO 2.9 Mediana para *n* impar

Encuentre la mediana para el conjunto de datos {6, 3, 8, 5, 3].

SOLUCIÓN

PASO 1 Los datos, clasificados en orden de tamaño, son 3, 3, 5, 6, y 8.

PARA SU INFORMACIÓN
El valor de $d(\tilde{x})$ es la profundidad de la mediana, NO es el valor de la mediana, $\tilde{x}$.

PASO 2 Profundidad de la mediana: $d(\tilde{x}) = \dfrac{n+1}{2} = \dfrac{5+1}{2} = 3$ (la "3ª posición).

PASO 3 La mediana es el tercer número desde cualquier extremo de los datos clasificados, es decir $\tilde{x} = \mathbf{5}$.

Observe que la mediana esencialmente separa el conjunto de datos clasificados en dos subconjuntos de igual tamaño (vea figura 2.16).

FIGURA 2.16
Mediana de {3, 3, 5, 6, 8}

Como sucede en el ejemplo 2.9, cuando n es impar, la profundidad de la mediana, $d(\tilde{x})$, siempre será un entero. No obstante, cuando n es par, la profundidad de la mediana, $d(\tilde{x})$, siempre será medio número, como se muestra en el ejemplo 2.10.

EJEMPLO 2.10 Mediana de *n* par

Encuentre la mediana de la muestra 9, 6, 7, 9, 10, 8.

SOLUCIÓN

PASO 1 Los datos, clasificados en orden de tamaño, son 6, 7, 8, 9, 9 y 10.

PASO 2 La profundidad de la mediana: $d(\tilde{x}) = \dfrac{n+1}{2} = \dfrac{6+1}{2} = 3.5$ (la posición "3.5ava").

PARA SU INFORMACIÓN
La mediana es el punto de en medio por cantidad.

PASO 3 La mediana está a la mitad entre los datos tercero y cuarto. Para hallar el número que está a la mitad entre cualesquier dos valores, sume los dos valores y divida la suma entre 2. En ese caso, sume el tercer valor (8) y el cuarto valor (9) y luego divida la suma (17) entre 2. La mediana es $\tilde{x} = \dfrac{8+9}{3} = \mathbf{8.5}$, un número que está a la mitad entre los dos números "medios" (vea figura 2.17). Observe que, de nuevo, la mediana separa el conjunto de datos clasificado en dos subconjuntos de igual tamaño.

FIGURA 2.17
Mediana de {6, 7, 8, 9, 9, 10}

INSTRUCCIONES DE TECNOLOGÍA: MEDIANA

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:     Calc > Column Statistics
Select:     Median
Enter:      Input variable: C1 > OK
```

Excel Introduzca los datos en la columna A y active una celda para la respuesta; luego continúe con:

```
Choose:     Insert Function, fx > Statistical > MEDIAN > OK
Enter:      Number 1: (A2:A6 or select cells)
```

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:     2nd > LIST > Math > 4:median(
Enter:      L1
```

Moda: la moda es el valor de x que se presenta con mayor frecuencia.

En el conjunto de datos del ejemplo 2.9, {3, 3, 5, 6, 8}, la moda es 3 (vea la figura 2.18).

FIGURA 2.18
Moda de {3, 3, 5, 6, 8}

En la muestra 6, 7, 8, 9, 9, 10, la moda es 9. En esta muestra, sólo el 9 se presenta más de una vez; en los datos del ejemplo 2.9, sólo el 3 se presenta más de una vez. Si dos o más valores de la muestra están empatados por la frecuencia más alta (número de veces que se presenta), decimos que **no hay moda.** Por ejemplo, en la muestra 3, 3, 4, 5, 5, 7, el 3 y el 5 aparecen igual número de veces. No hay un valor que aparezca con más frecuencia, por tanto, esta muestra no tiene moda.

Rango medio: es el número que está exactamente a la mitad entre el dato de valor más bajo, L, y el dato de valor más alto, H. Se encuentra al promediar los valores bajo y alto:

$$\text{rango medio} = \frac{\textit{valor bajo} + \textit{valor alto}}{2}$$

$$\text{rango medio} = \frac{L + H}{2} \qquad (2.3)$$

Para el conjunto de datos del ejemplo 2.9, {3, 3, 5, 6, 8}, $L = 3$ y $H = 8$ (vea la figura 2.19).

Por tanto,

$$\textit{rango medio} = \frac{L + H}{2}$$

$$= \frac{3 + 8}{2} = 5.5$$

FIGURA 2.19
Rango medio de {3, 3, 5, 6, 8}

Las cuatro medidas de tendencia central representan cuatro métodos diferentes de describir el centro. Estos cuatro valores pueden ser iguales, pero es más probable que sean diferentes.

Para los datos muestrales del ejemplo 2.10, la media, $\tilde{x}$, es 8.2; la mediana, $\tilde{x}$, es 8.5; la moda es 9, y el rango medio es 8. La relación entre ellos y con los datos se muestra en la figura 2.20.

FIGURA 2.20
Medidas de tendencia central para {6, 7, 8, 9, 9, 10}

CASO PRÁCTICO 2.11

"Promedio" significa cosas diferentes

Cuando se trata de conveniencia, pocas cosas pueden compararse con la maravillosa herramienta matemática que se denomina promediar. Con un promedio, se pueden tomar algunos números de cualquier tema y calcular un número que representará todo el conjunto.

Hay algo que debemos recordar; existen varias clases de medidas que por lo general se conocen como *promedios*, y cada una da una imagen diferente de los números que se le pide representar.

Tome un ejemplo. La tabla 2.11 ilustra los ingresos anuales de 10 familias.

TABLA 2.11

Ingresos anuales de 10 familias

$54 000	$39 000	$37 500	$36 750	$35 250	$31 500	$31 500	$31 500	$31 500	$25 500

¿Cuál sería el ingreso "típico" de este grupo? Hacer el promedio daría la respuesta, de modo que calculemos el ingreso típico al usar las clases de promediar más sencillas y que se usan con más frecuencia.

- *La media aritmética.* Ésta es la forma de promedio más común, obtenida al sumar los elementos del conjunto de datos y luego dividirlos entre el núme-

ro de elementos; para estos datos, la media aritmética es $35 400. La media es representativa del conjunto de datos, en el sentido de que la suma de las cantidades en las que los números más altos rebasan la media es exactamente igual que la suma de las cantidades en las que los números más bajos no llegan a la media.

Los ingresos más altos rebasan la media por un total de $25 650. Los ingresos más bajos no llegan a la media por un total de $25 650.

- *La mediana.* Como ya habrá observado el lector, seis familias ganan menos que la media y cuatro familias ganan más. Usted podría representar este grupo variado por el ingreso de la familia que está justo en el centro de todo el grupo. La mediana asciende a $33 375.
- *El rango medio*. Otro número que podría usarse para representar el promedio es el rango medio, que se obtiene al calcular la cantidad que está a la mitad entre los ingresos más alto y más bajo: $39 750.
- *La moda*. Entonces, tres clases de promedios, y no una familia en realidad tiene el ingreso comparable a cualquiera de ellos. Digamos que se desea representar el grupo al expresar el ingreso más frecuente. A esto se llama moda. El ingreso modal sería $31 500.

Existen cuatro promedios diferentes, cada uno de ellos válido, correcto e informativo a su manera. ¡Pero cómo difieren!

media aritmética	*mediana*	*rango medio*	*modo*
$35 400	$33 375	$39 750	$31 500

Y diferirían todavía más si sólo una familia del grupo fuera de millonarios □ o una que no tenga trabajo—. El valor grande de $54 000 (muy diferente de los otros valores) está sesgando los datos hacia valores más grandes. Este sesgo hace que la media y el rango medio se hagan mucho más grandes en valor.

Entonces tenemos tres lecciones. Primero, cuando vea o escuche de un promedio, averigüe cuál es el promedio. Sabrá entonces qué clase de imagen se le presenta. En segundo término, considere los números que se promedian para que pueda usted juzgar si el promedio que se utiliza es apropiado. Por último, no suponga que se busca una cuantificación matemática literal cada vez que alguien diga "promedio". No lo es. Es frecuente que digamos "la persona promedio" sin pensar que esto implique una media, mediana o moda. Todo lo que queremos expresar es la idea de que otras personas pueden pensar en forma muy similar al resto de nosotros.

Fuente: reimpreso con permiso de *Changing Times,* marzo de 1980. Copyright de The Kiplinger Washington Editors.

Ahora que hemos aprendido a calcular varias estadísticos muestrales, la siguiente pregunta es "¿Cómo expresar la respuesta final?".

Regla de redondeo: cuando se redondea una respuesta, convengamos en mantener un lugar decimal más en nuestra respuesta de lo que estaba presente en la información original. Para evitar acumulación de redondeos, redondee sólo la respuesta final, no los pasos intermedios. Es decir, evite usar un valor redondeado para hacer más cálculos. En nuestros ejemplos previos, los datos estaban formados de números enteros; por tanto, las respuestas que tengan valores decimales deben redondearse al décimo más cercano. En el *Manual de Soluciones del Estudiante* vea instrucciones específicas sobre cómo hacer el redondeo.

SECCIÓN 2.4 EJERCICIOS

2.57 Explique por qué es posible hallar la media para los datos de una variable cuantitativa pero no para una variable cualitativa.

2.58 El número de hijos, x, pertenecientes a cada una de ocho familias que se registran para natación es el siguiente: 1, 2, 1, 3, 2, 1, 5, 3. Encuentre la media, $\bar{x}$

2.59 Pida a su profesor que elabore un ejercicio donde se observe el efecto de equilibrio de la media. El problema debe basarse en los datos de la siguiente gráfica con un punto de datos en 10. Debe realizar las modificaciones en la gráfica hasta alcanzar una media de 1. Como modelo utilice la grafica que se muestra.

a. ¿Cuántos bloques fueron necesarios para balancear y obtener una media de 1?

b. ¿En qué valor están situados estos bloques?

2.60 La autopista interestatal 64 de Norteamérica se localiza entre Portsmouth, VA, en I-264 en el extremo este hasta St. Louis, MO, en I-270 en el extremo oeste. Al pasar por seis estados y cinco ciudades importantes, la I-64 cruza nueve carreteras más. El número de millas en cada estado es como sigue: Missouri: 16 millas, Illinois: 132 millas, Indiana: 124 millas, Kentucky: 191 millas, West Virginia: 183 millas, Virginia: 299 millas. (Un diagrama puede ser útil.)

Fuente: http://www.ihoz.com/I90.html

a. Encuentre la distancia media entre ciudades importantes a lo largo de la I-64.

b. Encuentre la distancia media entre cruceros con carreteras interestatales a lo largo de la I-64.

2.61 La autopista interestatal 29 cruza muchas otras carreteras cuando pasa por cuatro estados en la región central de Estados Unidos, alargándose desde el extremo sur de Kansas City, MO, en la I-35 hasta el extremo norte en Pembina, ND, en la frontera con Canadá.

Interestatal 29 de Estados Unidos

Estado	Millas	Número de intersecciones	Estado	Millas	Número de intersecciones
Missouri	123	37	South Dakota	252	44
Iowa	161	32	North Dakota	217	40

Fuentes: Rand McNally y http://www.ihoz.com/ilist.html

a. Encuentre la distancia media entre cruceros en Missouri.

b. Encuentre la distancia media entre cruceros en Iowa.

c. Encuentre la distancia media entre cruceros en Dakota del Norte.

d. Encuentre la distancia media entre cruceros en Dakota del Sur.

e. Encuentre la distancia media entre cruceros a lo largo de la U.S. I-29.

f. Encuentre la media de las cuatro medias encontradas al contestar las partes de la a a la d.

g. Compare las respuestas encontradas a las partes e y f. ¿Esperaba que fueran las mismas? Explique por qué son diferentes.

2.62 Encuentre la estatura media de un equipo de baloncesto: 73, 76, 72, 70 y 74 pulgadas.

2.63 Encuentre la mediana de la cantidad mediana pagada en Jim's Burgers si los sueldos por hora son \$4.25, \$4.15, \$4.90, \$4.25, \$4.60, \$4.50, \$4.75.

2.64 El siguiente ejercicio demuestra el efecto de equilibrio que puede tener un dato sobre la media y la mediana.

a. ¿Qué sucedería si el punto rojo se mueve a la extrema derecha? ¿Qué le pasa a la media? ¿Cómo se comportaría la mediana?

b. ¿Qué sucedería si el punto rojo se desplaza a la extrema izquierda? ¿Qué le pasa a la media? ¿Qué le sucedería a la mediana?

c. ¿Cuál medida de tendencia central, la media o la mediana, da un mejor sentido del centro cuando un resultado aislado está presente en los datos?

2.65 La cantidad de automóviles por apartamento propiedad de una muestra de inquilinos en un complejo habitacional grande es 1, 2, 1, 2, 2, 2, 1, 2, 3, 2. ¿Cuál es la moda?

2.66 Cada año, aproximadamente 160 universidades compiten en la Competencia Nacional de Canoas de Concreto. Cada equipo debe diseñar una canoa marinera hecha de una sustancia no conocida por su capacidad para flotar. Las canoas deben pesar entre 100 y 350 libras. Encuentre el rango medio.

Fuente: Reader's Digest, marzo de 2005.

2.67 a. Encuentre la media, mediana, moda y rango medio de los datos muestrales 9, 6, 7, 9, 10, 8.

b. Verifique y analice la relación entre las respuestas de la parte a, como se muestra en la figura 2.20 de la página 77.

2.68 Considere la muestra 2, 4, 7, 8, 9. Encuentre lo siguiente:

a. media, $\bar{x}$ b. mediana, $\tilde{x}$

c. moda d. rango medio

2.69 Considere la muestra 6, 8, 7, 5, 3, 7. Encuentre lo siguiente:

a. media, $\bar{x}$ b. mediana, $\tilde{x}$

c. moda d. rango medio

2.70 A quince estudiantes universitarios seleccionados al azar se les pidió dijeran el número de horas que durmieron la noche previa. Los datos resultantes son 5, 6, 6, 8, 7, 7, 9, 5, 4, 8, 11, 6, 7, 8, 7. Encuentre lo siguiente:

a. media, $\bar{x}$ b. mediana, $\tilde{x}$

c. moda d. rango medio

2.71 Una muestra aleatoria de 10 de los corredores de automóviles NASCAR de la Copa Nextel 2005 produjo las siguientes edades: 33, 48, 41, 29, 40, 48, 44, 42, 49, 28.

a. Encuentre la edad media para los 10 corredores NASCAR de la Copa Nextel 2005.

b. Encuentre la edad mediana para los 10 corredores NASCAR de la Copa Nextel 2005.

c. Encuentre el rango medio de edad para los 10 corredores NASCAR de la Copa Nextel 2005.

d. Encuentre la moda, si existe, para la edad de los 10 corredores NASCAR de la Copa Nextel 2005.

2.72 Una meta constante en la manufactura de lentes de contacto es mejorar las características que afecten la potencia del lente y la agudeza visual. Una de estas características comprende las herramientas con las que finalmente se manufacturan los lentes. Se examinaron los resultados de los lotes de desarrollo del proceso inicial para ver la característica crítica X. A continuación aparecen los datos resultantes:

0.026	0.027	0.024	0.023	0.034	0.035	0.035	0.033	0.034
0.033	0.032	0.038	0.041	0.041	0.021	0.022	0.027	0.032
0.023	0.023	0.024	0.017	0.023	0.019	0.027		

Fuente: Bausch & Lomb (no se menciona la variable ni los datos codificados, a solicitud de B&L)

a. Trace una gráfica de puntos y un histograma de los datos de la característica crítica X.

b. Encuentre la media para la característica crítica X.

c. Encuentre la mediana para la característica crítica X.

d. Encuentre el rango medio para la característica crítica X.

e. Encuentre la moda, si existe, para la característica crítica X.

f. ¿Qué característica de la distribución, como se muestra en las gráficas halladas en la parte a, parece poco común? ¿Dónde caen las respuestas halladas en las partes b, c, y d respecto a la distribución? Explique.

g. Identifique al menos una posible causa para esta situación aparentemente poco común.

2.73 Una medida del rendimiento de aerolíneas es el porcentaje total de vuelos a tiempo. Para enero de 2005, los porcentajes de llegadas a tiempo de vuelos

nacionales en los 31 aeropuertos más grandes de Estados Unidos fueron como sigue:

ATL	69.09	BWI	74.01	BOS	62.14

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: U.S. Department of Transportation, Bureau of Transportation Statistics

a. Encuentre el porcentaje medio de llegadas a tiempo para enero de 2005.
b. Encuentre la mediana del porcentaje de llegadas a tiempo para enero de 2005.
c. Construya un diagrama de tallo y hoja de los datos.
d. Describa la relación entre la media y la mediana y qué propiedades de los datos hacen que la media sea menor que la mediana.

(Retenga estas soluciones para usarlas en el ejercicio 2.99 de la página 92.)

2.74 El "promedio" es una estadística que generalmente se reporta. Esta pequeña parte de información puede ser muy informativa o muy engañosa, con la media y la mediana siendo las dos que más se reportan.

a. La media es una medida útil, pero puede ser engañosa. Describa una circunstancia cuando la media sea muy útil como promedio y una circunstancia cuando la media sea muy engañosa como promedio.
b. La mediana es una medida útil, pero puede ser engañosa. Describa una circunstancia cuando la mediana sea muy útil como promedio y una circunstancia cuando la mediana ser muy engañosa como promedio.

2.75 A todos los estudiantes de tercer grado de la Escuela Elemental Roth se les aplicó una prueba de resistencia y condiciones físicas. Resultaron los siguientes datos:

12	22	6	9	2	9	5	9	3	5	16	1	22
18	6	12	21	23	9	10	24	21	17	11	18	19
17	5	14	16	19	19	18	3	4	21	16	20	15
14	17	4	5	22	12	15	18	20	8	10	13	20
6	9	2	17	15	9	4	15	14	19	3	24	

a. Construya una gráfica de puntos.
b. Encuentre la moda.
c. Elabore una distribución de frecuencia agrupada usando grupos 1-4, 4-7, etc., y trace un histograma de la distribución.
d. Describa la distribución; específicamente, ¿la distribución es bimodal (alrededor de qué valores)?
e. Compare sus respuestas de los incisos a, c, y comente sobre la relación entre la moda y los valores modales en estos datos.
f. ¿Podría ocurrir la discrepancia encontrada en la comparación de la parte e cuando se use una distribución de frecuencia no agrupada? Explique.
g. Explique por qué, en general, la moda de un conjunto de datos no necesariamente nos da la misma información que los valores modales.

2.76 Es frecuente que a los consumidores se les advierta de no comer demasiado alimento alto en calorías, grasas y sodio por numerosas razones de saludo y condiciones físicas. *Nutrition Actino HealthLetter* publicó una lista de conocidas marcas de "perros calientes" bajos en grasas que comúnmente se les aplica etiqueta de "sin grasa", "reducidos en grasas", "bajo contenido de grasas", "para dieta", etc., junto con sus calorías, contenido de grasas y sodio. Todas las cantidades siguientes son para un "perro caliente":

Marca de "perro caliente"	Calorías	Grasa (g)	Sodio (mg)
Ball P	50	0	460
Butterball F	40	0	490

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: Nutrition ActionHealthLetter, "On the Links," julio/agosto 1998, p. 12-13.

a. Encuentre la media, mediana, moda y rango medio de las calorías, grasa y contenido de sodio de todas las salchichas descritas. Use una tabla para resumir sus resultados.
b. Construya una gráfica de puntos del contenido de grasa. Localice la media, mediana, moda y rango medio en la gráfica.
c. En el verano de 2005, el ganador del Famoso Concurso de Comer Perros Calientes el Cuatro de Julio de Nathan consumió 49 perros calientes en 12 minutos. Si se le hubiera servido la mediana del perro caliente, ¿cuántas calorías, gramos de grasa, y miligramos de sodio consumió en esa sola sentada? Si la tolerancia diaria recomendada por ingestión de sodio es de 2400 mg, ¿la rebasó el concursante? Explique.

2.77 Es probable que el número de carreras anotadas por equipos de las ligas mayores de béisbol sea influenciado porque el juego se desarrolle en su campo, como

equipo local, o en un estadio como equipo visitante. En un intento por medir las diferencias entre jugar como local o como visitante, se registró el número de carreras anotadas por juego por cada uno de los equipos de ligas mayores. La tabla siguiente resume los datos.

Equipo	Prom. carreras como local	Prom. carreras como visitante
Angels	4.83	5.49
Red Sox	6.38	5.33

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://mlb.mlb.com

a. Encuentre la media, mediana, máximo, mínimo, y rango medio de carreras anotadas por los equipos cuando jugaban como locales.

b. Encuentre la media, mediana, máximo, mínimo, y rango medio de carreras anotadas por los equipos cuando jugaban como visitantes.

c. Compare cada una de las medidas que encontró en las partes a y b. ¿Qué puede concluir?

2.78 ¿Todo aumenta cada año? A veces así parece. El porcentaje anual de aumento en el consumo de combustibles para motores, para los años 2002-2003, aparece en la tabla siguiente que publica el gobierno de Estados Unidos. Observe que el consumo no aumentó en todos los estados y no todos los estados lo reportaron.

1.8	16.9	1.8	0.1	– 2.8	– 0.4	– 2.6	– 0.8	– 6.3	2.7	1.1
0.1	– 4.6	– 0.4	0.2	0.9	– 2.8	1.5	3.6	10.8	2.3	0.5
– 1	– 1	2.9	1.6	– 1.2	– 0.1	5.5	0.2	8.9	2.1	0.5
0.1	2.9	– 0.5	3.7	1.1	– 1.3	0	1.3	– 2.4	0.9	0.1

Fuente: U.S. Department of Transportation, Federal Highway Administration

a. Explique el significado de valores negativos y positivos, valores grandes y pequeños, valores cercanos a cero, y valores no cercanos a cero.

b. Examine los datos de la tabla. ¿Qué anticipa usted que aspecto tendrá el "cambio de porcentaje"? ¿Cuál piensa usted que será el "cambio de porcentaje" medio? Justifique su estimación, sin ningún trabajo preliminar ni cálculos.

c. Si usted espera que haya un cambio muy pequeño, o no lo haya, ¿qué valor tendrá la media? Explique.

d. Construya un histograma del porcentaje de cambio.

e. Calcule el porcentaje medio de cambios en consumo de 2002 a 2003.

f. La Federal Highway Administration informó del aumento en porcentaje para todo Estados Unidos como 0.5586 de 1%. El valor calculado para la media en la parte e no es el mismo. Explique cómo es esto posible.

2.79 A los estudiantes les gusta participar en la "Batalla de los Sexos" cuando se trata de ver quién es mejor conductor de vehículos con licencia de manejo. Pero, ¿cuál género supera al otro en las carreteras? Los números pueden sorprenderle. A continuación vea el número de conductores en cada uno de los 18 estados seleccionados al azar.

Estado	Hombre	Mujer
KY	1,389,380	1,410,255
DE	286,144	298,992

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: U.S. Department of Transportation, Federal Highway Administration

a. ¿Las mujeres conductoras superan a los hombres? Estudie la tabla y vea si los datos parecen apoyar lo que usted piensa. Explique su respuesta inicial.

b. Defina la variable "razón H/M" como el número de conductores hombres con licencia dividido entre el número de conductoras con licencia en cada estado. Calcule la "razón H/M" para los estados de la muestra.

c. Si el valor de la "razón H/M" es cercano a 1.0, ¿qué significa esto? ¿Y si es más de 1.0? ¿O menos de 1.0? Explique.

d. Construya un histograma.

e. Describa la distribución mostrada en el histograma encontrado en la parte d.

f. Calcula el valor medio de la "razón H/M."

g. Explique el significado de valores de cada una de las colas del histograma.

h. Indique dos estados, no de la tabla precedente, que usted espera hallar cerca de cada cola de la distribución de H/M. Explique por qué piensa que estos estados tendrán razones altas o bajas.

i. Conteste las preguntas de las partes d y f usando los datos de todos los 51 renglones.

j. Compare los resultados encontrados en la parte i con los de las partes d y f.

k. ¿Qué tan precisas fueron sus predicciones para la parte h? Explique.

2.80 Supongamos que usted es responsable de la planificación del estacionamiento necesario para un nuevo complejo de departamentos de 256 unidades, y se le pide basar las necesidades en la estadística "el número promedio de vehículos por familia es 1.9."

a. ¿Cuál promedio (media, mediana, moda, rango medio) le será útil? Explique.

b. Explique por qué 1.9 no puede ser la mediana, la moda, o el rango medio para la variable "número de vehículos".

c. Si el propietario desea un estacionamiento con capacidad para 90% de los inquilinos dueños de vehículos, ¿para cuántos espacios debe planificar?

2.81 ¿En qué estados pagan más impuestos los residentes? ¿Y en cuáles pagan menos? Esto quizá depende de la variable empleada para medir la cantidad pagada en impuestos. En 2004 el Tax Policy Center informó de las siguientes estadísticas, acerca de los impuestos promedio anuales de 2002, y el porcentaje de impuesto personal pagado por persona por estado.

	Impuestos por personal	Lugar	% ingreso personal	Lugar
Hawaii	$2748	1	9.6	1
South Dakota	$1283	50	4.8	47
New Hampshire	$1478	45	4.4	50

Fuentes: Federation of Tax Administrators (2004) y U.S. Bureau of the Census y Bureau of Economic Analysis, http://taxpolicycenter.org/TaxFacts/TFDB/TFTemplate.cfm?Docid=309&Topic2id=90

a. Compare y contraste las variables "impuestos por persona" y "porcentaje de ingreso personal". ¿Cómo toma usted en cuenta las diferencias en lugares para Dakota del Sur y New Hampshire?

b. Con base en esta información, usando la cantidad de impuestos más alta y más baja pagada por persona, por estado, ¿cuál fue la cantidad "promedio" pagada por persona?

c. Con base en esta información, usando el porcentaje de ingreso más alto y bajo pagado por persona, por estado, ¿cuál fue el porcentaje "promedio" pagado por persona?

d. Explique por qué sus respuestas a las partes b y c son el único valor promedio que usted puede determinar a partir de la información dada. ¿Cuál es el nombre de este promedio?

2.82 El profesor y el grupo de alumnos de su grupo han hecho un trato sobre el examen que acaban de pasar y que está siendo calificado. Si el grupo alcanza una calificación media de 74 o más, no habrá trabajo para el próximo fin de semana; si la calificación es de 72 o menos, entonces no sólo habrá tarea como de costumbre sino que todo el grupo tendrá que presentarse el sábado, y hacer 2 horas de limpieza general de los terrenos de la escuela como proyecto de servicio a la comunidad. Hay 15 estudiantes en el grupo de usted. El maestro ha calificado los primeros 14 exámenes, y la calificación media es de 73.5. El examen de usted es el único que falta de calificar.

a. ¿Qué calificación debe usted obtener para que el grupo gane el trato?

b. ¿Qué calificación debe usted obtener para que el grupo no tenga que hacer el trabajo de servicio a la comunidad?

2.83 Comenzando con los datos 70 y 100, sume tres datos a la muestra para que la muestra tenga lo siguiente (justifique su respuesta en cada caso):

a. Media de 100

b. Mediana de 70

c. Moda de 87

d. Rango medio de 70

e. Media de 100 y una mediana de 70

f. Media de 100 y una moda de 87

g. Media de 100 y un rango medio de 70

h. Media de 100, una mediana de 70, y una moda de 87

2.84 Se muestran diferentes histogramas de frecuencias. ¿Qué puede concluir de los datos que se presentan?

2.5 Medidas de dispersión

Una vez que se ha localizado el "medio" mediante la aplicación de las medidas de tendencia central, nuestra búsqueda de información a partir de conjuntos de datos se dirige ahora a las medidas de dispersión. Las **medidas de dispersión** incluyen el *rango, varianza* y *desviación estándar.* Estos valores numéricos describen la cantidad de dispersión, o variabilidad, que se encuentra entre los datos: los datos agrupados de manera estrecha tienen valores relativamente pequeños, y aquellos datos que estén más dispersos tienen valores más grandes. La agrupación más cercana posible ocurre cuando los datos no tienen dispersión (todos los datos son del mismo valor); en esta situación, la medida de dispersión será cero. No hay límite sobre qué tan dispersos puedan ser los datos; por tanto, las medidas de dispersión pueden ser muy grandes. La medida de dispersión más sencilla es el rango.

Rango: es la diferencia en valor entre los datos de valor más alto, H, y los datos de valor más bajo, L:

$$\textit{rango} = \textit{valor alto} - \textit{valor bajo}$$

$$\text{rango} = H - L \tag{2-4}$$

La muestra 3, 4, 5, 6, 8 tiene un rango de $H - L = 8 - 3 = 5$. El rango de 5 nos dice que estos datos caen todos ellos dentro de un intervalo de 5 unidades (ver la figura 2.21).

FIGURA 2.21
Rango de {3, 3, 5, 6, 8}

Las otras medidas de dispersión que se van a estudiar en este capítulo son medidas de dispersión alrededor de la media. Para desarrollar una medida de dispersión alrededor de la media, contestemos primero a la pregunta: ¿qué tan lejos de la media está cada x?

Desviación desde la media: una desviación desde la media, $(x - \bar{x})$, es la diferencia entre el valor de x y la media, $\bar{x}$.

Cada valor individual de x se desvía de la media en una cantidad igual a $(x - \bar{x})$. Esta desviación $(x - \bar{x})$ es cero cuando x es igual a la media, $\bar{x}$. La desviación $(x - \bar{x})$ es positiva cuando x es mayor que $\bar{x}$ y negativa cuando x es menor que $\bar{x}$.

Considere la muestra 6, 3, 8, 5, 3. Usando la fórmula (2.1), $\bar{x} = \dfrac{\Sigma x}{n}$, encontramos que la media es 5. Cada desviación, $(x - \bar{x})$, se encuentra entonces al restar 5 de cada valor x:

Datos, x	6	3	8	5	3
Desviación, $x - \bar{x}$	1	-2	3	0	-2

FIGURA 2.22
Desviaciones desde la media

La figura 2.22 muestra las cuatro desviaciones desde la media.

Para describir el valor "promedio" de estas desviaciones, podríamos usar la desviación media, la suma de las desviaciones dividida entre n, $\frac{\sum(x - x)}{n}$. No obstante, como la suma de las desviaciones, $\Sigma(x - \bar{x})$, es exactamente cero, la desviación de la media también será cero. De hecho, siempre será cero, lo cual significa que no es una estadística útil. ¿Cómo ocurre esto, y por qué?

La suma de las desviaciones, $\Sigma(x - \bar{x})$, es siempre cero porque las desviaciones de x valores menores que la media (que son negativas) cancelan los valores de x mayores que la media (que son positivos). Podemos eliminar este efecto neutralizador si hacemos algo para que todas las desviaciones sean positivas. Esto se puede lograr de dos formas. Primero, usando el valor absoluto de la desviación, $|x - \bar{x}|$, podemos tratar cada desviación como su "tamaño" o distancia únicamente. Para nuestra ilustración obtenemos las siguientes *desviaciones absolutas.*

Datos, x	6	3	8	5	3
Valor absoluto de desviación, $\lvert x - \bar{x} \rvert$	1	2	3	0	2

Desviación media absoluta: es la media de los valores absolutos de las desviaciones desde la media:

$$\textit{desviación media absoluta} = \frac{\textit{(suma de (valores absolutos de desviaciones)}}{\textit{número}}$$

$$\textit{desviación media absoluta} = \frac{\sum \lvert \boldsymbol{x} - \bar{x} \rvert}{x} \qquad (2.5)$$

Para nuestro ejemplo, la suma de las desviaciones absolutas es
8 (1 + 2 + 3 + 0 + 2) y

$$\text{desviación media absoluta} = \frac{\sum \lvert x - \bar{x} \rvert}{n} = \frac{8}{5} = 1.6$$

Aun cuando esta medida particular de dispersión no se usa con frecuencia, es una medida de dispersión. Nos indica la "distancia" media a la que están los datos desde la media.

Una segunda forma de eliminar el efecto neutralizador positivo-negativo es elevar al cuadrado cada una de las desviaciones; el cuadrado de las desviaciones será de valores positivos (positivos o cero). El cuadrado de las desviaciones se usa para hallar la *varianza.*

PARA SU INFORMACIÓN
Consulte la página 90 para obtener una explicación de estos iconos.

Varianza muestral: la varianza muestral, s^2, es la media del cuadrado de las desviaciones, calculada usando $n - 1$ como divisor:

$$\text{varianza muestral:} \quad s \text{ cuadrada} = \frac{\text{suma de (cuadrado de desviaciones)}}{\text{número} - 1}$$

D

$$\mathbf{s^2} = \frac{\sum (x - x)^2}{n - 1} \qquad (2.6)$$

donde n es el tamaño muestral, es decir, el número de datos de la muestra.

La varianza de la muestra 6, 3, 8, 5, 3 se calcula en la tabla 2.12 usando la fórmula (2.6)

TABLA 2.12

Cálculo de varianza usando la fórmula (2.6)

Paso 1 Hallar $\sum x$	Paso 2 Hallar $\bar{x}$	Paso 3 Hallar toda $x - \bar{x}$	Paso 4 Hallar $\sum (x - \bar{x})^2$	Paso 5 Hallar s^2
6	$\bar{x} = \frac{\sum x}{n}$	$6 - 5 = 1$	$(1)^2 = 1$	$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$
3		$3 - 5 = -2$	$(-2)^2 = 4$	
8		$8 - 5 = 3$	$(3)^2 = 9$	
5	$\bar{x} = \frac{25}{5}$	$5 - 5 = 0$	$(0)^2 = 0$	$s^2 = \frac{18}{4}$
3		$3 - 5 = -2$	$(-2)^2 = 4$	
$\sum x = 25$	$\bar{x} = 5$	$\sum (x - \bar{x}) = 0$ ck	$\sum (x - \bar{x})^2 = 18$	$s^2 = \mathbf{4.5}$

Notas:

1. La suma de todos los valores de x se usa para hallar $\bar{x}$.
2. La suma de las desviaciones, $\sum (x - \bar{x})$, es siempre cero, siempre que se use el valor exacto de $\bar{x}$. Use este dato como prueba en sus cálculos, como se hizo en la tabla 2.12 (denotada por ck).
3. Si se usa un valor redondeado $\bar{x}$, entonces $\sum (x - \bar{x})$ no siempre será exactamente cero pero será razonablemente cerca de cero.
4. La suma del cuadrado de las desviaciones se encuentra al elevar al cuadrado cada una de las desviaciones y luego sumar el cuadrado de valores.

Para demostrar gráficamente lo que nos dicen las varianzas de conjuntos de datos, considere un segundo conjunto de datos: {1, 3, 5, 6, 10}. Nótese que los datos están más dispersos que los datos de la tabla 2.12. De conformidad con esto, su varianza calculada es mayor en $s^2 = 11.5$. En la figura 2.23 se ve una comparación gráfica ilustrativa de estas dos muestras juntas y sus varianzas.

FIGURA 2.23
Comparación de datos

Desviación muestral estándar: la desviación estándar de una muestra, s, es la raíz cuadrada positiva de la varianza:

desviación muestral estándar: $s =$ *raíz cuadrada de varianza muestral*

$$s = \sqrt{s^2} \tag{2.7}$$

Para las muestras de la figura 2.23, las desviaciones estándar son $\sqrt{4.5}$ o **2.1,** y $\sqrt{11.5}$, o **3.4.**

Nota: es frecuente que el numerador para la varianza muestral, $\Sigma(x - \bar{x})^2$, se denomine *suma de cuadrados de x* y se simbolice por SS(x). Así, la fórmula (2.6) se puede expresar como

$$\text{varianza muestral: } s^2 = \frac{\text{SS}(x)}{n - 1} \tag{2.8}$$

donde $\text{SS}(x) = \sum(x - \bar{x})^2$.

Las fórmulas para varianza se pueden modificar en otras formas para uso más fácil en diversas situaciones. Por ejemplo, supóngase que tenemos la muestra 6, 3, 8, 5, 2. La varianza para esta muestra está calculada en la tabla 2.13.

TABLA 2.13

Cálculo de varianza usando la fórmula (2.6)

Paso 1. Hallar $\sum x$	Paso 2. Hallar $\bar{x}$	Paso 3. Hallar toda $x - \bar{x}$	Paso 4. Hallar $\sum(x - \bar{x})^2$	Paso 5. Hallar s^2
6	$\bar{x} = \frac{\sum x}{n}$	$6 - 4.8 = 1.2$	$(1.2)^2 = 1.44$	$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$
3		$3 - 4.8 = -1.8$	$(-1.8)^2 = 3.24$	
8		$8 - 4.8 = 3.2$	$(3.2)^2 = 10.24$	
5	$\bar{x} = \frac{24}{5}$	$5 - 4.8 = 0.2$	$(0.2)^2 = 0.04$	$s^2 = \frac{22.80}{4}$
2		$2 - 4.8 = -2.8$	$(-2.8)^2 = 7.84$	
$\sum x = 24$	$\bar{x} = 4.8$	$\sum(x - \bar{x}) = 0$ ck	$\sum(x - \bar{x})^2 = 22.80$	$s^2 =$ **5.7**

La aritmética para este ejemplo se ha hecho más complicada porque la media contiene dígitos diferentes de cero a la derecha del punto decimal. No obstante, la "suma de cuadrados de *x*", el numerador de la fórmula (2.6), se puede rescribir de tal forma que $\bar{x}$ 5 no se incluya:

Suma de cuadrados de x

$$SS(x) = \sum x^2 - \frac{(\sum x)^2}{n} \tag{2.9}$$

PARA SU INFORMACIÓN
Consulte la página 90 para obtener una explicación de los iconos.

La combinación de las fórmulas (2.8) y (2.9) da la "fórmula breve" para la varianza muestral:

Varianza muestral, "fórmula breve"

$$s\ \text{cuadrada} = \frac{(\text{suma de } x^2) - \left[\dfrac{(\text{suma de } x)^2}{\text{número}}\right]}{\text{número} - 1}$$

$$\text{varianza muestral:} \quad s^2 = \frac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n - 1} \tag{2.10}$$

Las fórmulas (2.9) y (2.10) se denominan *breves* porque evitan los cálculos de $\bar{x}$. Los cálculos para SS(x), s^2, y s usando las fórmulas (2.9), (2.10), y (2.7) se efectúan como se ve en la tabla 2.14.

TABLA 2.14

Cálculo de desviación estándar usando el método breve

Paso 1. Calcular $\sum x$	Paso 2. Calcular $\sum x^2$	Paso 3. Calcular SS (x)	Paso 4. Calcular s^2	Paso 5. Calcular s
6	$6^2 = 36$	$SS(x) = \sum x^2 - \dfrac{(\sum x)^2}{}$	$s^2 = \dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{}$	$s = \sqrt{s^2}$
3	$3^2 = 9$			$s = \sqrt{5.7}$
8	$8^2 = 64$	$SS(x) = 138 - \dfrac{(24)^2}{5}$		$s = \mathbf{2.4}$
5	$5^2 = 25$		$s^2 = \dfrac{22.8}{4}$	
2	$2^2 = 4$	$SS(x) = 138 - 115.2$		
$\sum x = 24$	$\sum x^2 = 138$	$SS(x) = 22.8$	$s^2 = 5.7$	

La unidad de medida de la desviación estándar es la misma que la de los datos. Por ejemplo, si nuestros datos están en libras, entonces la desviación estándar, s, también estará en libras. La unidad de media para la varianza podría entonces ser considerada como *unidades cuadradas*. En nuestro ejemplo de libras, esto sería *libras cuadradas*. Como se ve, la unidad tiene muy poco significado.

INSTRUCCIONES DE TECNOLOGÍA: DESVIACIÓN ESTÁNDAR

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:     Calc > Column Statistics
Select:     Standard deviation
Enter:      Input variable: C1 > OK
```

Excel Introduzca los datos en la columna A y active una celda para la respuesta; luego continúe con:

```
Choose:     Insert Function, fx > Statistical > STDEV > OK
Enter:      Number 1: (A2:A6 or select cells)
```

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:      2nd > LIST > Math > 7:StdDev(
Enter:       L1
```

INSTRUCCIONES DE TECNOLOGÍA: ESTADÍSTICAS ADICIONALES

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:      Calc > Column Statistics
             Then one   at a time select the desired statistic
Select:      N total    Number of data in column
             Sum        Sum of the data in column
             Minimum    Smallest value in column
             Maximum    Largest value in column
             Range      Range of values in column
             Sum of squares     Sum of squared x-values, ∑x²
Enter:       Input variable: C1 > OK
```

Excel Introduzca los datos en la columna A y active una celda para la respuesta; luego continúe con:

```
Choose:      Insert Function, fx > Statistical > COUNT
                                               > MIN
                                               > MAX
             OR                 > All > SUM
                                      > SUMSQ
Enter:       Number 1: (A2:A6 or select cells)
For range, write a formula: Max ( )   Min ( )
```

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:      2nd    LIST    Math    5:sum(
                                    1:min(
                                    2:max(
Enter:       L1
```

Desviación estándar en su calculadora: casi todas las calculadoras tienen dos fórmulas para hallar la desviación estándar y calcular ambas impensadamente, pero esperando que el usuario decida cuál es correcta para los datos dados. ¿Cómo decidir?

La desviación muestral estándar se representa con la letra s y usa la fórmula "dividir entre $n-1$".

La desviación poblacional estándar se denota por σ y usa la fórmula "dividir entre n".

Cuando se tengan datos muestrales, siempre debe usar la fórmula s o "dividir entre $n-1$". Tener los datos poblacionales es una situación que probablemente nunca ocurra, como no sea en un ejercicio de libros de texto. Si no se sabe si se tienen datos muestrales o datos poblacionales, es buena "apuesta" a que son datos muestrales y entonces usar la fórmula s o "dividir entre $n-1$".

Fórmulas múltiples: los estadísticos tienen diversas fórmulas por conveniencia, es decir, conveniencia respecto a la situación. Los siguientes enunciados ayudarán al lector a decidir cuál fórmula usar:

1. Cuando trabaje en computadora y use software de estadística, por lo general primero se guarda todos los datos. La computadora maneja con facilidad operaciones repetidas, y puede "volver a examinar" los datos guardados con la frecuencia que sea necesaria para completar un procedimiento. Los cálculos para varianza muestral se efectuarán usando la fórmula (2.6), siguiendo el proceso que se ve en la tabla 2.12.
2. Cuando trabaje con una calculadora que tenga funciones estadísticas integradas, la calculadora debe efectuar las operaciones necesarias en cada uno de los datos cuando se introduzcan los valores (la mayor parte de las calculadoras de bolsillo no tienen capacidad para guardar datos.) A continuación, una vez introducidos los datos, los cálculos se efectuarán mediante las sumatorias apropiadas. Los cálculos para varianza muestral se harán usando la fórmula (2.10), siguiendo el procedimiento que se ve en la tabla 2.14.
3. Si usted hace los cálculos manualmente o con ayuda de una calculadora, pero sin usar funciones estadísticas, la fórmula más conveniente a usar dependerá de cuántos datos se disponga y qué tan fácil sea trabajar con valores numéricos.

Cuando una fórmula tiene formas múltiples, busque uno de los iconos siguientes:

se usa para identificar la fórmula más probable de usar por una computadora.

se usa para identificar la fórmula más probable de usar por una calculadora.

se usa para identificar la fórmula más probable para cálculos hechos manualmente.

D se usa para identificar la fórmula de "definición".

SECCIÓN 2.5 EJERCICIOS

2.85 En 2004, el Tax Policy Center publicó las estadísticas siguientes acerca del promedio anual de impuestos de 2002, y el ingreso personal pagado por persona por estado.

	Impuestos por persona	Lugar	Porcentaje de ingreso personal	Lugar
Hawaii	$2748	1	9.6	1
South Dakota	$1283	50	4.8	47
New Hampshire	$1478	45	4.4	50

Fuentes: Federation of Tax Administrators (2004) y U.S. Bureau of the Census y Bureau of Economic Analysis, http://taxpolicycenter.org/TaxFacts/TFDB/TF-Template.cfm?Docid=309&Topic2id=90

a. Encuentre el rango para la cantidad de impuestos pagados por persona.

b. Encuentre el rango para el porcentaje de ingreso personal pagado en impuestos por persona.

2.86 a. El dato $x = 45$ tiene un valor de desviación de 12. Explique el significado de esto.

b. El dato $x = 84$ tiene un valor de desviación de –20. Explique el significado de esto.

2.87 La suma $\sum(x - \bar{x})$ es siempre cero. ¿Por qué? Recuerde la definición de la media (p. 73) y vea si puede justificar este enunciado.

2.88 Todas las medidas de variación son no negativas en valor para todos los conjuntos de datos.

a. ¿Qué significa que un valor sea "no negativo"?

b. Describa las condiciones necesarias para que una medida de variación tenga el valor cero.

c. Describa las condiciones necesarias para que una medida de variación tenga el valor positivo.

2.89 Una muestra contiene los datos {1, 3, 5, 6, 10}.

a. Use la fórmula (2.6) para hallar la varianza.

b. Use la fórmula (2.10) para hallar la varianza.

c. Compare los resultados de las partes a y b.

2.90 Considere la muestra 2, 4, 7, 8, 9. Encuentre lo siguiente:

a. Rango

b. Varianza s^2, usando la fórmula (2.6)

c. Desviación estándar, s

2.91 Considere la muestra 6, 8, 7, 5, 3, 7. Encuentre lo siguiente:

a. Rango

b. Varianza s^2, usando la fórmula (2.6)

c. Desviación estándar, s

2.92 Dada la muestra 7, 6, 10, 7, 5, 9, 3, 7, 5, 13, encuentre lo siguiente:

a. Varianza s^2 usando la fórmula (2.6)

b. Varianza s^2 usando la fórmula (2.10)

c. Desviación estándar, s

2.93 A quince estudiantes universitarios seleccionados al azar se les pide indicar el número de horas que durmieron la noche anterior. Los datos resultantes son 5, 6, 6, 8, 7, 7, 9, 5, 4, 8, 11, 6, 7, 8, 7. Encuentre lo siguiente:

a. La varianza s^2, usando la fórmula (2.6)

b. Varianza s^2, usando la fórmula (2.10)

c. La desviación estándar, s

2.94 Una muestra aleatoria de 10 de los corredores NASCAR de la Copa Nextel 2005 produjo las siguientes edades: 33, 48, 41, 29, 40, 48, 44, 42, 49, 28.

a. Encuentre el rango.

b. Encuentre la varianza.

c. Encuentra la desviación estándar.

2.95 Sumar (o restar) el mismo número de cada valor de un conjunto de datos no afecta las medidas de variabilidad para ese conjunto de datos.

a. Encuentre la varianza de este conjunto de datos anuales de calefacción-grados-día: 6017, 6173, 6275, 6350, 6001, 6300.

b. Encuentre la varianza de este conjunto de datos (obtenida al restar 6000 de cada uno de los valores de la parte a): 17, 173, 275, 350, 1, 300.

2.96 Un aspecto de la belleza de un paisaje es su variabilidad. A continuación aparecen las elevaciones (en pies sobre el nivel del mar) de 12 ciudades seleccionadas al azar de la región de los Finger Lakes del norte del estado de Nueva York.

559	815	767	668	651	895
1106	1375	861	1559	888	1106

Fuentes: http://www.city-data.com

a. Encuentre la media.

b. Encuentre la desviación estándar.

2.97 A los reclutas de una academia de policía se les requirió someterse a una prueba que mide la capacidad para hacer ejercicio. La capacidad para hacer ejercicio (en minutos) se obtuvo de cada uno de 20 reclutas:

25	27	30	33	30	32	30	34	30	27
26	25	29	31	31	32	34	32	33	30

a. Trace una gráfica de puntos de los datos.

b. Encuentre la media.

c. Encuentre el rango.

d. Encuentre la varianza.

e. Encuentre la desviación estándar.

f. Usando la gráfica de puntos de la parte a, trace una línea que represente el rango. A continuación trace una línea que se inicie en la media con una longitud que represente el valor de la desviación estándar.

g. Describa cómo están relacionados la distribución de datos, el rango y la desviación estándar.

2.98 La revista *Better Roads* publicó el porcentaje de puentes interestatales y de propiedad del estado que estaban estructuralmente deficientes, o funcionalmente obsoletos (%SD/FO), para cada uno de los estados

de Estados Unidos en 2003. (Los porcentajes se expresan en forma decimal, por ejemplo 0.20 = 20%).

Estado	SD/FO*	Estado	SD/FO*	Estado	SD/FO*
AK	0.20	AL	0.22	AR	0.20

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: Better Roads, noviembre de 2003.

SD/FO = estructuralmente deficiente o funcionalmente obsoleto

a. Construya un histograma.
b. La variable "%SD/FO" ¿parece tener una distribución aproximadamente normal?
c. Calcule la media.
d. Encuentre la mediana.
e. Encuentre el rango.
f. Encuentre la desviación estándar.

Retenga estas soluciones para usarlas en el ejercicio 2.125 en la p. 105.)

2.99 Una medida de la operación de líneas aéreas se refiere a porcentajes totales de vuelos a tiempo. Para enero de 2005, los porcentajes de llegadas a tiempo de vuelos nacionales en 31 de los aeropuertos más grandes de Estados Unidos fueron como sigue:

ATL	69.1	BWI	74.0	BOS	62.1

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: U.S. Department of Transportation, Bureau of Transportation Statistics

a. Encuentre el rango y la desviación estándar para los porcentajes de llegadas a tiempo.
b. Trace líneas en el diagrama de tallo y hoja dibujado al contestar el ejercicio 2.73 que representa el rango y desviación estándar. Recuerde: la desviación estándar es una medida de la dispersión alrededor de la media.
c. Describa la relación entre la distribución de los datos, el rango, y la desviación estándar.

2.100 Considere estos dos conjuntos de datos:

Conjunto 1	46	55	50	47	52
Conjunto 2	30	55	65	47	53

Ambos conjuntos tienen la misma media, 50. Compare estas medidas para ambos conjuntos $\sum(x-\bar{x})$, $\sum|x-\bar{x}|$, $SS(x)$, y rango. Comente sobre el significado de estas comparaciones.

2.101 Comente sobre el enunciado: "La pérdida media para clientes en el First State Bank (que no estaba asegurado) fue de \$150. La desviación estándar de las pérdidas fue −\$125".

2.102 Empiece con $x = 100$ y sume cuatro valores x para hacer una muestra de cinco datos tales que:

a. $s = 0$
b. $0 < s < 1$
c. $5 < s < 10$
d. $20 < s < 30$

2.103 Cada una de dos muestras tiene una desviación estándar de 5. Si los dos conjuntos de datos se agrupan en un conjunto de 10 datos, ¿la nueva muestra tendrá una desviación estándar que sea menor, igual o mayor que la desviación estándar original de 5? Para justificar su respuesta, haga dos conjuntos de datos, cada uno con una desviación estándar de 5. Incluya los cálculos.

2.104 El ejercicio que se muestra en la siguiente ilustración compara las medias y las desviaciones estándar con sus correspondientes histogramas. ¿Qué puede indicar de la ilustración?

2.6 Medidas de posición

Las **medidas de posición** se utilizan para describir la posición que un dato específico posee en relación con el resto de los datos cuando están en orden por categorías. *Cuartiles* y *percentiles* son dos de las medidas de posición más populares.

Cuartiles: son valores de la variable que dividen los datos ordenados en cuartos; cada conjunto de datos tiene tres cuartiles. El *primer cuartil*, Q_1, es un número tal que a lo sumo 25% de datos son menores en valor que Q_1 y a lo suma 75% son mayores. El *segundo cuartil* es la mediana. El *tercer cuartil*, Q_3, es un número tal que a lo sumo 75% de los datos son menores en valor que Q_3 y a lo sumo 25% son mayores. (Vea la figura 2.24.)

FIGURA 2.24
Cuartiles

El procedimiento para determinar los valores de los cuartiles es el mismo que para percentiles y se muestra en la siguiente descripción de *percentiles*. Recuerde que sus datos deben estar clasificados de bajo (L) a alto (H).

Percentiles: son los valores de la variable que dividen un conjunto de datos clasificados en 100 subconjuntos iguales; cada conjunto de datos tiene 99 percentiles (vea la figura 2.25). El k-ésimo percentil, P_k, es un valor tal que a lo sumo k% de los datos son menores en valor que P_k y a lo sumo $(100 - k)$% de los datos son mayores (vea la figura 2.26).

FIGURA 2.25 Percentiles

FIGURA 2.26 *k*-ésimo percentil

Notas:

1. El primer cuartil y el 25avo percentil son lo mismo; es decir, $Q_1 = P_{25.}$ También, $Q_3 = P_{75}$.
2. La mediana, el segundo cuartil, y el 50avo percentil son todos lo mismo: $\tilde{x} = Q_2 = P_{50}$. Por tanto, cuando se nos pida hallar P_{50} o Q_2, usamos el procedimiento para hallar la mediana.

El procedimiento para determinar el valor de cualquier k-ésimo percentil (o cuartil) comprende cuatro pasos básicos como se indica en el diagrama de la figura 2.27. El ejemplo 2.12 se muestra el procedimiento.

FIGURA 2.27
Procedimiento para hallar P_k

EJEMPLO 2.12 Cálculo de cuartiles y percentiles

Usando la muestra de 50 calificaciones del examen final de estadística elemental que aparecen en la tabla 2.15, encuentre el primer cuartil, Q_1; el 58-avo percentil, P_{58}; y el tercer cuartil, Q_3.

TABLA 2.15

Calificaciones sin procesar para examen de estadística

60	47	82	95	88	72	67	66	68	98	90	77	86
58	64	95	74	72	88	74	77	39	90	63	68	97
70	64	70	70	58	78	89	44	55	85	82	83	
72	77	72	86	50	94	92	80	91	75	76	78	

SOLUCIÓN

PASO 1 Ordene los datos: puede ser formulada una lista ordenada (vea tabla 2.16), o puede usarse una gráfica que muestre los datos clasificados. La gráfica de puntos y el diagrama de tallo y hoja son útiles para este propósito. El diagrama de tallo y hoja es especialmente útil, porque da números de profundidad contados desde ambos extremos cuando se genera en computadora (vea figura 2.28). El paso 2 es el mismo para las tres estadísticas.

Encuentre Q_1:

PASO 2 Encuentre $\frac{nk}{100}$: $\frac{nk}{100} = \frac{(50)(25)}{100} = \mathbf{12.5}$

($n = 50$ y $k = 25$, porque $Q_1 = P_{25}$.)

TABLA 2.16

Datos clasificados: Calificaciones de examen

39	64	72	78	89
44	66	72	80	90
47	67	74	82	90
50	68	74	82	91
55	68	75	83	92
58	70	76	85	94
58	70	77	86	95
60	70	77	86	95
63	72	77	88	97
64	72	78	88	98

FIGURA 2.28 Final Calificaciones de examen final

PASO 3 Encuentre la profundidad de Q_1:$d(Q_1) = \mathbf{13}$ (porque 12.5 contiene una fracción, **B** es el siguiente entero más grande, 13.)

PASO 4 Encuentre Q_1:Q_1 es el 13avo valor, contando desde L (vea la tabla 2.16 o figura 2.28), $Q_1 = \mathbf{67}$

Encuentre P_{58}:

PASO 2 Encuentre entre $\frac{nk}{100}$: $\frac{nk}{100} = \frac{(50)(58)}{100} = \mathbf{29}$ ($n = 50$ y $k = 58$ para

PASO 3 Encuentre la profundidad de P_{58}: $d(P_{58}) = \mathbf{29.5}$ (como **A** = 29, un entero, sume 0.5 y use 29.5.)

PASO 4 Encuentre P_{58}: P_{58} es el valor que está a la mitad entre los valores de las 29ava y 30ava piezas de datos, contando desde L (ver la tabla 2.16 o la figura 2.28), y

$$P_{58} = \frac{77 + 78}{2} = \mathbf{77.5}$$

Por tanto, puede decirse que "a lo más, 58% de las calificaciones de examen son menores en valor que 77.5". Esto es equivalente a decir que "a lo más, 42% de las calificaciones de examen son mayores en valor a 77.5."

Técnica opcional: cuando k es mayor a 50, reste k de 100 y use $(100 - k)$ y ponga k en el paso 2. La profundidad se cuenta entonces desde el dato más alto, H.

Encuentre Q_3, usando la técnica opcional:

PASO 2 Encuentre $\frac{nk}{100}$: $\frac{nk}{100} = \frac{(50)(25)}{100} = \mathbf{12.5}$ ($n = 50$ y $k = 75$, donde $Q_3 = P_{75}$, y $k > 50$; use $100 - k = 100 - 75 = 25$.)

PASO 3 Encuentre la profundidad de Q_3 desde H

PASO 4 Encuentre la profundidad de Q_3: Q_3 es el 13avo valor, contando desde H (vea tabla 2.16 o figura 2.28), $Q_3 = \mathbf{86}$

Por tanto, puede decirse que "a lo sumo, 75% de las calificaciones de examen son menores en valor a 86." Esto también equivale a decir que "a lo sumo, 25% de las calificaciones de examen son mayores en valor a 86".

Una medida adicional de tendencia central, el *cuartil medio*, puede ahora definirse.

Cuartil medio: El valor número que está al centro entre el primer cuartil y el tercer cuartil.

$$\text{cuartil medio} = \frac{Q_1 + Q_3}{2} \tag{2.11}$$

EJEMPLO 2.13 Para hallar el cuartil medio

Encuentre el cuartil medio para el conjunto de 50 calificaciones de examen dadas en el ejemplo 2.12.

SOLUCIÓN

$Q_1 = 67$ y $Q_3 = 86$, como se ve en el ejemplo 2.12. Así,

$$\text{cuartil medio} = \frac{Q_1 + Q_3}{2} = \frac{67 + 86}{2} = \mathbf{76.5}$$

La mediana, el rango central, y el cuartil medio no son necesariamente el mismo valor. Cada uno es el valor central, pero por diferentes definiciones de "medio". La figura 2.29 resume la relación de estas tres estadísticas como se aplican a las calificaciones de 50 exámenes del ejemplo 2.12.

FIGURA 2.29
Calificaciones de examen final

Un *resumen de cinco números* es muy eficiente para describir un conjunto de datos. Es información fácil de obtenerse y es muy informativa para el lector.

Resumen de cinco números: el resumen de cinco números está compuesto de lo siguiente:

1. L, el valor más pequeño del conjunto de datos
2. Q_1, el primer cuartil (también llamado P_{25}, el 25avo percentil)
3. $\bar{x}$, la mediana
4. Q_3, el tercer cuartil (también llamado P_{75}, el 75avo percentil)
5. H, el valor más grande del conjunto de datos

El resumen de cinco números para el conjunto de 50 calificaciones de examen del ejemplo 2.12 es

39	67	75.5	86	98
L	Q_1	$\tilde{x}$	Q_3	H

Observe que estos cinco valores numéricos dividen el conjunto de datos en cuatro subconjuntos, con un cuarto de los datos en cada subconjunto. Del resumen de cinco números, podemos observar cuánto se dispersan los datos en cada uno de los cuartos. Ahora podemos definir una medida adicional de dispersión.

Rango de intercuartil: es la diferencia entre los cuartiles primero y tercero. Es el rango de 50% central de los datos.

El resumen de cinco números es todavía más informativo cuando se muestra en un diagrama trazado a escala. Un diagrama gráfico que logra esto se conoce como *diagrama de caja y bigotes*.

Diagrama de caja y bigotes: es una representación gráfica del resumen de cinco números. Los cinco valores numéricos (más pequeño, primer cuartil, mediana, tercer cuartil, y más grande) están ubicados en una escala ya sea vertical u horizontal.

La caja se usa para describir la mitad central de los datos que está entre dos cuartiles. Los bigotes son segmentos de recta que se usan para describir la otra mitad de los datos: un segmento de recta representa el cuarto de los datos que es menor en valor que el primer cuartil, y un segundo segmento de recta representa el cuarto de los datos que es mayor en valor que el tercer cuartil.

La figura 2.30 es un diagrama de caja y bigotes de las 50 calificaciones de examen.

FIGURA 2.30
Diagrama de caja y bigotes

INSTRUCCIONES DE TECNOLOGÍA: PERCENTILES

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:     Data > Sort . . .
Enter:      Sort column(s): C1  By column: C1
Select:     Store sorted data in: Column(s) of current worksheet
Enter:      C2 > OK
```

Una lista ordenada de datos se obtendrá en C2. Determine la posición de profundidad y localice el percentil deseado.

Excel Introduzca los datos en la columna A y active una celda para la respuesta; luego continúe con:

```
Choose:     Insert Function, fx > Statistical > PERCENTILE > OK
Enter:      Array:  (A2:A6 or select cells)
            k: K (desired percentile; ex. .95, .47)
```

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:     STAT > EDIT > 2:SortA(
Enter:      L1
Enter:      percentile × sample size (ex. .25 × 100)
Based on product, determine the depth position; then continue with:
Enter:      L1(depth position) > Enter
```

INSTRUCCIONES DE TECNOLOGÍA: RESUMEN DE CINCO NÚMEROS

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:      Stat > Basic Statistics > Display Descriptive Statistics . . .
Enter:       Variables: C1 > OK
```

Excel Introduzca los datos en la columna A; luego continúe con:

```
Choose:      Tools > Data Analysis* > Descriptive Statistics > OK
Enter:       Input Range:  (A2:A6 or select cells)
Select:      Labels in First Row (if necessary)
             Output Range
               Enter:  (B1 or select cell)
Select:      Summary Statistics > OK
To make output readable:
Choose:      Format > Column > Auto?t Selection
```

*Si el análisis de datos no se muestra en la herramienta menú, vea página 62.

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:      STAT > CALC > 1:1-VAR STATS
Enter:       L1
```

INSTRUCCIONES DE TECNOLOGÍA: DIAGRAMA DE CAJA Y BIGOTES

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:      Graph > Boxplot...  > One Y, Simple > OK
Enter:Graph variables:        C1
Optional:
Select:      Labels > Titles/Footnoes
Enter:       your title, footnotes > OK
Select:      Scale > Axes and Ticks
Select:      Transpose value and category scales  > OK > OK
```

Para gráficas de caja múltiples, introduzca un conjunto de datos adicional en C2; luego haga como se describe aquí y:

```
Choose:      Graph > Boxplot...  > Multiple Y's. Simple > OK
Enter:       Graph variables: C1 C2 > OK
Optional:    See above.
```

Excel Introduzca los datos en la columna A; luego continúa con:

```
Choose:      Tools > Data Analysis Plus* > BoxPlot > OK
Enter:       (A2:A6 or select cells)
```

Para editar la gráfica de caja, repase las opciones mostradas con histogramas de edición de la página 62.

*Data Analysis Plus es un conjunto de macros estadísticos para EXCEL.

TI-83/84 Plus Introduzca los datos en L1; luego continúe con:

```
Choose:      2nd > STAT PLOT >
             1:Plot1 . . .
Choose:      ZOOM > 9:ZoomStat > TRACE >>>
```

Si los puntos medios de clase están en L1 y las frecuencias en L2, haga como aquí se describe excepto para:

```
Enter:       Freq: L2
```

Para gráficas de caja múltiples, introduzca un conjunto de datos adicional en L2 y L3; haga como se describe aquí y:

```
Choose:      2nd > STAT PLOT > 2:Plot2 . . .
```

La posición de un valor específico se puede medir en términos de la media y desviación estándar usando la *calificación estándar*, comúnmente llamada *calificación z*.

Calificación estándar, o calificación *z*: es la posición que un valor particular de x tiene respecto a la media, medido en desviaciones estándar. La calificación z se encuentra con la fórmula

$$z = \frac{\text{valor} - \text{media}}{\text{desviación estándar}} = \frac{x - \bar{x}}{s} \tag{2.12}$$

EJEMPLO 2.14 **Para hallar calificaciones z**

Encuentre las calificaciones estándar para (a) 92 y (b) 72 respecto a una muestra de calificaciones de examen que tienen una calificación media de 75.9 y una desviación estándar de 11.1.

SOLUCIÓN

a. $x = 92, \bar{x} = 75.9, s = 11.1$. Así, $z = \dfrac{x - \bar{x}}{s} = \dfrac{92 - 75.9}{11.1} = \dfrac{16.1}{11.1} = \mathbf{1.45.}$

b. $x = 72, \bar{x} = 75.9, s = 11.1$. Así, $z = \dfrac{x - \bar{x}}{s} = \dfrac{72 - 75.9}{11.1} = \dfrac{-3.9}{11.1} = \mathbf{-0.35.}$

Esto significa que la calificación de 92 está a aproximadamente 1.5 desviaciones estándar arriba de la media, y la calificación de 72 está aproximadamente a un tercio de una desviación estándar debajo de la media.

Notas:

1. Típicamente, el valor calculado de z se redondea al centésimo más cercano.
2. Las calificaciones z típicamente varían en valor de alrededor de –3.00 a +3.00.

Debido a que la calificación z es una media de posición relativa respecto a la media, se puede usar para ayudarnos a comparar dos calificaciones bruta que provengan de poblaciones separadas. Por ejemplo, supongamos que el lector desea comparar una calificación que haya recibido en un examen con la calificación de una amiga en un examen comparable en su curso. Usted recibió una calificación bruta de 45 puntos; ella obtuvo 72 puntos. ¿Es mejor la calificación de ella? Necesitamos más información antes que podamos sacar una conclusión. Suponga que la media del examen que usted tomó era de 38 y la media del examen de ella era 65. Las calificaciones están ambas 7 puntos arriba de la media, pero todavía no podemos sacar una conclusión definitiva. La desviación estándar del examen que usted tomó era de 7 puntos, y era de 14 puntos sobre el examen de su amiga. Esto significa que la calificación de usted es 1 desviación estándar arriba de la media ($z = 1.0$), mientras que la calificación de su amiga está sólo 0.5 desviaciones estándar arriba de la media ($z = 0.5$). La calificación de usted tiene la posición relativa "mejor", de tal forma que se concluye que es mejor que la de su amiga. (De nueva cuenta, esto es hablando desde un punto de vista relativo.)

INSTRUCCIONES DE TECNOLOGÍA: COMANDOS ADICIONALES

MINITAB (Versión 14)

Introduzca los datos en C1; entonces:
Para ordenar los datos en orden ascendente y guardarlos en C2, continúe con:

```
Choose:     Data > Sort . . .
Enter:      Sort column(s): C1 By column: C1
Select:     Store sorted data in: Column(s) of current worksheet
Enter:      C2 > OK
```

Para formar una distribución de frecuencia no agrupada de datos enteros, continúe con:

```
Choose:     Stat > Tables > Tally Individual Variables
Enter:      Variables: C1
Select:     Counts > OK
```

Para imprimir datos en la ventana de sesión, continúe con:

```
Choose:     Data > Display Data
Enter:      Columns to display: C1 or C1 C2 or C1–C2 > OK
```

Excel

Introduzca los datos en la columna A; luego continúe con lo siguiente para ordenar los datos:

```
Choose:     Data > Sort
Enter:      Sort by: (A2:A6 or select cells)
Select:     Ascending or Descending
            Header row or No header row
```

TI-83/84 Plus

Introduzca los datos en L1; luego continúe con lo siguiente para ordenar los datos:

```
Choose:     2nd > STAT > OPS > 1:SortA(
Enter:      L1
```

Para formar una distribución de frecuencia de los datos en L1, continúe con:

```
Choose:     PRGM > EXEC > FREQDIST*
Enter:      L1 > ENTER
            LW BOUND = first lower class boundar
```

```
UP BOUND = last upper class boundary
WIDTH = class width (use 1 for ungrouped distribution)
```

*El programa 'FREQDIST' es uno de muchos programas disponibles para descargarse desde un sitio web. Vea instrucciones específicas en la página 42.

INSTRUCCIONES DE TECNOLOGÍA: GENERAR MUESTRAS ALEATORIAS

MINITAB (Versión 19) Los datos se ponen en C1:

```
Choose:     Calc > Random Data > {Normal, Uniform, Integer, etc.}
Enter:      Generate: K rows of data
            Store in column(s): C1
            Population parameters needed: (μ, σ, L, H, A, or B)
            > OK
            (Required parameters will vary depending on the distribution)
```

Excel

```
Choose:     Tools > Data Analysis > Random Number Generation > OK
Enter:      Number of Variables: 1
            Number of Random Numbers: (desired quantity)
Select:     Distribution: Normal, Integers, or others
Enter:      Parameters: (μ, σ, L, H, A, or B)
            (Required parameters will vary depending on the distribution.)
Select:     Output Range
Enter:      (A1 or select cell)
```

TI-83/84 Plus

```
Choose:     STAT > 1:EDIT
Highlight:  L1
Choose:     MATH > PRB > 6:randNorm(or5:randInt(
Enter:      μ, σ, # of trials or L, H, # of trials
```

INSTRUCCIONES DE TECNOLOGÍA: SELECCIONAR MUESTRAS ALEATORIAS

MINITAB (Versión 14) Los datos existentes de dónde seleccionar deben estar en C1; luego continúe con:

```
Choose:     Calc > Random Data > Sample from Columns
Enter:      Sample: K rows from column(s): C1
            Store samples in: C2
Select:     Sample with replacement (optional) > OK
```

Excel Los datos existentes de dónde seleccionar deben estar en la columna A; luego continúe con:

```
Choose:     Tools > Data Analysis > Sampling > OK
Enter:      Input range: (A2:A10 or select cells)
Select:     Labels (optional)
            Random
              Enter:   Number of Samples: K
            Output range:
              Enter: (B1 or select cell)
```

CASO PRÁCTICO 2.15 El 85avo percentil de límite de velocidad

PASANDO CON 85% DEL FLUJO

Para el iniciado, la "regla del 85avo percentil" parece extraña, poco ortodoxa, y hasta puede ser temible, pero este punto de referencia de límite de velocidad ha guiado a ingenieros de tráfico durante décadas e incluso es reconocido como política oficial en muchas jurisdicciones gubernamentales. La idea es que los límites de velocidad máxima deben establecerse de manera que 85% de los vehículos en un tramo particular de carretera estén en ese límite o abajo del mismo. Según políticas en California, los ingenieros de tráfico rutinariamente miden la rapidez con que circulan los automovilistas y luego establecen el límite en el 85avo percentil de la velocidad de tráfico.

"El razonamiento es que 85% de los conductores circulen razonablemente y 15% no lo hagan", dijo David Roseman, ingeniero de tráfico de Long Beach. "Por tanto debemos estar diseñando nuestras velocidades para dar cabida a conductores razonables. Tom Jones, ingeniero principal de tráfico para la ciudad de Los Ángeles, agrega "La regla del 85avo percentil fue establecida hace muchos años. Es un criterio de diseño [sic], pero no quiere decir que necesariamente sea bueno".

Los grupos que están a favor de la seguridad odian la regla del 85avo percentil, porque piensan que acelerar es un riesgo serio y creciente en carreteras. En efecto, la policía hace menos patrullajes de rutina del tránsito y las velocidades están aumentando, según estudios publicados por grupos de seguridad. Barbara Harsha, directora ejecutiva de la Governors Highway Safety Assn. en Washington, por ejemplo, se preocupa de que la regla del 85avo percentil pueda usarse para legitimar un aumento de velocidad inseguro. Cuando un tráfico congestionado no limita la velocidad, muchas secciones de las autopistas del sur de California tienen velocidades promedio de más de 80 mph, bastante más que el límite legal. Se hace muy poco caso de los límites fijados de 25 mph en calles residenciales, según estudios de tránsito en algunos sectores citadinos.

"Eso sólo muestra que los límites legales de velocidad son demasiado bajos", dice Chad Dornsife de la National Motorists Association, grupo que representa a personas generalmente no satisfechas y con frecuencia indignadas por los reglamentos de tránsito y aplicación de leyes policíacas. Dice que los límites bajos de velocidad incorrectamente establecidos en realidad aumentan el número de accidentes y cuestan vidas, porque estimulan velocidades desiguales y crean riesgos. Él dice, por ejemplo, que cuando Montana impuso límites de velocidad por primera vez, se duplicaron los accidentes mortales. Un problema secundario que cita Dornsife, respecto a los límites artificialmente bajos de velocidad, es que los intervalos de luz ámbar a veces están basados en límites establecidos, lo cual deja poco tiempo para que autos rápidos se detengan para el cambio de luz antes de llegar a un crucero. Eso, dice Dornsife, crea choques en cruceros. "Todas las generaciones que han pasado por esto no creen en la regla del 85avo percentil," agrega. "A la comunidad que aplica la ley no le gusta la regla del 85avo percentil porque aplican menos multas. A los nuevos ingenieros de tráfico ni siquiera se les enseña la regla del 85avo percentil."

Fuente: Ralph Vartabedian, columnista de *Los Angeles Times*, 9 de marzo, 2005, http://www.latimes.com/classified/automotive/highway1/la-hy-wheels9mar09,1,6721856.story?ctrack=2&cset=true

¿SABÍA USTED...?

Los automovilistas en el estado de Nueva York recorrieron un total de 135 046 000 000 millas en carreteras de ese estado en 2003. Eso es más de 5.4 millones de viajes alrededor de nuestro planeta en el ecuador. ¡Un largo recorrido!

SECCIÓN 2.6 EJERCICIOS

2.105 En la tabla de calificaciones de examen de la tabla 2.16 que se muestra en la página 94 busque lo siguiente.

a. Usando el concepto de profundidad, describa la posición de 91 en el conjunto de 50 calificaciones de examen en dos formas diferentes.

b. Encuentre P_{20} y P_{35}, para las calificaciones de examen de la tabla 2.16 en la página 94.

c. Encuentre P_{80} y P_{95} para las calificaciones de examen en la tabla 2.16.

2.106 A continuación están las calificaciones de la American College Test (ACT) obtenidas por los 25 miembros de un grupo egresados de una preparatoria local:

21	24	23	17	31	19	19	20	19	25	17	23	16
21	20	28	25	25	21	14	19	17	18	28	20	

a. Trace una gráfica de puntos de las calificaciones del ACT.

b. Usando el concepto de profundidad, describa la posición del 24 en el conjunto de 25 calificaciones del ACT en dos formas diferentes.

c. Encuentre P_5, P_{10}, y P_{20} para las calificaciones del ACT.

d. Encuentre P_{99}, P_{90}, y P_{80} para las calificaciones del ACT.

2.107 Los salarios anuales (en $100) del profesorado de jardín de niños y primaria, empleados en una de las escuelas elementales del distrito escolar, son los siguientes:

574	434	455	413	391	471	458	269	501
326	367	433	367	495	376	371	295	317

a. Trace una gráfica de puntos de los salarios.

b. Usando el concepto de profundidad, describa la posición del 295 en el conjunto de 18 salarios en dos formas diferentes.

c. Encuentre Q_1 para estos salarios.

d. Encuentre Q_3 para estos salarios.

2.108 Quince países se seleccionaron al azar de la lista de la *World Factbook 2004* de países, y la tasa de mortalidad infantil por 1000 nacimientos fue como sigue:

6.38	101.68	9.48	69.18	64.19	3.73	21.31	52.71
13.43	29.64	15.24	5.85	11.74	9.67	8.68	

Fuente: The World Factbook 2004

a. Encuentre los cuartiles primero y tercero para la tasa de mortalidad infantil por cada 1000 nacimientos.

b. Encuentre el cuartil medio.

2.109 Los siguientes datos son las producciones (en libras) de lúpulos:

3.9	3.4	5.1	2.7	4.4	7.0	5.6	2.6	4.8	5.6
7.0	4.8	5.0	6.8	4.8	3.7	5.8	3.6	4.0	5.6

a. Encuentre los cuartiles primero y cuarto de las producciones.

b. Encuentre el cuartil medio.

c. Encuentre y explique los percentiles P_{15}, P_{33}, y P_{90}.

2.110 Un estudio de investigación de la destreza manual requirió determinar el tiempo necesario para completar un trabajo. El tiempo para cada una de 40 personas con discapacidades se muestra a continuación (los datos están ordenados):

7.1	7.2	7.2	7.6	7.6	7.9	8.1	8.1	8.1	8.3	8.3	8.4	8.4	8.9
9.0	9.0	9.1	9.1	9.1	9.1	9.4	9.6	9.9	10.1	10.1	10.1	10.2	
10.3	10.5	10.7	11.0	11.1	11.2	11.2	11.2	12.0	13.6	14.7	14.9	15.5	

a. Encuentre Q_1.
b. Encuentre Q_2.
c. Encuentre Q_3.
d. Encuentre Q_{95}.
e. Encuentre el resumen de 5 números.
f. Trace el diagrama de caja y bigotes.

2.111 Elabore un diagrama de caja y bigotes para el conjunto de datos con el resumen de cinco números 42-62-72-82-97.

2.112 El U.S. Geological Survey recolectó datos de depositación atmosférica en las Montañas Rocallosas. Parte del proceso de muestreo era determinar la concentración de iones de amonio (en porcentajes). A continuación vea los resultados de las 52 muestras:

2.9	4.1	2.7	3.5	1.4	5.6	13.3	3.9	4.0
2.9	7.0	4.2	4.9	4.6	3.5	3.7	3.3	5.7
3.2	4.2	4.4	6.5	3.1	5.2	2.6	2.4	5.2
4.8	4.8	3.9	3.7	2.8	4.8	2.7	4.2	2.9
2.8	3.4	4.0	4.6	3.0	2.3	4.4	3.1	5.5
4.1	4.5	4.6	4.7	3.6	2.6	4.0		

a. Encuentre Q_1.
b. Encuentre Q_2.
c. Encuentre Q_3.
d. Encuentre el cuartil medio.
e. Encuentre Q_{30}.
f. Encuentre el resumen de 5 números.
g. Trace el diagrama de caja y bigotes.

2.113 El equipo varonil de baloncesto de la NCAA "Big Dance" mete velocidad cada mes de marzo. Si se ve el porcentaje de graduación de estos atletas, no obstante, se encuentra que muchos equipos no califican, según un estudio publicado en marzo de 2005. A con-

tinuación veamos los porcentajes de graduación para 64 de los equipos del torneo de 2005.

Porcentajes de graduación, equipos varoniles 2005, Torneo de Baloncesto de la NCAA División I

40	38	100	55	44	58	30	11	40	19	43	27	0
64	75	58	54	44	40	11	50	30	40	71	92	53
33	29	40	25	33	43	25	27	58	47	55	60	8
17	17	0	40	25	14	67	45	33	20	15	50	27
29	57	36	45	45	73	15	100	67	44	57	55	

Fuente: 2004 NCAA Graduation-Rates Report

a. Trace una gráfica de puntos del porcentaje de graduación.

b. Elabore un diagrama de tallo y hoja de estos datos.

c. Encuentre el resumen de 5 números y trace un diagrama de caja y bigotes.

d. Encuentre P_5 y P_{95}.

e. Describa la distribución de porcentajes de graduación, asegurándose de incluir información captada en los incisos a a d.

f. ¿Hay equipos cuyos porcentajes de graduación parecen ser muy distintos del resto? ¿Cuántos? ¿Cuáles? Explique.

2.114 El porcentaje de accidentes mortales en carreteras nacionales en 2003 en Estados Unidos fue el más bajo desde que se inició este registro hace 29 años, pero estos números son todavía escalofriantes. El número de personas fallecidas en accidentes de tránsito en vehículos de motor, por estado, incluyendo el Distrito de Columbia, en 2003 es como sigue:

1001	294	135	471	462	262	439	668	203	943
95	142	293	928	1283	293	1491	512	1193	600
1120	67	1453	894	657	368	1531	1577	3675	394
627	3169	834	207	871	127	105	104	309	848
4215	1603	441	649	1232	747	1277	968	69	165
632									

Fuente: Road & Travel Magazine, 2004

a. Trace una gráfica de puntos de los datos de pérdidas humanas.

b. Elabore un diagrama de tallo y hoja de estos datos.

c. Encuentre el resumen de cinco números y dibuje un diagrama de caja y bigotes. Describa cómo se manejan los tres datos de valores grandes.

d. Encuentre P_{10} y P_{90}.

e. Describa la distribución del número de fallecimientos por estado, asegurándose de incluir información que haya aprendido en los incisos a a d.

f. ¿Por qué no podría ser equitativo sacar conclusiones acerca del nivel relativo de seguridad de carreteras, en los 50 estados y el Distrito de Columbia, con base en estos datos.

2.115 ¿Llegan siempre a tiempo los vuelos de líneas aéreas? El público en general piensa que siempre llega con cierta demora, pero, ¿en realidad es así? La U.S. Bureau of Transportation Statistics lleva registros y periódicamente informa lo que averigua. A continuación aparecen los porcentajes de llegadas a tiempo en los 31 aeropuertos más importantes de Estados Unidos, para el periodo del 1 de enero de 2004 al 31 de octubre de 2004.

ATL	73.55	BOS	78.38	BWI	80.91

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: U.S. Department of Transportation, Bureau of Transportation Statistics

a. Elabore una gráfica de puntos acerca de datos de operación a tiempo.

b. Trace un diagrama de tallo y hoja de estos datos.

c. Encuentre el resumen de cinco números y trace un diagrama de caja y bigotes.

d. Encuentre P_{10} y P_{20}.

e. Describa la distribución del porcentaje de llegadas a tiempo, asegurándose de incluir información aprendida en las partes de la a a la d.

f. ¿Por qué sería más probable que se hablara de 80 o 90% como porcentajes elevados de operación en lugar de 80 o 90% como porcentajes medios?

g. ¿Existen aeropuertos cuyos porcentajes de llegadas a tiempo parezcan muy diferentes al resto? ¿Cuántos? ¿Cuáles son? Explique.

2.116 Los estadios de béisbol de las ligas mayores varían en antigüedad, estilo, número de asientos y en muchos otros aspectos, pero, para los jugadores, el tamaño del campo es lo más importante. Supóngase que acordamos medir el tamaño del campo, usando para ello la distancia de la placa del *home* a la cerca del jardín central. A continuación aparece la distancia (en pies) a la cerca del jardín central en los 30 estadios más importantes de las ligas mayores.

422	405	400	400	400	402	404	435	399	410
400	400	400	400	408	401	395	410	410	401
420	408	405	410	402	415	400	404	405	400

Fuente: http://mlb.mlb.com

a. Construya un histograma.

b. El rango del intercuartil se describe por los límites de 50% medios de los datos, Q_1 y Q_3. Encuentre el rango intercuartil.

c. ¿Existen algunos campos que parezcan ser considerablemente menores o mayores que otros?

d. ¿Hay una gran diferencia en el tamaño de estos 30 campos, según la distancia al jardín central? Justifique su respuesta con evidencia estadística.

2.117 ¿Qué propiedad necesita la distribución para que la mediana, el rango central y el cuartil medio sean del mismo valor?

2.118 Henry Cavendish, químico y físico inglés (1731-1810), realizó muchos de sus experimentos usando mediciones cuantitativas. Él fue el primero en medir con precisión la densidad de nuestro planeta. A continuación aparecen 29 mediciones (ordenadas, para comodidad del lector) de la densidad de la Tierra hechas por Cavendish en 1798 que para ello usó una balanza de torsión. La densidad se presenta como un múltiplo de la densidad del agua. (Las mediciones son en g/cm^3.)

4.88	5.07	5.10	5.26	5.27	5.29	5.29	5.30	5.34	5.34
5.36	5.39	5.42	5.44	5.46	5.47	5.50	5.53	5.55	5.57
5.58	5.61	5.62	5.63	5.65	5.68	5.75	5.79	5.85	

Fuente Los datos e información descriptiva están basados en material de "Do robust estimators work with real data?" por Stephen M. Stigler, *Annals of Statistics* 5(1977), 1055-1098.

a. Describa el conjunto de datos al calcular la media, mediana y desviación estándar.

b. Construya un histograma y explique cómo demuestra los valores de la estadística descriptiva de la parte a.

c. Encuentre el resumen de cinco números.

d. Construya un diagrama de caja y bigotes y explique cómo demuestra los valores de la estadística descriptiva de la parte c.

e. Con base en las dos gráficas, ¿qué "forma" es esta distribución de mediciones?

f. Suponiendo que las mediciones de densidad de la Tierra tengan una distribución aproximadamente normal, alrededor de 95% de los datos deben caer dentro de dos desviaciones estándar de la media. ¿Es cierto esto?

2.119 Encuentre la calificación z de las calificaciones de examen de 92 y 63 de un examen que tiene una media de 72 y desviación estándar de 12.

2.120 Una muestra tiene una media de 50 y una desviación estándar de 4.0. Encuentre la calificación z para cada valor de x:

a. $x = 54$ b. $x = 50$

c. $x = 59$ d. $x = 45$

2.121 Un examen produjo calificaciones con una calificación media de 74.2 y una desviación estándar de 11.5. Encuentre la calificación z para cada calificación x del examen:

a. $x = 54$ b. $x = 68$

c. $x = 79$ d. $x = 93$

2.122 Un examen aplicado en todo el país tiene una media de 500 y una desviación estándar de 100. Si su calificación estándar en este examen fue 1.8, ¿cuál fue su calificación del examen?

2.123 Una muestra tiene una media de 120 y una desviación estándar de 20.0. Encuentre el valor de x que corresponde a cada una de estas calificaciones estándar:

a. $z = 0.0$ b. $z = 1.2$

c. $z = -1.4$ d. $z = 2.05$

2.124

a. ¿Qué significa decir que $x = 152$ tiene una calificación estándar de $+1.5$?

b. ¿Qué significa decir que un valor particular de x tiene una calificación estándar de -2.1?

c. En general, ¿la calificación estándar es una medida de qué aspecto?

2.125 Considere el porcentaje de puentes interestatales y de propiedad del estado que estaban estructuralmente deficientes, o funcionalmente obsoletos (SD/FO), que aparecen en el ejercicio 2.98 de la página 91.

a. Omita los nombres de los estados y ordene los valores SD/FO en de menor a mayor, leyendo horizontalmente en cada renglón.

b. Construya una tabla resumen de cinco números y el correspondiente diagrama de caja y bigotes.

c. Encuentre el porcentaje del cuartil medio y el rango del intercuartil.

d. ¿Cuáles son las calificaciones z para California, Hawai, Nebraska, Oklahoma, y Rhode Island?

2.126 La evaluación de la American College Test (ACT) está diseñada para evaluar el desarrollo educacional general de estudiantes de preparatoria, así como su capacidad para completar trabajos a nivel de universidad. La tabla siguiente contiene la media y desviación estándar de calificaciones obtenidas por 1 171 460 estudiantes de preparatoria de los grupos egresados en 2004 que tomaron los exámenes del ACT.

2004	Inglés	Matemáticas	Lectura	Razonamiento en ciencias	Global
Media	20.4	20.7	21.3	20.9	20.9
Desviación estándar	5.9	5.0	6.0	4.6	4.8

Fuente: American College Testing

Convierta las siguientes calificaciones del ACT en calificaciones z para inglés y matemáticas. Compare el lugar entre los dos exámenes.

a. $x = 30$ b. $x = 23$ c. $x = 12$

d. Explique por qué las posiciones relativas en inglés y matemáticas cambiaron para las calificaciones del ACT de 30 y 12.

e. Si Jessica tuvo 26 en uno de los exámenes del ACT, ¿en cuál de ellos tendría la mejor calificación relativa posible? Explique por qué.

2.127 ¿Qué valor de x tiene la posición más alta respecto al conjunto de datos del cual proviene?

A: $x = 85$, donde la media es 72 y la desviación estándar es igual a 8.

B: $x = 93$, donde la media es 87 y la desviación estándar es igual a 5.

2.128 ¿Qué valor de x tiene la posición más baja respecto al conjunto de datos del cual proviene?

A: $x = 28.1$, donde $\bar{x} = 25.7$ y $s = 1.8$

B: $x = 39.2$, donde $\bar{x} = 34.1$ y $s = 4.3$

2.7 Interpretación y comprensión de la desviación estándar

La desviación estándar es una medida de variación (dispersión) en los datos. Se ha definido como un valor calculado con el uso de fórmulas. Aún así, el estudiante puede preguntarse qué es realmente y cómo se relaciona con los datos. Es una clase de medidor por el que podemos comparar la variabilidad de un conjunto de datos con la de otro. Esta "medida" particular puede comprenderse mejor al examinar dos enunciados que nos dicen cómo se relaciona la desviación estándar con los datos: la *regla empírica* y el *teorema de Chebyshev.*

La regla empírica y prueba de normalidad

Regla empírica: si una variable está normalmente distribuida, entonces (1) dentro de una desviación estándar de la media habrá aproximadamente 68% de los datos; (2) dentro de dos desviaciones estándar de la media, habrá aproximadamente 95% de los datos; y (3) dentro de tres desviaciones estándar de la media habrá aproximadamente 99.7% de los datos. (Esta regla aplica de manera específica a una distribución normal [en forma de campana], pero con frecuencia se aplica como guía interpretativa a cualquier distribución agrupada.)

La figura 2.31 muestra los intervalos de 1, 2 y 3 desviaciones estándar alrededor de la media de una distribución aproximadamente normal. Por lo general estas

FIGURA 2.31
Regla empírica

proporciones no se presentan de manera exacta en una muestra, pero los valores observados estarán cercanos cuando una muestra grande se tome de una población normalmente distribuida.

Si una distribución es aproximadamente normal, será casi simétrica y la media dividirá la distribución en dos (la media y la mediana son iguales en una distribución simétrica). Esto nos permite refinar la regla empírica, como se muestra en la figura 2.32.

FIGURA 2.32
Refinamiento de la regla empírica

La regla empírica se puede usar para determinar si un conjunto de datos está normalmente distribuido en forma aproximada. Demostremos esta aplicación al trabajar con la distribución de calificaciones de examen final que hemos estado usando en todo este capítulo. Se encontró que la media, $\bar{x}$, es 75.6, y la desviación estándar, s, fue 14.9. El intervalo de 1 desviación estándar debajo de la media, $\bar{x} - s$, a 1 desviación estándar arriba de la media, $\bar{x} + 2s$, es $75.6 - 14.9 = 60.7$ a $75.6 + 14.9 = 90.5$. Este intervalo (60.7 a 90.5) incluye 61, 62, 63, . . . ,89, 90. Con una inspección de los datos ordenados (tabla 2.16, p. 94), vemos que 35 de los 50 datos, es decir 70%, están dentro de 1 desviación estándar de la media. Además, $\bar{x} + 2s = 75.6 - (2)(149) = 75.6 - 29.8 = 45.8$ da $\bar{x} + 2s = 75.6 + 29.8 = 105.4$ el intervalo de 45.8 a 105.4. De los 50 datos, 48, o sea 96%, están dentro de dos desviaciones estándar de la media. Los 50 datos, o sea 100%, están incluidos dentro de 3 desviaciones estándar de la media (de 30.9 a 120.3). Esta información se puede poner en una tabla para comparación con los valores dados por la regla empírica (vea la tabla 2.17).

TABLA 2.17

Porcentajes observados contra la regla empírica

Intervalo	Porcentaje de regla empírica	Porcentaje encontrado
$\bar{x} - s$ a $\bar{x} + s$	$\approx$68	70
$\bar{x} - 2s$ a $\bar{x} + 2s$	$\approx$95	96
$\bar{x} - 3s$ a $\bar{x} + 3s$	$\approx$99.7	100

Los porcentajes encontrados están razonablemente cercanos a los pronosticados por la regla empírica. Al combinar esta evidencia con la forma del histograma (ver la figura 2.10, p. 60), podemos con seguridad decir que los datos de examen final están distribuidos normalmente en forma aproximada.

Hay otra forma de probar la normalidad, que es al trazar una gráfica de probabilidad (una ojiva dibujada en papel de probabilidad*) usando computadora o una calculadora de gráficas. Para nuestra ilustración, una gráfica de probabilidad de las calificaciones de examen final de estadística se ilustra en la figura 2.33. La prueba de normalidad, en este punto de nuestro estudio de estadística, es simplemente comparar la gráfica de los datos (la ojiva) con la recta trazada de la esquina inferior izquierda a la superior derecha de la gráfica. Si la ojiva es cercana a esta recta, se dice que la distribución es aproximadamente normal. La escala vertical empleada para construir la gráfica de probabilidad se ajusta de modo que la ojiva para una distribución exactamente normal trazará la recta. La ojiva de las calificaciones de examen sigue la recta de modo muy cercano, lo que sugiere que la distribución de las calificaciones de examen es aproximadamente normal.

FIGURA 2.33

Gráfica de probabilidad de calificaciones de un examen de estadística

Si el estudiante usa computadora, obtendrá información al determinar la normalidad. Esta información llega en forma de un valor p, y si su valor es mayor a

*En este asunto de probabilidad la escala vertical no es uniforme, se ha ajustado para considerar la forma de una distribución normal y sus porcentajes acumulativos.

0.05, se puede suponer que la muestra fue extraída de una distribución aproximadamente normal (si el valor $p \leq 0.05$, no normal). (El valor p se definirá en forma más completa en el capítulo 8, sección 8.5.)

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE NORMALIDAD

MINITAB (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:     Stat > Basic Statistics > Normality Test
Enter:      Variable: C1
            Title: your title > OK
```

Excel Excel usa una prueba de normalidad, no la gráfica de probabilidad.

Introduzca los datos en la columna A; luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Chi-Squared Test of Normality >
            OK
Enter:      Input Range: data (A1:A6 or select cells)
Select:     Labels (if column headings were used) > OK
```

Los valores esperados para una distribución normal están dados con relación a la distribución dada. Si el valor p es mayor a 0.05, entonces la distribución dada es aproximadamente normal.

TI-83/84 Plus Teclee los datos en L1; luego continúe con:

```
Choose:     Window
Enter:      at most the smallest data value, at least the largest data
            value, x scale, −5, 5, 1,1
Choose:     2nd > STAT PLOT > 1:Plot
```

Teorema de Chebyshev

En caso que los datos no muestren una distribución aproximadamente normal, el teorema de Chebyshev nos da información acerca de cuánto de los datos caerá dentro de intervalos centrados en la media para todas las distribuciones.

Teorema de Chebyshev: la proporción de cualquier distribución que se encuentre dentro de k desviaciones estándar de la media es al menos $1 - \frac{1}{k^2}$, donde k es cualquier número positivo mayor a 1. Este teorema aplica a todas las distribuciones de datos.

Este teorema indica que dentro de 2 desviaciones estándar de la media ($k = 2$), siempre se encontrará al menos 75% (es decir, 75% o más) de los datos:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0.75, \textbf{ al menos 75\%}$$

La figura 2.34 muestra una distribución agrupada que ilustra al menos 75%.

Si consideramos el intervalo encerrado por 3 desviaciones estándar en cualquier lado de la media ($k = 3$), el teorema dice que siempre encontraremos al menos 89% (es decir, 89% o más) de los datos:

$$1 - \frac{1}{k^2} = 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 0.89, \textbf{ al menos 89\%}$$

La figura 2.35 muestra una distribución agrupada que ilustra al menos 89%.

FIGURA 2.34 Teorema de Chebyshev $k = 2$

FIGURA 2.35 Teorema de Chebyshev con $k = 3$

Veamos de nuevo los resultados de la prueba de resistencia física aplicada a los estudiantes de tercer grado del ejercicio 2.45, página 70. Sus resultados de prueba aparecen a continuación en orden y se muestran en el histograma.

1	2	2	3	3	3	4	4	4	5	5	5	5	6	6	6
8	9	9	9	9	9	9	10	10	11	12	12	12	13	14	14
14	15	15	15	15	16	16	16	17	17	17	17	18	18	18	18
19	19	19	19	20	20	20	21	21	21	22	22	22	23	24	24

Algunas preguntas de interés son: ¿esta distribución satisface la regla empírica? ¿Se cumple como verdadero el teorema de Chebyshev? ¿Es esta distribución aproximadamente normal?

Para responder las dos primeras preguntas necesitamos hallar el porcentaje de datos en cada uno de los tres intervalos alrededor de la media. La media es 13.0, y la desviación estándar es 6.6.

media$\pm k$ (desv.est.)	Intervalo	Porcentaje encontrado	Empírico	Chebyshev
$13.0 \pm 1(6.6)$	6.4 a 19.6	$39/64 = 60.9\%$	68%	–
$13.0 \pm 2(6.6)$	−0.2 a 26.2	$64/64 = 100\%$	95%	Al menos 75%
$13.0 \pm 3(6.6)$	−6.8 a 32.8	$64/64 = 100\%$	99.70%	Al menos 89%

Se deja al lector la tarea de verificar los valores de la media, desviación estándar, los intervalos y los porcentajes.

Los tres porcentajes encontrados (60.9, 100, y 100) no se aproximan a los porcentajes de 68, 95, y 99.7 calculados en la regla empírica. Los dos porcentajes encontrados (100 y 100) concuerdan con el teorema de Chebyshev ya que son mayores a 75 y 89%. Recuerde, el teorema de Chebyshev se cumple para todas las distribuciones.

La prueba de normalidad, introducida en las páginas 108 y 109, da un valor p de 0.009, y junto con la distribución vista en el histograma y los tres porcentajes hallados, es razonable concluir que estos resultados de examen no están normalmente distribuidos.

SECCIÓN 2.7 EJERCICIOS

2.129 Las instrucciones para una tarea fácil incluyen el enunciado "La longitud debe ser no más de 25 palabras de 200". ¿Qué valores de x, número de palabras, satisfacen estas instrucciones?

2.130 La regla empírica indica que podemos esperar hallar la proporción de la muestra que está incluida entre lo siguiente:

a. $\bar{x} - s$ y $\bar{x} + s$ b. $\bar{x} - 2s$ y $\bar{x} - 2s$

b. $\bar{x} - 3s$ y $\bar{x} + s$

2.131 ¿Por qué es que la calificación z para un valor que pertenece a una distribución normal por lo general está entre −3 y +3?

2.132 La duración media de cierto neumático es 30 000 millas y la desviación estándar es 2 500 millas.

a. Si suponemos que las distancias están normalmente distribuidas, ¿aproximadamente qué porcentaje de estos neumáticos durará entre 22 500 y 37 500 millas?

b. Si no suponemos nada acerca de la forma de la distribución, ¿aproximadamente qué porcentaje de estos neumáticos tendrá una duración comprendida entre 22 500 y 37 500 millas?

2.133 El tiempo promedio de limpieza para el personal de una empresa de tamaño medio es 84.0 horas y la desviación estándar es 6.8 horas. Suponga que la regla empírica es apropiada.

a. ¿Qué proporción del tiempo tardará 97.6 horas o más en limpiar la planta?

b. ¿Dentro de qué intervalo es que el tiempo total de limpieza caerá 95% del tiempo?

2.134 a. ¿Qué proporción de una distribución normal es mayor a la media?

b. ¿Qué proporción está dentro de 1 desviación estándar de la media?

c. ¿Qué proporción es mayor que un valor que está 1 desviación estándar debajo de la media?

2.135 Usando una regla empírica, determine el porcentaje aproximado de una distribución normal que se espera caiga dentro del intervalo descrito.

a. Menor a la media

b. Más a 1 desviación estándar arriba de la media

c. Menos de 1 desviación estándar arriba de la media

d. Entre 1 desviación estándar debajo de la media y 2 desviaciones estándar arriba de la media

2.136 Según la regla empírica, casi todos los datos deben estar entre $(\bar{x} - 3s)$ y $(\bar{x} + 3s)$. El rango toma en cuenta todos los datos.

a. ¿Qué relación debe cumplirse (aproximadamente) entre la desviación estándar y el rango?

b. ¿Cómo se pueden usar los resultados de la parte a para calcular la desviación estándar en situaciones cuando se conoce el rango?

2.137 El teorema de Chebyshev garantiza que ¿cuál proporción de una distribución estará incluida entre lo siguiente?:

a. $\bar{x} - 2s$ y $\bar{x} + 2s$ b. $\bar{x} - 3s$ y $\bar{x} + 3s$

2.138 Según el teorema de Chebyshev, ¿qué proporción de una distribución estará dentro de $k = 4$ desviaciones estándar de la media?

2.139 El teorema de Chebyshev puede expresarse en una forma equivalente a la dada en la página 109. Por ejemplo, para decir "al menos 75% de los datos caerá dentro de 2 desviaciones estándar de la media" es equivalente a decir "al menos, 25% estará a más de 2 desviaciones estándar de la media".

a. A lo más, ¿qué porcentaje de una distribución estará 3 o más desviaciones estándar de la media?

b. A lo más, ¿qué porcentaje de una distribución estará 4 o más desviaciones estándar de la media?

2.140 Las calificaciones obtenidas por estudiantes en Estados Unidos son noticia, con frecuencia, y se saca toda clase de conclusiones con base en estas calificaciones. La ACT Assessment está diseñada para evaluar el desarrollo educacional general de estudiantes de preparatoria y su capacidad para completar un trabajo de nivel universitario. Una de las categorías probadas es el razonamiento en ciencias. La calificación ACT media para todos los egresados de preparatoria en 2004 en razonamiento en ciencias fue 20.9, con una desviación estándar de 4.6.

a. Según el teorema de Chebyshev, al menos ¿qué porcentaje de calificaciones ACT de egresados de preparatoria en razonamiento en ciencias está entre 11.7 y 30.1?

b. Si sabemos que las calificaciones del ACT están normalmente distribuidas, ¿qué porcentaje de calificaciones del ACT en razonamiento en ciencias está entre 11.7 y 30.1?

2.141 El primer día de clases en el semestre pasado, a 50 estudiantes se les preguntó la distancia entre sus casas y la universidad (a la milla más cercana.) A continuación veamos los datos resultantes:

6	5	3	24	15	15	6	2	1	3
5	10	9	21	8	10	9	14	16	16
10	21	20	15	9	4	12	27	10	10
3	9	17	6	11	10	12	5	7	11
5	8	22	20	13	1	8	13	4	18

a. Construya una distribución de frecuencia agrupada de los datos usando 1-4 como el primer grupo.

b. Calcule la media y la desviación estándar.

c. Determine los valores de $\bar{x} = 2s$, y determine el porcentaje de datos dentro de 2 desviaciones estándar de la media.

2.142 Una de las muchas cosas que la U.S. Census Bureau informa al público es el aumento en población para varias regiones geográficas dentro del país. El porcentaje de aumento en población para los 100 condados de más rápido crecimiento en Estados Unidos, del 1 de abril de 2000 al 1 de julio de 2003, es como sigue:

15.4	12.1	13.0	14.8	21.5

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.census.gov/counties/CO-EST2003-09.html

a. Calcule la media y desviación estándar.

b. Clasifique los datos en una lista ordenada.

c. Determine los valores de $\bar{x} \pm s$, $\bar{x} \pm 2s$, y $\bar{x} \pm 3s$, y determine el porcentaje de datos dentro de 1, 2, y 3 desviaciones estándar de la media.

d. Los porcentajes encontrados en la parte c, ¿concuerdan con la regla empírica? ¿Qué significa eso?

e. Los porcentajes encontrados en la parte c, ¿concuerdan con el teorema de Chebyshev? ¿Qué significa eso?

f. Construya un histograma y otra gráfica que usted elija. Las gráficas, ¿muestran una distribución que concuerde con sus respuestas en las partes d y e? Explique.

2.143 Cada año, a los aficionados del futbol colegial de la NCAA desean conocer quienes integrarán el próximo grupo de estudiantes de primer año, y que además serán jugadores. A continuación aparecen las estaturas (en pulgadas) de los mejores 100 jugadores de futbol de preparatoria para 2005, como les clasifica el analista reclutador Tom Lemming de ESPN.com:

75	70	71	75	76	76	70	72	70	75	75	68	73	75	74
74	73	75	77	75	72	77	73	73	72	71	78	79	80	74
73	79	78	77	73	74	74	72	73	75	68	72	72	73	72
70	76	73	74	76	74	76	74	78	75	77	77	78	74	73
76	70	76	77	77	70	73	75	76	73	75	76	78	75	71
78	75	76	77	78	75	76	74	74	79	73	74	76	71	74
71	76	76	74	76	76	76	72	76	73					

Fuente: ESPN.com

a. Construya un histograma y otra gráfica de su elección que muestre la distribución de estaturas.

b. Calcule la media y desviación estándar.

c. Ordene los datos en una lista clasificada.

d. Determine los valores de $\bar{x} \pm s$, $\bar{x} \pm 2s$, y $\bar{x} \pm 3s$, y determine el porcentaje de datos dentro de 1, 2, y 3 desviaciones estándar de la media.

e. Los porcentajes encontrados en la parte d, ¿concuerdan con la regla empírica?, ¿qué implica lo anterior? Explique.

f. Los porcentajes encontrados en la parte d, ¿concuerdan con el teorema de Chebyshev? ¿Qué significa eso?

g. ¿Las gráficas muestran una distribución que concuerde con sus respuestas en el inciso e)? Explique.

2.144 Cada año, a los aficionados al futbol colegial de la NCAA desean conocer el peso de los jugadores del grupo de novatos del año en curso. A continuación aparecen los pesos (en libras) de los mejores 100 jugadores de futbol colegial de Estados Unidos para 2005, como los clasifica el analista reclutador Tom Lemming de ESPN.com. Sólo porque estos datos fueron tomados de los mismos 100 jugadores de futbol del ejercicio 2.143 no significa que las distribuciones serán iguales. De hecho, son muy diferentes, como podremos ver.

207	220	218	215	215

Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: ESPN.com

a. Construya un histograma y otra gráfica de su elección que muestre la distribución de pesos.

b. Calcule la media y desviación estándar.

c. Ordene los datos en una lista clasificada.

d. Determine los valores de $\bar{x} \pm s$, $\bar{x} \pm 2s$, y determine el porcentaje de datos dentro de 1, 2, y 3 desviaciones estándar de la media.

e. Los porcentajes encontrados en la parte d, ¿concuerdan con la regla empírica? ¿Qué implica esto? Explique.

f. ¿Las gráficas muestran una distribución que concuerde con sus respuestas en la parte e? Explique.

2.145 La regla empírica indica que los intervalos de 1, 2, y 3 de desviación estándar alrededor de la media contendrán 68, 95, y 99.7%, respectivamente.

a. Utilice los comandos de computadora o calculadora de la página 101 para generar al azar una muestra de 100 datos de una distribución normal con media de 50 y desviación estándar de 10. Construya un histograma usando fronteras de clase que sean múltiplos de la desviación estándar de 10; es decir, use las fronteras de 10 a 90 en intervalos de 10 (vea los comandos que se indican en las páginas 61-62). Calcule la media y la desviación estándar usando los comandos que se encuentran en las páginas 74 y 88; luego inspeccione el histograma para determinar el porcentaje de los datos que caen dentro de cada una de los intervalos de 1, 2, y 3 de desviación estándar. ¿Qué tan cercanamente se comparan los tres porcentajes con los porcentajes expresados en la regla empírica?

b. Repita la parte a. ¿Obtuvo usted resultados semejantes a los de la parte a? Explique.

c. Considere repetir la parte a varias veces más. ¿Son semejantes los resultados cada vez? Si es así, ¿en qué forma?

d. ¿Qué concluye usted acerca de la verdad de la regla empírica?

2.146 El teorema de Chebyshev expresa que "al menos $1 - \frac{1}{k^2}$ de los datos de una distribución estarán dentro de k desviaciones estándar de la media.

a. Use los comandos de computadora de la página 101 para generar al azar una muestra de 100 datos de una distribución uniforme (no normal) que tenga un valor bajo de 1 y un valor alto de 10. Construya un histograma usando fronteras de clase de 0 a 11 en incrementos de 1 (vea los comandos en las páginas 61-62). Calcule la media y la desviación estándar usando los comandos que se hallan en las páginas 74 y 88; luego inspeccione el histograma para determinar el porcentaje de los datos que cayeron dentro de cada uno de los intervalos 1, 2, 3, y 4 desviaciones estándar. ¿Qué tan cercanamente se comparan estos porcentajes con los porcentajes indicados en el teorema de Chebyshev y en la regla empírica?

b. Repita el inciso a. ¿Obtuvo usted resultados semejantes a los de la parte a? Explique.

c. Considere repetir la parte a varias veces más. ¿Son semejantes los resultados cada vez? Si es así, ¿en qué forma son semejantes?

d. ¿Qué concluye usted acerca de la verdad del teorema de Chebyshev y la regla empírica?

2.8 El arte de la mentira estadística

"Existen tres clases de mentiras: mentiras, mentiras tremendas y estadísticas." Estas admirables palabras pronunciadas por Benjamin Disraeli (primer ministro inglés del siglo XIX) representan el punto de vista cínico de la estadística que tienen numerosas personas. Casi todos nos encontramos en el extremo del consumidor de la estadística y, en consecuencia, tenemos que "vivir con ellas".

Buena aritmética, malas estadísticas

Exploremos una indiscutible completa mentira estadística. Supongamos que un pequeño negocio emplea ocho personas que ganan entre \$300 y \$350 a la semana. El propietario del negocio se paga a sí mismo \$1250 a la semana y reporta al público en general que el salario promedio pagado a los empleados de su firma es \$430 por semana. Ese puede ser un ejemplo de buena aritmética, pero malas estadísticas. Es una mala representación de la situación porque sólo un empleado, el propietario, recibe más que el salario medio. El público pensará que la mayoría de los empleados perciben alrededor de \$430 por semana.

Engaño gráfico

Las representaciones gráficas pueden ser complicadas y engañosas. La escala de frecuencia (que suele ser el eje vertical) debería empezar en cero para presentar una imagen total. Por lo general, las gráficas que no empiezan en cero se usan para

ahorrar espacio pero esto también puede ser engañoso. Las gráficas en las que la escala de frecuencia empieza en cero tienden a destacar el tamaño de los números involucrados, mientras que las que están recortadas pueden tender a destacar la variación de los números sin considerar el tamaño real de los mismosnúmeros; la leyenda de la escala horizontal también puede ser engañosa. Es necesario inspeccionar con todo cuidado las presentaciones gráficas antes de sacar alguna conclusión del caso que se trate.

Los siguientes dos casos ejemplos prácticos que se muestran a continuación demostrarán parte de estas malas representaciones.

Mala representación sobrepuesta

CASO PRÁCTICO 2.16

Expresar lo que el lector espera/Malas noticias anticipadas

Courtesy of the Ithaca Times

Source: http://www.math.yorku.ca/SCS/Gallery/context.html

Esta "ingeniosa" cubierta gráfica, de *Ithaca Times* (7 de diciembre, 2000), tiene que ser la peor gráfica que aparece en primera plana. El artículo de la portada, "¿Por qué tienen que costar tanto los estudios universitarios?" presenta dos gráficas sobrepuestas sobre un fondo de un campo de la Universidad de Cornell. Las dos líneas quebradas representan "Colegiatura de Cornell" y la "Categoría de Cornell", con la colegiatura aumentando de manera constante y la categoría tambaleante y bajando. Se crea una imagen muy clara: los estudiantes obtienen menos, y pagan más.

Ahora veamos las dos gráficas por separado. Observe: (1) Las gráficas cubren dos periodos diferentes. (2) Las escalas verticales difieren. (3) La "mejor" mala representación viene de la impresión de que una "caída en categoría" representa una menor calidad de educación. ¿Un sexto lugar no sería mejor que el lugar 15?

POR LOS NÚMEROS: EN 35 AÑOS, LA COLEGIATURA EN CORNELL HA TOMADO UNA PARTE CADA VEZ MAYOR DE LA MEDIANA DEL INGRESO FAMILIAR DEL ESTUDIANTE (O DEL INGRESO FAMILIAR MEDIO DEL ESTUDIANTE)

ORDEN DE JERARQUÍA: EN 12 AÑOS, LA CATEGORÍA DE CORNELL EN *US NEWS WORLD REPORT* HA SUBIDO Y CAÍDO ERRÁTICAMENTE.

Fuente: http://www.math.yorku.ca/SCS/Gallery/Context.html

Lo que se concluye es que con la estadística, como sucede en los demás idiomas, se abusa. En manos de un descuidado, ignorante o inescrupuloso, la información estadística puede ser tan falsa como las "mentiras tremendas".

Escala truncada

CASO PRÁCTICO 2.17

Lo sencillo no siempre es lo mejor

Esta gráfica es clara y fácil de leer, pero, ¿representa la información que se muestra? Truncar escalas en gráficas lleva con frecuencia a impresiones visuales engañosas. Por ejemplo, en "Quejas en el de servicio del contratista", parece que "tarda demasiado" es dos veces más probable como motivo de queja que "es desordenado". Busquemos otras representaciones visuales erróneas.

QUEJAS EN EL SERVICIO DEL CONTRATISTA

Datos de USA Today, 9/5/2001.

SECCIÓN 2.8 EJERCICIOS

2.147 ¿Es posible que ocho empleados ganen entre \$300 y \$350, un noveno gane \$1250 a la semana, y que la media sea \$430? Verifique su respuesta.

2.148 La gráfica "Plan de gastos para el Día del Amor y la amistad" muestra una distribución de frecuencia relativa. Esta gráfica se califica como "gráfica con trampa complicada".

PLAN DE GASTOS PARA EL DÍA DEL AMOR Y LA AMISTAD

Datos de Darryl Haralson y Julie Zinder. USA Today. Margen de error ±3 puntos porcentuales.

a. Esta gráfica, ¿es de barras o es un histograma? Explique.

b. ¿Cómo es que esta gráfica viola el principio para dibujar histogramas?

2.149 "¿Qué está mal en esta imagen?" Ésta es la pregunta que uno debe hacerse al ver las gráficas del caso práctico 2.16 que se incluye en de la página 115.

a. Encuentre y describa al menos cuatro aspectos relacionados con aspectos acerca de la gráfica de la portada del *Ithaca Times* que se utilizan en forma están incorrectamente empleados.

b. Encuentre y describa al menos dos aspectos acerca de la gráfica de "Pecking Order" que se representan con errores.

2.150 a. Encuentre y describa al menos cuatro impresiones incorrectas creadas al truncar el eje horizontal en la gráfica del caso práctico 2.17 "Quejas en el servicio del contratista".

b. Vuelva a dibujar la gráfica de barras iniciando en "cero" la escala horizontal.

c. Comente sobre el efecto que su gráfica tenga sobre la impresión presentada.

2.151 Con mucha frecuencia, el mejor valor por el precio de un artículo era reportado por consumidores como uno de los atractivos cuando decidían dónde hacer sus compras en vacaciones. Cuando se les preguntaba, "¿Qué atrae a vacacionistas a las tiendas?" en la navidad en 2004, respondieron como sigue:

Qué	Porcentaje
Valor	76
Ubicación cómoda	68
Calidad	62
Selección	60

Fuente: USA Today y NPD Group

Elabore dos gráficas de barras para describir los datos de porcentajes. En la primera gráfica, en el eje vertical aplique una escala de 50 a 80; en la segunda, de 0 a 80. ¿Cuál es su conclusión respecto a la forma en la que se acumulan los porcentajes de las cuatro respuestas respuestas con base en las dos gráficas de barras?, ¿y qué recomendaría para mejorar las presentaciones?

2.152 Encuentre un artículo o un anuncio que incluya una gráfica que en alguna forma represente mal la información de estadísticas. Describa la forma en que esta gráfica representa los datos.

2.9 Media y desviación estándar de una distribución de frecuencias (opcional)

Cuando los datos muestrales se presentan en la forma de una distribución de frecuencias, necesitamos hacer una ligera adaptación a las fórmulas (2.1) y (2.10) para determinar la media, la varianza, y la desviación estándar.

EJEMPLO 2.18 Cálculos usando una distribución de frecuencia

TABLA 2.18

Distribución de frecuencia no agrupada

x	f
1	5
2	9
3	8
4	6
	$\sum f = 28$

Encuentre la media, la varianza, y la desviación estándar para los datos muestrales que se representan mediante por la distribución de frecuencias de la tabla 2.18.

Nota: esta distribución de frecuencias representa una muestra de 28 valores: cinco números 1, nueve 2, ocho 3 y seis 4.

Para calcular la media muestral $\bar{x}$ y la varianza muestral s^2 usando las fórmulas (2.1) y (2.10), necesitamos la suma de los 28 valores de x, $\sum x$, y la suma de los 28 valores de x^2, $\sum x^2$.

Las sumas, $\sum x$ y $\sum x^2$, podrían hallarse como sigue:

$$\sum x = 1\underbrace{+1+\cdots+1}_{\text{5 de ellas}} + 2\underbrace{+2+\cdots+2}_{\text{9 de ellas}} + 3\underbrace{+3+\cdots+3}_{\text{8 de ellas}} + 4\underbrace{+4+\cdots+4}_{\text{6 de ellas}}$$

$$= (5)(1) + (9)(2) + (8)(3) + (6)(4)$$

$$= 5 + 18 + 24 + 24 = \mathbf{71}$$

$$\sum x^2 = \underbrace{1^2+\cdots+1^2}_{\text{5 de ellas}} + \underbrace{2^2+\cdots+2^2}_{\text{9 de ellas}} + \underbrace{3^2+\cdots+3^2}_{\text{8 de ellas}} + \underbrace{4^2+\cdots+4^2}_{\text{6 de ellas}}$$

$$= (5)(1) + (9)(4) + (8)(9) + (6)(16)$$

$$= 5 + 36 + 72 + 96 = \mathbf{209}$$

No obstante, usaremos la distribución de frecuencias para determinar estas sumatorias al expandirla y para convertirla en una *tabla de extensiones*. Las extensiones xf y x^2f se forman al multiplicar horizontalmente las columnas, fila por fila, y luego sumar para hallar tres totales de columna. El objetivo La meta de la tabla de extensiones es obtener los estos tres totales de tres columnas (vea la tabla 2.19).

TABLA 2.19

Distribución de frecuencia no agrupada: extensiones xf y x^2f

x	f	xf	x^2f
1	5	5	5
2	9	18	36
3	8	24	72
4	6	24	96
	$\sum f = 28$	$\sum xf = 71$	$\sum x^2f = 209$

número de datos — *suma de x, usando frecuencias* — *suma de x^2, usando frecuencias*

Notas:

1. Las extensiones de la columna xf son los subtotales de los valores semejantes de x semejantes.
2. Las extensiones de la columna x^2f son los subtotales de los valores semejantes de x cuadradas semejantes.
3. Los totales de las tres columnas, Σf, Σxf, y Σx^2f, son los valores previamente conocidos como n, Σx, y Σx^2, respectivamente. Esto es, $\Sigma f = n$, el número de piezas de datos; $\Sigma xf = x$, la suma de los datos; y $\Sigma x^2f = x^2$, la suma del cuadrado de los datos.
4. Considere la f en las expresiones Σxf y Σx^2f como una indicación de que las sumas se obtuvieron con el uso de una distribución de frecuencia.
5. La suma de la columna x NO es un número significativo. La columna x contiene una lista de cada posible valor de x una vez, que no toma en cuenta los valores repetidos.

Para hallar la **media** de una distribución de frecuencia, modificamos la fórmula (2.1) de la página 73 para indicar el uso de la distribución de frecuencias:

Media de distribución de frecuencias

$$x\ barra = \frac{\textit{suma de toda x, usando frecuencias}}{\textit{número usando frecuencias}}$$

$$\bar{x} = \frac{\sum xf}{\sum f} \qquad (2.13)$$

El valor central de x para la distribución de frecuencia de la tabla 2.19 se encuentra usando la fórmula (2.13):

$$\text{media:} \qquad \bar{x} = \frac{\sum xf}{\sum f} = \frac{71}{28} = 2.536 = \mathbf{2.5}$$

Para hallar la **varianza** de la distribución de frecuencia, modificamos la fórmula (2.10) de la página 88 para indicar el uso de la distribución de frecuencia:

Varianza de la distribución de frecuencia

$$s\ cuadrada = \frac{(\text{suma de } x^2\text{, usando frecuencias}) - \left[\dfrac{(\text{suma de } x\text{, usando frecuencias})^2}{\text{número, usando frecuencias}}\right]}{\text{número, usando frecuencias} - 1}$$

$$s^2 = \frac{\sum x^2 f - \dfrac{(\sum xf)^2}{\sum f}}{\sum f - 1} \qquad (2.14)$$

La varianza de x para la distribución de frecuencias de la tabla 2.19 se encuentra mediante el uso de la fórmula (2.14):

$$\text{varianza:} \qquad s^2 = \frac{\sum x^2 f - \dfrac{(\sum xf)^2}{\sum f}}{\sum f - 1} = \frac{209 - \dfrac{(71)^2}{28}}{28 - 1} = \frac{28.964}{27} = 1.073 = \mathbf{1.1}$$

La **desviación estándar** de x para la distribución de frecuencia de la tabla 2.19 se encuentra con el uso de la fórmula (2.7), la raíz cuadrada positiva de la varianza.

$$\text{desviación estándar:} \quad s = \sqrt{s^2} = \sqrt{1.073} = 1.036 = \mathbf{1.0}$$

EJEMPLO 2.19 Cálculos usando frecuencias agrupadas

Encuentre la media, varianza, y desviación estándar de la muestra de 50 calificaciones de examen usando la distribución de frecuencia agrupada de la tabla 2.8 (p. 58).

SOLUCIÓN Usaremos una tabla de extensiones para hallar las tres sumatorias, como se realizó en la misma forma que hicimos en el ejemplo 2.18. Los puntos medios de clase se usarán como valores representativos para las clases.

El valor medio de x para la distribución de frecuencias de la tabla 2.20 (p. 120) se encuentra con el uso de la fórmula (2.13):

$$\text{varianza:} \qquad \bar{x} = \frac{\sum xf}{\sum f} = \frac{3780}{50} = \mathbf{75.6}$$

El valor central de x para la distribución de frecuencias de la tabla 2.20 (p. 120) se encuentra con el uso de la fórmula (2.14):

$$\text{varianza:} \qquad s^2 = \frac{\sum x^2 f - \dfrac{(\sum xf)^2}{\sum f}}{\sum f - 1} = \frac{296\ 600 - \dfrac{3780^2}{50}}{50 - 1} = \frac{10\ 832}{49}$$

$$= 221.0612 = \mathbf{221.1}$$

La desviación estándar de x para la distribución de frecuencias de la tabla 2.20 se encuentra con el uso de la fórmula (2.7):

$$\text{desviación estándar:} \qquad s = \sqrt{s^2} = \sqrt{221.0612} = 14.868 = \mathbf{14.9}$$

TABLA 2.20

Distribución de frecuencias de 50 calificaciones de examen

Número de grupo	Puntos medios de la clase o grupo, x	f	xf	x^2f
1	40	2	80	3 200
2	50	2	100	5 000
3	60	7	420	25 200
4	70	13	910	63 700
5	80	11	880	70 400
6	90	11	990	89 100
7	100	4	400	40 000
		$\sum f = 50$	$\sum xf = 3780$	$\sum x^2f = 296\ 600$

INSTRUCCIONES DE TECNOLOGÍA: ESTADÍSTICA DE LA DISTRIBUCIÓN DE FRECUENCIA

MINITAB (Versión 14)

Introduzca los puntos medios de grupo o valores de datos en C1 y las frecuencias correspondientes en C2; luego continúe con los siguientes comandos para obtener la tabla de extensiones:

```
Choose:      Calc > Calculator . . .
Enter:       Store result in variable: C3
             Expression: C1*C2 > OK
Repeat the preceding commands, replacing the variable with C4 and the ex-
pression with C1*C3.
Choose:      Calc > Column Statistics
Select:      Sum
Enter:       Input variable: C2
             Store result in: K1 > OK
Repeat preceding 'sum' commands, replacing variable with C3 and result with
K2.
Repeat preceding 'sum' commands, replacing variable with C4 and result with
K3.
Choose:      Data > Display data
Enter:       Columns to display: C1-C4 K1-K3 > OK
```

Para hallar la media, varianza y desviación estándar, respectivamente, continúe con:

```
Choose:      Calc > Calculator
Enter:       Store result in variable: K4
             Expression: K2/K1 > OK
Repeat preceding 'mean' commands, replacing variable with K5 and expression
with (K3-(K2**2/K1))/(K1-1).
Repeat preceding 'mean' commands, replacing variable with K6 and expression
with SQRT(K5) (select square root from functions).
Choose:      Data > Display data
Enter:       Columns to display: K4-K6 > OK
```

Excel Introduzca los puntos medios de grupo o valores de datos en la columna A y las frecuencias correspondientes en la columna B; active C1 o C2 (dependiendo de si se usan encabezados de columna); luego continúe con los siguientes comandos para obtener la tabla de extensiones:

```
Enter:        = A2*B2 (if column headings are used)
Drag:         Bottom right corner of C2 down to give other products
Activate D2 and repeat preceding commands, replacing the formula with =
A2*C2.
Activate the data in columns B, C, and D.
Choose:       AutoSum (sums will appear at the bottom of the columns)
```

Para hallar la media, active **E2**; luego continúe con:

```
Enter:        = (column C total/column B total) (ex. = C9/B9)
```

Para hallar la varianza, active **E3** y repita precediendo los comandos "media", sustituyendo la fórmula con **= (D9 - (C9^2/B9))/(B9–1).**

```
To find the standard deviation, activate E4 and repeat preceding 'mean' com-
mands, replacing the formula with = SQRT(E3).
```

TI-83/84 Plus Introduzca los puntos medios de grupo o valores de datos en L1 y las frecuencias en L2; luego continúe con:

```
Highlight:    L3
Enter:        L3 = L1*L2
Highlight:    L4
Enter:        L4 = L1*L3
Highlight:    L5(1) (?rst position in L5 column)
Enter:        L5(1) =sum(L2)     [∑f]
              [sum = 2nd LIST > MATH > 5:sum(]
              L5(2)  = sum(L3)   [∑xf]
              L5(3)  = sum(L4)   [∑x²f]
              L5(4)  = L5(2)/L5(1) [to ?nd mean]
              L5(5) = (L5(3) – ((L5(2))²/L5(1)))/(L5(1)–1)
              [to ?nd variance]
              L5(6) = 2nd √   (L5(5))
              [to ?nd standard deviation]
```

Si la tabla de extensiones no es necesaria, simplemente use:

```
Choose:       STAT > CALC > 1:1-VAR STATS
Enter:        L1, L2
```

SECCIÓN 2.9 EJERCICIOS (OPCIONALES)

2.153 En una encuesta se pidió, a quienes respondieron, que elaboraran una lista del "número de teléfonos" por familia, x; a continuación se ven los resultados como una distribución de frecuencia.

x	0	1	2	3	4
f	1	3	8	5	3

a. Complete la tabla de extensiones.

b. Encuentre las tres sumatorias, $\sum f$, $\sum xf$, $\sum x^2 f$, para la distribución de frecuencias.

c. Describa lo que representa cada uno de lo siguiente: $x = 4$, $f = 8$, $\sum f$, $\sum xf$.

d. Explique por qué (i) la "suma de la columna x" no tiene relación con la "suma de los datos," y (ii) la "Σxf" representa la "suma de los datos" representada por la distribución de frecuencia.

2.154 a. Encuentre la media de los datos mostrados en la distribución de frecuencias del ejercicio 2.153.

b. Encuentre la varianza para los datos mostrados en la distribución de frecuencias del ejercicio 2.153.

c. Encuentre la desviación estándar para los datos mostrados en la distribución de frecuencias del ejercicio 2.153.

2.155 Los dentistas pedriátricos dicen que el primer examen dental de un niño debe ser entre la edad de 6 meses y 1 año. Los años del primer examen para una muestra de niños se indican en la distribución:

Edad de primer examen dental, *x*	1	2	3	4	5
Número de niños, *f*	9	11	23	16	21

a. Encuentre la edad media del primer examen dental para estos niños.

b. Encuentre la edad mediana.

c. Encuentre la desviación estándar.

2.156 En una encuesta se preguntó a un grupo de médicos cuántos hijos habían tenido. Los resultados se resumen en esta distribución de frecuencia no agrupada:

Número de hijos	0	1	2	3	4	6
Número de médicos	15	12	26	14	4	2

Calcule la media muestral, varianza, y desviación estándar para la cantidad del número de niños hijos de los médicos.

2.157 Los aumentos de peso (en gramos) de los pollos que se alimentan con una dieta alta en proteínas fueron como sigue:

Aumento en peso	12.5	12.7	13.0	13.1	13.2	13.8
Frecuencia	2	6	22	29	12	4

a. Encuentre la media.

b. Encuentre la varianza.

c. Encuentre la desviación estándar.

2.158 Encuentre la media, varianza y desviación estándar de los datos mostrados en la siguiente distribución de frecuencia.

Clase	2–6	6–10	10–14	14–18	18–22
f	2	10	12	9	7

2.159 Encuentre la media, varianza y desviación estándar para esta distribución de frecuencia agrupada:

Fronteras de clase	3–6	6–9	9–12	12–15	15–18
f	2	10	12	9	7

2.160 La siguiente distribución de distancias en viajes suburbanos se obtuvo para una muestra de empleados de Mutual of Nebraska:

Distancia (millas)	Frecuencia	Distancia (millas)	Frecuencia
1.0–3.0	2	9.0–11.0	35
3.0–5.0	6	11.0–13.0	15
5.0–7.0	12	13.0–15.0	5
7.0–9.0	50		

Encuentre la media y la desviación estándar para las distancias en viajes suburbanos.

2.161 Un técnico de control de calidad seleccionó veinticinco cajas de 1 libra de un proceso de producción, y encontró la siguiente distribución en los pesos de las cajas de pesos (en onzas):

Peso	Frecuencia	Peso	Frecuencia
15.95–15.98	2	16.04–16.07	3
15.98–16.01	4	16.07–16.10	1
16.01–16.04	15		

Encuentre la media y la desviación estándar para esta distribución de peso.

2.162 Se ha encontrado que 35.2 millones de estadounidenses de 16 años o más pescan en nuestras aguas. Una muestra de pescadores en agua dulce produjo la siguiente distribución de edades:

Edad de los pescadores, *x*	15–25	25–35	35–45	45–55	55–65	65–75
# de pescadores, *f*	13	20	28	20	10	9

Encuentre la media y la desviación estándar para esta distribución.

2.163 La industria privada informa que más de 31,000 trabajadores no se presentaron al trabajo en 2005 por el síndrome de túnel carpiano (enfermedad de los nervios que causa dolor en brazos, muñecas y manos). La duración (en días) que se ausentaron los trabajadores como resultado de este problema varía grandemente.

Días de ausencia, *x*	0–10	10–20	20–30	30–40	40–50
Número de trabajadores, *f*	37	24	38	32	27

Encuentre la media y la desviación estándar para esta distribución.

2.164 El California Department of Education publica un informe anual de las calificaciones en el del Scholastic Aptitude Test (Examen de aptitud escolar) para estudiantes de diversos distritos escolares. La siguiente tabla de frecuencia muestra resultados de exámenes verbales para distritos escolares en 2003-2004 para el condado de Merced.

Distrito	Número examinado	Promedio verbal
Delhi Unified	34	434
Dos Palos Oro Loma Jt. Unified	48	431
Gustine Unified	37	482
Hilmar Unified	43	488
Le Grand Union High	28	369
Los Banos Unified	109	479
Merced Co. Office of Education	0	0
Merced Union High	534	450

Fuente: http://data1.cde.ca.gov/dataquest/SAT-I1.asp?cChoice=SAT1&cYear=2003-04&TheCount

a. ¿Qué significan las entradas 34 y 434 para Delhi Unified?

b. ¿Cuál es el total para las calificaciones de todos los estudiantes de Delhi Unified?

c. ¿Cuántos resultados de examen de estudiantes se muestran en esta tabla?

d. ¿Cuál es el total para todas las calificaciones de estudiantes mostradas en la tabla?

e. Encuentre el resultado medio de examen verbal del SAT.

2.165 Una muestra aleatoria de personas de todas las edades se tomó de la población de Estados Unidos, y las 75 edades resultantes (en años) aparecen en la tabla siguiente:

22	7	72	32	18	4	9	48	49	18	18	58	47	39	48
1	27	61	48	25	34	75	29	53	37	25	42	49	29	31
10	27	8	78	63	50	39	32	5	39	8	15	8	50	39
46	38	4	9	43	3	65	25	67	19	9	34	8	36	48
56	73	24	20	34	38	45	40	11	40	37	17	63	9	91

a. Construya una distribución de frecuencia agrupada para las edades usando puntos medios de clase de 0, 10, 20,90. Muestre en su tabla los puntos medios de clase y las cantidades de frecuencia asociadas.

b. Construya un histograma.

c. La variable edad, ¿parece tener una distribución aproximadamente normal?

d. Calcule la edad media.

e. Encuentre la edad mediana.

f. Encuentre el rango de edades.

g. Encuentre la desviación estándar de edades.

h. Compare los valores encontrados en las partes de la d a la g contra las correspondientes estadísticas calculadas usando los datos no agrupados y que se dan. Use el error porcentual en cada caso, y presente todos los resultados en una tabla para hacer el caso de usted.

2.166 La publicación USA Snapshot titulada "Monjas, un orden que envejece" informa que la edad media de 94 022 monjas católicas romanas en Estados Unidos es 65 años, y los porcentajes de monjas en Estados Unidos por grupo de edad son como sigue:

Menos de 50	51-70	Más de 70	Se negaron dar su edad
16%	42%	37%	5%

Esta información está basada en una encuesta de 1049 monjas católicas romanas. Suponga que la encuesta había resultado en la distribución de frecuencia siguiente (52 edades desconocidas):

Edad	20–30	30–40	40–50	50–60	60–70	70–80	80–90
Frecuencia	34	58	76	187	254	241	147

(Vea el histograma trazado en el ejercicio 2.42, p. 69.)

a. Encuentre la media, mediana, moda, y rango central para esta distribución de edades.

b. Encuentre la varianza y desviación estándar.

2.167 El número de programas que ve en televisión un aficionado a los deportes en una semana típica, se describió en la publicación USA Snapshot titulada "Aficionados encuentran deportes en diarios y TV" (21 de diciembre de 2004).

Programas	0	1–2	3–4	5–6	7	8 o más
Porcentaje	35.0%	24.8%	15.4%	11.4%	8.6%	4.8%

Esta información está basada en una encuesta de deportes de ESPN. Suponga que la encuesta resultó en la siguiente distribución de frecuencia.

Programas	0	1–2	3–4	5–6	7	8 o más
Frecuencia	44	31	19	14	11	6

a. ¿Cuántos fueron encuestados?

b. Trace un histograma de estos datos.

c. Encuentre el número medio de programas vistos por semana. (Use 8.5 "8 o más" como punto medio.)

d. Encuentre el número medio de programas vistos por semana.

e. Encuentre el número de moda de programas vistos por semana.

2.168 La USA Snapshot "Reembolso después de graduación" informa que los pasantes que solicitan préstamos promedian $16 500 de deuda al graduarse. La cantidad de sus deudas mensuales, con base en el porcentaje se muestra a continuación.

Deuda	Menos de $100	$100-$149	$150-$199	$200-$249	$250-$299	$300 o más
Porcentaje	17%	17%	17%	19%	10%	20%

Suponga que otra encuesta ha resultado en la distribución de frecuencia mostrada a continuación:

Deuda	Menos de $100	$100–$149	$150–$199	$200–$249	$250–$299	$300 o más
Frecuencia	125	158	127	175	100	165

a. ¿Cuántos fueron encuestados?

b. Trace un histograma de estos datos.

c. Encuentre la media de la distribución de frecuencia.

d. Encuentre la mediana de la distribución de frecuencia.

e. Encuentre la moda de la distribución de frecuencia.

2.169 Una muestra aleatoria de 250 personas que viven en el estado de Nueva York generó resultó en la siguiente distribución de edades:

Grupo de edades	Frecuencia	Grupo de edades	Frequency
Menos de 5 años	18	45–54 años	48
5–14 años	35	55–64 años	21
15–24 años	20	65–74 años	17
25–34 años	35	75–84 años	16
35–44 años	38	85 años y más	2

a. ¿Qué sucede si esta distribución es diferente de las distribuciones descritas en el texto?

b. ¿Cómo pueden los procedimientos de esta sección adaptarse para contener las clases de los extremos de esta distribución?

c. Trace un histograma de la distribución de edades.

d. Encuentre la edad media para las personas incluidas en esta muestra.

e. Encuentre la desviación estándar.

2.170 No se espera que un golfista profesional del Champions Tour juegue en todos los torneos que se presenten en el curso de una temporada. El número de torneos donde juegan cada uno de los 50 máximos ganadores de dinero del torneo de 2004 se muestra en la página 125.

a. Construya una distribución de frecuencia agrupada que muestre el número de torneos donde jugaron, usando intervalos de grupo de 9-11, 11-13, 29-31; los puntos medios de clase; y las cantidades de frecuencia asociadas.

b. Encuentre la media, varianza, y desviación estándar del número de torneos donde jugaron, con y sin usar la distribución agrupada.

c. Compare los dos conjuntos de respuestas que usted obtuvo en la parte b. ¿De qué porcentaje es el error en cada caso?

Jugador	Eventos	Jugador	Eventos	Jugador	Eventos	Jugador	Eventos	Jugador	Eventos
Craig Stadler	21	Bruce Lietzke	20	D. A. Weibring	25	Bruce Summerhays	28	Dana Quigley	30
Mark James	20	John Jacobs	28	David Eger	28	Dave Barr	28	Jay Sigel	28
Lonnie Nielsen	26	Bruce Fleisher	28	Jim Ahern	27	Doug Tewell	27	Pete Oakley	12
Hale Irwin	23	Bob Gilder	28	Jim Thorpe	26	Bobby Wadkins	26	Morris Hatalsky	27
Jerry Pate	27	Gary McCord	14	Graham Marsh	30	Joe Inman	26	Walter Hall	26
Don Pooley	21	Larry Nelson	25	Dave Stockton	21	Tom Jenkins	27	Hugh Baiocchi	26
Tom Kite	27	Fuzzy Zoeller	21	Allen Doyle	27	Vicente Fernandez	26	Peter Jacobsen	9
Jose Maria Canizares	26	Gary Koch	18	Ed Fiori	28	Mike McCullough	28	John Harris	25
John Bland	26	Mark McNulty	20	Rodger Davis	20			Keith Fergus	18
Gil Morgan	26	Andy Bean	28	Wayne Levi	27			Tom Purtzer	19
		Tom Watson	12					Des Smyth	27

Fuente: PGA Tour, Inc.

REPASO DEL CAPÍTULO

En retrospectiva

En este capítulo hemos presentado algunas de las técnicas más comunes de estadística descriptiva; hay demasiados tipos específicos de estadística que se emplean en casi todo campo especializado de estudio para poder analizarlos aquí. Hemos resumido los usos sólo de la estadística más universal. Específicamente, el lector ha visto varias técnicas gráficas básicas (gráficas de círculo y de barras, diagramas de Pareto, gráficas de puntos, diagramas de tallo y hoja, histogramas, y diagramas de caja y bigote) que se usan para presentar datos muestrales en forma gráfica. El lector también ha sido introducido a algunas de las medidas más comunes de tendencia central (media, mediana, moda, rango central, y cuartil medio), medidas de dispersión (rango, varianza y desviación estándar), así como medidas de posición (cuartiles, percentiles, y calificaciones z).

El lector debe saber ahora que un promedio puede ser cualquiera de cinco estadísticas diferentes, y debe entender las distinciones entre los diferentes tipos de promedios. El artículo "Promedio" significa cosas diferentes" que se presenta en el caso práctico 2.11 (pp. 77-78) presenta cuatro de los promedios que se estudian estudiados en este capítulo. Puede volver a leerlo ahora y encontrar que tiene más significado y es de más interés; será un tiempo bien empleado.

También debe entender el concepto de una desviación estándar; así como los aspectos básicos de la regla empírica y el teorema de Chebyshev con este propósito.

Los ejercicios de este capítulo (al igual que en otros) son muy importantes; reforzarán los conceptos estudiados antes que el estudiante continúe para aprender cómo usar estas ideas en capítulos posteriores. Una buena comprensión de las técnicas descriptivas presentadas en este capítulo es fundamental para el éxito en los capítulos finales.

Vocabulario y conceptos clave

ancho de clase (p. 56)
calificación z (p. 99)
clase (p. 56)
clase modal (p. 64)
cuartil (p. 93)
cuartil medio (p. 75)
datos cualitativos (p. 41)
datos cuantitativos (p. 44)
desviación desde la media (p. 84)
desviación estándar (pp. 87, 119)
diagrama de Pareto (p. 42)
diagrama de tallo y hoja (p. 46)

distribución (p. 44)
distribución de frecuencia (p. 55)
distribución de frecuencia agrupada (p. 56)
distribución de frecuencia no agrupada (p.56)
distribución de frecuencia relativa (p. 59)
distribución de frecuencias bimodal (p. 63)
distribución en forma de campana (p. 106)
distribución normal (pp.106-108)
distribución rectangular (p. 63)
distribución sesgada (p. 63)
frecuencia (p. 44)
frecuencia relativa (p. 59)
frontera de clase (p. 57)
gráfica de barras (p. 41)
gráfica de caja y bigote (p. 96)
gráfica de pastel (p. 141)
gráfica de puntos (p. 44)
histograma (p. 59)
histograma de frecuencias (p. 60)
histograma de frecuencias relativas (p. 69)
media (pp. 73, 118)
mediana (p. 74)
medida de dispersión (p. 84)
medida de posición (p. 92)
medida de tendencia central (p. 73)
moda (p. 76)
ojiva (p. 65)
percentil (p. 93)
profundidad (pp. 74, 94)
punto medio de clase (marca de clase) (p. 58)
rango (p. 84)
rango central (p. 76)
rango intercuartil (p. 96)
regla empírica (p. 106)
resumen de 5 números (p. 96)
sumatoria (p. 58)
teorema de Chebyshev (p. 109)
varianza (pp.86, 119)
x barra ($\bar{x}$) (p. 73)

Objetivos de aprendizaje

✓ Crear e interpretar diagramas gráficos, incluyendo gráficas de pastel, gráficas de barras, diagramas de Pareto, gráficas de puntos y diagramas de tallo y hoja.	EJ. 2.4, Ejer. 2.5, 2.13, 2.15, 2.19, 2.25, 2.27
✓ Comprender y ser capaz de describir la diferencia entre distribuciones de frecuencias tanto agrupadas como no agrupadas, frecuencia relativa y frecuencia relativa acumulativa.	pp. 55-56, 60, 64-65
✓ Identificar y describir las partes de una distribución de frecuencia: fronteras de clase, ancho de clase y un punto medio de clase.	EJ. 2.6, Ejer. 2.41, 2.43
✓ Crear e interpretar histogramas de frecuencia, histogramas de frecuencia relativa y ojivas.	pp. 60, 64-65, Ejer. 2.33, 2.36, 2.38
✓ Identificar las formas de las distribuciones.	pp. 63-65
✓ Calcular, describir y comparar las cuatro medidas de tendencia central: media, mediana, moda y rango central.	EJ. 2.11, Ejer. 2.67
✓ Comprender el efecto que tienen los resultados aislados en cada una de las cuatro medidas de tendencia central.	Ejer. 2.179, 2.180, 2.216
✓ Calcular, describir, comparar e interpretar las dos medidas de dispersión: rango y desviación estándar (varianza).	pp. 84-87, Ejer. 2.91, 2.97
✓ Calcular, describir e interpretar las medidas de posición: cuartiles, percentiles y calificaciones *z*.	EJ. 2.12, 2.14, Ejer. 2.109, 2.119, 2.200
✓ Crear e interpretar gráficas de caja.	Ejer. 2.114
✓ Comprender la regla empírica y el teorema de Chebyshev, así como ser capaz de evaluar y establecer una conformidad del conjunto de datos a estas reglas.	Ejer. 2.130, 2.137, 2.143
✓ Saber cuándo y dónde no usar ciertas estadísticas: gráficas y numéricas.	pp. 114-115, Ejer. 2.148, 2.149
✓ Calcular le media y desviación estándar para distribuciones de frecuencia no agrupadas y agrupadas. (Opcional.)	EJ. 2.18, 2.19, Ejer. 2.155, 2.159

Ejercicios del capítulo

2.171 "¿Quién cree en la regla de 5 segundos?" Casi todas las personas dicen que el alimento que cae al piso no es seguro para consumirse.

¿QUIÉN CREE EN LA REGLA DE 5 SEGUNDOS?

Datos de Anne R. Carey y Juan Thomassie, USA Today.

a. Elabore una gráfica de pastel que describa los porcentajes de adultos para cada respuesta.

b. Si se van a encuestar a 300 adultos, ¿qué frecuencias esperaría encontrar para cada respuesta en la gráfica "¿Usted se comería algún alimento que se hubiera caído al suelo?".

2.172 Las provisiones necesarias para un bebé durante su primer año pueden ser costosas, un promedio de $5000, como se observa en esta gráfica de barras dividida.

a. Construya una gráfica de pastel que muestre esta misma información.

b. Construya una gráfica de barras que muestre esta misma información.

c. Compare el aspecto de la gráfica de barras dividida con la gráfica de pastel que se presenta en la parte a y la gráfica de barras del inciso b. ¿Cuál representa mejor la relación entre varios costos de provisiones de bebé?

PRESUPUESTO PARA BEBÉ

Datos de Julie Snider, © 2005 USA Today.

2.173 Existen varios tipos de gráficas estadísticas a escoger cuando se representa un conjunto de datos. La "gráfica de barras dividida" que se muestra aquí es una alternativa de la gráfica de pastel.

¿CÓMO GASTARÁ LA DEVOLUCIÓN DE SUS IMPUESTOS?

Datos de Darryl Haralson y Jerry Mosemak, © 2004 USA Today

a. Construya una gráfica de pastel que muestre esta misma información.

b. Compare el aspecto de la gráfica de barras dividida y la gráfica de pastel que se realizó en el inciso a. ¿Cuál es más fácil de leer? ¿Cuál proporciona una representación más precisa de la información que se presenta?

2.174 Una de las formas en que los estudiantes pagan su colegiatura es mediante préstamos de dinero que deben ser pagados en última instancia, y la gráfica de barras dividida siguiente muestra la deuda mensual a la que muchos alumnos se enfrentan después de graduarse.

PAGO DE LA DEUDA CONTRATADA PARA EL PAGO DE ESTUDIOS

Datos de Justin Dickerson y Sam Ward. ©2004 USA Today.

a. Construya una gráfica de pastel que muestre esta misma información.

b. Elabore una gráfica de barras que muestre esta misma información.

c. Compare el aspecto de la gráfica de barras dividida que se muestra cuando la gráfica de pastel se muestra en el inciso a y la gráfica de que se muestra en el inciso b. ¿Cuál representa mejor la relación entre varias cantidades de deuda mensual?

2.175 Las 10 principales causas de muerte en Estados Unidos durante 2002 fueron publicadas por el National Center for Statistics and Analysis en un reportaje de enero de 2005. Se reportó un total de 2 443 387 defunciones registradas.

Causa de muerte	Número (10 000)
Enfermedad de Alzheimer's	5.9
Enfermedad respiratoria crónica	12.5
Diabetes	7.3
Enfermedad del corazón	69.7
Gripe/pulmonía	6.6
Neoplasmas malignos	55.7
Accidente automovilístico	4.4
Nefritis/nefrosis	4.1
Ataque cerebral	3.4
Golpe	16.3

Fuente: NHTSA's National Center for Statistics and Analysis

a. Construya un diagrama de Pareto de esta información.

b. Escriba un párrafo que describa lo que el diagrama de Pareto muestra en forma tan impresionante a su lector.

2.176 La U.S. Census Bureau publicó la siguiente distribución de edades de 2003 para habitantes del estado de Nueva York. La encuesta American Community Survey realizada en 2003 está limitada a la población de familias y excluye la población que vive en instituciones, dormitorios de universidades y otras viviendas colectivas.

Distribución de sexo y edad

Masculino	8 953 019	Menores de 5 años	1 205 816
Femenino	9 647 508	5-14 años	2 537 813
		15-24 años	2 353 665
		25-34 años	2 587 995
		35-44 años	2 991 609
		45-54 años	2 682 845
		55-64 años	1 897 521
		65-74 años	1 218 850
		75-84 años	857 177
		85 años o mayores	267 236

Fuente: U.S. Census Bureau

a. Construya una distribución de frecuencias relativas de los datos de género y edad.

b. Elabore una gráfica de barras para los datos del género.

c. Construya un histograma de los datos de edades.

d. Explique por qué motivo la gráfica trazada en la parte b no es un histograma, y la gráfica realizada en el inciso c es un histograma.

2.177 Identifique cada uno de lo siguiente como ejemplos de (1) variables de atributos (cualitativas) o (2) variables numéricas (cuantitativas).

a. Las calificaciones registradas por personas que toman su examen escrito para obtener licencia estatal para conducir.

b. Si un motociclista posee o no una licencia válida de motociclista.

c. El número de televisores que se han instalado en una casa.

d. La marca de jabón en pastilla que se usa en el cuarto de baño.

e. El valor de cupón de descuento que se utiliza en la compra de una caja de cereales.

2.178 Identifique cada uno de lo siguiente como ejemplos de (1) variables de atributos (cualitativas) o (2) variables numéricas (cuantitativas).

a. La cantidad de peso perdido el mes pasado por una persona que sigue una dieta estricta.

b. Los promedios de bateo de los jugadores de béisbol de las ligas mayores.

c. Las decisiones que toma el jurado en los juicios contra delitos graves.

d. El uso de filtros solares antes de tomar el sol (siempre, con frecuencia, a veces, rara vez, nunca).

e. La razón por la que un gerente no actuó contra el desempeño deficiente de un empleado.

2.179 Considere las muestras A y B. Observe que las dos muestras son las mismas excepto que el 8 de A ha sido sustituido por un 9 en B.

A:	2	4	5	5	7	8
B:	2	4	5	5	7	9

¿Qué efecto tiene cambiar el 8 a un 9 en cada una de las siguientes estadísticas?

a. Media b. Mediana c. Moda d. Rango central

e. Rango f. Varianza g. Desviación estándar

2.180 Considere las muestras C y D. Observe que las dos muestras son las mismas excepto por dos valores.

C:	20	60	60	70	90
D:	20	30	70	90	90

¿Qué efecto tiene cambiar los dos 60 a 30 y 90 en cada una de las siguientes estadísticas?

a. Media b. Mediana c. Moda d. Rango central

e. Rango f. Varianza g. Desviación estándar

2.181 Se dice que la preparación con un nuevo acelerador disminuye en más de 4% el tiempo de secado de pintura de látex. Se efectuaron varias pruebas de muestras con los siguientes porcentajes de disminución en el tiempo de secado:

5.2	6.4	3.8	6.3	4.1	2.8	3.2	4.7

a. Encuentre la media muestral.

b. Determine la desviación estándar de la muestra.

c. ¿Considera que estos porcentajes promedian 4 o más? Explique.

(Conserve estas soluciones para usarlas en el ejercicio 9.28, p.490.)

2.182 Se supone que la gasolina que se bombea desde el oleoducto de un proveedor tiene un octanaje de 87.5. En 13 días consecutivos, se tomó y analizó una muestra del octanaje, con los siguientes resultados:

88.6	86.4	87.2	88.4	87.2	87.6	86.8
86.1	87.4	87.3	86.4	86.6	87.1	

a. Encuentre la media muestral.

b. Determine la desviación estándar de la muestra.

c. ¿Considera que estas lecturas promedian 87.5? Explique.

(Conserve estas soluciones para usarlas en el ejercicio 9.56, p. 494.)

2.183 Los datos que se presentan a continuación son las edades de 118 delincuentes conocidos que cometieron robo de autos el año pasado en Garden City, Michigan:

11	14	15	15	16	16	17	18	19	21	25	36
12	14	15	15	16	16	17	18	19	21	25	39
13	14	15	15	16	17	17	18	20	22	26	43
13	14	15	15	16	17	17	18	20	22	26	46
13	14	15	16	16	17	17	18	20	22	27	50
13	14	15	16	16	17	17	19	20	23	27	54
13	14	15	16	16	17	18	19	20	23	29	59
13	15	15	16	16	17	18	19	20	23	30	67
14	15	15	16	16	17	18	19	21	24	31	
14	15	15	16	16	17	18	19	21	24	34	

a. Encuentre la media.

b. Encuentre la mediana.

c. Encuentre la moda.

d. Encuentre Q_1 y Q_3.

e. Encuentre P_{10} y P_{95}.

2.184 En mayo pasado se tomó una encuesta a 32 trabajadores del edificio 815 de la Eastman Kodak Com-

pany. A cada uno de ellos se le preguntó: "¿Cuántas horas vio televisión el día de ayer?" Los resultados fueron como sigue:

0	0	½	1	2	0	3	2½	0	0	1
1½	5	2½	0	2	2½	1	0	2	0	
2½	4	06	2½	0	½	1	1½	0	2	

a. Construya un diagrama de tallo y hoja.

b. Encuentre la media.

c. Encuentre la mediana.

d. Encuentre la moda.

e. Encuentre el rango central.

f. ¿Cuál medida de tendencia central representaría mejor al televidente promedio si usted estuviera tratando de describir a uno de éstos? Explique.

g. ¿Cuál medida de tendencia central describiría mejor el tiempo de televisión visto? Explique.

h. Determine el rango.

i. Encuentre la varianza.

j. Encuentre la desviación estándar.

2.185 La distancia de frenado en una superficie mojada se determinó para 25 autos, cada uno de los cuales corría a 30 millas por hora. Los datos (en pies) se muestran en el siguiente diagrama de tallo y hoja:

6	3 7 6 3 9
7	4 2 0 1 1 2 0
8	5 4 5 5 6
9	4 1 0 0 5
10	5 4

Encuentre la media y la desviación estándar de estas distancias de frenado.

2.186 El sitio Forbes.com publicó el EPS (beneficios por acción) de 2004 en dólares para 17 compañías de la industria bancaria.

Nombre	EPS ($)	Nombre	EPS ($
Astoria Financial	2.92	Popular	1.71
Banknorth Group	2.20	State Street	3.13
Bank of America	3.67	Synovus Finl	1.36
BB & T	2.61	UnionBanCal	4.70
Compass Bancshares	2.86	Wachovia	3.68
Golden West Finl	3.97	Wells Fargo	4.00
M & T Bank	5.74	Westcorp	3.71
National City	3.75	Zions Bancorp	4.36
North Fork Bancorp	1.83		

Fuente: http://www.forbes.com/lists/results.jhtml

a. Encuentre el EPS medio para los bancos.

b. Determine la mediana del EPS para los bancos.

c. Calcule el rango central del EPS para los bancos.

d. Escriba un ensayo que compare los resultados de las partes a, b y c.

e. Encuentre la desviación estándar del EPS para los bancos.

f. Halle el porcentaje de los datos que están dentro de 1 desviación estándar de la media.

g. Determine el porcentaje de los datos que estén dentro de 2 desviaciones estándar de la media.

h. Con base en los resultados anteriores, analice si usted piensa que los datos están normalmente distribuidos, e indique por qué.

2.187 La Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation, informó del número de reportes de mal manejo de equipaje presentados por 1000 pasajeros de aerolíneas durante octubre de 2004. El promedio de la industria fue 4.02.

Aerolínea	Reportes	Pasajeros	Reportes/1000
AirTran	2084	1 148 779	1.81
JetBlue	2295	1 057 510	2.17

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation

a. Defina los términos *población* y *variable* respecto a esta información.

b. Los números reportados (1.81, 2.17,...,12.21), ¿son datos o estadísticas? Explique.

c. El promedio, 4.02, ¿es un dato, una estadística, o un valor de parámetro? Explique por qué.

d. El "promedio de la industria", ¿es la media de las cantidades de reportes por 1000 en aerolíneas? Si no es así, explique en detalle cómo están relacionados los 19 valores de aerolíneas con el promedio de la industria.

2.188 Uno de los primeros científicos que analizó la densidad del nitrógeno fue Lord Raleigh. Él observó que la densidad del nitrógeno producido del aire parecía ser mayor que la del nitrógeno que se genera a partir de compuestos químicos. Sus conclusiones,

¿parecen ser justificadas aun cuando tiene tan pocos datos?

Las mediciones de Lord Raleigh, que primero aparecieron en *Proceedings, Royal Society* (Actas, Real Sociedad, Londres, 55, 1894, pp.340-344) aparecen a continuación. Los datos son la masa de nitrógeno que llena cierto frasco bajo presión y temperatura especificadas.

Atmosférica		Química	
2.31017	2.31010	2.30143	2.29940
2.30986	2.31028	2.29890	2.29849
2.31010	2.31163	2.29816	2.29889
2.31001	2.30956	2.30182	2.30074
2.31024		2.29869	2.30054

Fuente: http://exploringdata.cqu.edu.au/datasets/nitrogen.xls

a. Construya gráficas de puntos continuas de los dos conjuntos de datos, utilice para ello una escala común.
b. Calcule la media, mediana, desviación estándar, y los cuartiles primero y tercero para cada conjunto de datos.
c. Elabore gráficas de caja contiguas de los dos conjuntos de datos, usando una escala común.
d. Analice cómo se comparan estos dos conjuntos de datos. Este reducido par de conjuntos de datos ¿muestra evidencia convincente de una diferencia?

PARA SU INFORMACIÓN Las diferencias entre estos conjuntos de datos ayudaron a desencadenar el descubrimiento del gas argón.

2.189 Los golfistas que más dinero ganaron en el Torneo Nacional de 2004, junto con sus ganancias totales, aparecen a continuación:

Jugador	Dinero ($)	Jugador	Dinero ($)
Jimmy Walker	371 346	D. A. Points	332 815

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: PGA Tour. Inc.

a. Calcule la media y desviación estándar de lo que ganaron los golfistas del Torneo Nike.
b. Encuentre los valores de $\bar{x} - s$ y $\bar{x} + s$.
c. ¿Cuántas, de las 50 piezas de datos, tienen valores entre $\bar{x} - s$ y $\bar{x} + s$? ¿Qué porcentaje de la muestra es esto?
d. Encuentre los valores de $\bar{x} - 2s$ y $\bar{x} + 2s$.
e. ¿Cuántas de las 50 piezas de datos tienen valores entre $\bar{x} - 2s$ y $\bar{x} + 2s$? ¿Qué porcentaje de la muestra es esto?
f. Encuentre los valores de $\bar{x} - 3s$ y $\bar{x} + 3s$.
g. ¿Qué porcentaje de la muestra tiene valores entre $\bar{x} - 3s$ y $\bar{x} + 3s$?
h. Compare las respuestas encontradas en las partes e y g contra los resultados pronosticados por el teorema de Chebyshev.
i. Compare las respuestas encontradas en las partes c, e, y g contra los resultados pronosticados por la regla empírica. El resultado, ¿sugiere una distribución aproximadamente normal?
j. Verifique su respuesta al inciso i usando uno de los conjuntos de instrucciones de tecnología.
k. ¿Tiene sentido su respuesta al inciso j? Explique.

2.190 Pida a uno de sus maestros una lista de calificaciones del examen (15 a 25 calificaciones) que se aplicó a un grupo de alumnos.

a. Encuentre cinco medidas de tendencia central.
b. Determine tres medidas de dispersión.
c. Construya un diagrama de tallo y hoja. Este diagrama, ¿sugiere que las calificaciones están normalmente distribuidas?
d. Encuentre las siguientes medidas de ubicación: (i) Q_1 y Q_3, (ii) P_{15} y P_{60}, y (iii) la calificación z estándar para la calificación más alta.

2.191 Las longitudes (en milímetros) de 100 truchas cafés en el estanque 2-B del Criadero de Peces Tierra Feliz, el 15 de junio del año pasado, fueron como sigue:

15.0	15.3	14.4	10.4

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

a. Encuentre la media.
b. Encuentre la mediana.
c. Encuentre la moda.
d. Encuentre el rango central.
e. Encuentre el rango.
f. Encuentre Q_1 y Q_3.
g. Encuentre el cuartil medio.
h. Encuentre P_{35} y P_{64}.
i. Construya una distribución de frecuencia agrupada que use 10.0-10.5 como la primera clase.
j. Construya un histograma de la distribución de frecuencia.
k. Construya una distribución de frecuencia relativa acumulativa.
l. Construya una ojiva de la distribución de frecuencia relativa acumulativa.
m. Encuentre la media de la distribución de frecuencia. (Opcional)

n. Encuentre la desviación estándar de la distribución de frecuencia. (Opcional.)

2.192 El sistema nacional de carreteras está formado por carreteras interestatales y no interestatales. La Federal Highway Administration informó el número de millas de cada tipo en cada estado. A continuación aparece una muestra aleatoria de 20. Defina la "razón I/N" como el número de millas interestatales dividido entre el número de millas no interestatales.

Millas de carreteras interestatales y no interestatales por estado

Estado	Interestatal	No interestatal	Estado	Interestatal	No interestatal
NE	235	590	TN	1 073	2 171
FL	1 471	2 897	NJ	1 000	1 935
MA	367	924	LA	904	1 701
HI	55	291	TX	3 233	10 157
MT	1 192	2 683	OH	1 574	2 812
MN	912	3 060	IN	782	2 434
GA	1 245	3 385	NM	1 674	3 476
OK	930	2 431	NC	482	2 496
NV	1 019	2 743	AR	1 167	1 566
RI	71	198	DE	13	70

Fuente: Federal Highway Administration, U.S. Department of Transportation

a. Inspeccione los datos. ¿Cuál estima usted que es la razón "promedio" I/N?
b. Calcule la "razón I/N" para cada uno de los 20 estados de la lista.
c. Trace un histograma de la "razón I/N".
d. Calcule la "razón I/N" media para los 20 estados de la lista.
e. Use el número total de millas interestatales y no interestatales de 20 estados para calcular la "razón I/N" para los 20 estados combinados.
f. Explique por qué las respuestas a las partes d y e no son las mismas.
g. Calcule la desviación estándar para la "razón I/N" para los 20 estados de la lista.

2.193 El National Environmental Satellite, Data, and Information Service, U.S. Department of Commerce, publicó el area (millas cuadradas) y la población en el año 2000 para los 48 estados contiguos de Estados Unidos. Cuando estudie cuántas personas viven en un país tan grande y variado como Estados Unidos, quizá una variable más interesante a estudiar que la población de cada estado podría ser la densidad de población de cada estado dado que los 48 estados contiguos varían tanto en área. Defina "densidad" de un estado como la población del estado dividida entre su área.

Estado	Área (millas cuadradas)	Población
AL	51 610	4 447 100
AZ	113 909	5 130 632

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: U.S. Department of Commerce, http://www5.ncdc.noaa.gov/climatenormals/hcs/HCS_42.pdf

a. Mencione tres estados que usted piense estarán entre aquéllos con la más alta densidad. Justifique su selección.
b. Mencione tres estados que usted piense estarán entre aquellos con la más baja densidad. Justifique su selección.
c. Describa lo que usted piense del aspecto que tendrá la distribución de densidad. Incluya ideas de forma de distribución (normal, sesgada, etc.).
d. Usando los totales de los 48 estados, calcule la densidad general para los 48 estados contiguos. Usando la población y área de cada estado, calcule las densidades individuales para los 48 estados contiguos de Estados Unidos.
e. Calcule las medidas de tendencia central.
f. Construya un histograma.
g. Ordene por categorías los valores de densidad. Identifique los cinco estados con la densidad más alta y los cinco con la más baja.
g. Compare la distribución de información de densidad (respuestas a las partes e a la g) contra sus expectativas (respuestas a las partes a a la c). ¿Cómo lo hizo?

2.194 El volumen de árboles de navidad vendidos anualmente en Estados Unidos ha bajado en décadas recientes, según un informe del USDA National Agricultural Statistics Service. Los 50 estados informan de contribuciones de unos 25 millones de árboles de navidad anualmente a las ventas totales en Estados Unidos. Además, cada estado informa de su cosecha por condado. Los principales 20 condados producto en

Estados Unidos provienen de siete estados. El número de árboles vendidos por los principales 20 condados en 2002 aparece en la siguiente tabla. Este estudio se hace cada 5 años.

Número de árboles de navidad vendidos por condado (10,000 unidades)

42.8	25.4	84.8	36.5	16.7
21.3	87.6	65.4	140.0	15.2
41.2	17.2	15.0	103.0	22.1
25.6	20.3	259.0	64.2	19.1

Fuente: USDA National Agricultural Statistics Service

a. Calcule la media, mediana, y rango central para el número de árboles de navidad vendidos anualmente por los principales 20 condados productores.

b. Calcule la desviación estándar.

c. ¿Qué le dicen a usted las respuestas a las partes a y b acerca de la distribución para el número de árboles? Explique.

d. Observe que la desviación estándar es un número mayor a la media. ¿Qué significa eso en esta situación?

e. Trace una gráfica de puntos de los datos.

f. Localice los valores de las respuestas a las partes a y b para la gráfica de puntos trazada para la parte e.

g. Conteste de nuevo las partes c y d, usando la información aprendida de la gráfica de puntos.

2.195 A continuación aparecen los porcentajes de egresados de preparatoria por estado, de la clase 2003 que tomaron el American College Test.

0.73	0.07	0.16	0.10	0.52	0.62	0.69	0.70	0.12	0.76
0.32	0.05	0.60	0.69	0.73	0.15	0.12	0.74	0.16	0.73
0.27	0.30	1.00	0.67	0.34	0.15	0.08	0.33	0.63	0.80
0.73	0.41	0.21	0.88	0.08	0.80	0.06	0.67	0.69	0.07
0.15	0.22	0.66	0.69	0.06	0.64	0.34	0.11	0.62	0.12

Fuente: ACT Inc., The College Board

a. Examine los datos de la tabla. ¿Qué forma de distribución anticipa usted que producirán estos datos? Explique por qué eligió esa forma de distribución.

b. Construya el histograma de estos porcentajes.

c. Describa, en detalle, la distribución de su histograma de la parte b.

d. Compare su descripción de la parte c contra sus expectativas de la parte a. ¿Qué tan cerca está? ¿En qué no pensó de la parte a que es evidente ahora que ve la distribución real?

e. Encuentre el porcentaje medio.

f. ¿Dónde cae la media en la distribución? Localice la media en el histograma construido para la parte b. El porcentaje de la media, ¿es representativo de estos datos? Explique.

g. Encuentre la desviación estándar.

h. En el histograma construido para la parte b, localice los valores de 1 desviación estándar arriba y debajo de la media. ¿Cuánto de la distribución está entre estos valores?

i. ¿Por qué es tan grande la desviación estándar? Explique en detalle.

2.196 Las cantidades en dólares que aparecen enseguida son el promedio de lo que ganan por hora trabajadores de producción, o no supervisores, de nóminas de empresas privadas importantes no agrícolas. Investigue esta información, buscando cualquier configuración visual que pudiera existir. Encuentre estadísticas numéricas y gráficas por meses y por años. Describa todas las configuraciones que encuentre.

Año	Ene.	Feb.	Mar.
2001	14.48	14.54	14.58
2002	15.05	15.11	15.15

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.bls.gov/

a. Use los 5 años que se indican en el archivo.

b. Use los 11 años que se indican en el archivo.

2.197 ¿Quién comió las M&M? La tabla siguiente da las cantidades de colores y peso neto (en gramos) para una muestra de 30 bolsas de M&M. El peso neto anunciado es 47.9 gramos por bolsa.

Caso	Roja	Verde	Azul	Anaranjada	Amarilla	Café	Peso
1	15	9	3	3	9	19	49.79
2	9	17	19	3	3	8	48.98

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.math.uah.edu/stat/

Christine Nickel y Jason York, proyecto ST 687, verano 1998

Hay algo acerca de un caso de este conjunto de datos que es sospechosamente inconsistente con el resto de los datos. Encuentre la inconsistencia.

a. Construya dos gráficas diferentes para los pesos.

b. Calcule varias estadísticas numéricas para los datos del peso.

c. ¿Encontró algunas inconsistencias potenciales en las partes a y b? Explique.

d. Encuentre el número de M&M de cada bolsa.

e. Construya dos gráficas diferentes para el número de M&M por bolsa.

f. Calcule varias estadísticas numéricas para el número de M&M por bolsa.

g. ¿Qué inconsistencia encontró en las partes e y f? Explique.

h. Dé una explicación posible en cuanto a por qué la inconsistencia no aparece en los datos del peso, pero sí en los datos numéricos.

2.198 Para una distribución normal (o forma de campana), encuentre el rango de percentil que corresponde a:

a. $z = 2$ b. $z = -1$

c. Trace la curva normal, mostrando la relación entre la calificación z y los percentiles para las partes a y b.

2.199 Para una distribución normal (o forma de campana), encuentre la calificación z que corresponda al k-ésimo percentil:

a. $k = 20$ b. $k = 95$

c. Trace la curva normal, mostrando la relación entre la calificación z y los percentiles para las partes a y b.

2.200 Bill y Rob son buenos amigos, aun cuando estudian en diferentes escuelas en su ciudad. El sistema escolar de la ciudad utiliza una serie de exámenes físicos que aplica a todos los estudiantes de preparatoria. Después de completar los exámenes físicos, Bill y Rob están comparando sus calificaciones para ver quién estuvo mejor en cada evento. Necesitan ayuda.

	Sentadillas	"Lagartijas"	Carrera de relevos	Carrera de 50 yardas	Tiro de sóftbol
Bill	$z = -1$	$z = -1.3$	$z = 0.0$	$z = 1.0$	$z = 0.5$
Rob	61	17	9.6	6.0	179 ft
Media	70	8	9.8	6.6	173 ft
Desviación estándar	12	6	0.6	0.3	16 ft

Bill recibió los resultados de su prueba en calificaciones z, mientras que a Rob se le dieron calificaciones sin procesar. Como ambos entienden de calificaciones sin procesar, convierta las calificaciones z de Bill en unas sin procesar para hacer una comparación precisa.

2.201 Las gemelas Jean y Joan Wong están en quinto grado (diferentes secciones), y al grupo se le ha aplicado una serie de exámenes de pruebas. Si las calificaciones para estos exámenes están (en forma aproximada) normalmente distribuidas, ¿cuál muchacha tiene la calificación relativa más alta en cada una de las pruebas siguientes? Explique sus respuestas.

Prueba	Jean: Calificación z	Joan: Percentil
Condición física	2.0	99
Postura	1.0	69
Agilidad	1.0	88
Flexibilidad	−1.0	35
Fuerza	0.0	50

2.202 Es frecuente que las calificaciones alcanzadas por estudiantes en Estados Unidos sean noticia, y se sacan toda clase de conclusiones con base en estas calificaciones. El ACT Assessment está diseñado para evaluar el desarrollo educativo general de estudiantes de preparatoria y su capacidad para completar trabajos al nivel universitario. La tabla siguiente muestra la media y desviación estándar para las calificaciones de todos los egresados de preparatoria, en 2001 y en 2004, en los cuatro exámenes del ACT y sus calificaciones globales.

	Inglés	Matemáticas	Lectura	Razonamiento en ciencias	Global
2001					
Media	20.5	20.7	21.3	21.0	21.0
Desviación estándar	5.6	5.0	6.0	4.6	4.7
2004					
Media	20.4	20.7	21.3	20.9	20.9
Desviación estándar	5.9	5.0	6.0	4.6	4.8

Fuente: American College Testing

Con base en la información de la tabla:

a. Analice qué tan semejantes y diferentes entre sí son las cinco distribuciones respecto al valor central y dispersión.

b. Analice cualquier cambio en las calificaciones entre 2001 y 2004. Incluya en su respuesta aspectos específicos acerca de cómo ha cambiado, o no ha cambiado, cada distribución de prueba según el valor central y dispersión.

2.203 Es frecuente que las especificaciones de manufactura se basen en los resultados de muestras tomadas de lotes piloto satisfactorios. Los siguientes datos resultaron de una de estas situaciones, en la que ocho lotes piloto se completaron y muestrearon. Las dimensiones de partículas resultantes son en angstroms (donde 1 Å $= 10^{-8}$ cm):

3923	3807	3786	3710	4010	4230	4226	4133

a. Encuentre la media muestral.

b. Encuentre la desviación estándar muestral.

c. Suponiendo que la dimensión de la partícula tiene una distribución aproximadamente normal, determine la especificación de manufactura que limita 95% de las dimensiones de partículas (es decir, encuentre el intervalo de 95%, $\bar{x} \pm 2s$).

2.204 Delco Products, división de General Motors, produce un soporte que se usa como parte de un conjunto eléctrico de seguros de puertas. La longitud de este soporte se vigila constantemente. Una muestra de 30 soportes eléctricos de puertas tuvo las siguientes longitudes (en milímetros):

11.86	11.88	11.88	11.91	11.88	11.88	11.88	11.88	11.88	11.86
11.88	11.88	11.88	11.88	11.86	11.83	11.86	11.86	11.88	11.88
11.88	11.83	11.86	11.86	11.86	11.88	11.88	11.86	11.88	11.83

Fuente: Con permiso de Delco Products Division, GMC

a. Sin hacer cálculo alguno, ¿qué estimaría usted para la media muestral?

b. Construya una distribución de frecuencia no agrupada.

c. Trace un histograma de esta distribución de frecuencia.

d. Utilice la distribución de frecuencia y calcule la media muestral y desviación estándar.

e. Determine los límites del intervalo $\bar{x}\ 3s$ y marque este intervalo en el histograma.

f. Los límites de especificación del producto son 11.7-12.3. ¿Indica la muestra que la producción está dentro de estos requisitos? Justifique su respuesta.

2.205 A los estadounidenses les gustan las sopas, que siguen siendo uno de los alimentos preferidos para el refrigerio y como bocadillo. Los fabricantes proporcionan las calorías y el contenido de sodio en la etiqueta. Los datos para 40 latas y mezclas de varias porciones (8 onzas), casi todas las cuales eran variedades de bajo contenido en grasas, aparecen en la tabla siguiente.

Marca de sopa	Calorías	Sodio (mg)
Arrowhead Mills Red Lentil	100	230
Baxters Italian Bean & Pasta	80	430

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

a. Calcule la media y desviación estándar de calorías y contenido de sodio de las sopas que aparecen en la tabla.

b. Utilice sus respuestas a la parte a para probar el teorema de Chebyshev de que al menos 75% de las calorías y contenido de sodio de las sopas caerá dentro de $\pm$ 2 desviaciones estándar de la media. ¿Es éste el caso?

c. Encuentre los límites para una desviación estándar de $\pm$ 1 de la media para el contenido de sodio de las sopas. ¿Parece éste seguir la regla empírica? Explique

2.206 El gerente de la peluquería Jerry's recientemente pidió a sus últimos 50 clientes perforaran una tarjeta de tiempo cuando llegaran primero a la peluquería, y volvieron a perforarla cuando pagaran su cuenta de corte de pelo. El gerente utilizó los datos de las tarjetas para medir cuánto tiempo les tomó a él y sus peluqueros cortar el pelo, y con esta información programó sus intervalos de citas. Se tabularon los tiempos siguientes (en minutos):

a. Construya una gráfica de tallo y hoja de estos datos.

50	21	36	35	35	27	38	51	28	35
32	32	27	25	24	38	43	46	29	45
40	27	36	38	35	31	28	38	33	46
35	31	38	48	23	35	43	31	32	38
43	32	18	43	52	52	49	53	46	19

b. Calcule la media, mediana, moda, rango, rango central, varianza y desviación estándar de los tiempos de servicio de corte de pelo.

c. Construya una tabla de resumen de 5 números.

d. Según el teorema de Chebyshev, ¿al menos 75% de los tiempos de servicio de corte de pelo caerán entre cuáles dos valores? ¿Es cierto esto? Explique por qué o por qué no.

e. ¿Con qué separación recomendaría usted que Jerry programe sus citas, para mantener peluquería operando a un paso cómodo?

2.207 Cada año, los corredores de autos de categoría libre compiten para el premio NASCAR. Ganan puntos con base en carreras que terminen programadas en el circuito. Al finalizar la temporada de 2004, las posiciones se publicaron en NASCAR.com; los primeros 32 corredores se muestran en la siguiente tabla:

Corredor	Puntos	Corredor	Puntos
Kurt Busch	6506	Jimmie Johnson	6498

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: NASCAR

a. Trace una gráfica de puntos.

b. Calcule la media y desviación estándar de los puntos acumulados por corredores de la NASCAR.

c. Construya una tabla de resumen de 5 números y trace un diagrama de caja y bigotes.

d. Según el teorema de Chebyshev, ¿al menos 75% de los puntos caerán entre cuáles dos cantidades? ¿Es éste el caso?

e. Según la regla empírica, ¿aproximadamente 68% de los puntos caerán entre cuáles dos cantidades? ¿Es éste el caso?

f. Compare sus respuestas a las partes d y e contra los resultados pronosticados por la regla empírica. Su comparación, ¿sugiere que la distribución de puntos de la NASCAR se aproxima a la distribución normal? Explique.

g. (Opcional.) La temporada 2004 tuvo un total de 88 corredores que ganaron puntos. La lista completa está incluida en el archivo de datos para este ejercicio. Usando las 88 piezas de datos, conteste las preguntas de las partes a a la e.

2.208 La siguiente gráfica de puntos muestra el número de intento de pases lanzados por los mariscales de campo de 22 de los equipos de la NFL, que jugaron en una tarde particular de domingo.

a. Describa la distribución, incluyendo la forma en que se relacionan los puntos *A* y *B* con los otros.

b. Si se elimina el punto *A*, y quizás el punto *B*, ¿se diría que los datos restantes tienen una distribución aproximadamente normal? Explique.

c. Con base en la información acerca de distribuciones que nos dan el teorema de Chebyshev y la regla empírica, ¿qué tan típico es un evento que piensa usted que representa el punto *A*? Explique.

2.209 Comenzando con los datos de 70 y 85, sume tres datos a su muestra para que ésta tenga lo siguiente (justifique su respuesta en cada caso).

a. Una desviación estándar de 5

b. Una desviación estándar de 10

c. Una desviación estándar de 15

d. Compare sus tres muestras y la variedad de valores necesarios para obtener cada una de las desviaciones estándar pedidas.

2.210 Elabore un conjunto de 18 datos (considérelos como calificaciones de examen) para que la muestra satisfaga cada uno de estos conjuntos de criterios:

a. La media es 75, y la desviación estándar es 10.

b. La media es 75, el máximo es 98, el mínimo es 40, y la desviación estándar es 10.

Figura para el ejercicio 2.208

c. La media es 75, el máximo es 98, el mínimo es 40, y la desviación estándar es 15.

d. ¿Qué tan diferentes son los datos de la muestra de la parte b respecto los de la parte c?

2.211 Construya dos gráficas diferentes de los puntos (62.2), (74,14), (80,20), y (94,34).

a. En la primera gráfica, en todo el eje horizontal, trace intervalos iguales y márquelos como 62, 74, 80 y 94; trace intervalos iguales en todo el eje vertical y márquelos como 0, 10, 20, 30 y 40. Localice los puntos y enlácelos con segmentos de recta.

b. En la segunda gráfica, en todo el eje horizontal, trace intervalos igualmente espaciados y márquelos como 60, 65, 70, 75, 80, 85, 90 y 95; delimite el eje vertical en intervalos iguales y márquelos como 0, 10, 20, 30 y 40. Localice los puntos y enlácelos con segmentos de recta.

c. Compare el efecto que esa escala tiene en el aspecto de las gráficas de las partes a y b. Explique la impresión presentada por cada gráfica.

2.212 Cuando se realizó el estudio de Internet "Usted y la Internet" (sección 2.1, p. 39), parecía que la variable *x*, el número de actividades por Internet en una semana, tenía una distribución aproximadamente normal. Esa distribución está aproximada por esta distribución de frecuencia relativa:

Actividades en Internet/ semana, *x*	Frecuencia relativa	Actividades en Internet/ semana, *x*	Frecuencia relativa
1	0.01	8	0.14
2	0.03	9	0.11
3	0.05	10	0.08
4	0.09	11	0.05
5	0.10	12	0.04
6	0.14	13	0.03
7	0.13		

a. Seleccione una muestra aleatoria de tamaño 40 de esta representación de frecuencia relativa de la población de todos los usuarios de Internet.

b. Construya un histograma de la muestra obtenida en la parte a. No agrupe los datos. (Vea las instrucciones que siguen.)

MINITAB (Versión 14)

Introduzca los valores *x* en C1 y las correspondientes frecuencias relativas en C2; luego continúe con:

```
Choose:     Calc > Random Data > Discrete
Enter:      Generate: 40 rows of data
            Store in column(s): C3
            Values (of x) in: C1
            Probabilities in: C2 > OK
```

Excel

Introduzca los valores *x* en la columna A y las correspondientes frecuencias relativas en la columna B; luego continúe con:

```
Choose:     Tools > Data Analysis > Random Number
            Generation > OK
Enter:      Number of Variables: 1
            Number of Random Numbers: 40
            Distribution: Discrete
            Value & Prob. Input Range:
            (A2:B5 select data cells not labels)
Select:     Output Range
Enter:      (C1 or select cell)
```

c. Encuentre la media, mediana y la desviación estándar de la muestra obtenida en la parte a.

d. Repita las partes a-c tres veces más, asegurándose de conservar las respuestas para cada conjunto de datos juntos.

e. Describa las similitudes y diferencias entre las distribuciones mostradas en los cuatro histogramas.

f. Elabore una gráfica que muestre las estadísticas numéricas para cada una de las cuatro muestras y describa la variabilidad de una muestra a otra para cada estadística.

g. Haga una tabla que muestre las estadísticas numéricas a partir de la misma distribución. Escriba una frase que describa la variabilidad total entre estas cuatro muestras aleatorias.

2.213 Utilice computadora para generar una muestra aleatoria de 500 valores de una variable *x* normalmente distribuida con una media de 100 y una desviación estándar de 20. Construya un histograma de los 500 valores.

a. Utilice los comandos de computadora de la página 101 para generar al azar una muestra de 500 datos desde una distribución normal con una media de 100 y desviación estándar de 20. Construya un histograma usando fronteras de clase que sean múltiplos de la desviación estándar 20; esto es, use

fronteras de 20 a 180 en intervalos de 20 (vea comandos en las páginas 61-62).

Consideremos los 500 valores x encontrados en la parte a como una población.

b. Use los comandos de computadora de las páginas 101-102 para seleccionar al azar una muestra de 30 valores de la población encontrada en la parte a. Construya un histograma de la muestra con los mismos intervalos de clase usados en la parte a.

c. Repita la parte b tres veces.

d. Calcule diversos valores (media, mediana, máximo, mínimo, desviación estándar, etc.) que describan la población y cada una de las cuatro muestras. (Vea comandos en la p. 89.)

e. ¿Piensa usted que una muestra de 30 datos representa en forma adecuada una población? (Compare cada una de las cuatro muestras encontradas en las partes b y c contra la población.)

2.214 Repita el ejercicio 2.213 usando un tamaño diferente de muestra. Podría intentar unos pocos tamaños diferentes de muestra: $n = 10$, $n = 15$, $n = 20$, $n = 40$, $n = 50$, $n = 75$. ¿Qué efecto tiene aumentar el tamaño de la muestra sobre la efectividad de la muestra al describir la población? Explique.

2.215 Repita el ejercicio 2.213 usando poblaciones con distribuciones de diferente forma.

a. Utilice una distribución uniforme o rectangular. (Sustituya los subcomandos empleados en el ejercicio 2.213; en lugar de NORMAL use: UNIFORM con un bajo de 50 y un alto de 150, y use fronteras de clase de 50 a 150 en incrementos de 10.)

b. Use una distribución sesgada. (Sustituya los subcomandos empleados en el Ejercicio 2.213; en lugar de NORMAL use: POISSON 50 y use fronteras de clase de 20 a 90 en incrementos de 5.)

c. Use una distribución en forma de J. (Sustituya los subcomandos empleados en el ejercicio 2.213; en lugar de NORMAL use: EXPONENTIAL 50 y use fronteras de clase de 0 a 250 en incrementos de 10.)

d. La forma de la distribución de la población, ¿tiene un efecto en lo bien que una muestra de tamaño 30 representa la población? Explique.

e. ¿Qué efecto piensa que tiene cambiar el tamaño muestral sobre la efectividad de la muestra para describir la población? Intente unos pocos tamaños muestrales diferentes. ¿Los resultados concuerdan con sus expectativas? Explique.

2.216 *¡Resultados aislados!* ¿Con qué frecuencia se presentan? ¿Qué hacemos con ellos? Complete la parte a para ver la frecuencia con que se presentan. Luego complete la parte b para decidir qué hacer con los resultados aislados.

a. Use la tecnología de su selección para tomar muestras de diversos tamaños (10, 30, 100, 300 serían buenas opciones) de una distribución normal (una media de 100 y desviación estándar de 20 funcionan bien) y vea cuántos resultados aislados contiene una muestra generada en forma aleatoria. Es probable que le sorprenda. Genere 10 muestras de cada tamaño para un resultado más representativo. Describa sus resultados; en particular comente sobre la frecuencia de los resultados aislados en sus muestras.

MINITAB

```
Choose:     Calc > Random Data > Normal
Enter:      Generate    10      rows of data
            (Use n = 10, 30, 100, 300)
            Store in column(s): C1–C10
            Mean:       100
            Stand. Dev.: 20
Choose:     Graph > Boxplot > Multiple Y's Simple >
            OK
Enter:      Graph variables: C1–C10
Choose:     Data View
Select:     Interquartile range box
            Outlier symbols
```

En la práctica, deseamos hacer algo acerca de los puntos de datos que se descubra son resultados aislados. Primero este resultado aislado debe inspeccionarse: si hay alguna razón obvia de por qué es incorrecto, debe corregirse. (Por ejemplo, la estatura de 59 pulgadas de una mujer bien podría introducirse incorrectamente como 95 pulgadas, lo cual sería casi 8 pies de estatura y algo muy poco probable). Si el valor del dato puede corregirse, hágalo. De otra forma, debe ponderar la opción entre descartar buenos datos (incluso si son diferentes) y conservar datos erróneos. En este nivel, es probable que sea mejor hacer una nota acerca del resultado aislado y continuar usando la solución. Para ayudar a entender el efecto de eliminar un valor de resultado aislado, veamos este conjunto de datos generado en forma aleatoria de una distribución normal $N(100, 20)$.

b. Construya una gráfica de caja e identifique cualesquier resultados aislados.

74.2	84.5	88.5	110.8	97.6	100.2	116.4	78.3	154.8	144.7
110.6	93.7	113.3	96.1	86.7	97.3	102.8	91.8	58.5	120.1
102.8	82.5	107.6	91.1	95.7	98	98.4	81.9	58.5	118.1

c. Elimine el resultado aislado y construya una nueva gráfica de caja.

d. Describa su hallazgo y comente sobre por qué podría ser mejor y menos confuso no descartar resultados aislados cuando estudie introducción a la estadística.

2.217 La distribución de horas de crédito, por estudiante, tomada este semestre en cierta universidad fue como sigue:

Horas de crédito	Frecuencia	Horas de crédito	Frecuencia
3	75	15	400
6	150	16	1050
8	30	17	750
9	50	18	515
12	70	19	120
14	300	20	60

a. Trace un histograma de los datos.

b. Encuentre las cinco medidas de tendencia central.

c. Encuentre Q_1 y Q_3.

d. Encuentre P_{15} y P_{12}.

e. Encuentre las tres medidas de dispersión (rango, s^2 y s).

2.218 Un artículo en *Therapeutic Recreation Journal* informa de una distribución para la variable "número de desacuerdos persistentes". Sesenta y seis pacientes y su especialista de recreación terapéutica contestaron sí o no a una lista de problemas. El desacuerdo ocurre cuando el especialista y el paciente no responden de manera idéntica a un renglón de la lista. Se convierte en desacuerdo persistente si el renglón permanece en desacuerdo después de una segunda entrevista.

x	0	1	2	3	4	5	6	7	8	9	10	11
y	2	2	4	10	7	9	8	11	7	3	1	2

Fuente: Datos reimpresos con permiso de la National Recreation and Park Association, Alexandria, VA, de Pauline Petryshen y Diane Essex-Sorlie, "Persistent Disagreement Between Therapeutic Recreation Specialists and Patients in Psychiatric Hospitals," *Therapeutic Recreation Journal*, Vol. XXIV, tercer trimestre, 1990.

a. Trace una gráfica de puntos sobre estos datos muestrales.

b. Encuentre el número mediano de desacuerdos persistentes.

c. Encuentre el número medio de desacuerdos persistentes.

d. Encuentre la desviación estándar del número de desacuerdos persistentes.

e. Trace una recta vertical en la gráfica de puntos en la media.

f. Trace un segmento de recta horizontal en la gráfica de puntos cuya longitud represente la desviación estándar (empezando en la media).

2.219 *USA Today* (25 de octubre, 1994) informó en el USA Snapshot "Mystery of the remote" que 44% de las familias encuestadas nunca ponían fuera de su lugar el control remoto del televisor de su casa, 38% sí lo perdían de una a cinco veces por semana y 17% lo perdían más de cinco veces por semana; uno por ciento de las familias encuestadas no lo sabía. Suponga que usted toma una encuesta que resultó en los datos siguientes. Sea x el número de veces por semana que el control remoto del televisor de la familia se pierde.

x	0	1	2	3	4	5	6	7	8	9
f	220	92	38	21	24	30	34	20	16	5

a. Construya un histograma.

b. Encuentre la media, mediana, moda y rango central.

c. Encuentre la varianza y desviación central.

d. Encuentre Q_1, Q_3 y P_{90}.

e. Encuentre el cuartil central.

f. Encuentre el resumen de 5 números y trace un diagrama de caja y bigotes.

2.220 La siguiente tabla muestra la distribución de edades de jefes de familias:

Edad del jefe de familia	Número	Edad del jefe de familia	Número
20–25	23	50–55	48
25–30	38	55–60	39
30–35	51	60–65	31
35–40	55	65–70	26
40–45	53	70–75	20
45–50	50	75–80	16

a. Encuentre la edad media de los jefes de familias.

b. Encuentre la desviación estándar.

2.221 Se obtuvo la vida útil de 220 lámparas incandescentes de 60 watts, así como la distribución de frecuencia que se muestra en esta tabla:

Límites de clase	f	Límites de clase	f
500–600	3	1000–1100	57
600–700	7	1100–1200	23
700–800	14	1200–1300	13
800–900	28	1300–1400	7
900–1000	64	1400–1500	4

a. Construya un histograma de estos datos usando una escala vertical para las frecuencias relativas.

b. Encuentre la vida útil media.

c. Encuentre la desviación estándar de las vidas útiles.

2.222 El pago mensual por su automóvil, ¿le impide gastar en otras cosas? Más de 56% dicen "sí". La distribución de la cantidad gastada en pagos mensuales por el auto es como sigue: 32% gastan menos de $300, 43% gastan de $300 a $499, 17% gastan de $500 a $699, y 8% gastan $700 o más. Suponga que esta información se obtuvo de una muestra de 1000 personas que hacen pagos por sus autos. Use valores de $150, $400, $600 y $800 como puntos centrales de clase, y estime la media muestral y la desviación estándar para la variable *x*, cantidad gastada.

2.223 A continuación se muestran los beneficios por acción para 40 empresas de la industria de equipos de radio y de transmisión:

4.62	0.10	1.29	7.25	6.04	3.20	9.56	4.90	4.22	3.71
0.25	1.34	2.11	5.39	0.84	−0.19	3.72	2.27	2.08	1.12
1.07	2.50	2.14	3.46	1.91	7.05	5.10	1.80	0.91	0.50
5.56	1.62	1.36	1.93	2.05	2.75	3.58	0.44	3.15	1.93

a. Elabore una distribución de frecuencia y un histograma de frecuencias para estos datos.

b. ¿Qué clase de su distribución de frecuencia contiene la mediana?

Proyecto del capítulo

Usted y la Internet

Regresemos a la sección 2.1, "Usted y la Internet" (p. 39), como una forma de evaluar lo que hemos aprendido en este capítulo. Con base en los porcentajes indicados en la gráfica "Lo que hacen usuarios de la Internet", quienes respondieron pudieron escoger más de una actividad en Internet. Supongamos que a usted se le pregunta ¿cuántas de las actividades citadas seleccionaría como algo que usted hace?, y que a una muestra de estudiantes también se le preguntó acerca de sus actividades en Internet. ¿Sus respuestas diferirían de la de usted? ¿Diferirían de los 4000 que respondieron? "Trabajando en el contenido del capítulo 2" nos ayudará a contestar estas preguntas.

Trabajando en el contenido del capítulo 2

2.224 A los estudiantes de un curso de estadística que se ofrece en Internet se les preguntó en cuántas actividades diferentes en Internet participan en una semana típica. Los datos siguientes muestran el número de actividades:

6	7	3	6	9	10	8	9	9	6	4	9	4	9
4	2	3	5	13	12	4	6	4	9	5	6	9	
11	5	6	5	3	7	9	6	5	12	2	6	9	

a. Haga una lista de todos los tipos de tablas y gráficas mostradas en el capítulo 2 que serían apropiados para usar con el conjunto de los 40 datos de la lista.

b. ¿Qué tipos de gráficas no serían apropiados? Explique por qué.

c. Exhiba los datos usando cada una de las tablas y gráficas citadas en la parte a.

d. ¿Qué gráfica piensa usted que mejor representa los datos? Explique por qué.

e. Encuentre las cinco medidas de tendencia central para estos datos (media, mediana, moda, rango central y cuartil central).

f. Encuentre las tres medidas de dispersión para los datos (rango, varianza y desviación estándar).

g. Encuentre el valor de varias medidas de posición: P_5, P_{10}, Q_1, Q_3, P_{90} y P_{98}.

h. ¿En cuántas actividades diferentes de Internet participa usted durante una semana típica? Usando la media y desviación estándar calculada en las partes e y f, determine su calificación z. ¿Qué le dice esto acerca de usted mismo respecto al uso de Internet para estudiantes de estadística?

i. Use una gráfica de la parte c más al menos una medida de tendencia central y una medida de dispersión; también escriba una descripción del uso de Internet para estudiantes de estadística, el número de actividades de Internet por semana.

j. Según la regla empírica, si la distribución es normal, aproximadamente 68% del número de diferentes actividades en Internet que realizan estudiantes de estadística caerá entre cuáles dos valores? ¿Es esto cierto? ¿Por qué sí o por qué no?

k. Según el teorema de Chebyshev, ¿aproximadamente 75% del número de diferentes actividades en Internet que realizan estudiantes de estadística caerán dentro de cuáles dos valores? ¿Es esto cierto? ¿Por qué sí o por qué no?

l. La información muestral descrita en la gráfica "Lo que hacen usuarios de la Internet", de la sección 2.1, es diferente pero está relacionada con la información muestral con la que ha estado usted trabajando en las partes a-k. Describa los datos recolectados de la gráfica de la sección 2.1 y explique cómo difieren de los datos de la lista que aparecen en este ejercicio.

Para su estudio

2.225 a. Diseñe su propio estudio del uso de Internet. Defina una población específica de la que usted tomará muestras, describa su plan de muestreo, recolecte sus datos, y conteste a las partes c-l en "Trabajando en el contenido del capítulo 2", ejercicio 2.224.

b. Analice las diferencias y similitudes entre el uso de Internet descrito por la muestra de 40 estudiantes de estadística (dado en el ejercicio 2.224) y la muestra de usted.

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero; si no lo es, cambie las palabras en negrita con las que hagan que el enunciado sea siempre verdadero.

2.1 La **media** de una muestra siempre divide los datos en dos mitades (mitad más grande y mitad más pequeña en valor que sí misma).

2.2 Una medida de **tendencia central** es un valor cuantitativo que describe qué tan dispersos están los datos alrededor de un valor central.

2.3 La suma de los cuadrados de las desviaciones desde la media, $\sum(x-\bar{x})^2$, **a veces** será negativa.

2.4 Para cualquier distribución, la suma de las desviaciones desde la media es igual a **cero**.

2.5 La desviación estándar para el conjunto de valores 2, 2, 2, 2 y 2 es **2**.

2.6 En un examen, Juan calificó en el 50avo percentil y Jorge en el 25avo percentil; por tanto, la calificación del examen de Juan fue el **doble** de la de Jorge.

2.7 La frecuencia de una clase es el número de piezas de datos cuyos valores caen dentro de las **fronteras** de esa clase.

2.8 Se usan **distribuciones de frecuencia** en estadística para presentar en una forma concisa grandes cantidades de valores repetitivos.

2.9 La unidad de medida para la calificación estándar es siempre **desviaciones estándar**.

2.10 Para una distribución en forma de campana, el rango será aproximadamente igual a **6 desviaciones estándar**.

SEGUNDA PARTE: Aplicación de conceptos

2.11 Los resultados de un estudio de consumidores completado en la Corner Convenience Store se presentan en el histograma siguiente. Conteste cada una de las preguntas.

a. ¿Cuál es el ancho de clase?

b. ¿Cuál es el punto central de clase para la clase 31-61?

c. ¿Cuál es la frontera superior para la clase 61-91?

d. ¿Cuál es la frecuencia de la clase 1-31?

e. ¿Cuál es la frecuencia de la clase que contiene el valor de x más grande observado?

f. ¿Cuál es la frontera inferior de la clase con la frecuencia más grande?

g. ¿Cuántas piezas de datos se muestran en este histograma?

h. ¿Cuál es el valor de la moda?

i. ¿Cuál es el valor del rango central?

j. Estime el valor del 90avo percentil, P_{90}.

2.12 Una muestra de compras de varios clientes de la Corner Convenience Store resultó en los siguientes datos muestrales (x = número de artículos comprados por cliente):

x	1	2	3	4	5
f	6	10	9	8	7

a. ¿Qué representa el 2?

b. ¿Qué representa el 9?

c. ¿Cuántos clientes se usaron para formar esta muestra?

d. ¿Cuántos artículos fueron comprados por los clientes de esta muestra?

e. ¿Cuál es el número más grande de artículos comprados por un cliente?

Encuentre cada uno de lo siguiente (muestra fórmulas y trabajo):

f. Moda g. Mediana h. Rango central

b. Media j. Varianza k. estándar

2.13 Dado el conjunto de datos 4, 8, 9, 8, 6, 5, 7, 5, 8, encuentre cada una de las siguientes estadísticas muestrales:

a. Media b. Mediana c. Moda

d. Rango central e. Primer cuartil f. P_{40}

g. Varianza h. Desviación estándar

i. Rango

2.14 a. Encuentre la calificación estándar para el valor $x = 452$ respecto a su muestra, donde la media muestral es 500 y la desviación estándar es 32.

b. Encuentre el valor de x que corresponda a la calificación estándar de 1.2, donde la media es 135 y la desviación estándar es 15.

TERCERA PARTE: Para entender los conceptos

Conteste todas las preguntas.

2.15 La Corner Convenience Store dio seguimiento del número de clientes pagadores que tuvo diariamente, en las horas de mediodía, durante 100 días. Las estadísticas resultantes están redondeadas al entero más cercano:

media = 95 rango central = 93

mediana = 97 rango = 56

moda = 98 desviación

primer cuartil = 85 estándar = 12

tercer cuartil = 107

a. ¿La Corner Convenience Store sirvió a qué número de clientes pagadores durante las horas de medio, con más frecuencia que a cualquier otro número? Explique cómo determinó su respuesta.

b. ¿En cuántos días hubo entre 85 y 107 clientes pagadores durante las horas de mediodía? Explique cómo determinó su respuesta.

c. ¿Cuál fue el número máximo de clientes pagadores durante cualquier hora del mediodía? Explique cómo determinó su respuesta.

d. ¿Para cuántos de los 100 días estuvo el número de clientes pagadores dentro de 3 desviaciones estándar de la media ($x \pm 3s$)? Explique cómo determinó su respuesta.

2.16 Mr. VanCott inició su propio taller de máquinas hace varios años. Su negocio ha crecido y ha tenido mucho éxito en años recientes.

Actualmente emplea a 14 personas, incluyéndose él mismo, y paga los siguientes salarios anuales:

Propietario, Presidente	$80 000	Trabajador	$25 000
Gerente de finanzas	50 000	Trabajador	25 000
Gerente de producción	40 000	Trabajador	25 000
Mayordomo de taller	35 000	Trabajador	20 000
Trabajador	30 000	Trabajador	20 000
Trabajador	30 000	Trabajador	20 000
Trabajador	28 000	Trabajador	20 000

a. Calcule los cuatro "promedios": media, mediana, moda y rango central.

b. Trace una gráfica de puntos de los salarios y localice en ella cada uno de los cuatro promedios.

c. Suponga que usted fuera el escritor de artículos asignado a escribir el caso de esta semana acerca del taller de máquinas de Mr. VanCott, uno de una serie sobre pequeños negocios de la localidad que están prosperando. Usted planea entrevistar a Mr. VanCott, a su gerente de finanzas, al mayordomo del taller y a uno de sus trabajadores más recientes. ¿Qué promedio estadístico piensa usted que le dará cuando se le pregunte, "¿Cuál es el salario anual promedio pagado a los empleados aquí?" Explique por qué cada persona entrevistada tiene una perspectiva diferente y por qué este punto de vista puede hacer que cada uno cite un promedio estadístico diferente.

d. ¿Qué hay de la distribución de estos salarios que hace que los cuatro "valores promedio" sean tan diferentes?

2.17 Invente un conjunto de datos que contenga tres o más valores en los siguientes casos:

a. Donde la media sea 12 y la desviación estándar sea 0

b. Donde la media sea 20 y el rango sea 10

c. Donde la media, media y moda sean todas iguales

d. Donde la media, mediana y moda sean todas diferentes

e. Donde la media, mediana y la moda sean todas diferentes y la media sea la más grande y la moda sea la más pequeña

f. Donde la media, mediana y la moda sean todas diferentes y la media sea la más grande y la mediana sea la más pequeña

2.18 Un conjunto de exámenes fue calificado a máquina. Posteriormente se descubrió que deben sumarse 2 puntos a cada calificación. El estudiante A dijo, "La calificación media debe también aumentarse en 2 puntos". El estudiante B agregó, "La desviación estándar también debe aumentarse en 2 puntos". ¿Quién tiene razón? Justifique su respuesta.

2.19 El estudiante A dijo, "La desviación estándar y la varianza preservan la misma unidad de medida que los datos". El estudiante B no estuvo de acuerdo, diciendo "La unidad de medida para la varianza es una unidad de medición que no tiene sentido". ¿Quién tiene razón? Justifique su respuesta.

CAPÍTULO

3 Análisis descriptivo y presentación de datos bivariados

3.1 El chico ha crecido

3.2 Datos bivariados

3.3 Correlación lineal

3.4 Regresión lineal

AP/Wide World Photos

3.1 El chico ha crecido

MINNEAPOLIS El chico ha crecido y tiene el premio del jugador más valioso de la Asociación Nacional de Baloncesto (NBA) para demostrarlo. Kevin Garnett obtuvo 120 de los 123 votos de primer lugar para vencer al ganador en dos ocasiones Tim Duncan para el honor del lunes, tres días después que su equipo "Timberwolves" de Minnesota ganaron por primera vez una serie de eliminatorias. Los compañeros de equipo de Garnett asistieron a una atestada conferencia de prensa en la cancha de los Timberwolves, y él los elogió en repetidas ocasiones.

Jugando por toda la cancha, de centro a defensa, el jugador de 7 pies de estatura promedió 24.2 puntos, fue líder de la liga con 13.9 rebotes y 5.0 asistencias esta temporada, y sus estadísticas en eliminatorias son incluso mejores. Garnett se unió a Larry Bird como los únicos jugadores en promediar 20 puntos, 10 rebotes y cinco asistencias en cinco años consecutivos.

Apodado "El Chico", Garnett llegó al equipo de estrellas en su segunda temporada, y su éxito ayudó a alimentar la ola de jugadores a pasar de novatos a profesionales.

GARNETT GANA 120 DE 123 VOTOS DE PRIMER LUGAR
Associated Press, Lunes, 4 de mayo, 2004

Timberwolves de Minnesota, temporada regular 2003-2004

Jugador	Faltas personales por juego	Puntos por juego	Jugador	Faltas personales por juego	Puntos por juego
Garnett	2.5	24.2	Madsen	2.4	3.6
Cassell	3	19.8	Martin	1.4	3.4
Sprewell	1.2	16.8	McLeod	1.2	2.7
Szczerbiak	1.5	10.2	Goldwire	1.0	2.6
Hudson	1.1	7.5	Miller	1.9	2.5
Hoiberg	1.7	6.7	Johnson	2.4	1.9
Olowokandi	3.2	6.5	Lewis	0.7	1.1
Trent	1.9	5.6	Ebi	0.4	0.8
Hassell	2.5	5			

Fuente: http://sports.espn.go.com/nba/teams

¿Juega usted baloncesto, o al menos es aficionado? ¿Le parece que quienes encestan más puntos también cometen más faltas? Piense en aquellos que no anotan ningún punto, ¿incluso así cometen faltas? Ahora, aun cuando usted no juega baloncesto ni es aficionado sí sabe de relaciones. Piense en sí mismo; le parece que cuando joven en desarrollo ¿la medida de sus zapatos también aumentaba? ¿Hay alguna relación entre la estatura de una persona y la medida de sus zapatos? ¿Le parece que los estudiantes que estudian más obtienen mejores calificaciones? ¿Hay alguna relación entre horas estudiadas y calificaciones? ¿Le parece que los estudiantes que recorren más distancia en ir a la escuela también necesitan más tiempo para llegar a ésta? Cuando trabaje este capítulo 3 aprenderá a presentar dos datos variables de modo que se vea la relación entre ambos; desde este punto podrá determinar la fuerza de la relación, llamada *correlación*, y la ecuación de la recta empleada para pronósticos, llamada *análisis de regresión.* Una vez que complete los tres temas principales que acabamos de indicar, podrá investigar más la forma en que "El chico ha crecido" en el proyecto de capítulo de la página 199.

SECCIÓN 3.1 EJERCICIOS

3.1 Consulte los datos de los Timberwolves de la página 145 para contestar las preguntas siguientes:

a. ¿Hay una relación (patrón) entre las dos variables, puntos anotados por juego y número de faltas personales cometidas por juego? Explique por qué sí o por qué no.

b. ¿Piensa usted que es razonable (o posible) pronosticar el número de puntos anotados, con base en el número de faltas personales cometidas por juego para un jugador de los Timberwolves? Explique por qué sí o por qué no.

3.2 a. ¿Hay relación entre la estatura de una persona y la medida de sus zapatos cuando pasa de la infancia a la edad de 16 años? Cuando una variable aumenta, ¿también aumenta la otra? Explique sus respuestas.

b. ¿Hay una relación entre estatura y medida de calzado para personas mayores de 16 años? Las personas de más estatura ¿usan zapatos más grandes? Explique sus respuestas.

3.2 Datos bivariados

En el capítulo 2 vimos cómo presentar en forma gráfica y describir numéricamente datos muestrales para una variable. Ahora expandiremos estas técnicas para abarcar datos muestrales que comprenden dos variables pareadas.

Datos bivariados: son los valores de dos variables diferentes que se obtienen del mismo elemento poblacional.

Cada una de las dos variables puede ser *cualitativa* o *cuantitativa.* En consecuencia, tres combinaciones de tipos de variable pueden formar datos bivariados:

1. Ambas variables son cualitativas (atributos).
2. Una variable es cualitativa (atributo) y la otra es cuantitativa (numérica).
3. Ambas variables son cuantitativas (ambas numéricas).

En esta sección presentamos métodos tabulares y gráficos para exhibir cada una de estas combinaciones de datos bivariados.

Dos variables cualitativas

Cuando resultan datos bivariados de dos variables cualitativas (de atributo o categóricas), es frecuente que los datos se ordenen en una **tabulación cruzada** o **tabla de contingencia.** Veamos un ejemplo.

EJEMPLO 3.1 Construcción de tablas de tabulación cruzada

PARA SU INFORMACIÓN
$m = n$ (filas)
$n = n$ (columnas) para una tabla de contingencia $m \times n$.

Treinta estudiantes de nuestra universidad se identificaron y clasificaron al azar según dos variables: género (M/F) y especialidad (artes liberales, administración de empresas, tecnología), como se muestra en la tabla 3.1. Estos 30 datos bivariados pueden resumirse en una tabla de tabulación cruzada de 2×3, donde las dos filas representan los dos géneros, masculino y femenino, y las tres columnas representan las tres categorías de especialidad de artes liberales (LA), administración de empresas (BA), y tecnología (T). La entrada en cada celda se encuentra al determinar cuántos estudiantes caben en cada categoría. Adams es masculino (M) y artes liberales (LA) y se clasifica en la celda de la primera fila, primera columna. Vea la marca total roja de la tabla 3.2. Los otros 29 estudiantes están clasificados (totalizados, en negro) en forma semejante.

La tabla resultante de tabulación cruzada de 2×3 (de contingencia), tabla 3.3, muestra la frecuencia para cada categoría cruzada de las dos variables junto con los totales de fila y columna, llamados *totales marginales* (o *marginales)*. El total de los totales marginales es el *gran total* y es igual a n, el *tamaño muestral.*

TABLA 3.1

Géneros y especialidad de 30 estudiantes universitarios

Nombre	Género	Especialidad	Nombre	Género	Especialidad	Nombre	Género	Especialidad
Adams	M	LA	Feeney	M	T	McGowan	M	BA
Argento	F	BA	Flanigan	M	LA	Mowers	F	BA
Baker	M	LA	Hodge	F	LA	Ornt	M	T
Bennett	F	LA	Holmes	M	T	Palmer	F	LA
Brand	M	T	Jopson	F	T	Pullen	M	T
Brock	M	BA	Kee	M	BA	Rattan	M	BA
Chun	F	LA	Kleeberg	M	LA	Sherman	F	LA
Crain	M	T	Light	M	BA	Small	F	T
Cross	F	BA	Linton	F	LA	Tate	M	BA
Ellis	F	BA	Lopez	M	T	Yamamoto	M	LA

TABLA 3.2

Tabulación cruzada de género y especialidad (total)

	Especialidad		
Género	LA	BA	T
M	||||| (5)	|||||	(6)	||||| || (7)
F	|||||	(6)	|||| (4)	|| (2)

TABLA 3.3

Tabulación cruzada de género y especialidad (total)

	Especialidad			
Género	LA	BA	T	Total de fila
M	5	6	7	18
F	6	4	2	12
Total de columna	11	10	9	30

Es frecuente que las tablas de contingencia muestren porcentajes (frecuencias relativas). Estos porcentajes pueden basarse en toda la muestra o en las clasificaciones de la submuestra (fila o columna).

Porcentajes basados en el gran total (toda la muestra)

Las frecuencias de la tabla de contingencia mostrada en la tabla 3.3 pueden convertirse fácilmente en porcentajes del gran total, si se divide la frecuencia entre el gran total y el resultado se multiplica por 100. Por ejemplo, 6 se convierte en 20%

$\left[\left(\frac{6}{30}\right) \times 100 = 20\right]$. Ver tabla 3.4.

De la tabla de porcentajes del gran total, podemos ver fácilmente que 60% de la muestra eran hombres, 40% mujeres, 30% estudiaban una especialización en tecnología, y así sucesivamente. Estas mismas estadísticas (valores numéricos que describen resultados muestrales) se pueden mostrar en una gráfica de barras (vea la figura 3.1).

TABLA 3.4

Tabulación cruzada de género y especialidad (frecuencias relativas; % del gran total)

	Especialidad			
Género	LA	BA	T	Total de fila
M	17%	20%	23%	60%
F	20%	13%	7%	40%
Total de columna	37%	33%	30%	100%

FIGURA 3.1 Gráfica de barras

La tabla 3.4 y la figura 3.1 muestran la distribución de estudiantes hombres de artes liberales, estudiantes mujeres de artes liberales, estudiantes hombres de administración de empresas, etcétera, respecto a toda la muestra.

Porcentajes basados en totales de fila

Las frecuencias de la misma tabla de contingencia, tabla 3.3, pueden expresarse como porcentajes de los totales de fila (o género) al dividir cada entrada de fila entre el total de la fila y multiplicar por 100 los resultados. La tabla 3.5 está basada en totales de fila.

De la tabla 3.5 vemos que 28% de los estudiantes hombres se especializaban en artes liberales, mientras que 50% de las estudiantes mujeres se especializaban en artes liberales. Estas mismas estadísticas se muestran en la gráfica de barras de la figura 3.2.

TABLA 3.5

Tabulación cruzada de género y especialidad (% de totales de fila)

	Especialidad			
Género	LA	BA	T	Total de fila
M	28%	33%	39%	100%
F	50%	33%	17%	100%
Total de columna	37%	33%	30%	100%

FIGURA 3.2 Gráfica de barras

La tabla 3.5 y la figura 3.2 muestran, por separado, la distribución de las tres especialidades para estudiantes hombres y mujeres.

Porcentajes basados en totales de columna

Las frecuencias de la misma tabla de contingencia, tabla 3.3, pueden expresarse como porcentajes de los totales de columna (o especialidad) al dividir cada entrada de columna entre el total de la columna y multiplicar por 100 los resultados. La tabla 3.6 está basada en totales de columna.

De la tabla 3.6 vemos que 45% de los estudiantes de la especialidad de artes liberales eran hombres, mientras que 55% de los estudiantes de artes liberales eran mujeres. Estas mismas estadísticas se muestran en la gráfica de barras de la figura 3.3.

TABLA 3.6

Tabulación cruzada de género y especialidad (% de totales de columna)

	Especialidad			
Género	LA	BA	T	Total de fila
M	45%	60%	78%	60%
F	55%	40%	22%	40%
Total de columna	100%	100%	100%	100%

FIGURA 3.3 Gráfica de barras

La tabla 3.6 y la figura 3.3 muestran, por separado, la distribución de estudiantes hombres y mujeres para cada especialidad.

INSTRUCCIONES DE TECNOLOGÍA: TABLA DE TABULACIÓN CRUZADA

MINITAB (Versión 14) Introduzca en C1 los valores categóricos sencillos de variable de fila, y en C2 los correspondientes valores categóricos sencillos de variable de columna; luego continúe con:

```
Choose:     Stat > Tables > Cross Tabulation and Chi-Square
Enter:      Categorical variables: For rows: C1 For columns: C2
Select:     Counts
            Row Percents
            Column Percents
            Total Percents > OK
```

Sugerencia: los cuatro subcomandos de que se dispone para 'Display' (Exhibir) se pueden usar juntos; no obstante, la tabla resultante será mucho más fácil de leer si se usa un subcomando a la vez.

Excel Usando encabezados o títulos de columna, introduzca en la columna A los valores categóricos de variable de fila y, en la columna B, los correspondientes valores categóricos de variable de columna; luego continúe con:

```
Choose:     Data > Pivot Table and PivotChart Report . . .
Select:     Microsoft Excel list or database > Next
Enter:      Range: (A1:B5 or select cells) > Next
Select:     Existing Worksheet
Enter:      (C1 or select cell) > Finish
Drag:       Headings to row or column (depends on preference)
            One heading into data area*
```

*Para otras sumatorias, haga doble clic de "Count of" en la caja de área de datos; luego continúe con:
Seleccione: **Resume por: Cuenta > Opciones**
Muestra datos como: % de fila o % de columna o % de total > OK

TI-83/84 Plus Los datos categóricos deben ser codificados numéricamente primero; use 1, 2, 3, para las diversas variables de columna y 1, 2, 3, para las diversas variables de columna. Introduzca en L1 los valores numéricos de variable de fila y, en L2, los correspondientes valores numéricos de variable de columna; luego continúe con:

```
Choose:     PRGM > EXEC > CROSSTAB*
Enter:      ROWS: L1 >ENTER
            COLS: L2 >ENTER
```

La tabla de tabulación cruzada que muestra frecuencias se guarda en la matriz [A], la tabla de tabulación cruzada que muestra porcentajes de fila está en la matriz [B], los porcentajes de columna en la matriz [C], y los porcentajes basados en el gran total en la matriz [D]. Todas las matrices contienen totales marginales. Para ver las matrices, continúe con:

```
Choose:     MATRX > NAMES
Enter:      1:[A] or 2:[B] or 3:[C] or 4:[D] > ENTER
```

*El programa 'CROSSTAB' es uno de numerosos programas que se pueden descargar del sitio web Duxbury. En la página 42 vea instrucciones específicas.

Una variable cualitativa y una cuantitativa

Cuando resultan datos bivariados de una variable cualitativa y una cuantitativa, los valores cuantitativos se ven como muestras separadas, con cada conjunto identificado por niveles de la variable cualitativa. Cada muestra se describe usando las técnicas del capítulo 2, y los resultados se exhiben juntos para fácil comparación.

EJEMPLO 3.2 Construcción de comparaciones juntas

Se midió la distancia necesaria para detener un automóvil de 3000 libras de peso en pavimento mojado, para comparar la capacidad de frenado de tres diseños de la superficie de rodadura de neumáticos (vea tabla 3.7). Los neumáticos de cada uno de los diseños fueron probados repetidas veces en el mismo automóvil en un pavimento mojado controlado.

TABLA 3.7

Distancias de frenado (en pies) de tres diseños de superficie de rodadura

Diseño A ($n = 6$)			Diseño B ($n = 6$)			Diseño C ($n = 6$)		
37	36	38	33	35	38	40	39	40
34	40	32	34	42	34	41	41	43

El diseño de la superficie de rodadura es una variable cualitativa con tres niveles de respuesta, y la distancia de frenado es una variable cuantitativa. La distribución de las distancias de frenado para el diseño A de superficie de rodadura ha de compararse con la distribución de distancias de frenado para cada uno de los otros diseños de superficie de rodadura. Esta comparación se puede hacer con técnicas numéricas y gráficas. Algunas de las opciones existentes se muestran en la figura 3.4, tabla 3.8 y la tabla 3.9.

FIGURA 3.4 Gráfica de puntos y diagrama de caja y bigotes usando una escala común

TABLA 3.8

Resumen de 5 números para cada diseño

	Diseño A	Diseño B	Diseño C
Alto	40	42	43
Q_3	38	38	41
Mediano	36.5	34.5	40.5
Q_1	34	34	40
bajo	32	33	39

TABLA 3.9

Media y desviación estándar para cada diseño

	Diseño A	Diseño B	Diseño C
Media	36.2	36.0	40.7
Desviación estándar	2.9	3.4	1.4

INSTRUCCIONES DE TECNOLOGÍA: GRÁFICAS DE CAJA Y GRÁFICAS DE PUNTOS JUNTAS

MINITAB (Versión 14) Teclee los valores numéricos en C1 y las correspondientes categorías en C2; luego continúe con:

```
Choose:     Graph > Boxplot... > One Y, With Groups > OK
Enter:      Graph variables: C1 Categorical variables: C2 > OK
```

Los comandos de MINITAB para construir gráficas de puntos juntas para datos en esta forma aparecen en la página 49.

Si los datos para las diversas categorías están en columnas separadas, use los comandos de MINITAB para gráficas de caja múltiples de la página 98. Si se hacen necesarias gráficas de puntos juntas para los datos de esta forma, continúe con:

```
Choose:     Graph > Dotplots
Select:     Multiple Y's, Simple > OK
Enter:      Graph variables: C1 C2 > OK
```

Excel Los comandos de Excel para construir una gráfica de caja sencilla aparecen en la página 98.

TI-83/84 Plus Los comandos de la TI-83/84 para construir gráficas de caja múltiples aparecen en la página 99.

Los comandos de la TI-83/84 para construir gráficas de puntos múltiples aparecen en la página 49.

Gran parte de la información aquí presentada también se puede demostrar usando muchas otras técnicas estadísticas, por ejemplo diagramas de tallo y hoja o histogramas.

Restringiremos nuestra discusión de este capítulo a técnicas descriptivas para la forma más básica de análisis de correlación y regresión, que es el caso lineal bivariado.

Dos variables cuantitativas

Cuando los datos bivariados son el resultado de dos variables cuantitativas, se acostumbra expresar matemáticamente los datos como **pares ordenados** (x, y), donde x es la **variable de entrada** (a veces llamada **variable independiente**) y y es la **variable de salida** (a veces llamada **variable dependiente**). Se dice que los datos

están *ordenados* porque un valor, x, siempre se escribe primero. Se llaman *pareados* porque para cada valor de x siempre hay un valor correspondiente de y de la misma fuente. Por ejemplo, si x es la estatura y y es el peso, entonces una estatura y un peso correspondiente se registran para cada persona. La variable de entrada x se mide o controla para pronosticar la variable de salida y. Supongamos que unos médicos investigadores están probando un nuevo medicamento al describir diferentes dosis y observar los tiempos de recuperación de sus pacientes. El investigador puede controlar la cantidad de medicamento prescrita, de modo que la cantidad de medicamento se designa como x. En el caso de estatura y peso, cualquiera de las variables podría tratarse como entrada y la otra como salida, dependiendo de la pregunta que se formule. No obstante, se obtienen diferentes resultados del análisis de regresión, dependiendo de la selección que se haga.

En problemas que se refieren a dos variables cuantitativas, presentamos gráficamente los datos muestrales en un *diagrama de dispersión.*

Diagrama de dispersión: es una gráfica de todos los pares ordenados de datos bivariados en un sistema de ejes de coordenadas. La variable de entrada, x, se localiza en el eje horizontal, y la variable de salida, y, se localiza en el eje vertical.

Nota: cuando construya un diagrama de dispersión, es conveniente construir escalas para que el rango de los valores y, en todo el eje vertical, sea igual o ligeramente más corto que el rango de los valores x en todo el eje horizontal. Esto crea una "ventana de datos" que es aproximadamente cuadrada.

EJEMPLO 3.3 Construcción de un diagrama de dispersión

En el curso de educación física de Mr. Chamberlain se tomaron varias notas. La siguiente muestra es el número de "lagartijas" y "sentadillas" hechas por 10 estudiantes seleccionados al azar:

(27, 30) (22, 26) (15, 25) (35, 42) (30, 38)
(52, 40) (35, 32) (55, 54) (40, 50) (40, 43)

La tabla 3.10 muestra estos datos muestrales y la figura 3.5 muestra un diagrama de dispersión de los datos.

TABLA 3.10

Datos para "lagartijas" y "sentadillas"

Estudiante	1	2	3	4	5	6	7	8	9	10
"Lagartijas", x	27	22	15	35	30	52	35	55	40	40
"Sentadillas, y	30	26	25	42	38	40	32	54	50	43

El diagrama de dispersión del curso de educación física de Mr. Chamberlain muestra un patrón definido. Observe que cuando aumentó el número de "lagartijas" también aumentó el de "sentadillas".

FIGURA 3.5
Diagrama de dispersión

CASO PRÁCTICO 3.4 Escuelas del noroeste de Ohio y cómo se clasifican

Durante mucho tiempo se ha sabido que la capacidad de un estudiante, para aprobar exámenes de suficiencia de cuarto grado del estado, está estrechamente relacionada con el nivel de ingreso de la familia del estudiante.

La gráfica siguiente muestra la forma en que operaron escuelas elementales individuales en los exámenes de suficiencia de marzo de 2000, en cuarto grado de matemáticas y lectura, y si las escuelas trabajaron mejor o peor de lo que podría pronosticarse con base en el nivel de pobreza de los estudiantes de esa escuela.

El porcentaje de niños que recibieron un refrigerio gratis o a precios bajos se empleó como medida de pobreza.

Puntuaciones de pobreza pronosticadas

Cada una de las 2,025 escuelas elementales de Ohio analizadas por The Blade está representada en esta gráfica como un punto. Los puntos se localizaron en la gráfica con base en el nivel de pobreza de cada escuela, comparada con su porcentaje total de aprobaciones en el examen de suficiencia de lectura de cuarto grado en el estado.

Fuente: Reimpreso con permiso de *The (Toledo) Blade,* 5 de agosto, 2001

Con base en el párrafo precedente y en la gráfica, las dos variables que se examinan en este ejemplo son "nivel de pobreza en la escuela" y "porcentaje de aprobación". El diagrama de dispersión muestra claramente que existe un patrón. Observe que los puntos están todos agrupados y con pendiente hacia abajo. Por tanto, cuando aumentó el nivel de pobreza (lea el eje *x* de izquierda a derecha), disminuyó el porcentaje de aprobados.

INSTRUCCIONES DE TECNOLOGÍA: DIAGRAMA DE DISPERSIÓN

MINITAB (Versión 14)

Introduzca en C1 valores de la variable *x* y en C2 los correspondientes valores de la variable *y*; luego continúe con:

```
Choose:     Graph > ScatterPlot. . . > Simple > OK
Enter:      Y variables: C2  X variables: C1
Select:     Labels > Titles/Footnotes
Enter:      Title: your title > OK > OK
```

Excel

Introduzca valores de la variable *x* en la columna A y en la columna B los valores correspondientes de la variable *y*; luego continúe con:

```
Choose:     Chart Wizard  > XY(Scatter)  >  1st picture (usually) > Next
Enter:      Data Range:  (A1:B12 or select cells(if necessary))  >  Next
Choose:     Titles
Enter:      Chart title: your title; Value(x) axis: title for x axis;
            Value(y) axis: title for y axis*  >
            Finish
```

*Para eliminar líneas de cuadrícula:

```
Choose:     Gridlines
Unselect:   Value(Y) axis: Major Gridlines  > Finish
```

Para editar el diagrama de dispersión, siga los comandos básicos de edición que se muestran para un histograma en la página 62.
Para cambiar la escala, haga doble clic en el eje; luego continúe con:

```
Choose:     Scale
Unselect:   any Auto boxes
Enter:      new values > OK
```

TI-83/84 Plus

Introduzca valores de la variable *x* en L1 y los valores correspondientes de la variable *y* en L2; luego continúe con:

```
Choose:     2nd  > STATPLOT > 1:Plot1
Choose:     ZOOM > 9:ZoomStat
            > TRACE  > > >
            or
            WINDOW
            Enter: at most lowest x value,
            at least highest x value,
            x-scale, — y-scale, at least
            highest y value, y-scale,1
            TRACE > > >
```

SECCIÓN 3.2 EJERCICIOS

3.3 En una encuesta nacional de 500 viajeros de negocios y 500 en viaje de placer, a cada una se les preguntó dónde les gustaría "más espacio."

	En un avión	Cuarto de hotel	En todo lugar
Negocios	355	95	50
De placer	250	165	85

a. Exprese la tabla como porcentajes del total.

b. Exprese la tabla como porcentajes de los totales de fila. ¿Por qué prefiere uno que la tabla se exprese en esta forma?

c. Exprese la tabla como porcentajes de los totales de columna. ¿Por qué prefiere uno que la tabla se exprese en esta forma?

3.4 La gráfica "Perspectiva para viajeros de negocios" muestra dos gráficas de pastel, cada una con cuatro secciones. Esta misma información podría estar representada en la forma de una tabla de contingencia de 2×4 de dos variables cualitativas.

PERSPECTIVA PARA VIAJEROS DE NEGOCIOS

Datos de Darryl Haralson y Marcy E. Mullins, USA TODAY; *Fuente:* Encuesta de Carlson Wagonlit Travel de 1500 viajeros de negocios y gerentes de viajes. Margen de error ± puntos porcentuales.

a. Identifique la población y mencione las dos variables.

b. Construya la tabla de contingencia usando entradas de porcentajes con base en totales de fila.

3.5 La gráfica "La edad perfecta" muestra los resultados desde una tabla de contingencia de 9×2 para una variable cualitativa y una cuantitativa.

"LA EDAD PERFECTA"

Datos de Cindy Hall y Genevieve Lynn, USA TODAY. *Fuente:* IRC Research para Walt Disney. © 1998 USA TODAY, reimpreso con permiso.

a. Identifique la población y mencione las variables cualitativas y cuantitativas.

b. Construya una gráfica de barras que muestre las dos distribuciones lado a lado.

c. ¿Parece haber una gran diferencia entre los géneros sobre este tema?

3.6 Según la ley de designación del Sistema Nacional de Autopistas de 1995, a los estados se les permite establecer sus propios límites de velocidad en autopistas. Casi todos los estados elevaron estos límites. Los límites máximos de velocidad en 2005 en cada estado, en autopistas interestatales para autos y camiones, aparecen en la tabla siguiente (en millas por hora).

Estado	Autos	Camiones	Estado	Autos	Camiones	Estado	Autos	Camiones
AL	70	70	LA	70	70	OH	65	55
AK	65	65	ME	65	65	OK	75	75
AZ	75	75	MD	65	65	OR	65	65
AR	70	65	MA	65	65	PA	65	65
CA	70	55	MI	70	55	RI	65	65
CO	75	75	MN	70	70	SC	70	70
CT	65	65	MS	70	70	SD	75	75
DE	65	65	MO	70	70	TN	70	70
FL	70	70	MT	75	65	TX	75	75
GA	70	70	NE	75	75	UT	75	75
HI	55	55	NV	75	75	VT	65	65
ID	75	70	NH	65	65	VA	65	65
IL	65	55	NJ	65	65	WA	70	60
IN	65	60	NM	75	75	WV	70	70
IA	65	65	NY	65	65	WI	65	65
KS	70	70	NC	70	70	WY	75	75
KY	65	65	ND	75	75			

Fuente: The National Motorists Association, http://www.motorists.com/issues/speed/StateSpeeds.html

a. Construya una tabla de tabulación cruzada de las dos variables, tipo de vehículo y límite máximo de velocidad en una autopista interestatal. Exprese los resultados en frecuencias, mostrando totales marginales.

b. Exprese la tabla de contingencia que obtuvo usted en la parte a en porcentajes basados en el gran total.

c. Trace una gráfica de barras que muestre los resultados de la parte b.

d. Exprese la tabla de contingencia que obtuvo usted en la parte a en porcentajes basados en el total marginal para límite de velocidad.

e. Trace una gráfica de barras que muestre los resultados de la parte b.

PARA SU INFORMACIÓN Si usa computadora o calculadora, intente los comandos de la tabla de tabulación cruzada que aparecen en la página 150.

3.7 Se realizó una encuesta de nivel estatal para investigar la relación entre preferencias de televidentes de ABC, CBS, NBC, PBS o FOX de informativos y sus afiliaciones a un partido político. Los resultados se muestran en forma tabular:

Estación de televisión

Afiliación política	ABC	CBS	NBC	PBS	FOX
Demócrata	203	218	257	156	226
Republicano	421	350	428	197	174
Otro	156	312	105	57	90

a. ¿Cuántos televidentes fueron interrogados?

b. ¿Por qué son datos bivariados? Mencione las dos variables. ¿Qué tipo de variable es cada una?

c. ¿Cuántos televidentes prefieren ver CBS?

d. ¿Qué porcentaje de la encuesta era de republicanos?

e. ¿Qué porcentaje de los demócratas preferían ABC?

f. ¿Qué porcentaje de televidentes era de republicanos y preferían PBS?

3.8 Considere la tabla de contingencia siguiente, que presenta los resultados de una encuesta de publicidad acerca del uso de crédito por clientes de Martan Oil Company.

	Número de compras en gasolinera el año pasado					
Método preferido de pago	0-4	5-9	10-14	15-19	≥20	Suma
De contado	150	100	25	0	0	275
Tarjeta de Cía. Petrolera	50	35	115	80	70	350
Tarjeta de crédito nacional o bancaria	50	60	65	45	5	225
Suma	250	195	205	125	75	850

a. ¿Cuántos clientes fueron interrogados?

b. ¿Por qué son datos bivariados? ¿Qué tipo de variable es cada una?

c. ¿Cuántos clientes preferían usar tarjeta de crédito de una compañía petrolera?

d. ¿Cuántos clientes hicieron 20 o más compras el año pasado?

e. ¿Cuántos clientes preferían usar tarjeta de crédito de una compañía petrolera e hicieron entre cinco y nuevo compras el año pasado?

f. ¿Qué significa el 80 en la cuarta celda de la segunda fila?

3.9 La tasa de desempleo en enero de 2005 para estados del este y el oeste de Estados Unidos fue como sigue:

Este	4.7	4.1	4.8	3.5	3.9	5.0	5.1	4.4
Oeste	4.1	5.8	4.9	4.3	4.3	6.4	5.5	3.4

Fuente: U.S. Bureau of Labor Statistics

Muestre estas tasas como dos gráficas de puntos que usen la misma escala; compare medias y medianas.

3.10 ¿Qué efecto tiene la cantidad mínima en la tasa de interés que se ofrece en certificados de depósito (CD) a tres meses? A continuación aparecen tasas de rendimiento anunciadas, *y*, para un depósito mínimo de \$500, \$1000, \$2000, \$2500, \$5000 o \$10 000, *x*. (Nótese que *x* está en \$100 y que *y* es el porcentaje de rendimiento anual.)

Depósito mínimo	Tasa	Depósito mínimo	Tasa	Depósito mínimo	Tasa
10	2.81	10	2.52	20	2.38
10	2.70	25	2.49	100	2.37
50	2.68	50	2.49	25	2.35
10	2.71	5	2.48	10	2.32
50	2.66	5	2.42	10	2.30
20	2.65	10	2.37	5	2.27
5	2.62	10	2.38	100	2.27
25	2.60	20	2.38		

Fuente: Bankrate.com, 10 de marzo, 2005

a. Elabore una gráfica de puntos de las tasas asociadas con cada uno de los seis diferentes requisitos de depósito mínimo, usando una escala común.

b. Elabore un resumen de 5 números y una gráfica de caja de los seis conjuntos de datos. Use la misma escala que para gráficas de caja.

c. Describa cualesquiera diferencia que vea entre los seis conjuntos de datos.

PARA SU INFORMACIÓN Si usa computadora o calculadora para el ejercicio 3.10, intente los comandos de la página 152.

3.11 ¿Puede pronosticarse la estatura de una mujer a partir de la estatura de su madre? A continuación aparecen estaturas de algunas parejas madre-hija; *x* es la estatura de la madre y *y* es la de la hija.

x	63	63	67	65	61	63	61	64	62	63	
y	63	65	65	65	64	64	63	62	63	64	
x	64	63	64	64	63	67	61	65	64	65	66
y	64	64	65	65	62	66	62	63	66	66	65

a. Trace dos gráficas de puntos usando la misma escala y mostrando los dos conjuntos de datos lado a lado.

b. ¿Qué se puede concluir al ver los dos conjuntos de datos como conjuntos separados en la parte a? Explique.

c. Trace un diagrama de dispersión de estos datos como pares ordenados.

d. ¿Qué se puede concluir al ver los datos presentados como pares ordenados? Explique.

3.12 La siguiente tabla contiene las estaturas (en metros), pesos (en kilogramos) y edades de los jugadores de los dos equipos que jugaron en la final de la Copa Mundial de 2002: Brasil y Alemania.

	Brasil			Alemania		
Jugador	Estatura	Peso	Edad	Estatura	Peso	Edad
1	1.93	86	28	1.88	88	33
2	1.95	85	29	1.90	87	33

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/johnson

Fuente: http://worldcup.espnsoccernet.com/index

a. Compare cada una de las tres variables —estatura, peso y edad— usando ya sea una gráfica de puntos o un histograma (use la misma escala).

b. Con base en lo que se ve en las gráficas de la parte a, ¿puede detectar una diferencia importante entre los dos equipos respecto a estas tres variables? Explique.

c. Explique por qué los datos, como se usan en la parte a, no son bivariados.

3.13 Considere las dos variables de la estatura y peso de una persona. ¿Cuál variable, estatura o peso, usaría usted como variable de entrada cuando estudie la relación entre ellas? Explique por qué.

3.14 Trace un eje de coordenadas y grafique los puntos (0,6), (3,5), (3,2), (5,0) para formar un diagrama de dispersión. Describa el patrón que muestran los datos en este diagrama.

3.15 ¿Da resultados estudiar para un examen?

a. Trace un diagrama de dispersión del número de horas estudiadas, *x*, comparado con la calificación de examen recibida, *y*.

x	2	5	1	4	2
y	80	80	70	90	60

b. Explique lo que pueda concluir con base en el patrón de datos que se muestran en el diagrama de

dispersión trazado en la parte a. (Retenga estas soluciones para usarlas en el ejercicio 3.58, p. 185.)

3.16 Consulte el caso práctico 3.4, p. 154, "Escuelas del noroeste de Ohio y cómo se clasifican" para contestar las siguientes preguntas:

a. ¿Cuáles son las dos variables empleadas?

b. El diagrama de dispersión, ¿sugiere una relación entre las dos variables? Explique.

c. ¿Qué conclusión, si la hay, se saca del aspecto del diagrama de dispersión?

3.17 Por lo general, los pediatras usan gráficas de crecimiento para observar el crecimiento de un niño. Considere la gráfica de crecimiento que sigue:

a. ¿Cuáles son las dos variables mostradas en la gráfica?

b. ¿Qué información representa el par ordenado (3, 87)?

c. Describa la forma en que el pediatra podría usar esta gráfica y qué tipos de conclusiones podrían basarse en la información mostrada por la gráfica.

3.18 a. Trace un diagrama de dispersión que muestre estatura, x, y peso, y, para el equipo de futbol de Brasil de la Copa Mundial usando los datos del ejercicio 3.12.

b. Trace un diagrama de dispersión que muestre estatura, x, y, peso, y para el equipo de futbol de Alemania de la Copa Mundial usando los datos del ejercicio 3.12.

c. Explique por qué los datos, como se usan en las partes a y b, son datos bivariados.

PARA SU INFORMACIÓN Si usa computadora o calculadora para el ejercicio 3.10, intente los comandos de la página 155.

3.19 Los datos siguientes muestran el número de horas, x, estudiado para un examen y la calificación recibida, y (y se mide en decenas, es decir, $y = 8$ significa que la calificación, redondeada a los 10 puntos más cercanos, es 80). Trace el diagrama de dispersión. (Retenga esta solución para usarla en el ejercicio 3.38, p. 170.)

x	2	3	3	4	4	5	5	6	6	6	7	7	7	8	8
y	5	5	7	5	7	7	8	6	9	8	7	9	10	8	9

3.20 Un psicólogo experimental dice que cuanto más edad tenga un niño, dará respuestas menos irrelevantes durante un experimento controlado. Para investigar esta aseveración, se recolectaron los siguientes datos. Trace un diagrama de dispersión. (Retenga esta solución para usarla en el ejercicio 3.39, p. 170.)

Edad, x	2	4	5	6	6	7	9	9	10	12
Respuestas irrelevantes, y	12	13	9	7	12	8	6	9	7	5

3.21 La siguiente tabla indica los porcentajes de estudiantes que reciben un refrigerio gratis o a precio bajo, comparados con los porcentajes que aprobaron la parte de lectura de un examen estatal. Los resultados son para el Condado Sandusky, Ohio, y se informaron en *The Blade*, periódico de Toledo, el 5 de agosto de 2001. El Condado de Sandusky tiene una combinación de 13 escuelas rurales y urbanas.

Escuela	% refrigerio gratis/ (precio bajo)	% aprueban lectura	Escuela	% refrigerio gratis/ (precio bajo)	% aprueban lectura
1	29	66	8	47	58
2	29	59	9	29	88
3	23	62	10	17	68
4	60	53	11	22	60
5	57	53	12	38	47
6	50	57	13	15	62
7	49	54			

Construya un diagrama de dispersión de estos datos. (Retenga esta solución para usar en el ejercicio 3.34,

3.22 Se seleccionó una muestra de 15 estudiantes de clase superior que viven en suburbios y que, cuando se inscribieron, se les pidió calcularan la distancia, x (la milla más cercana), y el tiempo y (los 5 minutos más cercanos), que necesitan para ir a diario a clase (ver tabla siguiente). Construya un diagrama de dispersión que describa estos datos.

Distancia, x	Tiempo, y	Distancia, x	Tiempo, y
18	20	2	5
8	15	15	25
20	25	16	30
5	20	9	20
5	15	21	30
11	25	5	10
9	20	15	20
10	25		

3.23 Los estadios de béisbol varían en edad, estilo, tamaño y en muchas otras formas. Los aficionados podrían considerar el tamaño del estadio en términos del número de asientos, mientras que los jugadores podrían medirlo por la distancia (en pies) de la placa del *home* a la barda del jardín central.

Asientos	CF	Asientos	CF	Asientos	CF
40 000	422	49 166	400	43 000	400
45 050	400	45 200	410	42 000	435
33 871	420	44 321	400	56 500	410
43 368	405	57 545	408	40 800	404
40 625	400	48 500	402	38 127	399
48 678	408	49 625	402	42 531	410
43 662	400	43 500	401	56 133	405
48 876	400	50 381	415	55 777	410
47 000	405	42 059	404	50 062	401
50 516	400	56 000	395	38 902	400

CF = distancia de la placa del home a la barda del jardín central

Fuente: http://mlb.mlb.com

a. ¿Hay una relación entre estas dos dimensiones para el "tamaño" de los 30 estadios de béisbol de ligas mayores?

b. ¿Qué cree usted que encontrará? ¿Los campos más grandes tienen más asientos? ¿Los campos más pequeños tienen más asientos? ¿No hay relación entre el tamaño del campo y el número de asientos? ¿Hay una fuerte relación entre el tamaño del campo y el número de asientos? Explique.

c. Construya un diagrama de dispersión.

d. Describa lo que dice el diagrama de dispersión, incluyendo una reacción a la respuesta de la parte b.

3.24 Casi todos los estadounidenses adultos viajan en auto. Pero, ¿tiene usted idea de cuántos tienen licencia en cada estado de Estados Unidos? La siguiente tabla indica el número de conductores, hombres y mujeres, que tienen licencia en cada uno de los 15 estados seleccionados al azar.

Conductores con licencia, por estado × (100 000)

Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
2.77	2.78	59.5	54.07	9.92	9.96
37.1	39.46	1.94	1.85	30.13	30.02
19.5	20.16	7.76	7.12	9.95	10.03
13.19	13.41	15.45	15.76	20.56	21.49
4.41	3.94	6.41	6.31	4.87	4.81

Fuente: Federal Highway Administration, U.S. Department of Transportation

a. ¿Espera usted hallar una relación lineal (línea recta) entre el número de conductores y el de conductoras con licencia, por estado? ¿Qué tan convincente piensa usted que es esta relación? Describa.

b. Construya un diagrama de dispersión usando x para el número de conductores, así como y para el de conductoras.

c. Compare el diagrama de dispersión contra sus expectativas de la parte a. ¿Cómo lo hizo? Explique.

d. ¿Existen puntos de datos que parecen separados del patrón creado por el resto de los pares ordenados? Si están alejados del conjunto de datos, ¿cambiarían los resultados? ¿Qué hizo que estos puntos estuvieran separados de los otros pero aún son parte del patrón extendido? Explique

e. (Opcional) La muestra, ¿proporcionó suficiente información para entender la relación entre las dos variables de esta situación? Explique.

3.25 ¿Las personas son ahora más fuertes que antes? ¿Puede usted correr más rápido? Comparemos los rendimientos de ganadores de medalla de oro en Juegos Olímpicos en el siglo pasado, como forma para decidir. Las distancias (en pulgadas) para ganadores de medalla de oro en salto de longitud, salto de altura y lanzamiento de disco se dan en la tabla siguiente. El evento del año está codificado, con 1900 = 0.

Año	Salto de longitud	Salto de altura	Lanzamiento de disco
−4	249.75	71.25	1147.5
0	282.875	74.8	1418.9

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.ex.ac.uk/cimt/data/olympics/olymindx.htm

a. Localice los datos para cada evento en un diagrama de dispersión por separado usando el año, x.

b. Describa la forma de la distribución. Para cada diagrama de dispersión, ¿la relación entre año y rendimiento parecen seguir una línea recta?

c. ¿Cómo contestan los tres diagramas de dispersión la pregunta: ¿son las personas más fuertes hoy? Explique.

d. En cada uno de los tres diagramas de dispersión, trace una recta que parezca seguir mejor el patrón de puntos de 1896 a 2004. Use esta recta como ayuda para pronosticar el rendimiento de ganadores de medallas de oro olímpico para cada evento en los juegos de Beijing en 2008.

e. Investigue la relación entre salto de altura y salto de longitud con ayuda de un diagrama de dispersión. Describa lo que encuentre.

3.26 La siguiente tabla indica la estatura (en pulgadas), el peso (en libras), y la fecha de nacimiento de los miembros del equipo de futbol profesional de los Rinos Furiosos de Rochester 2004.

Jugador	Estatura	Peso	DOB
1	68	160	12/7/1978
2	71	170	2/2/1970

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.RhinosSoccer.com

a. ¿Es cierto que los jugadores más altos pesan más? ¿Qué espera usted que indicará un diagrama de dispersión para un equipo de futbol?

b. Construya un diagrama de dispersión de la estatura, x, contra el peso, y.

c. El diagrama de dispersión de la parte b, ¿apoya su idea de la parte a? Explique por qué sí o por qué no.

d. ¿Es cierto que a medida que envejecen los jugadores, tienden a pesar más? ¿Qué espera usted que indique un diagrama de dispersión para un equipo de futbol?

e. Construya un diagrama de dispersión de la edad, x, contra el peso, y. (*Nota:* necesitará convertir fecha de nacimiento a edad en años. Las soluciones dadas están basadas en edades al 1 de enero, 2006.)

f. El diagrama de dispersión de la parte e, ¿apoya su idea de la parte d? Explique por qué sí o por qué no.

g. Si fuera usted a buscar edades de los jugadores al 1 de enero de 2008, ¿qué efecto tendría esto en el diagrama de dispersión construido en la parte e? Explique.

h. (Opcional) Construya el diagrama de dispersión de edad, x, contra peso, y, con base en edades al 1 de enero de 2008.

3.27 Ronald Fisher, estadista inglés (1890-1962), recolectó medidas para una muestra de 150 lirios. De interés eran cinco variables: especie, ancho de pétalos (PW), longitud de pétalo (PL), ancho de sépalo (SW), y longitud de sépalo (SL) (todo en mm). Los sépalos son las hojas más exteriores que cubren la flor antes que abra. La meta del experimento de Fisher era producir una función sencilla que pudiera usarse para clasificar flores correctamente. En la tabla siguiente se da una muestra aleatoria de su conjunto completo de datos.

Tipo	PW	PL	SW	SL	Tipo	PW	PL	SW	SL
0	2	15	35	52	1	24	51	28	58
2	18	48	32	59	1	19	50	25	63
1	19	51	27	58	0	1	15	31	49
0	3	13	35	50	1	23	59	32	68
0	3	15	38	51	2	13	44	23	63
2	12	44	26	55	2	15	42	30	59
1	20	64	38	79	1	25	57	33	67
2	15	49	31	69	1	21	57	33	67
2	15	45	29	60	0	2	15	37	54
2	12	39	27	58	1	18	49	27	63
1	22	56	28	64	1	17	45	25	49
1	13	52	30	67	1	24	56	34	63
0	2	14	29	44	0	2	14	36	50
2	16	51	27	60	2	10	50	22	60
0	5	17	33	51	0	2	12	32	50

a. Construya un diagrama de dispersión de la longitud de pétalo, x, y ancho de pétalo, y. Use diferentes símbolos para representar las tres especies.*

b. Construya un diagrama de dispersión de la longitud de sépalo, x, y el ancho de sépalo, y. Use símbolos diferentes para representar las tres especies.

c. Explique lo que describen los diagramas de dispersión de las partes a y b.

Veamos qué tan bien es que una muestra aleatoria representa los datos de los cuales fue seleccionada.

d. Además del hecho de que los diagramas de dispersión de las partes a y b tienen menos datos, comente sobre las similitudes y diferencias entre las distribuciones mostradas para 150 datos y para los 30 datos seleccionados al azar.

*In addition to using the commands on page 155, use:

For MINITAB:	Data display:	For each: Select: Group
	Group variable:	Select: Type
For TI-83-84:	**Enter different groups into separate x, y columns. Use a separate Stat Plot and "Mark" for each group.**	

3.28 Los eclipses totales de sol en realidad tienen lugar casi con la misma frecuencia que los eclipses totales de luna, pero son visibles en una trayectoria mucho más angosta. El ancho de la trayectoria y la duración varían considerablemente de un eclipse al siguiente. La siguiente tabla muestra la duración (en segundos) y el ancho de trayectoria (en millas) de 44 eclipses solares totales medidos en el pasado y los proyectados hasta el año 2010:

Fecha	Duración(s)	Ancho (mi)
1950	73	83
1952	189	85

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: The World Almanac and Book of Facts 1998, p. 296

a. Trace un diagrama de dispersión que muestre duración, y, y ancho de trayectoria, x, para los eclipses solares totales.

b. ¿Cómo describiría usted este diagrama?

3.3 Correlación lineal

El objetivo principal del **análisis de correlación lineal** es medir la fuerza de una relación lineal entre dos variables. Examinemos algunos diagramas de dispersión que demuestran diferentes relaciones entre entrada, o variables independientes, x, y salida, o variables dependientes, y. Si cuando x aumenta no hay cambio definido en los valores de y, decimos que **no hay correlación**, o no hay relación entre x y y. Si cuando x aumenta hay un cambio en los valores de y, entonces hay una correlación. La correlación es **positiva** cuando y tiende a aumentar y **negativa** cuando y tiende a disminuir. Si los pares ordenados (x, y) tienden a seguir una trayectoria de línea recta, hay una correlación lineal. La precisión del cambio en y cuando x aumenta determina la fuerza de la **correlación lineal**. Los diagramas de dispersión de la figura 3.6 demuestran estas ideas.

Se presenta una correlación lineal perfecta cuando todos los puntos caen exactamente en toda una recta, como se ve en la figura 3.7. La correlación puede ser positiva o negativa, dependiendo de si y aumenta o disminuye cuando x aumenta. Si los datos forman una recta horizontal o vertical, no hay correlación porque una variable no tiene efecto en la otra, como se ilustra en la figura 3.7.

FIGURA 3.6 Diagramas de dispersión y correlación

FIGURA 3.7 Pares ordenados que forman una recta

FIGURA 3.8 No hay correlación lineal

Los diagramas de dispersión no siempre aparecen en una de las formas mostradas en las figuras 3.6 y 3.7. A veces sugieren relaciones diferentes a las lineales, como se observa en la figura 3.8. Parece haber un patrón definido; no obstante, las dos variables no están relacionadas linealmente y por tanto no hay correlación lineal.

El **coeficiente de correlación lineal**, r, es la medida numérica de la fuerza de la relación lineal entre dos variables. El coeficiente refleja la consistencia del efecto que un cambio en una variable tiene sobre la otra. El valor del coeficiente de correlación lineal nos ayuda a contestar la pregunta: ¿hay correlación lineal entre las dos variables bajo consideración? El coeficiente de correlación lineal, r, siempre tiene un valor entre -1 y $+1$. Un valor de $+1$ significa una correlación positiva perfecta, y un valor de -1 muestra una correlación negativa perfecta. Si cuando x aumenta hay un aumento general del valor de y, entonces r será positivo en valor. Por ejemplo, se esperaría un valor positivo de r para la edad y estatura de niños porque a medida que envejecen tienen más estatura. Del mismo modo, considere la antigüedad, x, y valor de reventa, y, de un automóvil. A medida que el auto envejece, su valor de reventa disminuye. Puesto que cuando x aumenta, y disminuye, la relación resulta en un valor negativo para r.

El valor de r está definido por la **fórmula de momento de producto de Pearson:**

Fórmula de definición

D

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y} \tag{3.1}$$

Notas:

1. s_x y s_y son las desviaciones estándar de las variables x y y.
2. El desarrollo de esta fórmula se estudia en el capítulo 13.

Para calcular r, usaremos una fórmula alternativa, la (3.2), que es equivalente a la (3.1). Como cálculos preliminares, calcularemos por separado tres sumas de cuadrados y luego los sustituimos en la fórmula (3.2) para obtener r.

Fórmula de cálculo

$$\text{coeficiente de correlación lineal} = \frac{\text{suma de cuadrados para } xy}{\sqrt{(\text{suma de cuadrados para } x)(\text{suma de cuadrados para } y)}}$$

$$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}} \tag{3.2}$$

PARA SU INFORMACIÓN
SS(x) es el numerador de la varianza

Recuerde el cálculo de SS(x) de la fórmula (2.9) para la varianza muestral (p. 87)

$$\text{suma de cuadrados para } x = \text{suma de } x^2 - \frac{(\text{suma de } x)^2}{n}$$

$$SS(x) = \sum x^2 - \frac{(\sum x)^2}{n} \tag{2.9}$$

También podemos calcular:

$$\text{suma de cuadrados para } y = \text{suma de } y^2 - \frac{(\text{suma de } y)^2}{n}$$

$$SS(y) = \sum y^2 - \frac{(\sum y)^2}{n} \tag{3.3}$$

$$\text{suma de cuadrados para } xy = \text{suma de } xy - \frac{(\text{suma de } x)(\text{suma de } y)}{n}$$

$$SS(xy) = \sum xy - \frac{\sum x \sum y}{n} \tag{3.4}$$

EJEMPLO 3.5 Cálculo del coeficiente de correlación lineal, *r*

Encuentre el coeficiente de correlación lineal para los datos de (lagartijas)/(sentadillas) del ejemplo 3.3 (p. 153).

SOLUCIÓN Primero, construimos una tabla de extensiones (tabla 3.11) con una lista de todos los pares de valores (x, y) para ayudarnos a hallar x^2, xy, y y^2 para cada par y los cinco totales de columna.

TABLA 3.11

Tabla de extensiones para hallar cinco sumatorias

Estudiante	Lagartijas, x	x^2	Lagartijas, y	y^2	xy
1	27	729	30	900	810
2	22	484	26	676	572
3	15	225	25	625	375
4	35	1 225	42	1 764	1 470
5	30	900	38	1 444	1 140
6	52	2 704	40	1 600	2 080
7	35	1 225	32	1 024	1 120
8	55	3 025	54	2 916	2 970
9	40	1 600	50	2 500	2 000
10	40	1 600	43	1 849	1 720
	$\sum x = 351$	$\sum x^2 = 13\ 717$	$\sum y = 380$	$\sum y^2 = 15\ 298$	$\sum xy = 14\ 257$
	suma de x	*suma de x^2*	*suma de y*	*suma de y^2*	*suma de xy*

En segundo término, para completar los cálculos preliminares, sustituimos las cinco sumatorias (los cinco totales de columna) de la tabla de extensiones en las fórmulas (2.9), (3.3) y (3.4), y calculamos las tres sumas de cuadrados:

$$\mathrm{SS}(x) = \sum x^2 - \frac{(\sum x)^2}{n} = 13\ 717 - \frac{(351)^2}{10} = 1396.9$$

$$\mathrm{SS}(y) = \sum y^2 - \frac{(\sum y)^2}{n} = 15\ 298 - \frac{(380)^2}{10} = 858.0$$

$$\mathrm{SS}(xy) = \sum xy - \frac{\sum x \sum y}{n} = 14\ 257 - \frac{(351)(380)}{10} = 919.0$$

PARA SU INFORMACIÓN

Los valores de Σ y SS se harán necesarios para regresión en la sección 3.4. Asegúrese de guardarlos.

En tercer término, sustituimos las tres sumas de cuadrados en la fórmula (3.2) para hallar el valor del coeficiente de correlación:

$$r = \frac{\mathrm{SS}(xy)}{\sqrt{\mathrm{SS}(x)\mathrm{SS}(y)}} = \frac{919.0}{\sqrt{(1396.9)(858.0)}} = 0.8394 = \mathbf{0.84}$$

Nota: típicamente, r se redondea al centésimo más cercano.

El valor del coeficiente de correlación lineal nos ayuda a contestar la pregunta ¿hay correlación lineal entre las dos variables bajo consideración? Cuando el valor calculado de r es cercano a cero, concluimos que hay poca o ninguna correlación lineal. Cuando el valor calculado de r cambia de 0.0 hacia +1.0 o −1.0, esto indica una cada vez más fuerte correlación lineal entre las dos variables. Desde un punto de vista gráfico, cuando calculamos r, estamos midiendo qué tan bien es que una recta describe el diagrama de dispersión de pares ordenados. Cuando el valor de r cambia de 0.0 hacia +1.0 o −1.0, los puntos de datos crean un patrón que se acerca más a una recta.

PARA SU INFORMACIÓN

Vea cómo opera esto en el ejercicio 3.29, página 169.

INSTRUCCIONES DE TECNOLOGÍA: COEFICIENTE DE CORRELACIÓN

MINITAB (Versión 14) Introduzca los datos de la variable x en C1 y los datos correspondientes de la variable y en C2; luego continúe con:

```
Choose:     Stat > Basic Statistics > Correlation...
Enter:      Variables: C1 C2 > OK
```

Excel Introduzca los datos de la variable x en la columna A y los datos correspondientes de la variable y en la columna B, active una celda para la respuesta; luego continúe con:

```
Choose:     Insert function, fx > Statistical > CORREL > OK
Enter:      Array 1: x data range
            Array 2: y data range > OK
```

TI-83/84 Plus Introduzca los datos de la variable x en L1 y los datos correspondientes de la variable y en L2; luego continúe con:

```
Choose:     2nd > CATALOG > DiagnosticOn* > ENTER > ENTER
Choose:     STAT > CALC > 8:LinReg(a + bx)
Enter:      L1, L2
```

*Debe seleccionarse DiagnosticOn para que aparezcan r y r^2. Una vez fijada, omita este paso.

Para entender el coeficiente de correlación lineal

El siguiente método creará (1) un significado visual para la correlación, (2) un significado visual para lo que está midiendo el coeficiente lineal, y (3) un cálculo para r. El método es rápido y en general da una estimación razonable cuando la "ventana de datos" es aproximadamente cuadrada.

Nota: esta técnica de cálculo no sustituye al cálculo de r. Es muy sensible a la "dispersión" del diagrama. No obstante, si la "ventana de datos" es aproximadamente cuadrada, esta aproximación será útil como cálculo mental o prueba.

FIGURA 3.9
La ventana de datos

FIGURA 3.10
Enfoque en el patrón

Procedimiento

1. Construya un diagrama de dispersión de sus datos, asegurándose de asignar una escala a los ejes de modo que la gráfica resultante tenga una "ventana de datos" aproximadamente cuadrada, como lo demuestra en la figura 3.9 el marco en verde claro. La ventana puede no ser la misma región como lo determinan las acotaciones de las dos escalas, mostradas como rectángulo verde en la figura 3.9.
2. Ponga dos lápices en su diagrama de dispersión. Manteniéndolos paralelos, muévalos a una posición de manera que estén tan juntos como sea posible pero que tengan todos los puntos del diagrama de dispersión entre ellos. (Vea la figura 3.10.)
3. Visualice una región rectangular que esté limitada por los dos lápices y que termine un poco fuera de los puntos del diagrama de dispersión. (Vea la parte sombreada de la figura 3.10.)

FIGURA 3.11
Búsqueda de *k*

4. Calcule el número de veces que el rectángulo sea más largo que su ancho. Una forma fácil de hacerlo es marcar mentalmente cuadrados en el rectángulo. (Vea la figura 3.11.) A este número desígnelo como k.
5. El valor de r puede calcularse como $\pm\left(1 - \dfrac{1}{k}\right)$.
6. El signo asignado a r está determinado por la posición general de la longitud de la región rectangular. Si se encuentra en una posición creciente, r será positiva; si está en una posición decreciente, r será negativa (vea la figura 3.12). Si el rectángulo está en posición ya sea horizontal o vertical, entonces r será cero, cualquiera que sea la relación entre longitud y ancho.

FIGURA 3.12
(a) Posición creciente; (b) posición decreciente

Usemos este método para calcular el valor del coeficiente de correlación lineal para la relación entre el número de lagartijas y sentadillas. Como se ve en la figura 3.13, encontramos que el rectángulo es aproximadamente 3.5 veces más largo que su ancho, es decir, $k \approx 3.5$, y el rectángulo se encuentra en una posición creciente. Por tanto, nuestro cálculo para r es

$$r \approx +\left(1 - \frac{1}{3.5}\right) \approx \mathbf{+0.70}$$

FIGURA 3.13
Lagartijas contra sentadillas para 10 estudiantes

Causalidad y variables ocultas

Cuando tratamos de explicar el pasado, entender el presente y estimar el futuro, los juicios acerca de una causa y efecto son necesarios debido a nuestro deseo de imponer orden en nuestro entorno.

La *relación de causa y efecto* es más bien sencilla. Uno puede enfocarse en una situación, el *efecto* (por ejemplo una enfermedad o problema social), y tratar de determinar su *causa(s)*, o puede empezar con una *causa* (condiciones no sanitarias o pobreza) y discutir su *efecto(s)*. Para determinar la causa de algo, nos preguntamos **por qué** ocurrió. Para determinar el efecto, nos preguntamos **qué** pasó.

Variable oculta: es una variable no incluida en un estudio pero que tiene un efecto sobre las variables del estudio y hace parecer que esas variables están relacionadas.

Un buen ejemplo es la fuerte relación positiva mostrada entre la cantidad en pérdidas o daños causados por un incendio y el número de bomberos que combaten el incendio. El "tamaño" del incendio es la variable oculta; "ocasiona" la "cantidad" de daños y el "número" de bomberos.

Si hay una fuerte correlación lineal entre dos variables, entonces una de las siguientes situaciones puede ser verdadera acerca de la relación entre las dos variables:

1. Hay una relación directa de causa y efecto entre las dos variables.
2. Hay una relación inversa de causa y efecto entre las dos variables.
3. La relación entre ambas puede ser causada por una tercera variable.
4. La relación entre ambas puede ser causada por las interacciones de varias otras variables.
5. La relación aparente puede ser estrictamente una coincidencia.

Recuerde que una fuerte correlación no necesariamente implica causalidad.

A continuación veamos algunas dificultades que deben evitarse:

1. En una relación directa de causa y efecto, un aumento (o reducción) en una variable ocasiona un aumento (o reducción) en la otra. Supongamos que hay una fuerte relación positiva entre peso y estatura. El aumento en peso, ¿*causa* un aumento en estatura? No necesariamente. O, para decirlo de otro modo, ¿la disminución en peso *causa* una disminución en estatura? Intervienen otras muchas variables posibles, por ejemplo género, edad y tipo de cuerpo. Estas otras variables se denominan *variables ocultas.*
2. En el caso práctico 3.4 (p. 154), existió una correlación negativa entre el porcentaje de estudiantes que recibieron refrigerio gratis o a precio bajo y el porcentaje de estudiantes que aprobaron el examen de suficiencia de matemáticas. ¿Debemos contener los refrigerios gratuitos para que más estudiantes aprueben el examen de matemáticas? Una tercera variable es la motivación para esta relación, es decir, el nivel de pobreza.
3. No relacione de *correlación* a *causa:* sólo porque todas las personas que se mudan a la ciudad envejecen no significa que la ciudad *cause* envejecimiento. La ciudad puede ser un factor, pero no se puede basar un argumento en la correlación.

CASO PRÁCTICO 3.6 Primas de seguros de vida

© STOCK IMAGE/Alamy

Un coeficiente alto de correlación lineal, *r*, ¿implica que los datos son lineales en su naturaleza? La edad problema del asegurado, y sus primas mensuales de seguro de vida para no fumadores, parecen estar altamente correlacionados si vemos la gráfica siguiente. Cuando aumenta la edad problema, la prima mensual del seguro aumenta para cada uno de los géneros.

TABLA 3.12

Primas mensuales de seguro de vida para no fumadores

	$100 000		$250 000		$500 000	
Edad problema	Hombres ($)	Mujeres ($)	Hombres ($)	Mujeres ($)	Hombres ($)	Mujeres ($)
30	7.96	6.59	11.96	9.13	19.25	12.46
35	8.05	6.56	11.96	9.13	19.57	12.46
40	9.63	7.79	15.22	10.89	23.19	16.47
45	13.14	9.80	22.40	15.44	35.87	24.03
50	18.44	12.42	33.69	21.10	53.81	33.38
55	26.01	15.75	49.22	29.37	87.59	48.06
60	37.10	20.83	74.59	42.05	137.38	69.87

Fuente: http://www.reliaquote.com/termlife/default.asp; con acceso el 11 de marzo, 2005

Todas las primas presentadas son para las mejores clasificaciones de no fumadores de cada línea de transporte.

Consideremos la edad problema del asegurado y la prima mensual para hombres para una póliza de 100 000 dólares. El coeficiente de correlación calculado para esta clase específica de seguro resulta en un valor de $r = 0.932$. Típicamente, un valor de r cercano a 1.0 indica una relación lineal bastante fuerte; pero, espere. ¿Tenemos una relación lineal? Sólo un diagrama de dispersión nos lo puede decir.

El diagrama de dispersión muestra claramente un patrón que no es de línea recta. No obstante, el coeficiente de correlación fue muy alto. Es el patrón prolongado de los datos que produce una r calculada tan grande. La lección de este ejemplo es que siempre debemos iniciar con un diagrama de dispersión cuando consideremos una correlación lineal. El coeficiente de correlación sólo indica un lado del caso.

SECCIÓN 3.3 EJERCICIOS

3.29 La siguiente pantalla muestra un simulador que contiene diagramas de dispersión para diversos coeficientes de correlación.

a. Comenzando en $r = 0$, mueva la regla a la derecha hasta $r = 1$. Explique lo que está ocurriendo a los diagramas de dispersión correspondientes.

b. Comenzando en $r = 0$, mueva la regla a la izquierda hasta $r = 1$. Explique lo que está ocurriendo a los diagramas de dispersión correspondientes.

3.30 ¿Cómo se interpretan los resultados de un estudio de correlación que reportó un coeficiente de correlación lineal de −1.34?

3.31 ¿Cómo se interpretan los resultados de un estudio de correlación que reportó un coeficiente de correlación lineal de +0.3?

3.32 Explique por qué tiene sentido que un conjunto de datos tenga un coeficiente de correlación de cero cuando los datos muestran un patrón bien definido, como en la figura 3.9 (p. 163).

3.33 ¿Da resultado estudiar para un examen? El número de horas estudiadas, x, se compara con la calificación de examen recibida, y:

x	2	5	1	4	2
y	80	80	70	90	60

a. Complete los cálculos preliminares: extensiones, cinco sumas, SS(x), SS(y), y SS(xy).

b. Encuentre r.

3.34 La siguiente tabla contiene los porcentajes de estudiantes que reciben refrigerios gratis o a precio bajo, comparados con los porcentajes de quienes aprobaron la parte de lectura del examen estatal. Los resultados son para el Condado de Sandusky, Ohio, y fueron informados en *The Blade,* un periódico de Toledo, el 5 de agosto de 2001. El Condado de Sandusky es una combinación de 13 escuelas rurales y urbanas. (Algunos datos son como en el ejercicio 3.21, p. 159.)

Escuela	% refrigerio gratis/ (precio bajo)	% aprueban lectura	Escuela	% refrigerio gratis/ (precio bajo)	% aprueban lectura
1	29	66	8	47	58
2	29	59	9	29	88
3	23	62	10	17	68
4	60	53	11	22	60
5	57	53	12	38	47
6	50	57	13	15	62
7	49	54			

Encuentre: a. SS(x) b. SS(y) c. SS(xy) d. r

3.35 Muchas organizaciones ofrecen precios "especiales" de suscripción de revistas a sus socios. La American Federation of Teachers no es la excepción; veamos a continuación algunos precios que ofrecen para sus socios.

Revista	Precio normal	Su precio
Cosmopolitan	$29.97	$18.00
Sports Illustrated	$78.97	$39.75
Ebony	$20.00	$14.97
Rolling Stone	$23.94	$11.97
Martha Stewart Living	$24.95	$20.00

Fuente: American Federation of Teachers

a. Construya un diagrama de dispersión con "Su precio" como la variable dependiente, y, y "Precio normal" como la variable independiente, x.

Encuentre:

b. SS(x) c. SS(y) d. SS(xy)

e. Momento del producto de Pearson, r

3.36 Calcule el coeficiente de correlación para cada uno de lo siguiente:

3.37 Los manatíes nadan cerca de la superficie del agua y es frecuente que tengan problemas con muchos botes de motor en Florida. Considere la gráfica que sigue.

a. ¿Cuáles son los dos grupos de sujetos que se comparan?

b. ¿Cuáles son las dos variables que se usan para hacer la comparación?

c. ¿Qué conclusión se puede sacar con base en esta gráfica de dispersión?

d. ¿Qué podría hacer usted si fuera guardia de fauna en Florida?

3.38 a. Use el diagrama de dispersión que dibujó en el ejercicio 3.19 (p. 159), para calcular r para los datos muestrales sobre el número de horas estudiadas y la calificación de examen.

b. Calcule r.

3.39 a. Use el diagrama de dispersión que dibujó en el ejercicio 3.20 (p. 159), para calcular r para los datos muestrales sobre el número de respuestas irrelevantes y la edad del niño.

b. Calcule r.

PARA SU INFORMACIÓN ¿Ha intentado usar comandos de correlación en su computadora o calculadora?

3.40 Una compañía de mercadotecnia deseaba determinar si el número de anuncios comerciales por televisión estaba correlacionado linealmente con las ventas de su producto. Los datos, obtenidos de cada una de varias ciudades, se ven en la tabla siguiente.

Ciudad	A	B	C	D	E	F	G	H	I	J
Comerciales, x	12	6	9	15	11	15	8	16	12	6
Ventas unitarias, y	7	5	10	14	12	9	6	11	11	8

3.41 El simulador siguiente compara coeficientes de correlación con sus diagramas de dispersión. Después de varias rondas de práctica usando "New Plots", (Nuevas gráficas), explique su método de compararlos.

3.42 El simulador muestra prácticas para construir diagramas de dispersión para comparar con coeficientes de correlación determinados.

a. Después de colocar sólo 2 puntos, ¿cuál es el valor calculado de r para cada diagrama de dispersión? ¿Por qué?

b. ¿Cuál diagrama de dispersión encontró usted más fácil de construir?

3.43 Las bebidas para deportes son muy populares en la cultura actual en todo el mundo. La tabla siguiente menciona 10 productos diferentes que se pueden comprar en Inglaterra y los valores para tres variables: costo por porción (en peniques), energía por porción (en kilocalorías), y carbohidratos por porción (en gramos).

Bebida para deportes	Costo	Energía	Carbohidratos
Lucozade Sport RTD 330ml pouch/can	72	92	21.1
Lucozade Sport RTD 500ml bot.	79	140	32
Lucozade Sport RTD 650ml sports bot.	119	182	41.6
POWERade 500ml bot.	119	120	30
Gatorade Sports 750ml	89	188	45
Science in Sport Go Electrolyte (500ml)	99	160	40
High Five Isotonic electrolyte (750ml)	99	220	55
Isostar powder (per litre) 5l tub	126	320	77
Isostar RTD 500ml bot.	99	150	35
Maxim Electrolyte (per litre) 2kg bag	66	296	75

Nota: el costo es en peniques (p), 0.01 de libra inglesa, con valor de $0.0187 el 28 de marzo de 2005.

La energía se mide en kilocalorías; los carbs (carbohidratos) se miden en gramos.

Fuente: http://www.simplyrunning.net/articles/sports_drinks.htm

a. Trace un diagrama de dispersión usando x = carbs/porción y y = energía/porción.

b. ¿Parece haber relación lineal?

c. Calcule el coeficiente de correlación lineal, r.

d. ¿Qué parece decirnos este valor de correlación? Explique.

e. Repita las partes a-d usando x = costo/porción y y = energía/porción. (Retenga estas soluciones para usarlas en el ejercicio 3.59, p.185.)

3.44 La National Adoption Information Clearinghouse da seguimiento y publica información relacionada con adopciones en Estados Unidos. La tabla siguiente contiene el número de niños adoptados en cada uno de los 16 estados identificados al azar para 1991 y 2001.

Estado	1991	2001	Estado	1991	2001	Estado	1991	2001
DE	190	225	IA	1518	1116	WY	425	412
NV	779	764	NJ	2382	2384	AL	1939	1857
MI	4758	6274	AR	1678	1698	ID	879	1048
SC	1471	1648	HI	592	766	WA	2603	2748
GA	2330	3499	TN	751	2633	VT	350	407
AK	898	616						

Fuente: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services, 2004, http://naic.acf.hhs.gov/pubs/s_adopted/index.cfm

¿Hay una correlación lineal entre los datos de 1991 y 2001? Use estadísticas gráficas y numéricas para apoyar su respuesta.

3.45 La autopista 95, la más larga de las interestatales de norte a sur en Estados Unidos, tiene 1907 millas de largo y va de Houlton, Maine, en la frontera con Canadá hasta el extremo norte de Miami, Florida, como la US 1 en el extremo sur. Atraviesa 15 estados de la costa este; el número de millas y el de intersecciones en cada uno de los estados es éste.

Estado	FL	GA	SC	NC	VA	MD	DE	
Cruceros	73	19	39	44	51	38	9	
Millas	381	112	201	183	178	110	26	
Estado	PA	NJ	NY	CT	RI	MA	NH	ME
Cruceros	16	28	12	68	26	44	4	52
Millas	58	44	29	118	47	97	17	306

Fuente: Rand McNally y http://www.ihoz.com/I90.html

Usando los 15 datos:

a. Construya un diagrama de dispersión con el número de cruceros como la variable dependiente, y, y millas como la variable independiente, x.

b. ¿Parece haber un patrón lineal en los datos? ¿Parece razonable el patrón para las variables? Explique por qué sí o por qué no.

c. Calcule el coeficiente de correlación lineal, r.

d. El valor de r ¿parece razonable comparado con el patrón demostrado en el diagrama de dispersión? Explique.

Elimine la CT (118, 68) de los datos, y luego:

e. ¿Qué pasa con el punto de datos de Connecticut que lo hace diferente? ¿Es comprensible por qué es diferente?

f. ¿Qué efecto pareció tener la remoción del punto de datos de Connecticut en el diagrama?

g. Calcule el coeficiente de correlación lineal, r.

h. ¿Qué efecto tuvo la remoción del punto de Connecticut sobre el valor de r? ¿Cómo se compara esto con el efecto que usted anticipó? Explique.

3.46 Las compañías productoras de cine gastan millones de dólares para producir películas con la esperanza de atraer millones de personas a los cines. El éxito de una película puede medirse en muchas formas, dos de las cuales son los recibos de taquillas y el número recibido de nominaciones al Oscar. A continuación aparece una lista de diez cines en 2005 y sus correspondientes "tarjetas de reporte". Cada película se midió con base en su presupuesto de costos (en millones de dólares), sus recibos de taquillas (en millones de dólares), y el número recibido de nominaciones al Oscar.

Película	Presupuesto	Taquilla	Nominaciones
The Aviator	110	82.3	11
Finding Neverland	24	42.5	7
Million Dollar Baby	30	44.9	7
Ray	35	74.7	6
Sideways	16	52.8	5
Hotel Rwanda	17	14.2	3
Vera Drake	8.5	2.8	3
Eternal Sunshine of the Spotless Mind	20	34.1	2
Being Julia	10	5.1	1
Maria Full of Grace	3	6.5	1

Fuente: USA Today, 8 de febrero 2005, "Guía rápida de películas para premios"

a. Trace un diagrama de dispersión usando x = presupuesto y y = taquilla.

b. ¿Parece haber una relación lineal?

c. Calcule el coeficiente de correlación lineal, r.

d. ¿Qué parece decirnos este valor de correlación? Explique.

e. Repita las partes a-d usando x = taquilla y y = nominaciones.

3.47 El sistema nacional de autopistas consta de autopistas interestatales y no interestatales. A continuación aparecen 15 estados seleccionados al azar y su correspondiente número de millas de autopistas interestatales y no interestatales.

Sistema nacional de autopistas, número de millas, octubre 2005

Estado	Interestatal	No interestatal	Estado	Interestatal	No interestatal
AL	905	2715	NE	482	2496
VT	320	373	UT	940	1253
NH	235	589	TX	3233	10157
RI	71	197	OK	930	2431
AZ	1167	1565	WV	549	1195
IA	782	2433	AK	1082	1030
WI	745	3404	GA	1245	3384
NY	1674	3476			

Fuente: U.S. Department of Transportation

a. Construya un diagrama dispersión usando millas x = interestatales y y = no interestatales.

b. Describa el patrón indicado, incluyendo cualesquiera características no usuales.

c. Calcule el coeficiente de correlación.

d. Remueva Texas de los datos y repita las partes a-c.

e. Compare las respuestas halladas en la parte d con las halladas en las partes a y c, incluyendo comentarios acerca de qué efectos tuvo sobre el coeficiente de correlación la remoción de Texas de los datos.

3.48 Los jugadores, equipos y aficionados de la NBA están interesados en ver que sus mejores anotadores encesten muchos puntos, pero, al mismo tiempo, el número de faltas personales que cometen tiende a limitar su tiempo de juego. Por cada equipo, la tabla siguiente menciona el número de minutos jugados por partido (min/G) y el número de faltas personales cometidas por juego (PF/G) por el principal anotador durante la temporada de 2003-2004.

Equipo	Min/G	PF/G
Bulls	35.14	2.01
Lakers	37.65	2.71

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/johnson

Fuente: NBA.com

a. Construya un diagrama de dispersión.

b. Describa el patrón resultante. ¿Hay algunas características no usuales?

c. Calcule el coeficiente de correlación, r.

d. ¿Parece razonable el valor del coeficiente de correlación?

3.49 Al ver las pólizas de seguros de la tabla del caso práctico 3.6, podemos ver que a medida que una persona envejece, la prima de seguro aumenta. Es probable que el lector anticipe esto, pero veamos más de cerca una de las situaciones específicas mencionadas.

a. Calcule el coeficiente de correlación, r, para las variables edad problema (x) y prima mensual para $250 000 para hombres.

b. Trace un diagrama de dispersión de los datos del seguro para hombres a los $250 000 basados en la edad (x).

c. ¿Parecen los datos tener un patrón lineal? Explique.

d. Explique cómo es que un patrón de datos no lineal puede tener un alto coeficiente de correlación lineal.

e. Explique por qué debe haber anticipado este patrón no lineal.

f. (Opcional) Investigue una o más de las otras cinco columnas de primas de seguro que conteste las partes a-e para cada una.

3.50 En numerosas comunidades hay una fuerte correlación positiva entre la cantidad de helados vendidos en un mes determinado y el número de ahogados que ocurren en ese mes. ¿Significa esto que los helados ocasionan ahogamientos? Si no es así, ¿puede usted dar una explicación alternativa para la fuerte asociación? Escriba unas pocas frases que aborden estas preguntas.

3.51 Explique por qué uno debe esperar hallar una correlación positiva entre el número de carros de bomberos que responden a una llamada de incendio y la cantidad de pérdidas materiales en el incendio. ¿Significa esto que las pérdidas serían menos costosas si se despacharan menos carros de bomberos? Explique.

3.4 Regresión lineal

Aun cuando el coeficiente de correlación mide la fuerza de una relación lineal, no nos dice nada acerca de la relación matemática entre las dos variables. En la sección 3.3 se encontró que el coeficiente de relación para los datos de lagartijas/sentadillas era de 0.84 (vea pp. 164-165). Esto, además del patrón del diagrama de dispersión, implica que hay una relación lineal entre el número de lagartijas y el de sentadillas que hace un estudiante. No obstante, el coeficiente de correlación no nos ayuda a pronosticar el número de sentadillas que una persona puede hacer con base en saber que puede hacer 28 lagartijas. El **análisis de regresión** encuentra la ecuación

de la recta que mejor describe la relación entre las dos variables. Un uso de esta ecuación es hacer predicciones. Hacemos uso de estas predicciones regularmente, por ejemplo, predecimos el éxito que un estudiante tendrá en la universidad con base en sus resultados de preparatoria y predecir la distancia necesaria para detener un auto con base en su velocidad. En general, el valor exacto de y no se puede pronosticar y nos satisface saber que las predicciones son razonablemente cercanas.

La relación entre dos variables será una expresión algebraica que describe la relación matemática entre x y y. A continuación veamos algunos ejemplos de diversas relaciones posibles, llamadas *modelos* o **ecuaciones de predicción:**

Lineal (línea recta):	$\hat{y} = b_0 + b_1 x$
Cuadrática:	$\hat{y} = a + bx + cx^2$
Exponencial:	$\hat{y} = a(b^x)$
Logarítmica:	$\hat{y} = a \log_b x$

Las figuras 3.14, 3.15 y 3.16 muestran patrones de datos bivariados que parecen tener una relación, mientras que en la figura 3.17 las variables no parecen estar relacionadas.

FIGURA 3.14
Regresión lineal con pendiente positiva

FIGURA 3.15
Regresión lineal con pendiente negativa

FIGURA 3.16
Regresión curvilínea (cuadrática)

FIGURA 3.17
No hay relación

Si un modelo de recta parece apropiado, la recta de mejor ajuste se encuentra con el uso del **método de mínimos cuadrados.** Supongamos que $\hat{y} = b_0 + b_1 x$ es la ecuación de una recta, donde $\hat{y}$ representa el **valor pronosticado de** y que corresponde a un valor particular de x. El **criterio de mínimos cuadrados** requiere que encontremos las constantes b_0 y b_1 tales que $\Sigma(y - \hat{y})^2$ sea tan pequeña como sea posible.

La figura 3.18 muestra la distancia de un valor observado de y desde un **valor pronosticado de** $\hat{y}$. La longitud de esta distancia representa el valor $(y - \hat{y})$ (mostrado como un segmento de recta roja en la figura 3.18). Nótese que $(y - \hat{y})$ es positiva cuando el punto (x, y) está arriba de la recta y negativa cuando (x, y) está debajo de la recta.

La figura 3.19 muestra un diagrama de dispersión con lo que parece ser la **recta de mejor ajuste**, junto con 10 valores individuales $(y - \hat{y})$. (Se indican valores positivos en rojo; negativos, en verde.) La suma de los cuadrados de estas diferencias se minimiza (se hace tan pequeño como es posible) si la recta es en realidad la recta de mejor ajuste.

La figura 3.20 ilustra los mismos puntos de datos que la figura 3.19. Los 10 valores individuales de $(y - \hat{y})$ se localizan con una recta que definitivamente no es la recta de mejor ajuste. [El valor de $\Sigma(y - \hat{y})^2$ es 149, mucho mayor que 23 de la figura 3.19.] Toda recta diferente trazada que pase por este conjunto de 10 puntos resultará en un valor diferente para $\Sigma(y - \hat{y})^2$. Nuestro trabajo es hallar aquella recta que haga de $\Sigma(y - \hat{y})^2$ el mínimo valor posible.

FIGURA 3.18 Valores observados y pronosticados de *y*

FIGURA 3.19 La recta de mejor ajuste

$\sum (y - \hat{y})^2 = (-1)^2 + (+1)^2 + \ldots + (+1)^2 = 23.0$

FIGURA 3.20 Recta que no es de mejor ajuste

$\sum (y - \hat{y})^2 = (-6)^2 + (-4)^2 + \ldots + (+6)^2 = 149.0$

La ecuación de la recta de mejor ajuste está determinada por su **pendiente (b_1)** y su **ordenada en el origen (b_0).** (Vea un texto de geometría analítica y realice un repaso de los conceptos de pendiente y ordenada en el origen de una recta.) Los valores de las constantes(pendiente y ordenada en el origen) que satisfacen el criterio de mínimos cuadrados se encuentran usando las fórmulas siguientes:

Fórmula de definición

D

$$\text{pendiente:} \quad b_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \tag{3.5}$$

Usaremos un equivalente matemático de la fórmula (3.5) para la pendiente, b_1, que utiliza las sumas de cuadrados encontrados en los cálculos preliminares para correlación:

Fórmula computacional

$$\text{pendiente:} \quad b_1 = \frac{\text{SS}(xy)}{\text{SS}(x)} \tag{3.6}$$

Nótese que el numerador de la fórmula (3.6) es la fórmula SS(xy) (3.4) (p. 164) y el denominador es la fórmula (2.9) (p. 87) de los cálculos de coeficiente de correlación. Así, si previamente ha calculado el coeficiente de correlación lineal usando el procedimiento indicado en las páginas 164-165, puede fácilmente hallar la pendiente de la recta de mejor ajuste. Si r no fue calculada antes, elabore una tabla semejante a la tabla 3.11 (p. 165) y complete los cálculos preliminares necesarios.

Para la ordenada en el origen, tenemos:

Fórmula computacional

$$\text{ordenada en el origen } y = \frac{(\textit{suma de } y) - [(\textit{pendiente})(\textit{suma de } x)]}{\textit{número}}$$

$$b_0 = \frac{\sum y - (b_1 \cdot \sum x)}{n} \tag{3.7}$$

Fórmula computacional alternativa

$$\text{ordenada en el origen } y = y \textit{ barra} - (\textit{pendiente} \cdot x \textit{ barra})$$

$$b_0 = \bar{y} - (b_1 \cdot \bar{x}) \tag{3.7a}$$

Consideremos ahora los datos del ejemplo 3.3 (p. 153) y el problema de pronosticar el número de sentadillas que haga un estudiante con base en el número de lagartijas. Deseamos hallar la recta de mejor ajuste, $\hat{y} = b_0 + b_1x$. Los cálculos preliminares ya se han completado en la tabla 3.11 (p. 165). Para calcular la pendiente, b_1, usando la fórmula (3.6), recuerde que SS(xy)=919.0 y SS(x)=1396.9. Por tanto,

$$\text{pendiente:} \quad b_1 = \frac{\text{SS}(xy)}{\text{SS}(x)} = \frac{919.0}{1396.9} = 0.6579 = \mathbf{0.66}$$

Para calcular la ordenada en el origen, b_0, usando la fórmula (3.7), recuerde que $\sum x = 351$ y $\sum y = 380$ de la tabla de extensiones. Tenemos

$$\text{ordenada en el origen } y\text{:} \quad b_0 = \frac{\sum y - (b_1 \cdot \sum x)}{n} = \frac{380 - (0.6579)(351)}{10}$$

$$= \frac{380 - 230.9229}{10} = 14.9077 = \mathbf{14.9}$$

Al colocar los dos valores recién encontrados en el modelo $\hat{y} = b_0 + b_1x$,, obtenemos la ecuación de la recta de mejor ajuste:

$$\boldsymbol{\hat{y} = 14.9 + 0.66x}$$

Notas:

1. Recuerde conservar al menos tres lugares decimales extra cuando haga los cálculos, para asegurar una respuesta precisa.
2. Cuando haga redondeo de los valores calculados de b_0 y b_1, siempre conserve al menos dos cifras significativas en la respuesta final.

Ahora que conocemos la ecuación para la recta de mejor ajuste, tracemos la recta en el diagrama de dispersión para que podamos ver la relación entre la recta y los datos. Necesitamos dos puntos para trazar la recta sobre el diagrama. Seleccione dos valores x convenientes, uno cerca de cada extremo del dominio ($x = 10$ y $x = 60$ son buenas opciones para esta ilustración), y encontrar sus valores y correspondientes.

Para $x = 10$: $\hat{y} = 14.9 + 0.66x = 14.9 + 0.66(10) = 21.5$; **(10, 21.5)**

Para $x = 60$: $\hat{y} = 14.9 + 0.66x = 14.9 + 0.66(60) = 54.5$; **(60, 54.5)**

Estos dos puntos, (10, 21.5) y (60, 54.5), se localizan entonces en el diagrama de dispersión (usamos un signo + de color morado para distinguirlo de los puntos de datos) y se traza la recta de mejor ajuste (indicada en rojo en la figura 3.21).

Hay algunos datos adicionales que necesitamos analizar acerca del método de mínimos cuadrados.

1. La pendiente, b_1, representa el cambio pronosticado en y por aumento unitario en x. En nuestro ejemplo, donde $b_1 = 0.66$, si un estudiante puede hacer 10 lagartijas adicionales (x), pronosticamos que podría hacer aproximadamente $7(0.66 \times 10)$ sentadillas adicionales (y).
2. La ordenada en el origen es el valor de donde la recta de mejor ajuste cruza el eje y. (Cuando la escala vertical está ubicada arriba de $x = 0$, la ordenada en el origen se ve fácilmente en el diagrama de dispersión, mostrada como un signo + verde

FIGURA 3.21
Recta de mejor ajuste para lagartijas contra sentadillas

en la figura 3.21.) Primero, no obstante, al interpretar b_0, se debe considerar si $x = 0$ es un valor x realista antes de concluir que se pronosticaría $\hat{y} = b_0$ si $x = 0$. Para predecir que si un estudiante no hizo lagartijas, es probablemente incorrecto que todavía haría alrededor de 15 sentadillas ($b_0 = 14.9$). En segundo término, el valor de x de cero puede estar fuera del dominio de los datos en el que se basa la recta de regresión. Al predecir y con base en un valor x, compruebe estar seguro que el valor x se encuentre dentro del dominio de los valores x observados.

3. La recta de mejor ajuste siempre pasará por el *centroide*, el punto $\bar{x}, \bar{y}$). Cuando trace la recta de mejor ajuste en su diagrama de dispersión, use este punto como prueba. Para nuestra ilustración,

$$\bar{x} = \frac{\sum x}{n} = \frac{351}{10} = 35.1, \qquad \bar{y} = \frac{\sum y}{n} = \frac{380}{10} = 38.0$$

Vemos que la recta de mejor ajuste pasa por $(\bar{x}, \bar{y}) = (35.1, 38.0)$, como se muestra en ⊕ verde de la figura 3.21.

Trabajemos otro ejemplo para aclarar los pasos que intervienen en el análisis de regresión.

EJEMPLO 3.7 Cálculo de la ecuación de la recta de mejor ajuste

En una muestra aleatoria de ocho muchachas universitarias, a cada una se le preguntó su estatura (a la pulgada más cercana) y su peso (a las 5 libras más cercanas). Los datos obtenidos aparecen en la tabla 3.13. Encuentre una ecuación para predecir el peso de una estudiante con base en su estatura (la ecuación de la recta de mejor ajuste), y dibújela en el diagrama de dispersión de la figura 3.22.

TABLA 3.13

Estaturas y pesos de muchachas universitarias

	1	2	3	4	5	6	7	8
Estatura, x	65	65	62	67	69	65	61	67
Peso, y	105	125	110	120	140	135	95	130

SOLUCIÓN Antes de comenzar a buscar la ecuación para la recta de mejor ajuste, a veces es útil trazar el diagrama de dispersión, lo cual da una idea visual de la relación entre las dos variables. El diagrama de dispersión para los datos de estaturas y pesos de muchachas universitarias, en la figura 3.22, indica que el modelo lineal es apropiado.

FIGURA 3.22
Diagrama de dispersión

Para hallar la ecuación para la recta de mejor ajuste, primero necesitamos completar los cálculos preliminares, como se ve en la tabla 3.14. Los otros cálculos preliminares incluyen hallar SS(x) de la fórmula (2.9) y SS(xy) de la fórmula (3.4):

TABLA 3.14

Cálculos preliminares necesarios para hallar b_1 y b_0

Estudiante	Estatura, x	x^2	Peso, y	xy
1	65	4225	105	6825
2	65	4225	125	8125
3	62	3844	110	6820
4	67	4489	120	8040
5	69	4761	140	9660
6	65	4225	135	8775
7	61	3721	95	5795
8	67	4489	130	8710
	$\sum x = 521$	$\sum x^2 = 33\ 979$	$\sum y = 960$	$\sum xy = 62\ 750$

$$\text{SS}(x) = \sum x^2 - \frac{(\sum x)^2}{n} = 33\ 979 - \frac{(521)^2}{8} = 48.875$$

$$\text{SS}(xy) = \sum xy - \frac{\sum x \sum y}{n} = 62\ 750 - \frac{(521)(960)}{8} = 230.0$$

En segundo término, necesitamos hallar la pendiente y la ordenada en el origen usando las fórmulas (3.6) y (3.7):

$$\text{pendiente:} \quad b_1 = \frac{SS(xy)}{SS(x)} = \frac{230.0}{48.875} = 4.706 = \mathbf{4.71}$$

$$y\text{-ordenada en el origen: } b_0 = \frac{\sum y - (b_1 \cdot \sum x)}{n} = \frac{960 - (4.706)(521)}{8} = -186.478 = \mathbf{-186.5}$$

Así, la ecuación de la recta de mejor ajuste es $\hat{y} = \mathbf{-186.5 + 4.71}x$.

Para trazar la recta de mejor ajuste en el diagrama de dispersión, necesitamos localizar dos puntos. Sustituya dos valores de x, por ejemplo 60 y 70, en la ecuación para la recta de mejor ajuste y obtenga dos valores de $\hat{y}$: correspondientes.

$$\hat{y} = -186.5 + 4.71x = -186.5 + (4.71)(60) = -186.5 + 282.6 = 96.1 \approx 96$$

$$\hat{y} = -186.5 + 4.71x = -186.5 + (4.71)(70) = -186.5 + 329.7 = 143.2 \approx 143$$

Los valores (60, 96) y (70, 143) representan dos puntos (designados por un signo + rojo en la figura 3.23) que hacen posible que tracemos la recta de mejor ajuste.

FIGURA 3.23
Diagrama de dispersión con recta de mejor ajuste

Nota: en la figura 3.23, $(\bar{x}, \bar{y}) = (65.1, 120)$, también está sobre la recta de mejor ajuste. Es el símbolo $\oplus$. Use $(\bar{x}, \bar{y})$ como prueba de su trabajo.

¿SABÍA USTED...?

Una recta de regresión

En la Exposición Internacional de Londres, en 1884, Sir Francis Galton instaló un laboratorio en el que a las personas les pagaba 3 peniques por medirles la cabeza. Galton estaba interesado en predecir la inteligencia humana y a la persona que le pagaba le daría su opinión de su inteligencia. Después de la exposición, el laboratorio se cambió al Museo de Londres, donde Galton continuó recolectando datos acerca de características humanas como estatura, peso y fuerza. Galton elaboró gráficas bidireccionales de estaturas para padres e hijos, que en última instancia llevó a la pendiente de la recta de regresión.

Para hacer predicciones

Una de las numerosas razones para buscar una ecuación de regresión es hacer predicciones. Una vez que se haya establecido una relación lineal y se conozca el valor de la variable x de entrada, podemos predecir un valor de y, $\hat{y}$. Considere la ecuación $\hat{y} = -186.5 + 4.71x$ que relaciona la estatura y el peso de muchachas universitarias. Si una estudiante universitaria en particular mide 66 pulgadas de estatura, ¿qué peso pronostica usted que tendrá? El valor pronosticado es

$$\hat{y} = -186.5 + 4.71x = -186.5 + (4.71)(66) = -186.5 + 310.86$$
$$= 124.36 \approx 124 \text{ lb}$$

El lector no debe esperar que este valor pronosticado sea exacto; más bien, es el peso promedio que se esperaría para todas las estudiantes universitarias que midan 66 pulgadas de estatura.

Cuando se hagan predicciones con base en la recta de mejor ajuste, observe las siguientes restricciones:

1. La ecuación debe usarse para hacer predicciones sólo acerca de la población de la cual se extrajo la muestra. Por ejemplo, usar nuestra relación entre la estatura y el peso de muchachas universitarias sería cuestionable para predecir el peso de atletas profesionales dada su estatura.
2. La ecuación debe usarse sólo dentro del dominio muestral de la variable de entrada. Sabemos que los datos demuestran una tendencia lineal dentro del dominio de los datos en x, pero no sabemos cuál es la tendencia fuera de este intervalo. En consecuencia, los pronósticos pueden tener un alto riesgo fuera del dominio de los datos en x. Como podrá verse, en el ejemplo 3.7 no tiene sentido predecir que una estudiante universitaria de estatura cero pesará −186.5 libras. No se use una estatura fuera del dominio muestral de 61 a 69 pulgadas para predecir el peso. En ocasiones podría usarse la recta de mejor ajuste para calcular valores fuera del intervalo de dominio de la muestra. Esto puede hacerse, pero con precaución y sólo para valores cercanos al intervalo del dominio.
3. Si la muestra se tomó en 2006, no espere que los resultados sean válidos en 1929 o se cumplan en 2010. Las mujeres de hoy día pueden ser diferentes de las de 1929 y de las de 2010.

INSTRUCCIONES DE TECNOLOGÍA: RECTA DE MEJOR AJUSTE

MINITAB (Versión 14)

Introduzca los valores x en C1 y los valores y correspondientes en C2; luego, para obtener la ecuación para la recta de mejor ajuste, continúe con:

Method 1—

```
Choose:      Stat > Regression > Regression . . .
Enter:       Response (y): C2
             Predictors (x): C1 > OK
```

Para trazar el diagrama de dispersión con la recta de mejor ajuste sobrepuesta en los puntos de datos, LOS AJUSTES deben haberse seleccionado previamente; luego continúe con:

```
Choose:      Graph > Scatterplot
Select:      With Regression > OK
Enter:       Y variable: C2  X variable: C1
Select:      Labels > Titles/Footnotes
Enter:       Title: your title > OK > OK
```

OR

Method 2—

```
Choose:      Stat > Regression > Fitted Line Plot
Enter:       Response (Y): C2
             Response (X): C1
Select:      Linear
Select:      Options
Enter:       Title: your title > OK > OK
```

Excel Introduzca los datos de la variable x en la columna A y los datos de la variable y correspondientes en la columna B; luego continúe con:

```
Choose:     Tools > Data Analysis > Regression > OK
Enter:      Input Y Range: (B1:B10 or select cells)
            Input X Range: (A1:A10 or select cells)
Select:     Labels (if necessary)
            Output Range
            Enter: (C1 or select cell)
            Line Fits Plots > OK
```

Para hacer legible la salida; continúe con:

```
Choose:     Format > Column > Auto?t Selection
```

Para formar la ecuación de regresión, la ordenada en el origen está ubicada en el cruce de las columnas de la ordenada y los coeficientes, mientras que la pendiente está situada en el cruce de las columnas de la variable x y los coeficientes.

Para trazar la recta de mejor ajuste en el diagrama de dispersión, active la gráfica; luego continúe con:

```
Choose:     Chart > Add Trendline > Linear > OK
```

(Este comando también funciona con los comandos Excel para el diagrama de dispersión de la p. 155)

TI-83/84 Plus Introduzca los datos de la variable x en L1 y los datos de la variable y correspondientes en L2; luego continúe con:

Si sólo se desea la ecuación:

```
Choose:     STAT >CALC >8:LinReg(a + bx)
Enter:      L1, L2*
```

*Si se desean la ecuación y la gráfica en el diagrama de dispersión, use:

```
Enter:   L1, L2, Y1†
```

luego continúe con los mismos comandos para un diagrama de dispersión como se ve en la página 155.

†Para introducir Y1, use:

```
Choose:   VARS > Y-VARS > 1:Function > 1:Y1 > ENTER
```

Para entender la recta de mejor ajuste

El método siguiente creará (1) un significado visual para la recta de mejor ajuste, (2) un significado visual para lo que la recta de mejor ajuste está describiendo, y (3) una estimación para la pendiente y ordenada en el origen de la recta de mejor ajuste. Al igual que con la aproximación de r, las estimaciones de la pendiente y ordenada en el origen de la recta de mejor ajuste deben usarse sólo como estimación mental o prueba.

Nota: esta técnica de estimación *no sustituye* a los cálculos para b_1 y b_0.

Procedimiento

1. En el diagrama de dispersión de los datos, trace una recta que parezca la recta de mejor ajuste. (*Sugerencia:* si traza una recta paralela y a la mitad entre los dos lápices descritos en la sección 3.3 en la página 166 [figura 3.10], tendrá una estimación razonable para la recta de mejor ajuste.) Los dos lápices limi-

tan la "trayectoria" demostrada por los pares ordenados, y la recta que está en el centro de esta trayectoria aproxima la recta de mejor ajuste. La figura 3.24 muestra los lápices y la recta estimada resultante para el ejemplo 3.7.

FIGURA 3.24
Estimación de la recta de mejor ajuste para los datos de estudiantes universitarias

2. Esta recta se puede usar ahora para aproximar la ecuación. Primero, localizamos cualesquier dos puntos (x_1, y_1) y (x_2, y_2) en toda la recta y determinamos sus coordenadas. Dos de estos puntos, circulados en la figura 3.24, tienen las coordenadas (59, 85) y (66, 125). Estos dos pares de coordenadas se pueden usar ahora en la fórmula siguiente para estimar la pendiente b_1:

   estimación de la pendiente, b_1: $$b_1 \approx \frac{y_2 - y_1}{x_2 - x_1} = \frac{125 - 85}{66 - 59} = \frac{40}{7} = 5.7$$

3. Usando este resultado, las coordenadas de uno de los puntos, y la fórmula siguiente, podemos determinar una estimación para la ordenada en el origen, b_0:

   estimación de la ordenada en el origen, b_0:
   $$b_0 \approx y - b_1 \cdot x = 85 - (5.7)(59) = 85 - 336.3 = -251.3$$
   Así, b_0 es aproximadamente −250.

4. Ahora podemos escribir la ecuación estimada para la recta de mejor ajuste:
   $$\hat{y} = -250 + 5.7x$$

Ésta debe servir como estimación burda. La ecuación real calculada usando todos los pares ordenadas fue $\hat{y} = -186.5 + 4.71x$.

CASO PRÁCTICO 3.8 Contracción del concreto

CONTRACCIÓN POR DESECACIÓN

La contracción por desecación se define como la contracción de una mezcla endurecida de concreto debida a la pérdida de agua capilar. Esta contracción produce un aumento en esfuerzo de tracción, que puede llevar a agrietamiento, pandeo interno y desviación externa, antes que el concreto se someta a alguna clase de carga. Todo el

concreto de cemento Portland experimenta contracción por desecación, o cambio en volumen hidráulico, a medida que el concreto envejece. El cambio en volumen hidráulico en concreto es muy importante para el ingeniero en el diseño de una estructura. La contracción por desecación depende de diversos factores.

Estos factores incluyen las propiedades de los componentes, proporciones de los componentes, manera de mezclarlos, cantidad de humedad mientras se cura, entorno en el secado, y dimensiones del elemento. La contracción por desecación ocurre principalmente debido a la reducción de agua capilar por evaporación y el agua en la pasta de cemento. Cuando más alta sea la cantidad de agua en el cemento fresco, mayores son los efectos de la contracción por desecación.

La influencia de las propiedades del concreto sobre la contracción por desecación depende de la proporción de agua y contenido de los materiales cementosos, contenido de agregado, y contenido total de agua. El contenido total de agua es el más importante de éstos. La relación entre la cantidad del contenido de agua del concreto fresco y la contracción por desecación es lineal. El aumento del contenido de agua en uno por ciento aumentará aproximadamente en tres por ciento la contracción por desecación.

Fuente: http://www.engr.psu.edu/ce/concrete_clinic/expansionscontractions/dryshrinkage/dryingshrinkage.htm

Los datos siguientes se obtuvieron del sitio web donde se publicó el artículo precedente.

Debido a que éstos son datos bivariados, nuestra primera consideración es bosquejar un diagrama de dispersión. El contenido de agua será la variable independiente y se grafica en todo el eje x; la contracción será la variable dependiente y se grafica en todo el eje y.

Nótese que las dos variables parecen tener una relación lineal, como se explica en el artículo. Hay un patrón prolongado desde la esquina inferior izquierda a la esquina superior derecha del diagrama de dispersión. A medida que se aumentó el contenido de agua, también aumentó la contracción.

TABLA 3.15

Relación entre contenido de agua y contracción por desecación

Contenido de agua (kg/m)^3	Contracción (10^{-6})	Contenido de agua (kg/m)^3	Contracción (10^{-6})	Contenido de agua (kg/m)^3	Contracción (10^{-6})
202	380	178	280	187	340
210	360	178	420	191	340
220	400	193	350	210	450
231	390	204	440	225	460
242	580	220	450	232	350
167	255	166	240		

Si se calcula la recta de mejor ajuste, podrían hacerse predicciones de contracción con base en el contenido de agua. La recta de mejor ajuste es $y = -166.4 + 2.69x$.

Veamos la ordenada en el origen y pendiente para ver qué significa cada una respecto a esta información de contracción de concreto. La ordenada en el origen corresponde a $x = 0$. En este caso, $x = 0$ significa que no hay contenido de agua y el "concreto" todavía es cemento seco sin mezcla. Considere la ordenada en el origen, −166.4, como que es el valor en el eje y que localiza (verticalmente) la recta de mejor ajuste para que pase por los puntos de datos. La pendiente, 2.69, en este ejemplo es la cantidad de contracción, $2.69 \times 10^{-6} = 0.00000269$, por una unidad de aumento en contenido de agua. Esto indica que por cada kilogramo adicional de agua por metro cúbico de concreto, hay un aumento de 0.00000269 unidades de contracción. Para ver esto en un diagrama de dispersión, use las mismas unidades empleadas para cálculos [sin hacer caso del factor (10^{-6}); entonces, por cada unidad de aumento horizontal, deben verse 2.69 unidades de cambio verticalmente. Esto podría ser fácil de ver si la relación 2.69 a 1 se rescribe como 26.9/10. Esto ahora nos dice que por cada 10 kilogramos adicionales de agua/metro cúbico de mezcla, hay un aumento de casi 27 unidades de contracción.

SECCIÓN 3.4 EJERCICIOS

3.52 Trace un diagrama de dispersión para estos datos:

x	1	2.5	3	4	5	1.5
y	1.5	2.2	3.5	3	4	2.5

¿Se justifica usar las técnicas de regresión lineal sobre estos datos para hallar la recta de mejor ajuste? Explique.

3.53 Trace un diagrama de dispersión para estos datos:

x	2	12	4	6	9	4	11	3	10	11	3	1	13	12	14	7	2	8
y	4	8	10	9	10	8	8	5	10	9	8	3	9	8	8	11	6	9

¿Se justifica usar las técnicas de regresión lineal sobre estos datos para hallar la recta de mejor ajuste? Explique.

3.54 Se seleccionaron al azar 24 países de la lista de 2004 del *The World Factbook*. Se recolectaron datos acerca del porcentaje de cada país que está debajo de la línea de pobreza y las expectativas de vida en esos países.

País	Debajo de línea de pobreza	Expectativa de vida
Afganistan	23.0	42.46
Albania	30.0	77.06

Fuente: The World Factbook, 2004, http://www.cia.gov/cia/publications/

a. Construya un diagrama de dispersión del porcentaje de países debajo de la línea de pobreza, x, y la expectativa de vida, y.

b. ¿Parece que estas dos variables están correlacionadas?

c. ¿Se justifica el uso de técnicas de regresión lineal en estos datos? Explique.

3.55 Las fórmulas para hallar la pendiente y la ordenada en el origen de la recta de mejor ajuste usan sumatorias, Σ, y sumas de cuadrados, SS( ). Es importante saber la diferencia. Respecto al ejemplo 3.5 (p. 164):

a. Encuentre tres pares de valores: Σx^2, SS(x); Σy^2, SS(y), y Σxy, SS(xy).

b. Explique la diferencia entre los números para cada par de números.

3.56 Demuestre que la fórmula (3.7a) es equivalente a la fórmula (3.7) (p. 175).

3.57 Los valores de x empleados para hallar puntos para graficar la recta $\hat{y} = 14.9 + 0.66x$ en la figura 3.21 (p. 177) son arbitrarios. Suponga que se eligieron $x = 20$ y $x = 50$.

a. ¿Cuáles son los valores $\hat{y}$ correspondientes?

b. Localice estos dos puntos en la figura 3.21. ¿Están estos puntos en la recta de mejor ajuste? Explique por qué sí o por qué no.

3.58 ¿Es útil estudiar para un examen? El número de horas estudiado, x, se compara con la calificación de examen recibida, y:

x	2	5	1	4	2
y	80	80	70	90	60

a. Encuentre la ecuación para la recta de mejor ajuste.

b. Trace la recta de mejor ajuste en el diagrama de dispersión de los datos obtenidos en el ejercicio 3.15 (p. 158).

c. Con base en lo que se ve en respuestas a las partes a y b, ¿da resultado estudiar para un examen? Explique.

3.59 ¿Cuál es la relación entre carbohidratos consumidos y energía liberada en una bebida deportiva? ¿Alguna vez se ha preguntado si hay una relación? Usemos la bebida deportiva del ejercicio 3.43 en la página 171 para investigar la relación.

a. En el ejercicio 3.43 se trazó un diagrama de dispersión usando x = carbohidratos/(porción) y y = energía/(porción). Revise el diagrama de dispersión (si no lo trazó antes, hágalo ahora), y describa por qué piensa usted que hay o no hay una relación lineal.

b. Encuentre la ecuación para la recta de mejor ajuste.

c. Usando la ecuación hallada en la parte b, estime la cantidad de energía que uno puede esperar ganar al consumir 40 gramos de carbohidratos.

d. Usando la ecuación hallada en la parte b, estime la cantidad de energía que uno puede esperar ganar al consumir 65 gramos de carbohidratos.

3.60 Un estudiante utilizó regresión lineal para ayudarse a entender su cuenta telefónica mensual. La recta de mejor ajuste fue $\hat{y} = 23.65 + 1.28x$, donde x es el número de llamadas de larga distancia hechas durante un mes, y y es el costo total de la cuenta telefónica para un mes. En términos del número de llamadas de larga distancia y costo:

a. Explique el significado de la ordenada en el origen, 23.65.

b. Explique el significado de la pendiente, 1.28.

3.61 Para el ejemplo 3.7 (p. 177) y el diagrama de dispersión de la figura 3.23 de la página 179:

a. Explique cómo puede verse la pendiente de 4.71.

b. Explique por qué la ordenada en el origen de −186.5 no puede verse.

3.62 Si a todos los estudiantes del curso de educación física de Mr. Chamberlain, de las páginas 153 y 164, que pueden hacer 40 lagartijas se les pide hacer tantas sentadillas como sea posible:

a. ¿Cuántas sentadillas espera usted que haga cada uno?

b. ¿Podrán todos hacer el mismo número?

c. Explique el significado de la respuesta a la parte a.

3.63 Se llevó a cabo un estudio para investigar la relación entre el precio de reventa, y (en cientos de dólares), y la antigüedad, x (en años), de automóviles semicompactos de lujo fabricados en Estados Unidos. Se determinó que la ecuación de la recta de mejor ajuste fue $\hat{y} = 185.7 - 21.52x$.

a. Encuentre el valor de reventa de uno de estos autos cuando tiene 3 años de uso.

b. Encuentre el valor de reventa de uno de estos autos cuando tiene 6 años de uso.

c. ¿Cuál es el promedio anual de disminución del precio de reventa de estos autos?

3.64 Se efectuó un estudio para investigar la relación entre el costo, y (en decenas de miles de dólares), por unidad de equipo manufacturado y el número de unidades producidas por lote, x. La ecuación resultante para la recta de mejor ajuste fue $\hat{y} = 7.31 - 0.01x$, con x siendo observada para valores entre 10 y 200. Si un lote de producción se programó para producir 50 unidades, ¿qué costo por unidad se pronosticaría?

3.65 La Federal Highway Administration anualmente informa sobre impuestos estatales del combustible para automotores. Con base en el último reporte, en miles de dólares, la cantidad de recibos se puede estimar usando la ecuación: Recibos = −5359 + 0.9956 recolecciones.

a. Si un estado recolectó $500 000, ¿de cuánto estima usted que serán los recibos?

b. Si un estado recolectó $1 000 000, ¿de cuánto estima usted que serán los recibos?

c. Si un estado recolectó $1 500 000, ¿de cuánto estima usted que serán los recibos?

3.66 Se terminó un estudio de los hábitos de dar propina de clientes de un restaurante. Los datos para dos de las variables, x, la cantidad de la cuenta del restaurante, y y, la cantidad dejada como propina para quienes les sirvieronse utilizaron para construir un diagrama de dispersión. ¿Qué se espera que revele el diagrama de dispersión?

a. ¿Se espera que las dos variables muestren una relación lineal? Explique.

b. ¿Qué sugerirá el diagrama de dispersión acerca de una correlación lineal? Explique.

c. ¿Qué valor se espera para la pendiente de la recta de mejor ajuste? Explique.

d. ¿Qué valor se espera para la ordenada en el origen de la recta de mejor ajuste? Explique.

Los datos se usan para determinar la ecuación para la recta de mejor ajuste: $\hat{y} = 0.02 + 0.177x$.

e. ¿Qué representa la pendiente de esta recta cuando se aplica a esta situación? ¿Tiene sentido un valor de 0.177? Explique.

f. ¿Qué representa la ordenada en el origen de esta recta cuando se aplica a esta situación? ¿Tiene sentido un valor de 0.02? Explique.

g. Si la cuenta del restaurante fue de $30, ¿qué pronosticaría la recta de mejor ajuste para la propina?

h. Usando la recta de mejor ajuste, pronostique la propina para una cuenta de $31. ¿Cuál es la diferencia entre esta cantidad y la de $30 de la parte g? ¿Tiene sentido esta diferencia? ¿En dónde la ve en la ecuación para la recta de mejor ajuste?

3.67 Considere la figura 3.24 de la página 182. La ordenada en el origen de la gráfica es −250, no aproximadamente 80, como podría leerse de la figura. Explique por qué.

3.68 La rapidez de pasos (número de pasos por segundo) es importante para el corredor serio. La rapidez de pasos está estrechamente relacionada con la velocidad, y la meta de un corredor es alcanzar la óptima rapidez de pasos. Como parte de un estudio, investigadores midieron la rapidez de pasos a siete diferentes velocidades para 21 de las mejores corredoras. Los valores del promedio de rapidez de pasos para estas mujeres y las velocidades de la prueba se indican en la tabla siguiente.

Velocidad, x (ft/sec)	15.86	16.88	17.50	18.62	19.97	21.06	22.11
Rapidez de pasos, y	3.05	3.12	3.17	3.25	3.36	3.46	3.55

Fuente: R. C. Nelson, C. M. Brooks, and N. L. Pike, Comparación biomecánica de de corredores hombres y mujeres, en P. Milvy (ed.), *The Marathon: Physiological, Medical, Epistemological, and Psychological Studies* (Academia de Ciencias de Nueva York, 1977), pp. 793-807

a. Construya un diagrama de dispersión.

b. ¿Parece ser lineal la relación entre las dos variables?

c. Encuentre la ecuación de la recta de mejor ajuste.

d. Interprete la pendiente de la ecuación de la parte c. En otras palabras, ¿cuáles son las "unidades" de la pendiente?

e. Trace la recta de mejor ajuste en el diagrama de dispersión.

f. Usando la recta trazada en la parte e, pronostique la rapidez de pasos promedio si la velocidad es 19 pies por segundo.

g. ¿Cuál es la rapidez de pasos si la velocidad es cero? Interprete sus resultados. ¿Tienen sentido los resultados? Explique.

PARA SU INFORMACIÓN ¿Todavía no ha intentado usar comandos de computadora o calculadora?

3.69 Considere los datos de muchachas universitarias del ejemplo 3.7 y la recta de mejor ajuste. Cuando estime la recta de mejor ajuste a partir de un diagrama de dispersión, la selección para los dos puntos (x_1, y_1) y (x_2, y_2) a usarse es un tanto arbitraria. Cuando se usan puntos diferentes, resultarán valores ligeramente diferentes para b_0 y b_1, pero deben ser casi iguales.

a. ¿Qué puntos en el diagrama de dispersión (figura 3.24, p. 182) se usaron para estimar la pendiente y ordenada en el origen del ejemplo de la página 182? ¿Cuáles fueron las estimaciones resultantes?

b. Utilice los puntos (61, 95) y (67, 130) y encuentre los valores aproximados de la pendiente y ordenada en el origen.

c. Compare los valores hallados en la parte b con los descritos en la parte a. ¿Qué tan semejantes en valor son éstos?

d. Compare ambos conjuntos de estimaciones contra los valores reales de pendiente y ordenada en el origen encontrados en el ejemplo 3.7 en las páginas 177-179. Trace ambas rectas estimadas de mejor ajuste en el diagrama de dispersión de la figura 3.23. ¿Qué tan útiles piensa usted que podrían ser los valores estimados? Explique.

3.70 Los autos de lujo son muy bonitos cuando uno puede darse el lujo de tenerlos. Si no es así, quizá se debe considerar un modelo más asequible que "se sienta como de lujo", según Mitch McCullough, que evalúa de 60 a 70 vehículos anualmente como editor de New Car Test Drive (http://www.nctd.com).

Modelos de lujo	Costo ($1000)
Mercedes-Benz SLK320	46
Chevrolet Corvette	45
BMW 330i	35
Lexus ES 330	32
Lincoln Town Car	42
Lexus RX 330 SUV	36
Lincoln Aviator SUV	41
Porsche Cayenne S SUV	56
Land Rover Range Rover SUV	73
Cadillac Escalade SUV	53
Alternativa "semejante"	**Costo ($1000)**
Chrysler Crossfire	34
Nissan 350Z	27
Infiniti G35	30
Hyundai XG350L	26
Mercury Grand Marquis LS	30
Nissan Murano	28
Ford Explorer Eddie Bauer	34
Infiniti FX35	34
Volkswagen Touareg V8	43
GMC Yukon	36

Fuente: Reader's Digest, junio 2004

a. ¿Se espera que las dos variables muestren una relación lineal? Explique.

b. Construya un diagrama de dispersión usando x = costo de lujo y y = "semejante".

c. ¿Parece haber un patrón lineal? Explique.

d. Calcule la ecuación para la recta de mejor ajuste.

e. Use la ecuación hallada en la parte d para estimar el costo de un vehículo "semejante" comparable a uno de lujo de $40 000. Explique el significado de su respuesta.

f. Use la ecuación hallada en la parte d para estimar el costo de un vehículo "semejante" comparable a uno de lujo de $60 000. Explique el significado de su respuesta.

3.71 Los golfistas profesionales tienen un dilema clásico en golf: "haz un tiro largo para exhibirte, uno corto para ganar dinero". Es frecuente que el juego en corto (en el "green") lo que determina si ganan un torneo. El 7 de enero de 2005, en un artículo de *USA Today* titulado "En corto, la meta de Durant es mejorar", se publicó

una tabla que indicaba los porcentajes de victorias para los jugadores del torneo PGA de golfistas profesionales en la temporada de 2004, para llegar a los "greens" desde varias distancias.

Yardas	Distancia media	Porcentaje de victorias
Más o menos 200	213	44
176–200	188	53
151–175	163	61
126–150	138	68
101–125	113	72
76–100	88	78
75 o menos	63	85

Fuente: PGA Tour Shotlink

Usando las distancias medias en yardas como variable independiente, x, y el porcentaje de victorias como la variable dependiente, y:

a. Construya un diagrama de dispersión.

b. ¿Parece haber una correlación lineal? Justifique su respuesta.

c. Calcule el coeficiente de correlación lineal, r.

d. Interprete el coeficiente de correlación hallado en la parte c. Comente sobre su dirección y fuerza.

e. ¿Parece haber una relación lineal? Justifique su respuesta.

f. Calcule la ecuación de la recta de mejor ajuste.

g. Grafique la recta de mejor ajuste sobre el diagrama de dispersión.

h. Pronostique el porcentaje promedio de victorias para un golfista profesional si llegó hasta el "green" desde una distancia de 90 yardas.

3.72 Los siguientes datos son una muestra de las edades, x (años), y los precios, y ($\times$\$1000), para autos usados Honda Accord que se publicaron en AutoTrader.com el 10 de marzo, 2005:

x	y	x	y	x	y
3	24.9	7	11.9	6	16.4
7	9.0	6	15.2	4	21.2
5	17.8	2	25.9	3	24.9
4	29.2	2	26.9	5	20.0
6	15.7	4	23.8	7	13.6
3	24.9	5	19.3	5	18.8
2	25.7	4	21.9		

Fuente: http://autotrader.com

a. Trace un diagrama de dispersión.

b. Calcule la ecuación de la recta de mejor ajuste.

c. Grafique la recta de mejor ajuste sobre el diagrama de dispersión.

d. Pronostique el precio pedido promedio para todos los autos Honda Accord de 5 años de uso. Obtenga esta respuesta en dos formas: usando la ecuación de la parte b y usando la recta trazada en la parte c.

e. ¿Puede pensarse en algunas variables ocultas potenciales para esta situación? Explique cualquier papel posible que pudieran desempeñar.

3.73 Los equipos de béisbol ganan y pierden juegos. Muchos aficionados piensan que el promedio de carreras limpias admitidas (ERA) de un equipo tiene un efecto importante en ganar juegos. Durante la temporada de 2004, los 30 equipos de béisbol de las ligas mayores registraron el siguiente número de juegos ganados al tiempo que generaron estos promedios de ERA.

Juegos ganados	ERA	Juegos ganados	ERA	Juegos ganados	ERA
96	3.74	98	4.18	63	4.76
105	3.75	67	4.24	80	4.81
89	3.81	92	4.28	70	4.81
93	4.01	72	4.29	83	4.91
92	4.03	91	4.29	67	4.91
87	4.03	67	4.33	72	4.93
92	4.05	86	4.45	51	4.98
71	4.09	89	4.53	58	5.15
83	4.10	101	4.69	76	5.19
91	4.17	78	4.70	68	5.54

Fuente: http://mlb.mlb.com

a. ¿Piensa usted que los equipos con mejores porcentajes ERA tienen más juegos ganados? (Cuando menor ERA, menos carreras anotaba el otro equipo.)

b. Si esto es cierto, ¿qué aspecto tendrá el patrón sobre el diagrama de dispersión? Sea específico.

c. Construya un diagrama de dispersión de estos datos.

d. ¿El diagrama de dispersión sugiere que los equipos tienden a ganar más juegos cuando su porcentaje de carreras admitidas (ERA) es más bajo? Explique.

e. Calcule la recta de mejor ajuste usando x = ERA y y = número de juegos ganados.

f. En promedio, ¿cómo resulta afectado el número de juegos ganados por un aumento de 1 en el ERA? Explique cómo determinó este número.

g. Sus conclusiones ¿parecen apoyar la idea de que los equipos con mejores porcentajes ERA tendrán más juegos ganados? Justifique su respuesta.

3.74 La autopista Interestatal 90, la más larga de este a oeste en Estados Unidos, mide 3112 millas de largo, de Boston, Massachusetts, a la I-93 en la parte oriental de Seattle, Washington, en la costa del Pacífico. Cruza 13 estados del norte; el número de millas y número de cruceros en cada uno de estos estados aparecen a continuación.

Estado	WA	ID	MT	WY	SD	MN	WI
Núm. de cruceros	57	15	83	23	61	52	40
Millas	298	73	558	207	412	275	188

Estado	IL	IN	OH	PA	NY	MA
Núm. de cruceros	19	21	40	14	48	18
Millas	103	157	244	47	391	159

Fuente: Rand McNally and http://www.ihoz.com/I90.html

a. Construya un diagrama de dispersión.

b. Encuentre la ecuación de la recta de mejor ajuste usando x = millas y y = cruceros.

c. Usando la ecuación encontrada en la parte b, estime el número promedio de cruceros por milla en toda la I-90.

3.75 Al observar las primas de seguro que aparecen en el caso práctico 3.6, se ve fácilmente que los hombres pagan primas de seguro más altas que las mujeres de la misma edad. ¿Hay un patrón consistente para estas primas más altas? Para descubrir cualquier patrón en las primas de seguro de $250 000 para hombre/(mujer) del caso práctico 3.6:

a. Trace un diagrama de dispersión de las primas de seguros para hombres (y) contra mujeres (x). ¿El diagrama muestra una relación lineal? Explique.

b. Calcule el coeficiente de correlación lineal, r, para las variables. ¿Hay una fuerte relación lineal? ¿Por qué?

c. Calcule la ecuación de la recta de mejor ajuste.

d. Pronostique la prima mensual para un hombre que es de la misma edad que una mujer cuya prima mensual es $15.00.

e. Con base en las respuestas precedentes, ¿qué conclusión se puede sacar acerca de la relación entre estas primas de seguro para hombres y mujeres? ¿Qué papel desempeña la pendiente de la recta de mejor ajuste al describir la relación?

3.76 El éxito de un golfista profesional puede medirse en varias formas. El resultado final es probablemente cuánto dinero gana un golfista en un año determinado, pero a los golfistas también se les asigna un lugar mundial por puntos por cada evento en el que participen. A continuación está una tabla combinada, extractada del sitio web del torneo PGA (http://www.pgatour.com), donde aparecen los 20 jugadores que ganaron más dinero en la temporada de 2004, así como su clasificación mundial al finalizar 2004, número de eventos donde jugaron, cantidad de dinero que ganaron, y promedio de puntos por evento.

Lugar	Jugador	Eventos en 2004	Dinero ganado en 2004	Promedio de puntos*
1	Vijay Singh	29	10,905,166	12.97
2	Tiger Woods	19	5,365,472	11.90

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/johnson

*Clasificación Oficial Mundial de Golf. Esta estadística es el número promedio de puntos ganados por evento en las últimas 104 semanas. Estos puntos se otorgan con base en la posición final de un torneo, así como el grado de dificultad de éste. Los puntos valen inicialmente el doble de su valor original y se reduce en forma gradual en este periodo de 2 años. Hay ocho periodos de 13 semanas, y los puntos bajan en 0.25x sobre su valor en cada periodo.

Fuente: PGA TOUR, Inc.

a. Trace un diagrama de dispersión con "dinero 2004" como la variable dependiente, y, y "lugar" Como la variable de pronóstico, x.

b. El diagrama de dispersión de la parte a ¿sugiere que será útil una regresión lineal? Explique.

c. Calcule la ecuación de mejor ajuste.

d. Trace la recta de mejor ajuste en el diagrama de dispersión que obtuvo en la parte a. Explique el papel de una pendiente negativa para este par de variables.

e. ¿Ve usted una potencial variable oculta? Explique su posible papel.

f. Trace un diagrama de dispersión con "dinero 2004" como la variable dependiente, y, y "promedio de puntos" como la variable de pronóstico, x.

g. El diagrama de dispersión de la parte f ¿sugiere que será útil una regresión lineal? Explique.

h. Calcule la ecuación de mejor ajuste.

i. Trace la recta de mejor ajuste en el diagrama de dispersión que obtuvo en la parte f.

j. ¿Ve usted una potencial variable oculta? Explique su posible papel.

k. Trace un diagrama de dispersión con "dinero 2004" como la variable dependiente, y, y "eventos 2004" como la variable de pronóstico, x.

l. El diagrama de dispersión de la parte k ¿sugiere que será útil una regresión lineal? Explique.

m. Calcule la ecuación de mejor ajuste.

n. Trace la recta de mejor ajuste sobre el diagrama de dispersión que obtuvo en la parte k.

o. La línea de mejor ajuste ¿es útil para pronosticar el dinero de 2004 con base en el número de eventos jugados en 2004? Explique.

3.77 La Office of Aviation Enforcement & Proceedings, U.S.Department of Transportation, informó acerca del número de quejas que pasajeros de líneas aéreas (en miles) presentaron por mal manejo de equipaje durante octubre de 2004. El promedio de la industria fue 4.02 quejas por 1000 pasajeros.

Línea aérea	Quejas	Pasajeros
AirTran	2,084	1148.8
JetBlue	2,295	1057.5

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/jonhson

Fuente: Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation

a. Trace un diagrama de dispersión con el número de quejas como la variable dependiente, y, y el número de pasajeros (en miles) como la variable de pronóstico, x.

b. El diagrama de dispersión de la parte a ¿sugiere que será útil una regresión lineal? Explique.

c. Calcule la ecuación de mejor ajuste.

d. Trace la recta de mejor ajuste sobre el diagrama de dispersión que obtuvo en la parte a. ¿Qué tan bien se ajusta la recta a los datos del patrón? Explique.

3.78 Con frecuencia se publican clasificaciones Nielsen en periódicos de circulación nacional. El número de televidentes (en millones) de cada programa se da con la correspondiente clasificación Nielsen. Con el número de televidentes como x y la clasificación como y, los datos publicados en *USA Today* (7 de febrero, 2002) para el segmento de 7 a 10 p.m. resultó en un coeficiente de correlación de 0.99 y la ecuación de regresión $\hat{y} = 0.12 + 0.6x$. Explique cómo es que la pendiente y la ordenada en el origen demuestran que el número de televidentes, x, debe tener el más grande impacto en la predicción de la clasificación Nielsen.

3.79 La gráfica siguiente muestra la relación entre tres variables: número de conductores con licencia, número de vehículos registrados, y el tamaño de la población residente para Estados Unidos de 1961 a 2003. Estudie la gráfica y conteste estas preguntas:

a. ¿Parece razonable que la recta de Población y la recta de Conductores corran casi paralelas entre sí y que la recta de Población esté arriba de la de Conductores? Explique qué significa para ellas que sean paralelas. ¿Qué significaría si no fueran paralelas?

b. Las rectas de Conductores y de Vehículos a Motor se cruzan. ¿Qué significa esto? ¿Cuándo se cruzan las rectas, y qué representa el punto de intersección?

c. Explique la relación entre vehículos de motor y conductores antes de 1973.

Fuente: U.S. Dept. of Transportation–Federal Highway Administration

d. Explique la relación entre vehículos a motor y conductores después de 1973.

e. ¿Pronostica el lector que los conductores rebasarán a vehículos a motor después de 2003? ¿Por qué sí o por qué no?

f. Usando los años 1982 y 2000, estime las pendientes de la recta de Vehículos a Motor y la de Conductores. Compare y contraste las pendientes halladas.

3.80 El coeficiente de correlación y la pendiente de la recta de mejor ajuste están relacionadas por definición.

a. Verifique este enunciado.

b. Describa en qué forma la relación entre coeficiente de correlación y pendiente se puede ver en las estadísticas que describen un conjunto particular de datos.

c. Demuestre que $b_1 = r(s_y/s_x)$. Comente sobre esta relación.

REPASO DEL CAPÍTULO

En retrospectiva

Para resumir lo que acabamos de aprender: hay una diferencia distintiva entre el propósito del análisis de regresión y el propósito de correlación. En análisis de regresión, buscamos una relación entre las variables. La ecuación que representa esta relación puede ser la respuesta que se desea, o puede ser el medio para la predicción que se desea. En análisis de correlación, medimos la fuerza de la relación lineal entre dos variables.

Los casos prácticos del texto muestran una variedad de usos para las técnicas de correlación y regresión. Estos ejemplos merecen la pena de ser leídos otra vez. Cuando parece que los datos bivariados caen en toda una recta en el diagrama de dispersión, sugieren una relación lineal. Pero esto no es prueba de causa y efecto. Evidentemente, si un jugador de baloncesto comete demasiadas faltas personales, no estará anotando más puntos. Los jugadores en problemas de faltas están "arriesgándose al fracaso" sin probabilidades de anotación. También parece razonable que cuanto más tiempo de juego tengan, más puntos anotarán y más faltas cometerán. Así, existirán una correlación positiva y una relación de regresión positiva entre estas dos variables. En este caso, el tiempo es una variable oculta.

Los métodos lineales bivariados que hemos estudiado hasta aquí se han presentado como una primera y descriptiva mirada. Por necesidad, más detalles deben esperar hasta hacer más trabajo de desarrollo. Después de completar este capítulo, el estudiante debe tener una comprensión básica de datos bivariados, cómo se diferencian de sólo dos conjuntos de datos, cómo presentarlo, qué son correlación y análisis de regresión y cómo se usan cada uno de ellos.

Vocabulario y conceptos clave

datos bivariados (p. 146)
relación de causa y efecto (p. 167)
coeficiente de correlación lineal (p. 163)
tabla de contingencia (p. 147)
correlación (p. 162)
análisis de correlación (p. 162)
tabulación cruzada (p. 147)
variable dependiente (pp. 152, 162)
variable independiente (pp. 152, 162)
variable de entrada (p. 152)
criterio de mínimos cuadrados (p. 174)
recta de mejor ajuste (p. 174)
correlación lineal (p. 162)
regresión lineal (p. 173)
variable oculta (p. 167)
método de mínimos cuadrados (p. 174)
correlación negativa (p. 163)
par ordenado (p. 152)
variable de salida (p. 152)
momento de producto de Pearson, r, (p. 163)
correlación positiva (p. 163)
valor pronosticado (p. 174)
ecuación de predicción (p. 174)
regresión (p. 173)
análisis de regresión (p. 174)
diagrama de dispersión (p. 153)
pendiente, b_1 (p. 175)
ordenada en el origen, b_0 (p. 175)

Objetivos de aprendizaje

- ✓ Entender y tener aptitud de presentar y describir datos en la forma de dos variables cualitativas, en formato de tabla de contingencia y de gráficas apropiadas. — EJ. 3.1, pp. 146-149, Ejer. 3.83
- ✓ Comprender y tener aptitud de presentar y describir datos en la forma de una variable cualitativa y una variable cuantitativa, en formato de tabla y gráficas apropiadas. — EJ. 3.2, pp. 151-152, Ejer. 3.09, 3.10
- ✓ Entender y ser capaz de presentar y describir la relación entre dos variables cuantitativas usando un diagrama de dispersión. — EJ. 3.3, AP EJ. 3.4, pp. 152-155, Ejer. 3.15
- ✓ Entender y ser capaz de explicar una relación lineal. — pp. 162-163
- ✓ Calcular, describir e interpretar un coeficiente de correlación. — pp. 162-164, EJ. 3.5, Ejer. 3.33
- ✓ Calcular, describir e interpretar una recta de mejor ajuste. — EJ. 3.7
- ✓ Definir y entender la diferencia entre correlación y causalidad. — pp. 167-168, Ejer. 3.50, 3.51
- ✓ Determinar y explicar posibles variables ocultas y sus efectos en una relación lineal. — pp. 167-168, Ejer. 3.50, 3.51
- ✓ Entender y ser capaz de explicar la pendiente de la recta de mejor ajuste respecto al contexto en que se presenta. — Ejer. 3.60, 3.66
- ✓ Entender y ser capaz de explicar la ordenada en el origen de la recta de mejor ajuste respecto al contexto en que se presenta. — Ejer. 3.60, 3.66
- ✓ Crear un diagrama de dispersión con la recta de mejor ajuste trazada sobre él. — Ejer. 3.58
- ✓ Calcular valores de predicción basados en la recta de mejor ajuste. — pp. 179-180, Ejer. 3.63
- ✓ Entender y ser capaz de explicar lo que son valores de predicción. — pp. 179-180, 173-175
- ✓ Entender que las predicciones deben hacer sólo para valores dentro del dominio muestral y que debe tenerse cuidado para valores fuera de ese dominio. — pp. 179-180

Ejercicios del capítulo

3.81 El temor al dentista (o al sillón del dentista) es una emoción sentida por muchas personas de todas las edades. Se realizó una encuesta a 100 personas de cinco grupos de edades acerca de este temor, y éstos fueron los resultados:

	Elemental	Secundaria	Preparatoria	Universidad	Adulto
Temor	37	28	25	27	21
Sin temor	63	72	75	73	79

a. Encuentre los totales marginales.

b. Exprese las frecuencias como porcentajes del gran total.

c. Exprese las frecuencias como porcentajes de los totales marginales de cada uno de los grupos de edades.

d. Exprese las frecuencias como porcentajes de aquellos que temen y los que no temen.

e. Trace una gráfica de barras con base en grupos de edades.

3.82 La gráfica siguiente de "escondite de dinero para emergencias" indica en porcentajes las distribuciones para la cantidad que ambos géneros han ahorrado para emergencias.

a. Identifique la población, las variables y el tipo de variables.

b. Construya una gráfica de barras que muestre las dos distribuciones consecutivas.

c. ¿Estas distribuciones parecen diferir para los géneros? Explique.

"ESCONDITE DE DINERO PARA EMERGENCIAS"

Entre trabajadores de 25-64 años, 62% de hombres y 53% de mujeres tienen ahorros apartados para emergencias.

	Hombres	Mujeres
Menos de un mes de ingreso	12%	18%
1 a menos de 3 meses	31%	24%
3 a menos de 6 meses	21%	29%
6 o más meses de ingreso	36%	26%
No sabe	0%	3%

Datos de Anne R. Carey y Grant Jerding, USA TODAY; Fuente: Merryl Linch. © 1998 USA TODAY reimpreso con permiso.

3.83 Seis razas de perros han sido populares en Estados Unidos en los últimos años. La siguiente tabla indica las razas y el número de registros presentados en el American Kennel Club en 2003 y 2004.

Razas	2003	2004
De presa (Labrador)	144 896	146 692
De presa (Dorado)	52 520	52 550
Pastor alemán	43 938	46 046
Pachón	45021	44 555
Yorkshire Terriers	38 246	43 522
Salchicha	39 468	40 770

Fuente: American Kennel Club, http://www.akc.org/reg/dogreg_stats.cfm

a. Se da una tabulación cruzada de las dos variables, año (columnas) y raza de perro (filas). Determine los totales marginales.

b. Exprese la tabla de contingencia de la parte a en porcentajes basados en el gran total.

c. Trace una gráfica de barras que muestre los resultados de la parte b.

d. Exprese la tabla de contingencia de la parte a en porcentajes basados en el total marginal para el año.

e. Trace una gráfica de barras que muestre los resultados de la parte b.

3.84 ¿Cuándo fue la última vez que visitó al médico? Esa pregunta se formuló para la encuesta resumida en la siguiente tabla.

		Fecha de última consulta		
		Menos de 6 meses	6 meses a menos de 1 año	1 año o más
Edad	Menos de 28 años	413	192	295
	28-40	574	208	218
	Mayor de 40	653	288	259

a. Encuentre los totales marginales.

b. Exprese las frecuencias como porcentajes del gran total.

c. Exprese las frecuencias como porcentajes de totales marginales de cada grupo de edades.

d. Exprese las frecuencias como porcentajes de cada periodo.

e. Trace una gráfica de barras con base en el gran total.

3.85 Parte del control de calidad es dar seguimiento a lo que está ocurriendo. La siguiente tabla de contingencia muestra el número de piezas fundidas rechazadas el mes pasado, clasificadas por su causa y el turno de trabajo durante el que ocurrieron.

	Primer turno	Segundo turno	Tercer turno
Con arena	87	110	72
Mala colocación	16	17	4
Alabeo	12	17	16
Corazón roto	18	16	33
Pieza rota	17	12	20
Otros	8	18	22

a. Encuentre los totales marginales.

b. Exprese los números como porcentajes del gran total.

c. Exprese los números como porcentajes del total marginal de cada uno de los turnos.

d. Exprese los números como porcentajes de cada tipo de rechazo.

e. Trace una gráfica de barras con base en los turnos.

3.86 Determine si cada una de las siguientes preguntas requiere análisis de correlación o análisis de regresión para obtener una respuesta.

a. ¿Hay correlación entre las calificaciones que obtiene un estudiante de secundaria y las calificaciones que alcanzó en la universidad?

b. ¿Cuál es la relación entre el peso de un paquete y el costo de enviarlo por correo de primera clase?

c. ¿Hay una correlación lineal entre la estatura y la medida de los zapatos de una persona?

d. ¿Cuál es la relación entre el número de horas-trabajador y el número de unidades de producción completadas?

e. La calificación obtenida en cierta prueba de aptitud ¿está relacionada linealmente con una capacidad de la persona para realizar cierto trabajo?

3.87 La propietaria de un automóvil registra el número de galones de gasolina, x, necesario para llenar el tanque de gasolina y el número de millas recorridas, y, entre llenados del tanque.

a. Si ella hace un análisis de correlación de los datos, ¿cuál sería su propósito y cuál sería la naturaleza de sus resultados?

b. Si ella hace un análisis de regresión de los datos, ¿cuál sería su propósito y cuál sería la naturaleza de sus resultados?

3.88 Estos datos se generaron usando la ecuación $y = 2x + 1$.

x	0	1	2	3	4
y	1	3	5	7	9

Un diagrama de dispersión de los datos resulta en cinco puntos que caen perfectamente en una recta. Encuentre el coeficiente de correlación y la ecuación de la recta de mejor ajuste.

3.89 Considere este conjunto de datos bivariados:

x	1	1	3	3
y	1	3	1	3

a. Trace un diagrama de dispersión.

b. Calcule el coeficiente de correlación.

c. Calcule la recta de mejor ajuste.

3.90 Empiece con el punto (5,5) y sume al menos cuatro pares ordenados, (x, y), para hacer un conjunto de pares ordenados que presente las siguientes propiedades. Demuestre que su muestra satisface los requisitos.

a. La correlación de x y y es 0.0.

b. La correlación de x y y es + 1.0.

c. La correlación de x y y es − 1.0.

d. La correlación de x y y es entre −0.2 y 0.0.

e. La correlación de x y y es entre + 0.5 y + 0.7.

3.91 Se traza un diagrama de dispersión que muestra los datos para x y y, dos variables normalmente distribuidas. Los datos caen dentro de los intervalos $20 \leq x \leq 40$ y $60 \leq 100$. ¿Dónde esperaría usted hallar los datos en el diagrama de dispersión si:

a. el coeficiente de correlación es 0.00

b. el coeficiente de correlación es 0.33

c. el coeficiente de correlación es 0.8

d. el coeficiente de correlación es −0.3

e. el coeficiente de correlación es −0.8

3.92 Empiece con el punto (5,5) y sume al menos cuatro pares ordenados, (x, y), para hacer un conjunto de pares ordenados que presente las siguientes propiedades. Demuestre que su muestra satisface los requisitos.

a. La correlación de x y y es entre + 0.9 y + 1.0, y la pendiente de la recta de mejor ajuste es 0.5.

b. La correlación de x y y es entre + 0.5 y + 0.7, y la pendiente de la recta de mejor ajuste es 0.5.

c. La correlación de x y y es entre −0.7 y −0.9, y la pendiente de la recta de mejor ajuste es −0.5.

d. La correlación de x y y es entre + 0.5 y + 0.7, y la pendiente de la recta de mejor ajuste es −1.0.

3.93 Los equipos de béisbol de las ligas mayores con frecuencia dicen que firman jugadores con capacidad deportiva comparable a los confines de su estadio local, pensando que como la mitad de sus juegos son como locales, esto será una ventaja. Si éste fuera el caso, entonces parecería que los equipos tienden a anotar más carreras en su estadio local que como visitantes. A continuación aparece una lista del número promedio

de carreras anotadas como local (Prom carreras Loc) y el número promedio de carreras anotadas como visitante (Prom carreras Vis).

Prom carreras Loc	Prom carreras Vis
4.83	5.49
6.38	5.33

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/johnson

Source: http://mlb.mlb.com

a. En promedio, ¿los equipos anotan más carreras en su estadio local o como visitantes? ¿Qué piensa usted?

b. Si no hay relación entre x, promedio de carreras como locales, y y, promedio de carreras como visitantes, ¿cuál espera usted que sea el patrón en un diagrama de dispersión?

c. Si tienen una relación, ¿cuál espera usted que sea el patrón en un diagrama de dispersión?

d. Construya un diagrama de dispersión.

e. El diagrama de dispersión ¿parece apoyar sus respuestas a las partes b y c? Explique por qué sí o por qué no.

3.94 Se realizó un estudio biológico de un pececillo llamado albur* de nariz negra. Se registraron la longitud, y (en milímetros), y la edad, x (al año más cercano).

*Visite: http://www.dnr.state.oh.us/dnap/rivfish/bndace.html

x	0	3	2	2	1	3	2	4	1	1
y	25	80	45	40	36	75	50	95	30	15

a. Trace un diagrama de dispersión de estos datos.

b. Calcule el coeficiente de correlación.

c. Encuentre la ecuación de la recta de mejor ajuste.

d. Explique el significado de las respuestas a las partes a-c.

3.95 De la lista de países de *The World Factbook* 2004, se seleccionaron 24 de ellos al azar. Las expectativas de vida para hombres y mujeres se registraron para esos países.

País	Expectativa de vida hombres	Expectativa de vida mujeres
Albania	74.37	80.02
American Samoa	72.05	79.41

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/johnson

Fuente:The World Factbook, 2004, http://www.cia.gov/cia/publications/factbook/geos/ve.html

a. Construya un diagrama de dispersión de la expectativa de vida para hombres, x, y la expectativa de vida para mujeres, y.

b. ¿Parece que estas dos variables están correlacionadas?

c. Encuentre la ecuación de la recta de mejor ajuste?

d. ¿Qué representa el valor numérico de la pendiente?

3.96 El chirriar de grillos es un sonido bienvenido en una noche de verano. De hecho, esos grillos pueden darnos la temperatura. En el libro *The Song of Insects*, George W. Pierce, un maestro de física de Harvard, presentó datos reales que relacionan el número de chirridos por segundo, x, para grillos de franjas y la temperatura en °F, y. La tabla siguiente da datos reales de grillos y temperatura. Parece que el número de chirridos representa un promedio, porque se da al décimo más cercano.

x	y	x	y	x	y
20.0	88.6	15.5	75.2	15.0	79.6
16.0	71.6	14.7	69.7	17.2	82.6
19.8	93.3	17.1	82.0	16.0	80.6
18.4	84.3	15.4	69.4	17.0	83.5
17.1	80.6	16.2	83.3	14.4	76.3

Fuente: George W. Pierce, *The Song of Insects*, Harvard University Press, 1948

a. Trace un diagrama de dispersión del número de chirridos por segundo, x, y la temperatura del aire, y.

b. Describa el patrón mostrado.

c. Encuentre la ecuación para la recta de mejor ajuste.

d. Usando la ecuación de la parte c, encuentre las temperaturas que corresponden a 14 y 20 chirridos, que son los límites aproximados para el dominio del estudio.

e. Para este estudio, ¿parece razonable para este estudio el rango de valores de temperatura limitado por valores de temperatura y hallado en la parte d? Explique.

f. La siguiente vez que salga usted donde haya grillos que chirrían en una noche de verano y se encuentre sin termómetro, sólo cuente el número de chirridos y podrá dar la temperatura. Si la cuenta es de 16, ¿qué temperatura piensa que hay?

3.97 Los lagos son cuerpos de agua rodeados por tierras y pueden incluir mares. La tabla siguiente es una lista de áreas y profundidades máximas de 32 lagos en todo el mundo.

a. Trace un diagrama de dispersión que muestre el área, x, y la profundidad máxima, y, de los lagos.

b. Encuentre el coeficiente de correlación lineal entre área y profundidad máxima. ¿Qué implica el valor de esta correlación lineal?

Lago	Área (millas cuad.)	Prof. máx. (ft)
Mar Caspio	143 244	3 363
Superior	31 700	1 330

••• Los datos para este ejercicio están en: http://latinoamerica.cengage.com/jonhson

3.98 Las poblaciones de fauna silvestre son vigiladas con fotografías aéreas. El número de animales y sus ubicaciones respecto a áreas habitadas por seres humanos son información útil. A veces es posible observar las características físicas de los animales. La longitud de un lagarto puede estimarse en forma bastante precisa desde fotografías aéreas, no así su peso. Los datos siguientes son las longitudes, x (en pulgadas), y pesos, y (en libras), de lagartos capturados en la región central de Florida y pueden usarse para pronosticar el peso de un lagarto con base en su longitud.

Peso	Longitud	Peso	Longitud	Peso	Longitud
130	94	38	72	44	61
51	74	366	128	106	90
640	147	84	85	84	89
28	58	80	82	39	68
80	86	83	86	42	76
110	94	70	88	197	114
33	63	61	72	102	90
90	86	54	74	57	78
36	69				

Fuente: http://exploringdata.cqu.edu.au/stories.htm alligatr

a. Construya un diagrama de dispersión para la longitud, x, y el peso, y.

b. ¿Parece que el peso de un lagarto se puede pronosticar a partir de su longitud? Explique.

c. ¿Es lineal la relación?

d. Explique por qué la recta de mejor ajuste, como se describe en este capítulo, no es adecuada para estimar el peso basado en la longitud.

e. Encuentre el valor del coeficiente de correlación lineal.

f. Explique por qué el valor de r puede ser tan alto para un conjunto de datos que es tan obviamente no lineal en su naturaleza.

3.99 Los productores de caña de azúcar están interesados en la relación entre los acres de tierras cosechadas y la producción total de caña de azúcar (toneladas) de estos acres. Los datos siguientes son para la cosecha de 2001 de 14 condados productores de caña de azúcar seleccionados al azar en Luisiana.

Acres	Producción	Acres	Producción
33 700	940 000	20 200	590 000
15 200	460 000	33 800	1 020 000
14 400	440 000	20 500	585 000
2 300	65 000	33 100	1 020 000
30 200	830 000	8 000	200 000
13 100	380 000	41 100	1 130 000
29 600	860 000	17 900	570 000

Fuente: http://www.usda.gov/nass/graphics/county01/data/

a. Estos valores de datos tienen muchos ceros que aparecen. Cambie acres cosechados a cientos (100) de acres y producción a miles (1000) de toneladas de producción antes de continuar.

b. Construya un diagrama de dispersión de acres cosechados, x, y toneladas de producción, y.

c. ¿Parece lineal la relación entre las variables? Explique.

d. Encuentre la ecuación para la recta de mejor ajuste.

e. ¿Cuál es la pendiente para la recta de mejor ajuste? ¿Qué representa la pendiente? Explique lo que significa para el productor de caña de azúcar.

3.100 Relativamente pocas personas en viaje de trabajo usan sistemas de transporte colectivo cuando visitan grandes ciudades. La recompensa podría ser

sustancial, tanto en tiempo como en dinero, si saben cómo usar esos sistemas, como se hizo notar el 28 de diciembre, 2004, en el artículo de *USA Today* "Mass transit could save business travelers big bucks". *USA Today* reunió la siguiente información acerca de los sistemas ferroviarios de más movimiento en Estados Unidos.

Ciudad	Estaciones	Vehículos	Vías (millas)
Atlanta	38	252	193
Baltimore	14	100	34
Boston	53	408	108
Chicago	144	1190	288
Cleveland	18	60	42
Los Angeles	16	102	34
Miami	22	136	57
New York	468	6333	835
Philadelphia	53	371	102
San Francisco	43	669	246
Washington	86	950	226

Fuente: USA Today, 28 de diciembre, 2004

Suponga que se está promoviendo un sistema de transporte colectivo para una ciudad, donde usted ha recibido el cargo de elaborar información estadística (gráfica y numérica) acerca de la relación entre las siguientes tres variables: número de estaciones, número de vagones, y número de millas de vías. Se le proporcionaron los datos precedentes.

a. Empiece por inspeccionar los datos dados. ¿Observa cualquier cosa poco común acerca de los datos? ¿Hay algún valor que parezca bastante diferente del resto? Explique.

b. Su supervisor sugiere que usted elimine los datos para Nueva York. Haga un caso para que eso sea aceptable. Incluya algunas gráficas preliminares y estadísticas calculadas para justificar la eliminación de estos valores.

Usando los datos de las otras 10 ciudades:

c. Construya un diagrama de dispersión usando millas de vías como la variable independiente, x, y el número de estaciones como la variable dependiente, y.

d. ¿Hay evidencia de una relación lineal entre estas dos variables? Justifique su respuesta.

e. Encuentre la ecuación de la recta de mejor ajuste para la parte c.

f. Interprete el significado de la ecuación para la recta de mejor ajuste. ¿Qué le dice?

g. Construya un diagrama de dispersión usando millas de vías como la variable independiente, x, y el número de vagones como la variable dependiente, y.

h. ¿Hay evidencia de una relación lineal entre estas dos variables? Justifique su respuesta.

i. Encuentre la ecuación de la recta de mejor ajuste para la parte g.

j. Interprete el significado de la ecuación para la recta de mejor ajuste. ¿Qué le dice?

k. Construya un diagrama de dispersión usando el número de estaciones como la variable independiente, x, y el número de vehículos como la variable dependiente, y.

l. ¿Hay evidencia de una relación lineal entre estas dos variables? Justifique su respuesta.

m. Encuentre la ecuación de la recta de mejor ajuste para la parte k.

n. Interprete el significado de la ecuación para la recta de mejor ajuste. ¿Qué le dice?

o. La ciudad está recibiendo propuestas iniciales para un sistema de transporte colectivo de 50 millas de vías. Con base en las respuestas halladas en las partes c-n, ¿cuántas estaciones y cuántos vehículos serán necesarios para el sistema? Justifique sus respuestas.

p. Si alguien desea un estimado para el número de estaciones y vagones necesarios para un sistema de 100 millas, no deben sólo duplicar los resultados en la parte o. Explique por qué no.

q. Con base en las respuestas halladas en las partes c-n, ¿cuántas estaciones y cuántos vagones serán necesarios para un sistema de 100 millas? Justifique sus respuestas.

3.101 Las cigarras son insectos voladores que comen plantas. Una especie particular, las cigarras de 13 años (*Magicicada*), pasa cinco etapas juveniles en guaridas bajo tierra. Durante los 13 años en esos lugares, las cigarras crecen de aproximadamente el tamaño de una hormiga pequeña a casi el tamaño de una cigarra adulta. Cada 13 años, los animales emergen entonces de sus guaridas como adultos. La tabla siguiente presenta tres especies diferentes de estas cigarras de 13 años y

sus correspondientes pesos corporales como adultos (BW), en gramos, y longitud de alas (WL), en milímetros.

Especies	BW	WL	Especies	BW	WL
tredecula	0.15	28	*tredecula*	0.18	29
tredecim	0.29	32	*tredecassini*	0.21	27
tredecim	0.17	27	*tredecula*	0.15	30
tredecula	0.18	30	*tredecula*	0.17	27
tredecim	0.39	35	*tredecassini*	0.13	27
tredecim	0.26	31	*tredecassini*	0.17	29
tredecassini	0.17	29	*tredecassini*	0.23	30
tredecassini	0.16	28	*tredecim*	0.12	22
tredecassini	0.14	25	*tredecula*	0.26	30
tredecassini	0.14	28	*tredecula*	0.19	30
tredecassini	0.28	25	*tredecassini*	0.20	30
tredecim	0.12	28	*tredecula*	0.14	23

Fuente: http://insects/ummz.lsa.umich.edu

a. Construya un diagrama de dispersión de los pesos corporales, x, y las correspondientes longitudes de alas, y. Use un símbolo diferente para representar los pares ordenados para cada especie.

b. Describa lo que muestra el diagrama de dispersión respecto a la relación y especies.

c. Calcule el coeficiente de correlación, r.

d. Encuentre la ecuación par la recta de mejor ajuste.

e. Suponga que el peso corporal de una cigarra es 0.20 gramos. ¿Qué longitud de alas pronosticaría usted? ¿Qué especie piensa usted que podría ser esta cigarra?

3.102 El Old Faithful (Viejo Fiel) del Parque Nacional de Yellowstone ha sido durante mucho tiempo una gran atracción para turistas. Entender la duración de sus erupciones y el tiempo entre éstas es necesario para pronosticar el tiempo para la siguiente erupción. Las variables del conjunto de datos del Old Faithful son como sigue: fecha: se tomó un índice de la fecha de observación (días 1, 2 y 3 se dan aquí); duración: la duración de una erupción del géiser, en minutos; y tiempo entre erupciones: el tiempo hasta la siguiente erupción, en minutos.

Día 1		Día 2		Día 3	
Duración	Interrupción	Duración	Interrupción	Duración	Interrupción
4.4	78	4.3	80	4.5	76
3.9	74	1.7	56	3.9	82
4.0	68	3.9	80	4.3	84
4.0	76	3.7	69	2.3	53
3.5	80	3.1	57	3.8	86
4.1	84	4.0	90	1.9	51
2.3	50	1.8	42	4.6	85
4.7	93	4.1	91	1.8	45
1.7	55	1.8	51	4.7	88
4.9	76	3.2	79	1.8	51
1.7	58	1.9	53	4.6	80
4.6	74	4.6	82	1.9	49
3.4	75	2.0	51	3.5	82

Fuente: http://comp.uark.edu/~jtubbs/Biostat/Labs/Oldfaithful/oldfaithful.html

a. Construya un diagrama de dispersión de las 39 duraciones, x, y tiempo entre erupciones, y. Use un símbolo diferente para representar los pares ordenados para cada día.

b. Describa el patrón exhibido por los 39 pares ordenados.

c. Los datos para los días individuales ¿muestran el mismo patrón entre sí y como el conjunto total de datos?

d. Con base en la información del diagrama de dispersión, si la última erupción del Old Faithful duró 4 minutos, ¿cuánto pronosticaría usted que necesitamos esperar hasta que se inicie la siguiente erupción?

e. Encuentre la recta de mejor ajuste para los datos citados en la tabla.

f. Con base en la recta de mejor ajuste, si la última erupción del Old Faithful duró 4 minutos, ¿cuánto pronosticaría usted que necesitamos esperar hasta que se inicie la siguiente erupción?

g. ¿Qué efecto piensa usted que el patrón distintivo mostrado en el diagrama de dispersión tiene sobre la recta de mejor ajuste? Explique.

h. Compare los resultados encontrados en la parte h contra los resultados de las partes a-g. Analice sus conclusiones.

3.103 a. Verifique, algebraicamente, que la fórmula (3.2) para calcular r es equivalente a la fórmula de definición (3.1).

b. Verifique, algebraicamente, que la fórmula (3.6) es equivalente a la fórmula (3.5).

3.104 Esta ecuación da una relación que existe entre b_1 y r:

$$r = b_1 \sqrt{\frac{SS(x)}{SS(y)}}$$

a. Verifique la ecuación para estos datos:

x	4	3	2	3	0
y	11	8	6	7	4

b. Verifique esta ecuación usando las fórmulas (3.2) y (3.6).

Proyecto del capítulo

El chico ha crecido

Como una forma de evaluar las técnicas estadísticas para datos bivariados que hemos aprendido en este capítulo, regresemos a la sección 3.1 "El chico ha crecido", en la página 145. Para cualquier jugador de baloncesto, el número de puntos anotados por juego y el número de faltas personales cometidas por juego son de interés. ¿Podría existir una relación clara y definida entre estas dos variables, y, si es así, por qué?

Trabajando en el contenido del capítulo 3

3.105

Timberwolves de Minnesota, temporada regular 2003-2004

Jugador	Faltas personales por juego	Puntos por juego	Jugador	Faltas personales por juego	Puntos por juego
Garnett	2.5	24.2	Madsen	2.4	3.6
Cassell	3	19.8	Martin	1.4	3.4
Sprewell	1.2	16.8	McLeod	1.2	2.7
Szczerbiak	1.5	10.2	Goldwire	1.0	2.6
Hudson	1.1	7.5	Miller	1.9	2.5
Hoiberg	1.7	6.7	Johnson	2.4	1.9
Olowokandi	3.2	6.5	Lewis	0.7	1.1
Trent	1.9	5.6	Ebi	0.4	0.8
Hassell	2.5	5			

Fuente: http://sports.espn.go.com/nba/teams

a. Construya un diagrama de dispersión, usando puntos anotados por juego, y, y número de faltas personales cometidas por juego, x. Explique por qué piensa que hay o no hay una relación.

b. ¿Están correlacionadas las dos variables de puntos anotados por juego y el número de faltas personales cometidas por juego? Use el coeficiente de correlación para justificar su respuesta.

c. Exprese la relación entre las dos variables de puntos totales anotados, y, y número de faltas personales cometidas, x, como una ecuación lineal.

d. Usando los resultados de la parte c, si un jugador de los Timberwolves de Minnesota cometió dos faltas en un juego, ¿cuántos puntos esperaría usted que anotara?

e. Si el jugador de la parte d cometió una tercera falta personal, ¿cuántos puntos extra esperaría usted que él anote?

f. ¿Cómo se relaciona la pendiente para la recta de mejor ajuste contra el número de puntos adicionales esperados cuando el jugador comete una falta personal extra?

g. Los resultados precedentes ¿muestran una relación de causa y efecto entre los puntos totales anotados y el número de faltas personales cometidas? Explique.

h. El entrenador de un equipo ¿debe instruir a un jugador para que cometa una falta personal extra para que anote más puntos? Explique.

i. Mencione al menos una posible variable oculta para la situación precedente.

Suponga que la investigación precedente debe expandirse para incluir una variable adicional, "minutos jugados por juego".

j. Describa la relación que piensa usted que existe entre las variables "minutos jugados por juego" y "puntos anotados por juego". Explique por qué.

k. Describa la relación que piensa usted que existe entre las variables "minutos jugados por juego" y "número de faltas personales cometidas por juego". Explique por qué.

l. Podría "minutos jugados por juego" ser una variable oculta para el trabajo completado en las partes a-h? Explique.

3.106 a. La situación descrita en el ejercicio 3.105 sólo ocurrió con los Minnesota Timberwolves durante la temporada regular 2003-2004. Use la Internet (buscar por nombre de equipo) para obtener las estadísticas por equipo en esa temporada, respecto al equipo favorito de usted de baloncesto intercolegial o profesional, o vea al entrenador de un equipo local de secundaria o preparatoria.

b. Conteste las mismas preguntas formuladas en el ejercicio 3.105 para su equipo seleccionado.

c. Analice las diferencias y semejanzas entre los Minnesota Timberwolves y el equipo seleccionado por usted. Considere otras variables ocultas.

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

3.1 El análisis de **correlación** es un método de obtener la ecuación que representa la relación entre dos variables.

3.2 El coeficiente de correlación lineal se emplea para determinar la **ecuación que representa** la relación entre dos variables.

3.3 Un coeficiente de correlación de **cero** significa que las dos variables están perfectamente correlacionadas.

3.4 Siempre que la pendiente de la recta de regresión sea cero, el **coeficiente de correlación** también será cero.

3.5 Cuando r es positiva, b_1 siempre será **negativa.**

3.6 La **pendiente** de la recta de regresión representa la cantidad de cambio que se espera tenga lugar en y cuando x aumenta en una unidad.

3.7 Cuando el valor calculado de r es positivo, el valor calculado de b_1 será **negativo.**

3.8 Los coeficientes de correlación están entre **0 y + 1.**

3.9 El valor que se predice se denomina **variable de entrada.**

3.10 La recta de mejor ajuste se emplea para predecir el **valor promedio de *y*** que se puede esperar ocurra a un valor determinado de x.

SEGUNDA PARTE: Aplicación de conceptos

3.11 Consulte el diagrama de dispersión siguiente.

a. Relacione las descripciones de la columna 2 con los términos de la columna 1.

____población	(a) la potencia para un automóvil
____muestra	(b) todos los autos construidos en 2005 en Estados Unidos
____variable de entrada	(c) el kilometraje EPA para un auto
____variable de salida	(d) los automóviles de 2005 con rendimientos mostrados en el diagrama de dispersión

b. Encuentre el tamaño muestral.

c. ¿Cuál es el mínimo valor reportado para la variable de salida?

d. ¿Cuál es el máximo valor reportado para la variable de entrada?

e. ¿El diagrama de dispersión sugiere un coeficiente de correlación positivo, negativo o cero?

f. ¿Cuáles son las coordenadas del punto Q?

g. ¿La pendiente de la recta de mejor ajuste será positiva, negativa o cero?

h. ¿La ordenada en el origen para la recta de mejor ajuste será positiva, negativa o cero?

3.12 Un grupo de investigación presenta un coeficiente de correlación de dos variables. ¿Qué puede usted concluir de esta información?

3.13 Para los datos bivariados, las extensiones, y los totales indicados en la tabla, encuentre lo siguiente:

a. SS(x)

b. SS(y)

c. SS(xy)

d. El coeficiente de correlación, r

e. La pendiente, b_1

f. La ordenada en el origen, b_0

g. La ecuación de la recta de mejor ajuste

x	y	x^2	xy	y^2
2	6	4	12	36
3	5	9	15	25
3	7	9	21	49
4	7	16	28	49
5	7	25	35	49
5	9	25	45	81
6	8	36	48	64
28	49	124	204	353

TERCERA PARTE: Entendiendo los conceptos

3.14 Se aplicó una prueba para medir la capacidad en matemáticas de los habitantes de cierta población. Algunos de ellos se sorprendieron de hallar que los resultados de su prueba y las medidas de su calzado estaban fuertemente correlacionados. Explique por qué una fuerte correlación positiva no debe ser una sorpresa.

3.15 El estudiante A recolectó un conjunto de datos bivariados y calculó r, el coeficiente de correlación lineal. Su valor fue −1.78. El estudiante A proclamó que no había correlación entre las dos variables porque el valor de r no estaba entre −1.0 y +1.0. El estudiante B alegó que −1.78 era imposible y que sólo valores de r cercanos a cero implicaban que no hay correlación. ¿Quién tiene razón? Justifique su respuesta.

3.16 El coeficiente de correlación lineal, r, es un valor numérico que va de −1.0 a +1.0. Escriba una oración o dos que describa el significado de r para cada uno de estos valores:

a. −0.93

b. +0.89

c. −0.03

d. +0.08

e. −2.3

3.17 Forme un conjunto de tres o más pares ordenados tales que:

a. $r = 0.0$

b. $r = +1.0$

c. $r = -1.0$

d. $b_1 = 0.0$

Trabajando con sus propios datos

Cada semestre, nuevos estudiantes ingresan al entorno universitario y el lector puede preguntarse cómo será el estudiantado este semestre. Como estudiante de estadísticas principiante, acaba de terminar de estudiar tres capítulos de técnicas básicas de técnicas descriptivas; puede usar algunas de estas técnicas para describir algunas características del alumnado de su universidad.

A Información de una sola variable

1. Defina la población a estudiar.
2. Seleccione una variable por definir. (Puede definir su propia variable, o puede usar una de las variables de la tabla siguiente* si no le es posible recolectar sus propios datos. Pida ayuda a su profesor.
3. Recolecte 35 piezas de información para su variable.
4. Construya un diagrama de tallo y hoja de su información. (Asegúrese de aplicarle leyendas.)
5. Calcule el valor de la medida de tendencia central que piense que responde mejor a la pregunta: ¿Cuál es el valor promedio de su variable? Explique por qué seleccionó esta medida.
6. Calcule la media muestral para sus datos (a menos que haya empleado la media de la pregunta 5).
7. Calcule la desviación estándar muestral para sus datos.
8. Encuentre el valore del 85avo percentil, P_{85}.
9. Construya un diagrama gráfico (que no sea de tallo y hoja) que crea que "mejor" muestra sus datos. Explique por qué la gráfica presenta mejor sus datos.
10. Escriba un párrafo de resumen que describa lo que haya encontrado.

B Datos de dos variables (bivariados)

1. Defina la población a estudiar.
2. Seleccione y defina dos variables cuantitativas que producirán información divariada. (Puede definir sus propias variables, o puede usar dos de las variables de la tabla siguiente si no le es posible recolectar sus propios datos. Pida ayuda a su profesor.)
3. Recolecte 15 pares ordenados de datos.
4. Construya un diagrama de dispersión de sus datos. (Asegúrese de aplicar leyendas.)
5. Con el uso de una tabla para ayudar con la organización, calcule las extensiones x^2, xy, y y^2, y las sumas de x, y, x^2, xy y y^2.
6. Calcule el coeficiente de correlación lineal, r.
7. Calcule la ecuación de la recta de mejor ajuste.
8. Trace la recta de mejor ajuste sobre su diagrama de dispersión.
9. Escriba un párrafo de resumen que describa lo que haya encontrado.

*La tabla de datos de la página 203 se recolectó el primer día de clase del último semestre. Puede usarla como fuente de sus datos si no le es posible recolectar sus propios datos.

Variable *A*: género del estudiante (masculino/femenino)

Variable *B:* edad del estudiante en su último cumpleaños

Variable *C:* número de horas crédito terminadas hacia licenciatura

Variable *D:* "¿Tiene trabajo (tiempo completo/parcial)?" (sí/no)

Variable *E:* número de horas trabajadas la semana pasada, si D = sí

Variable *F:* sueldos (antes de impuesto) percibidos la semana pasada, si D = sí

PARA SU INFORMACIÓN

La computadora seleccionará su muestra aleatoria (vea p. 101).

Estudiante	*A*	*B*	*C*	*D*	*E*	*F*
1	M	21	16	No		
2	M	18	0	Sí	10	34
3	F	23	18	Sí	46	206
4	M	17	0	No		
5	M	17	0	Sí	40	157
6	M	40	17	No		
7	M	20	16	Sí	40	300
8	M	18	0	No		
9	F	18	0	Sí	20	70
10	M	29	9	Sí	8	32
11	M	20	22	Sí	38	146
12	M	34	0	Sí	40	340
13	M	19	31	Sí	29	105
14	M	18	0	No		
15	M	20	0	Sí	48	350
16	F	27	3	Sí	40	130
17	M	19	10	Sí	40	202
18	F	18	16	Sí	40	140
19	M	19	4	Sí	6	22
20	F	29	9	No		
21	F	21	0	Sí	20	80
22	F	39	6	No		
23	M	23	34	Sí	42	415
24	F	31	0	Sí	48	325
25	F	22	7	Sí	40	195
26	F	27	75	Sí	20	130
27	F	19	0	No		
28	M	22	20	Sí	40	470
29	F	60	0	Sí	40	390
30	M	25	14	No		
31	F	24	45	No		
32	M	34	4	No		
33	M	29	48	No		
34	M	22	80	Sí	40	336
35	M	21	12	Sí	26	143
36	F	18	0	No		
37	M	18	0	Sí	13	65
38	M	40	64	Sí	40	390
39	F	31	0	Sí	40	200
40	F	32	0	Sí	40	270
41	F	37	0	Sí	24	150
42	F	35	0	Sí	40	350
43	M	21	72	Sí	45	470
44	F	27	0	Sí	40	550
45	F	42	47	Sí	37	300
46	F	41	21	Sí	40	250
47	M	36	0	Sí	40	400
48	M	25	16	Sí	40	480
49	F	18	0	Sí	45	189
50	M	22	0	Sí	40	385
51	F	42	34	Sí	40	244
52	M	25	60	Sí	60	503
53	M	39	32	Sí	40	500
54	M	29	13	Sí	39	375
55	M	19	18	Sí	51	201
56	M	25	0	Sí	48	500
57	F	18	0	No		
58	M	32	68	Sí	44	473
59	F	21	0	No		
60	F	26	0	Sí	40	320
61	M	24	11	Sí	45	330
62	F	19	0	Sí	40	220
63	M	19	0	Sí	10	33
64	F	35	59	Sí	25	88
65	F	24	6	Sí	40	300
66	F	20	33	Sí	40	170
67	F	26	0	Sí	52	300
68	F	17	0	Sí	27	100
69	M	25	18	Sí	41	355
70	M	24	0	No		
71	M	21	0	Sí	30	150
72	M	30	12	Sí	48	555
73	F	19	0	Sí	38	169
74	M	32	45	Sí	40	385
75	M	26	90	Sí	40	340
76	M	20	64	Sí	10	45
77	M	24	0	Sí	30	150
78	M	20	14	No		
79	M	21	70	Sí	40	340
80	F	20	13	Sí	40	206
81	F	33	3	Sí	32	246
82	F	25	68	Sí	40	330
83	F	29	48	Sí	40	525
84	F	40	0	Sí	40	400
85	F	36	3	Sí	40	300
86	F	35	0	Sí	40	280
87	F	28	0	Sí	40	350
88	F	27	9	Sí	40	260
89	F	26	3	Sí	40	240
90	F	23	9	Sí	40	330
91	M	41	3	Sí	23	253
92	M	39	0	Sí	40	110
93	M	21	0	Sí	40	246
94	F	32	0	Sí	40	350
95	F	48	58	Sí	40	714
96	F	26	0	Sí	32	200
97	F	27	0	Sí	40	350
98	F	52	56	Sí	40	390
99	F	34	27	Sí	8	77
100	F	49	3	Sí	24	260

CAPÍTULO

4 Probabilidad

4.1 Estadística y los dulces

4.2 Probabilidad de eventos

4.3 Probabilidad condicional de eventos

4.4 Reglas de probabilidad

4.5 Eventos mutuamente excluyentes

4.6 Eventos independientes

4.7 ¿Existe relación entre eventos mutuamente excluyentes y la independencia?

© Rachel Epstein/The Image Works

4.1 Estadística y los dulces

¿De dónde vienen todos estos dulces tan coloridos?

¿Sabía usted que tienen 21 colores?

¿Sabía usted que la idea para los Dulces Sencillos de Chocolate "M&M's" nació en el "telón de fondo" de la guerra civil española? Cuenta la leyenda que en un viaje a España, Forrest Mars Sr. encontró soldados que comían bolitas de chocolate cubiertas de una capa azucarada dura para evitar que se derritieran. Mr. Mars se inspiró en este concepto y regresó a casa e inventó la receta para los Dulces Sencillos de Chocolate "M&M's".

La clase de estadística había comenzado y el maestro estaba hablando de porcentajes, proporciones y probabilidad, y en qué forma son semejantes pero diferentes. De pronto una estudiante dijo que escuchó que el grupo del semestre anterior hizo una lección usando, y comiendo, chocolates M&M's; ella preguntó si el grupo de este año haría algo semejante. La conversación pronto se enfocó por entero en los chocolates M&M's, sus combinaciones de color y el porcentaje de cada color. A los 24 miembros del grupo se les pidió que calcularan el porcentaje de cada color que ellos pensaban estaba contenido en estas pequeñas bolsas de color café de los Dulces Sencillos de Chocolate M&M's. Se les dijo que habría un premio para la persona cuyo cálculo fuera el más cercano al número real. Cada estudiante escribió los porcentajes y los entregó; a su vez, los estudiantes recibieron una pequeña bolsa café. "Ah, ¡esto es esa lección!". "Sí" dijo el maestro, "y antes que abran esas bolsas, *debemos* tener un plan". Cada estudiante debía contar el número de chocolates M&M's de cada color en su bolsa y anotar las seis cantidades; a continuación podrían determinarse los totales del grupo. En la tabla 4.1 aparece la distribución de cantidades resultante.

TABLA 4.1

Colores de M&M's por cantidad

Color	Cantidad
Café	91
Amarillo	112
Rojo	102
Azul	151
Naranja	137
Verde	99
	692

Los totales del grupo se convirtieron a porcentajes (tabla 4.2), y a cada estudiante se le pidió determinar los seis porcentajes que observaran en su propia bolsa de chocolates M&M's.

La discusión que siguió se centró en la variación que había de una bolsa a la otra, con algunos estudiantes bastante sorprendidos de ver tanta variación. Varias bolsas no tenían nada o sólo una pastilla de un color, y unas pocas bolsas tenían una proporción más bien grande de sólo uno o dos colores. ¿Alguna vez había usted observado algunos de estos extremos cuando abría una bolsa de chocolates M&M's?

TABLA 4.2

Colores de M&M's en porcentajes

Color	Porcentaje
Café	13.2
Amarillo	16.2
Rojo	14.7
Azul	21.8
Naranja	19.8
Verde	14.3
	100.0

Los porcentajes reportados en la tabla 4.2 son los de cada color hallados en esta muestra de 692 bolsas M&M's. Los porcentajes se comportan en forma muy semejante a números de probabilidad, pero la pregunta que se hace en probabilidad es diferente. En la ilustración precedente, estamos tratando la información como datos muestrales y describiendo los resultados que encontramos. Si ahora pensamos en términos de una probabilidad, vamos a dar un giro y tratar todo el conjunto de las 692 bolsas de M&M's como si fuera la lista completa de posibilidades, y hacer preguntas acerca de la semejanza de ciertos eventos cuando se selecciona una bolsa de M&M's de todo el conjunto de 692 bolsas.

Por ejemplo, supongamos que se vacían las 692 bolsas de M&M's en un gran tazón y mezclamos perfectamente los chocolates. Ahora considere la pregunta "Si al azar se selecciona un chocolate del tazón, ¿cuál es la probabilidad de que sea de color naranja?" Esperamos que el lector piense así: seleccionados al azar significa que cada chocolate M&M's tiene la misma probabilidad de ser elegido y, como hay 137 chocolates color naranja en el tazón, la probabilidad de seleccionar un chocolate de color naranja M&M's es 137/692, es decir 0.198

Ya antes hemos visto este número 0.198, sólo que se expresaba como 19.8%. Los porcentajes y los números de probabilidad son "lo mismo, pero diferentes." (Es probable que ya antes y en algún lugar usted haya oído esto.) Los números tienen el mismo valor y se comportan con las mismas propiedades; no obstante, la orientación de la situación y las preguntas hechas son diferentes, como veremos en la sección 4.2.

Después de completar el capítulo 4, tendremos oportunidad de investigar más a fondo "Estadística y los dulces" en la sección del Proyecto del capítulo 4.

SECCIÓN 4.1 EJERCICIOS

4.1 a. Si compró una bolsa de chocolates M&M's, ¿qué color de M&M's esperaría ver más? ¿Qué color menos? ¿Por qué?

b. Si compró una bolsa de chocolates M&M's, ¿esperaría hallar los porcentajes mencionados en la tabla 4.2? Si no es así, ¿por qué y qué esperaría?

4.2 a. Construya una gráfica de barras que muestre los porcentajes de la tabla 4.2 obtenidos a partir de los 692 chocolates M&M's.

b. Con base en su gráfica, ¿qué color de chocolates M&M's hubo con más frecuencia? ¿Cómo se muestra esto en su gráfica?

c. Con base en su gráfica, ¿qué color de chocolates M&M's hubo con menos frecuencia? ¿Cómo se muestra esto en su gráfica?

4.3 Si recibiera una pequeña bolsa de 40 chocolates M&M's, usando los porcentajes de la tabla 4.2 ¿cuántos de cada color "esperaría" encontrar?

4.4 *¿Tablas malas?* Así como hay gráficas malas (como se ve en la sección 2.8), hay tablas malas, es decir, tablas engañosas y difíciles de leer. Un grupo llamado Madres Contra Conductores Borrachos (MADD, por sus siglas en inglés) presentó la siguiente tabla referente a 6764 muertos en accidentes de tránsito que ocurrieron en 2002.

Días festivos 2002	Total muertos en tránsito	Total muertes relacionadas con alcohol
Víspera de año nuevo	118	45
Día de año nuevo	165	94
Días festivos de año nuevo	575	301
Domingo de super tazón	147	86
Día de san Patricio	158	72
Conmemoración de los caídos	491	237
Cuatro de julio	683	330
Fin de semana de día del trabajo	541	300

Halloween	268	109
Día de gracias	543	255
Día de gracias-año nuevo	4019	1561
Navidad	130	68
Víspera de año nuevo (2002)	123	57

Fuente: Mothers Against Drunk Driving (MADD), http://www.infoplease.com/ipa/A0777960.html

a. Los totales de columna no están incluidos porque serían valores que carecen de sentido. Examine la tabla y explique por qué.

b. Seleccione los días festivos apropiados que no se traslapan (columna 1) y verifique el número total de 6764 muertos en accidentes de tránsito para 2002.

c. Usando los días festivos seleccionados en la parte b, encuentre el número total de muertos en accidentes de tránsito relacionados con alcohol en días festivos en 2002.

d. Describa cómo organizaría esta tabla para hacerla que tenga sentido.

4.5 Utilice ya sea la tabla de números aleatorios (apéndice B), calculadora o computadora (vea p. 101) para simular lo siguiente:

a. Tirar 50 veces un dado; exprese sus resultados como frecuencias relativas.

b. Tirar al aire una moneda 100 veces; exprese sus resultados como frecuencia relativa.

4.6 Utilice ya sea la tabla de números aleatorios (apéndice B), calculadora o computadora (vea p.101) para simular la selección aleatoria de 100 números de un solo dígito, 0 al 9.

a. Haga una lista de los 100 dígitos.

b. Elabore una distribución de frecuencia relativa de los 100 dígitos.

c. Elabore un histograma de frecuencia relativa de la distribución en la parte b.

4.2 Probabilidad de eventos

Ahora estamos listos para definir lo que significa *probabilidad.* En forma específica, hablamos de "la probabilidad de que ocurra cierto evento".

Probabilidad de que ocurra un evento: es la frecuencia relativa con la que puede esperarse que el evento ocurra.

La probabilidad de un evento puede obtenerse en tres formas diferentes: (1) *empíricamente,* (2) *teóricamente* y (3) *subjetivamente.*

El método **empírico** fue ilustrado por los chocolates M&M's y sus porcentajes en la sección 4.1 y podría llamarse *probabilidad experimental* o *empírica.* Esta probabilidad es la *frecuencia relativa observada* con la que ocurre un evento. En nuestro ejemplo M&M's, observamos que 137 de los 692 chocolates M&M's eran de color naranja. La probabilidad empírica observada para el suceso de color naranja fue 137/692, es decir 0.198.

El valor asignado a la probabilidad del evento A como resultado de experimentación se puede hallar por medio de la fórmula:

Probabilidad empírica (observada): *P'*(A)

En palabras: $$\textit{probabilidad empírica } A = \frac{\textit{número de veces que } A \textit{ ocurrió}}{\textit{número de intentos}}$$

En álgebra: $$P'(\mathrm{A}) = \frac{n(\mathrm{A})}{n} \qquad (4.1)$$

Notación para probabilidad empírica: cuando el valor asignado a la probabilidad de un evento resulta de datos experimentales o empíricos, identificaremos la probabilidad del evento con el símbolo $P'(\ )$.

El método **teórico** para obtener la probabilidad de un evento usa un espacio muestral. Un espacio muestral es una lista de todos los posibles resultados del experimento bajo consideración. Cuando se utiliza este método, el espacio muestral debe contener puntos muestrales igualmente probables. Por ejemplo, el espacio muestral para el hecho de tirar un dado es {1, 2, 3, 4, 5, 6}. Cada resultado (es decir, número) es igualmente probable. Un evento es un subconjunto del espacio muestral. Por tanto, la probabilidad de un evento A, $P(\mathrm{A})$, es la razón entre el número de puntos que satisfacen la definición del evento A, $n(\mathrm{A})$, y el número de puntos muestrales de todo el espacio muestral, $n(\mathrm{S})$. Esto es,

Probabilidad teórica (esperada): *P*(A)

En palabras: $$\textit{probabilidad teórica de A} = \frac{\textit{número de veces que A ocurre en espacio muestral}}{\textit{número de elementos en espacio muestral}}$$

En álgebra: $$P(\mathrm{A}) = \frac{n(\mathrm{A})}{n(\mathrm{S})} \tag{4.2}$$

Notas:

1. Cuando el valor asignado a la probabilidad de un evento resulta de una fuente teórica, identificaremos la probabilidad del evento con el símbolo $P(\ )$.
2. El símbolo primo *no se usa* con probabilidades teóricas; se usa sólo para probabilidades empíricas.

EJEMPLO 4.1 Un dado

© Royalty-Free/CORBIS

© Digital Vision/Getty Images

Seis posibles resultados de un tiro.

Considere el caso de tirar un dado. Defina el evento A como el suceso de un número "mayor de 4". En un solo tiro de un dado, hay seis resultados posibles, haciendo $n(\mathrm{S}) = 6$. El evento "mayor de 4" está satisfecho por el suceso de un 5 o un 6; así, $n(\mathrm{A}) = 2$. Suponiendo que el dado es simétrico y que cada número tiene igual probabilidad de presentarse, la probabilidad de A es $\frac{2}{6}$, o sea $\frac{1}{3}$.

EJEMPLO 4.2 Un par de dados

Un par de dados (uno blanco, uno negro) se tira una vez, y se observa el número de puntos que aparezca en cada dado. El espacio muestral se ilustra en formato de tabla:

¿SABÍA USTED...?

¿Leche en tu té?

A fines de la década de 1920 en una reunión para tomar el té en una tarde de verano en Cambridge, Inglaterra, una de las invitadas dijo que el té sabe diferente dependiendo de si el té se vierte en la leche o la leche se vierte en el té, cosa que pareció ridícula. Después de muchas bromas, un hombre, Ronald A. Fisher, propuso una forma científica de probar la hipótesis de la invitada: combinar la leche y el té de los dos modos, luego ofrecerle uno de cada uno, dos a la vez en orden aleatorio, para que los identificara. Otros rápidamente se le unieron y ayudaron en la prueba: ella correctamente identificó 10 consecutivos. ¿Qué cree usted al respecto? ¿Podría ella decir la diferencia?

Representación de tabla

$n(S) = 36$

Ha de considerarse la suma de sus puntos. Una lista de las posibles "sumas" forma un espacio muestral, S = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} y $n(S) = 11$. No obstante, los elementos de este espacio muestral no son igualmente probables; por tanto, este espacio muestral no puede usarse para hallar probabilidades teóricas —debemos usar el espacio muestral de 36 puntos que se indica en la tabla precedente—. Al usar el espacio muestral de 36 puntos, el espacio muestral está por completo formado por puntos muestrales igualmente probables, y las probabilidades para las sumas de 2, 3, 4, etcétera, se pueden hallar con toda facilidad. La suma de 2 representa {(1,1)}, donde el primer elemento del par ordenado es el resultado para el dado blanco y el segundo elemento del par ordenado es el resultado para el dado negro. La suma de 3 representa {(2,1), (1,2)}; y la suma de 4 representa {(1,3), (3,1),(2,2)]; y así sucesivamente. Por tanto, podemos usar la fórmula (4.2) y el espacio muestral de 36 puntos para obtener las probabilidades para las 11 sumas.

$$P(2) = \frac{n(2)}{n(S)} = \frac{1}{36},\ P(3) = \frac{n(3)}{n(S)} = \frac{2}{36},\ P(4) = \frac{n(4)}{n(S)} = \frac{3}{36}$$

y así sucesivamente.

Cuando un experimento de probabilidad pueda ser considerado como una secuencia de eventos, un *diagrama de árbol* es con frecuencia una forma muy útil de presentar el espacio muestral.

EJEMPLO 4.3 Uso de diagramas de árbol

Ha de seleccionarse al azar una familia con dos hijos, y deseamos hallar la probabilidad de que la familia seleccionada tenga un hijo de cada género. Debido a que siempre habrá un hijo primogénito y un segundo hijo, usaremos un diagrama de árbol para mostrar los posibles arreglos de género, haciendo viable que determinemos la probabilidad. Comencemos por determinar la secuencia de eventos involucrados, primogénito y segundo hijo en este caso. Usemos el árbol para presentar los posibles resultados del primer evento (indicado en color café en la figura 4.1) y luego sumamos segmentos de ramas para mostrar los posibles resultados para el segundo evento (indicado en color naranja en la figura 4.1).

Notas:

1. Los dos segmentos de ramas que representan B y G para el segundo hijo deben trazarse a partir de cada uno de los resultados para el hijo primogénito, creando así el aspecto de "árbol".

FIGURA 4.1
Representación de diagrama de árbol de familia con dos hijos

2. Hay cuatro ramas; cada rama se inicia en la "raíz de árbol" y continúa a un "extremo" (formada por dos segmentos de ramas cada uno), que muestra un posible resultado.

Debido a que los segmentos de ramas son igualmente probables, suponiendo iguales probabilidades de género, las cuatro ramas son entonces igualmente probables. Esto significa que necesitamos sólo la cantidad de ramas para usar la fórmula 4.2 para hallar la probabilidad de que la familia tenga un hijo de cada género. Las dos ramas centrales, (B,G), y (G,B), representan el evento de interés, de modo que $n(\text{A}) = n(\text{uno de cada uno}) = 2$, mientras que $n(\text{S}) = 4$ porque hay un total de cuatro ramas. Así,

$$P(\text{uno de cada género en la familia de dos hijos}) = \frac{2}{4} = \frac{1}{2} = 0.5$$

Ahora consideremos seleccionar una familia de tres hijos y hallar la probabilidad de "al menos un niño" en esa familia. De nuevo la familia puede ser considerada como una secuencia de tres eventos —hijo primogénito, segundo hijo y tercer hijo—. Para crear un diagrama de árbol de esta familia, necesitamos sumar un tercer conjunto de segmentos de ramas a nuestro diagrama de árbol para una familia de dos hijos. Los segmentos verdes de ramas representan el tercer hijo (vea figura 4.2).

FIGURA 4.2
Representación de diagrama de árbol de familia con tres hijos

De nuevo, como los segmentos de ramas son igualmente probables, suponiendo igual probabilidad de género, las ocho ramas son entonces igualmente probables. Esto significa que sólo necesitamos la cantidad de ramas para usar la fórmula 4.2 para hallar la probabilidad de que la familia tenga al menos un hijo. Las siete ramas superiores tienen todas ellas uno o más hijos, el equivalente de "al menos uno".

$$P(\text{al menos un hijo en una familia de tres hijos}) = \frac{7}{8} = 0.875$$

Consideremos otra pregunta antes de salir de este ejemplo. ¿Cuál es la probabilidad de que el tercer hijo de esta familia de tres hijos sea niña? La pregunta es fácil en realidad; la respuesta es 0.5, porque hemos supuesto igual probabilidad del otro género. No obstante, si vemos los tres diagramas de la figura 4.2, hay dos formas de ver la respuesta. Primero, si vemos sólo los segmentos de ramas para el tercer hijo, vemos que uno de dos es para una niña en cada conjunto, o sea $\frac{1}{2}$, o 0.5. Del mismo modo, si vemos todo el diagrama de árbol, el último hijo es una niña en cuatro de las ocho ramas, es decir, $\frac{4}{8}$, o 0.5.

Cuando una pregunta de probabilidad contenga información acerca de los eventos en forma del número de elementos por conjunto, el porcentaje de cada conjunto, o la probabilidad de los diversos eventos, con frecuencia un *diagrama de Venn* es una forma muy útil de representar el espacio muestral.

EJEMPLO 4.4 Uso de diagramas de Venn

Un cliente afortunado en el lote de autos usados de Charlie tendrá que seleccionar al azar una llave de un barril de llaves; el barril contiene las llaves de todos los autos del lote, que a su vez tiene un inventario de 80 autos, 38 de ellos de modelos extranjeros, 50 son compactos y 22 son modelos compactos extranjeros. El diagrama de Venn que se ve en la figura 4.3 resume el inventario de Charlie. Nótese que algunos de los 38 modelos extranjeros son compactos y otros no lo son. Lo mismo es con respecto a modelos compactos; algunos son extranjeros y otros no lo son. Por tanto, cuando se descompone esta clase de información, es necesario empezar con lo más específico. En este caso, 22 autos son extranjeros y compactos que están representados por la región central del diagrama de Venn. De aquí, se puede determinar cuántos autos son extranjeros pero no compactos y cuántos son compactos pero no extranjeros. Vea la figura 4.3.

FIGURA 4.3
Representación del diagrama de Venn del inventario de autos usados de Charlie

Usted es el cliente afortunado que ha ganado la oportunidad de sacarse un auto gratis del lote de autos usados de Charlie, y está a punto de sacar una de las 80 llaves. ¿Cuál es la probabilidad de que gane un auto compacto que no sea extranjero? Al ver el diagrama de Venn, los autos extranjeros están dentro del círculo azul; por

tanto, los no extranjeros están fuera del círculo azul. El evento de interés es que el auto, junto con no extranjero, debe ser compacto (dentro del círculo rojo), que, con base en la figura 4.3, podemos determinar que es 28 de estos autos. Con la fórmula (4.2) encontramos que

$$P(\text{compacto no extranjero}) = \frac{28}{80} = 0.35$$

De manera práctica, el diagrama de Venn funciona igualmente bien si la información se hubiera dado en porcentajes o probabilidades. El diagrama parece igual excepto en que los valores se convierten ya sea en probabilidades o porcentajes. Para estar seguros que se haya cubierto todo el espacio muestral, la suma de todas las regiones debe ser exactamente 1.0 para que la leyenda sea correcta.

Nota: a veces es útil colocar una moneda en el círculo que representa un evento cuando se ve un evento que "no" ocurrió. En el diagrama de Venn de la figura 4.3, una moneda de 25¢ de dólar puesta en el círculo de "modelos extranjeros" hace visibles todos los modelos no extranjeros.

Siempre se da especial atención al espacio muestral. Al igual que la población estadística, el espacio muestral debe estar bien definido. Una vez definido el espacio muestral, el trabajo restante es más fácil.

En general, una probabilidad **subjetiva** resulta de un juicio personal. El servicio meteorológico local a veces asigna una probabilidad al evento "precipitación". Por ejemplo, "hay 20% de probabilidad de lluvia para hoy," o "hay 70% de nieve para mañana". En estos casos, el único método que hay para asignar posibilidades es el juicio personal. Estas asignaciones de probabilidad se denominan *probabilidades subjetivas.* La precisión de éstas depende de la capacidad del individuo para evaluar correctamente una situación.

Propiedades de números de probabilidad

Si la probabilidad es empírica, teórica o subjetiva, deben cumplirse las siguientes propiedades.

Propiedad 1

En palabras: "Una probabilidad es siempre un valor numérico entre cero y uno."

En álgebra: $0 \leq$ cada $P(\text{A}) \leq 1$

Notas acerca de la propiedad 1:

1. La probabilidad es 0 si el evento no puede ocurrir.
2. La probabilidad es 1 si el evento ocurre cada vez.
3. De otro modo, la probabilidad es un número fraccionario entre 0 y 1.

Propiedad 2

En palabras: "La suma de las probabilidades de todos los resultados de un experimento es igual a exactamente uno."

En álgebra $\sum_{\text{todos los resultados}} P(\text{A}) =$

Nota acerca de la propiedad 2: la lista de "todos los resultados" debe ser un conjunto de eventos que no se traslapen (mutuamente excluyente) que incluya todas las posibilidades (todo incluido).

Notas acerca de números de probabilidad:

1. La probabilidad representa una frecuencia relativa.
2. $P(\mathrm{A})$ es la razón entre el número de veces que un evento puede esperarse que ocurra y el número de intentos.
3. El numerador de la razón de probabilidad debe ser un número positivo o cero.
4. El denominador de la razón de probabilidad debe ser un número positivo (mayor a cero).
5. El número de veces que un evento puede esperarse que ocurra en n intentos es siempre menor o igual al número total de intentos, n.

¿Cómo están relacionadas las probabilidades empíricas y teóricas?

Considere el ejemplo de lanzar un dado y definir el evento A como la aparición de un "1". Un dado común y corriente tiene seis lados igualmente probables, de modo que la probabilidad teórica del evento A es $P(\mathrm{A}) = \frac{1}{6}$

¿Qué significa esto?

¿Espera ver un "1" en cada intento de seis tiros? Explique. Si no es así, ¿qué resultados espera? Si fuéramos a lanzar el dado varias veces y dar seguimiento a la proporción del tiempo en que se presenta el evento A, observaríamos una probabilidad empírica para el evento A. ¿Qué valor esperaría usted observar para $P'(\mathrm{A})$? Explique. ¿Cómo están relacionadas las dos probabilidades $P(\mathrm{A})$ y $P'(\mathrm{A})$? Explique.

Para tener una idea de esta relación, efectuemos un experimento.

EJEMPLO 4.5

Demostración-Ley de números grandes

El experimento se compone en 20 intentos. Cada uno de los intentos del experimento consistirá en lanzar un dado seis veces y recordar el número de veces que sale "1". Efectúe 20 tiros.

Cada fila de la tabla 4.3 muestra los resultados de un intento; efectuamos 20 intentos, de modo que hay 20 filas. La columna 1 contiene el número de unos (números 1) observados en cada intento (conjunto de seis tiros); la columna 2, la frecuencia relativa observada para cada intento; y la columna 3, la frecuencia relativa acumulada cuando se complete cada intento.

La figura 4.4a muestra la fluctuación (arriba y abajo) de la probabilidad observada, $P'(\mathrm{A})$ (tabla 4.3, columna 2), acerca de la probabilidad teórica, $P(\mathrm{A}) = \frac{1}{6}$, mientras que la figura 4.4 b muestra la fluctuación de la frecuencia relativa acumulativa (tabla 4.3, columna 3) y cómo se hace más estable. De hecho, la frecuencia relativa acumulativa se hace relativamente cercana a la probabilidad teórica o esperada, $\frac{1}{6}$, o $0.166\overline{6} = 0.167$

TABLA 4.3

Resultados experimentales de lanzar seis veces un dado en cada intento

Intento	Columna 1: número de (1) observado 4	Columna 2: frecuencia relativa 5	Columna 3: frecuencia relativa acumulativa	Intento	Columna 1: número de (1) observado 4	Columna 2: frecuencia relativa 5	Columna 3: frecuencia relativa acumulativa
1	1	1/6	1/6 = 0.17	11	1	1/6	10/66 = 0.15
2	2	2/6	3/12 = 0.25	12	0	0/6	10/72 = 0.14
3	0	0/6	3/18 = 0.17	13	2	2/6	12/78 = 0.15
4	1	1/6	4/24 = 0.17	14	1	1/6	13/84 = 0.15
5	0	0/6	4/30 = 0.13	15	1	1/6	14/90 = 0.16
6	1	1/6	5/36 = 0.14	16	3	3/6	17/96 = 0.18
7	2	2/6	7/42 = 0.17	17	0	0/6	17/102 = 0.17
8	2	2/6	9/48 = 0.19	18	1	1/6	18/108 = 0.17
9	0	0/6	9/54 = 0.17	19	0	0/6	18/114 = 0.16
10	0	0/6	9/60 = 0.15	20	1	1/6	19/120 = 0.16

FIGURA 4.4
Fluctuaciones halladas en el experimento de lanzar un dado

(a) Frecuencia relativa

(b) Frecuencia relativa acumulativa

Una gráfica acumulativa como la que se ve en la figura 4.4b demuestra la idea de un promedio a largo plazo y con frecuencia se conoce como la *ley de números grandes.*

Ley de números grandes: cuando aumenta el número de veces que se repite un experimento, la razón entre el número de sucesos exitosos y el número de intentos tenderá a aproximarse a la probabilidad teórica del resultado de un intento individual.

La ley de números grandes nos dice que cuanto más grande sea el número de intentos experimentales, n, se espera que la probabilidad empírica, $P'(\mathrm{A})$, será más cercana a la probabilidad verdadera o teórica, $P(\mathrm{A})$. Este concepto tiene muchas aplicaciones. El experimento precedente de lanzar un dado es un ejemplo en el que podemos fácilmente comparar resultados reales contra lo que esperamos ocurra; nos dio una oportunidad de verificar lo afirmado por la ley de números grandes.

El ejemplo 4.6 es una ilustración en la que vivimos con los resultados obtenidos de grandes conjuntos de datos cuando la expectativa teórica es desconocida

EJEMPLO 4.6

Usos de probabilidades empíricas

La clave para establecer primas adecuadas para seguros de vida es usar la probabilidad de que los asegurados vivirán 1, 2 o 3, años, etcétera, a partir del momento en que compran sus pólizas. Estas probabilidades se derivan de estadísticas reales de vida y muerte y por tanto son probabilidades empíricas. Son publicadas por el gobierno y son extremadamente importantes para la industria de seguros de vida.

Probabilidades como posibilidades

Las probabilidades pueden ser expresadas, y se expresan, en diversas formas; vemos y escuchamos muchas de ellas en las noticias casi todos los días. Las posibilidades son una forma de expresar probabilidades al expresar el número de formas en que un evento puede ocurrir en comparación con el número de formas en que no puede ocurrir. El enunciado de que "es cuatro veces más probable que llueva mañana a que no llueva" es un enunciado de probabilidad y se expresa como posibilidades: las posibilidades son 4 a 1 a favor a que haya lluvia mañana" (también se escribe 4:1).

La relación entre posibilidades y probabilidad se muestra enseguida.

Si las posibilidades a favor de un evento A son ***a* a *b*** (o ***a*:*b***), entonces

1. Las posibilidades contra el evento A son ***b* a *a*** (o ***b*:*a***)
2. La probabilidad del evento A es $\boldsymbol{P(\mathrm{A}) = \dfrac{a}{a+b}}$
3. La probabilidad de que el evento A no ocurra es $\boldsymbol{P(\text{no A}) = \dfrac{a}{a+b}}$

Para ilustrar esta relación, considere el enunciado "Las posibilidades a favor de que haya lluvia mañana son 4 a 1". Usando la notación precedente, $a = 4$ y $b = 1$.

Por tanto, la probabilidad de que llueva mañana es $\frac{4}{4+1}$, o $\frac{4}{5} = 0.8$. Las posibilidades contra lluvia de mañana son 1 a 4 (o 1:4), y la probabilidad de que no haya lluvia mañana es $\frac{1}{4+1}$, o $\frac{1}{5} = 0.2$

CASO PRÁCTICO 4.7

Tratando de vencer las posibilidades

Numerosos jóvenes aspiran a hacerse atletas profesionales. Sólo unos pocos lo logran, como se indica en la siguiente gráfica. Por cada 2400 jugadores universitarios de baloncesto de alto rendimiento, sólo 64 forman un equipo profesional; eso se traduce a una probabilidad de sólo 0.027 (64/2400).

POSIBILIDADES EN CONTRA PARA LLEGAR AL SIGUIENTE NIVEL

Las posibilidades que hay en contra para que un jugador de baloncesto de alto rendimiento en secundarias de Estados Unidos llegue a un equipo profesional son 2344 a 1, con base en números de 1989:

Información de Julie Stacey, © 1990 USA Today.

Hay muchos otros datos específicos, interesantes, ocultos en esta información. Por ejemplo, muchos estudiantes de secundaria sueñan en convertirse en jugadores profesionales de baloncesto, pero, de acuerdo con estos números, la probabilidad de que su sueño se convierta en realidad es de sólo 0.000427 (64/150000).

Una vez que un jugador haya llegado a un equipo universitario de baloncesto, podría estar muy interesado en las posibilidades de llegar a jugador de alto rendimiento. De los 3800 jugadores que a un equipo universitario, 2400 son jugadores de alto rendimiento, mientras que 1400 no juegan a este nivel. Por tanto, si un jugador ha llegado a un equipo universitario, las posibilidades de que juegue como de alto rendimiento son 2400 a 1400, lo cual se reduce a 12 a 7.

El jugador universitario de alto rendimiento que está jugando está interesado en sus posibilidades de llegar al siguiente nivel. Vemos que de los 2400 jugadores universitarios de alto rendimiento, sólo 64 llegan a equipos profesionales, mientras que 2336 no llegan; así, las posibilidades contra él para que llegue al siguiente nivel son 2336 a 64, lo cual se reduce a 73 a 2. Las posibilidades están fuertemente contra él para que llegue al siguiente nivel.

SECCIÓN 4.2 EJERCICIOS

4.7 Si usted lanza un dado 40 veces y 9 de los tiros resultan en un "5", ¿qué probabilidad empírica se observó para el evento "5"?

4.8 Explique por qué una probabilidad empírica, una proporción observada, y una frecuencia relativa son en realidad tres nombres diferentes para lo mismo.

4.9 Millones de personas viajan en ferrocarril todos los años. La Asociación Nacional de Pasajeros de Ferrocarril proporciona las siguientes cantidades de viajes en 2004.

Sistema ferroviario	Viajeros (millones)
Sistema Amtrak	25.0
Corredor Noroeste	14.2
Suburbano + Oeste	10.8

Fuente: National Association of Railroad Passengers http://www.infoplease.com/ipa/A0855824.html

a. ¿Qué porcentaje de pasajeros de ferrocarril usaron el sistema Amtrak en 2004?

b. Si uno de estos pasajeros ha de ser entrevistado, ¿cuál es la probabilidad de que él haya viajado en el sistema Amtrak en 2004 si es seleccionado al azar?

c. Explique la diferencia y la relación entre preguntas y respuestas a las partes a y b.

4.10 El Webster Aquatic Center ofrece varios niveles de lecciones de natación todo el año. Las lecciones vespertinas de lunes y miércoles de marzo de 2005 incluyeron clases desde bebés a adultos. El número en cada clasificación aparece en la tabla siguiente.

Tipos de lección de natación	Núm. de participantes
Bebés	15
Bebé muy pequeño	12
Renacuajos	12
Nivel 2	15
Nivel 3	10
Nivel 4	6
Nivel 5	2
Nivel 6	1
Adultos	4
Total	77

Si se selecciona al azar un participante, encuentre la probabilidad de lo siguiente:

a. El participante está en bebés muy pequeños.

b. El participante está en la lección para adultos.

c. El participante está en una lección de nivel 2 a nivel 6.

4.11 En septiembre de 2004, la American Payroll Association publicó los resultados de su encuesta nacional de semana de nómina 2004. Una de las preguntas inquiría acerca del ingreso familiar anual.

Ingreso familiar anual	Número	Porcentaje
Menos de $15 000	423	1.9%
$15 001–$30 000	2225	9.8%
$30 001–$50 000	5394	23.9%
$50 001–$75 000	5772	25.5%
$75 001–$100 000	4730	20.9%
$100 001–$150 000	3065	13.6%
Más de $150 000	984	4.4%

Fuente: American Payroll Association, http://www.AmericanPayroll.org

Suponga que una de las personas que respondieron a la encuesta ha de ser seleccionado al azar para una entrevista de seguimiento. Encuentre la probabilidad de los siguientes eventos.

a. El ingreso familiar del encuestado es $50 000 o menos.

b. El ingreso familiar del encuestado es $75 001 o más.

c. El ingreso familiar del encuestado es entre $30 000 y $100 000.

d. El ingreso familiar del encuestado es al menos $100 001.

4.12 El U.S. Department of Transportation publica anualmente el número de quejas de consumidores contra las principales aerolíneas por categoría. A continuación aparecen las cifras para 2002.

Categoría de queja	Número de quejas	Categoría de queja	Número de quejas
Problemas en vuelo	2031	Sobreventa	454
Servicio a clientes	1715	Tarifas	523
Equipaje	1421	Incapacidad	477
Reservaciones/venta de boletos/abordar	1159	Publicidad	68
Devoluciones	1106	Otras	322

Fuente: Office of Aviation Enforcement & Proceedings, U.S. Department of Transportation, Air Travel Consumer Report, http://www.infoplease.com/ipa/A0198353.html

Si una de estas quejas se selecciona al azar para evaluación de seguimiento, ¿cuál es la probabilidad de que la queja sea:

a. relacionada con problemas en vuelo?

b. acerca del servicio a clientes o equipaje?

c. relativa a las reservaciones/boletos/abordar o devoluciones o sobreventa?

d. que no sea de equipaje?

4.13 The Weather Underground, Inc., proporciona un planeador de viajes en Internet. Las estadísticas de condiciones atmosféricas que publica están basadas en fechas buscadas de 1975 a la actualidad. Imaginemos que el lector está planeando un viaje del 1 de marzo al 15 de marzo y obtuvo la siguiente información de pronóstico del clima de este planeador de viajes.

Seattle, WA
La temperatura alta promedio es 52°F ~ el rango histórico es de 33°F a 70°F
La temperatura baja promedio es 39°F ~ el rango histórico es de 20°F a 53°F
0% de probabilidad de día bochornoso (temp. más de 90°F) (0 de 390 días en récord histórico).
9% de posibilidad de un día caluroso (temp. más de 60°F) (37 de 390 días en récord histórico).
La precipitación promedio diaria es 0.23 ~ el rango histórico es de 0.00 a 1.47
48% de posibilidad de un día con precipitación (188 de 390 días en récord histórico).
Promedio de cobertura con humo significa mayormente nublado
71% de posibilidad de un día nublado (75 de 105 días en registro histórico).

San Diego, CA
La temperatura alta promedio es 66°F ~ el rango histórico es de 56°F a 85°F
La temperatura baja promedio es 53°F ~ el rango histórico es de 44°F a 62°F
0% de probabilidad de día caluroso (temp. superior a 90°F) (0 de 390 días en récord histórico).
95% de posibilidad de un día caluroso (temp. superior a 60°F) (371 de 390 días en récord histórico).
La precipitación promedio diaria es 0.12 ~ el rango histórico es de 0.00 a 1.95
27% de posibilidad de un día con precipitación (104 de 390 días en récord histórico).
Promedio de cobertura con humo significa parcialmente nublado
35% de posibilidad de un día nublado (37 de 105 días en registro histórico).

Fuente: The Weather Underground, Inc., http://www.wunderground.com/tripplanner/index.asp

a. Las probabilidades indicadas son la frecuencia relativa del evento con base en registros históricos. Verifique la probabilidad (posibilidad) de un día caluroso en Seattle. En San Diego.

b. Verifique la probabilidad (posibilidad) de un día con precipitación en Seattle. En San Diego.

c. Cuando usted viaje, planea usar la mitad de sus días sólo para "descansar y asolearse un poco." Con base en la información precedente, ¿qué ciudad es la mejor opción para días calurosos (pero no bochornoso) y soleados (no nublados)? Haga un caso para su respuesta.

4.14 Los dos entrenadores de futbol profesional que ganaron más juegos durante sus carreras fueron Don Shula y George Halas. Los equipos de Shula (Potros y Delfines) ganaron 347 juegos (récord de 347-173-6) y empataron 6 de los 526 juegos que dirigió, mientras que el equipo de Halas (Osos) ganó 324 juegos (récord de 324-151-31) y empató 31 de los 506 juegos que dirigió.

Fuente: Pro Football Hall of Fame (Salón de la Fama del Futbol Profesional)

Suponga que se lanza a un depósito una película de cada juego que cada uno de estos hombres dirigió, y se mezclan. Usted selecciona una película al azar del depósito y la carga en un proyector. ¿Cuál es la probabilidad de que la película que haya seleccionado muestre lo siguiente?

a. Un juego empatado

b. Un juego perdido

c. Uno de los equipos de Shula gane un juego

d. El equipo de Halas gane un juego

e. Uno de los equipos de Shula pierda un juego

f. El equipo de Halas pierda un juego

g. Uno de los equipos de Shula juegue para empatar

h. El equipo de Halas juegue para empatar

i. Un juego dirigido por Halas

j. Un juego dirigido por Shula

4.15 Un número de un solo dígito ha de seleccionarse al azar. Haga una lista del espacio muestral.

4.16 Se lanza un solo dado. ¿Cuál es la probabilidad de que el número en su cara superior sea lo siguiente?

a. Un 3

b. Un número impar

c. Un número menor a 5

d. Un número no mayor de 3

4.17 Se lanza un par de dados. En el ejemplo 4.2, la probabilidad de cada una de las posibles sumas se discutió y se encontraron tres de las probabilidades, $P(2)$, $P(3)$, y $P(4)$. Encuentre la probabilidad para cada una de las sumas restantes de dos dados: $P(5)$, $P(6)$, $P(7)$, $P(8)$, $P(9)$, $P(10)$, $P(11)$ y $P(12)$.

4.18 Se lanzan dos dados. Encuentre las probabilidades en las partes b-e. Use el espacio muestral dado en el ejemplo 4.2 (pp. 208-209).

a. ¿Por qué es que el conjunto {2, 3, 4, . . . 12} no es un espacio muestral?

b. P(dado blanco es un número non)

c. P(la suma es 6)

d. P(ambos dados muestran números nones)

e. P(el número del dado negro es mayor que el número del dado blanco)

4.19 Tome dos dados (uno blanco y otro de color) y láncelos 50 veces, registrando los resultados como pares ordenados [(blanco,color); por ejemplo, (3,5) representa 3 en el dado blanco y 5 en el dado de color]. (Podría simular estos 50 tiros usando una tabla de números aleatorios o una computadora.) Luego calcule cada probabilidad observada:

a. P'(punto blanco es un número non)

b. P'(la suma es 6)

c. P'(ambos dados muestran número non)

d. P'(número en dado de color es mayor que el número en el dado blanco)

e. Explique por qué estas respuestas y las respuestas encontradas en el ejercicio 4.18 de líneas antes no son exactamente iguales.

4.20 Use una tabla de números aleatorios o una computadora para simular el lanzamiento de un par de dados 100 veces.

a. Haga una lista de cada tiro como par ordenado y la suma.

b. Elabore una distribución de frecuencia no agrupada y un histograma de las sumas.

c. Describa cómo se comparan estos resultados con lo que usted esperaba que ocurriera cuando se lanzan dos dados.

INSTRUCCIONES DE TECNOLOGÍA: SIMULE DADOS

MINITAB (Versión 14)

```
Choose:  Calc > Random Data > Integer
Enter:   Generate: 100
         Store in column(s): C1 C2
         Minimum value: 1  Maximum value: 6 > OK
Choose:  Calc > Calculator
Enter:   Store result in variable: C3
         Expression:  C1 + C2  >  OK
Choose:  Stat > Tables > Tally Individual Variables
Enter:   Variable: C3
Select:  Counts  >  OK
```

Use los comandos del MINITAB de la página 61 para construir un histograma de frecuencia de los datos en C3. (Use las posiciones de Binning>midpoint y midpoint 2:12/1 si es necesario.)

Excel

Introduzca 1, 2, 3, 4, 5, 6 en la columna A, marque C1: **Die1;** D1: **Die2;** E1: **Dados,** y active B1.

```
Choose:  Format > Cells > Number > Number
Enter:   Decimal places: 8 > OK
Enter:   1/6 in B1
Drag:    Bottom right corner of B1 down for 6 entries
Choose:  Tools > Data Analysis > Random Number Genera-
         tion > OK
Enter:   Number of Variables: 2
         Number of Random Numbers: 100
         Distribution: Discrete
         Value and Probability Input Range:  (A1:B6 or
         select cells)
Select:  Output Range
Enter:   (C2 or select cells)  >  OK
```

Active la celda **E2**

```
Enter:   = C2 + D2  >  Enter
Drag:    Bottom right corner of E2 down for 100 entries
Choose:  Data > Pivot and PivotChart Table Report . . .
Select:  Microsoft Excel list or database  >  Next
Enter:   Range: (E1:E101 or select cells) > Next
Select:  Existing Worksheet
Enter:   (F1 or select cell)
Choose:  Layout
Drag:    "Dice" heading into both row & data areas
```

Haga doble clic en "suma de dados" en la caja de área de datos; luego continúe con:

```
Choose:  Summarize by: Count > OK > OK > Finish
```

Marque la columna J como "sumas" e introduzca los números 2, 3, 4, . . . 12 en ella. Use los comandos del histograma Excel de las páginas 61-62 con la columna E como el rango de entrada y la columna J como el rango de depósito.

TI-83/84 Plus

```
Choose:  MATH > PRB > 5:randInt(
Enter:   1,6,100)
Choose:  STOO  > 2nd L1

Repeat preceding for L2.

Choose:  STAT > EDIT > 1:Edit
Highlight:        L3
Enter:   L3 =L1 + L2
Choose:  2nd  > STAT PLOT
> 1:Plot1
Choose:  WINDOW
Enter:   −.5, 12.5, 1,
−10, 40, 10,1
Choose:  TRACE > > >
```

4.21 Las 12 cartas de caras (4 comodines, 4 reinas y 4 reyes) se sacan de un "monte" regular de barajas, y luego una carta se selecciona de este conjunto de cartas de caras. Haga un espacio muestral para este experimento.

4.22 Un experimento consiste en sacar una canica de una caja que contiene una mezcla de canicas rojas, amarillas y verdes. Hay al menos dos canicas de cada color.

a. Haga una lista del espacio muestral.

b. ¿Podemos estar seguros de que cada resultado del espacio muestral de la parte a sea igualmente probable? Explique.

c. Si se sacan dos canicas de la caja, haga una lista del espacio muestral.

d. ¿Son igualmente probables los resultados del espacio muestral de la parte c? Explique

4.23 Una caja contiene un billete de cada uno de lo siguiente: \$1, \$5, \$10 y \$20.

a. Se selecciona uno al azar; haga una lista del espacio muestral.

b. Se sacan dos billetes al azar (sin reposición); haga una lista del espacio muestral como un diagrama de árbol.

4.24 Se lanzan al aire tres monedas, y se registra el número de caras observadas. Encuentre la probabilidad para cada uno de los posibles resultados: 0C, 1C, 2C y 3C.

4.25 Un grupo de archivos en una clínica médica clasifica a los pacientes por género y por tipo de diabetes (tipo 1 o tipo 2). Las agrupaciones pueden mostrarse como sigue. La tabla da el número de cada clasificación.

	Tipo de diabetes	
Género	1	2
Masculino	30	15
Femenino	35	20

a. Presente la información en esta tabla de 2×2 como un diagrama de Venn usando "tipo 1" y "masculino" como los dos eventos mostrados en círculos. Explique la forma en que el diagrama de Venn y la tabla dada de 2×2 dan la misma información.

Si un archivo se selecciona al azar, encuentre la probabilidad de lo siguiente:

b. El individuo seleccionado es femenino.

c. El individuo seleccionado tiene diabetes tipo 2.

4.26 Unos investigadores han estado interesados durante largo tiempo en la relación que hay entre fumar cigarrillos y el cáncer de pulmón. La siguiente tabla muestra los porcentajes de mujeres adultas en un reciente estudio.

	Fuma	No fuma
Tiene cáncer	0.06	0.03
No tiene cáncer	0.15	0.76

a. Presente la información de esta tabla de 2×2 como un diagrama de Venn usando "fuma" y "tiene cáncer" como los dos eventos mostrados como círculos. Explique cómo el diagrama de Venn y la tabla dada de 2×2 dan la misma información.

Supongamos que una mujer adulta se selecciona al azar de esta población particular. Calcule la probabilidad de lo siguiente:

b. Ella fuma y tiene cáncer.

c. Ella fuma.

d. Ella no tiene cáncer.

e. Ella no fuma y no tiene cáncer.

f. Ella tiene cáncer si fuma.

g. Ella no tiene cáncer, sabiendo que no fuma.

4.27 Un establecimiento de piezas de repuesto vende piezas nuevas y usadas. Sesenta por ciento de las piezas en existencia son usadas. Sesenta y uno por ciento son usadas o defectuosas. Si 5% de las piezas del establecimiento son defectuosas, ¿qué porcentaje es de piezas usadas y defectuosas? Resuelva usando un diagrama de Venn.

4.28 Los dirigentes de un sindicato informan que 60% de los trabajadores en una gran fábrica pertenecen al sindicato, 90% ganan más de $12 por hora, y 40% pertenecen al sindicato y ganan más de $12 por hora. ¿Cree usted en estos porcentajes? Explique. Resuelva usando un diagrama de Venn.

4.29 Sea x la clasificación de éxito de un nuevo programa televisivo. La tabla siguiente contiene las probabilidades subjetivas asignadas a cada x para un nuevo programa particular por tres diferentes críticos de medios. ¿Cuáles de estos tres conjuntos de probabilidades son inapropiados porque violan una regla básica de probabilidad? Explique.

	Juez		
Clasificación de éxito, x	A	B	C
Altamente exitoso	0.5	0.6	0.3
Exitoso	0.4	0.5	0.3
No exitoso	0.3	-0.1	0.3

4.30 Un ingeniero de transportes, a cargo de un nuevo sistema de control de tránsito, expresa la probabilidad subjetiva de que el sistema funciona correctamente 99 veces con tanta frecuencia como cuando funciona mal.

a. Con base en esta idea, ¿cuál es la probabilidad de que el sistema funcione correctamente?

b. Con base en esta idea, ¿cuál es la probabilidad de que el sistema funcione mal?

4.31 a. Explique lo que significa este enunciado: "Cuando se lanza un solo dado, la probabilidad de un 1 es $\frac{1}{6}$."

b. Explique lo que significa este enunciado: "Cuando una moneda se lanza al aire una vez, hay una probabilidad de 50-50 de que salga cola".

4.32 La siguiente gráfica muestra la ley de números grandes y también permite ver si una persona tiene poderes psíquicos. Repita las simulaciones al menos 50 veces, intentando entre escoger ya sea una carta roja o una negra de un "monte" de barajas.

a. ¿Qué proporción de tiempo adivinó usted correctamente?

b. A medida que se hicieron más intentos, ¿empezaron a estabilizarse sus proporciones? Si es así, ¿a qué valor? ¿Tiene esto sentido para el experimento? ¿Por qué?

c. ¿Cómo podría una persona saber si tiene percepción extrasensorial?

4.33 Una computadora genera (de manera aleatoria) pares de enteros. El primer entero es entre 1 y 5, inclusive, y el segundo es entre 1 y 4, inclusive.

a. Represente el espacio muestral como un diagrama de árbol.

b. Haga una lista de sus resultados como pares ordenados, con x como el primer entero y y como el segundo entero.

4.34 Use una computadora (o tabla de números aleatorios) para simular el experimento descrito en el ejercicio 4.33; x es un entero de 1 a 5, y y es un entero de 1 a 4. Genere una lista de 100 valores x y 100 valores y todos ellos aleatorios.

a. Encuentre la frecuencia relativa para $x = 2$.

b. Encuentre la frecuencia relativa para $y = 3$.

c. Encuentre la frecuencia relativa para el par ordenado (2,3).

4.35 Un experimento consta de dos intentos. El primero es lanzar al aire una moneda de 1 centavo y observar si cae con cabeza o cola hacia arriba; el segundo es lanzar un dado y observar un 1, 2, 3, 4, 5 o 6.

a. Construya el espacio muestral usando un diagrama de árbol.

b. Haga una lista de sus resultados como pares ordenados, con el primer elemento representando la moneda y el segundo al dado.

4.36 Use una computadora (o tabla de números aleatorios) para simular 200 intentos del experimento descrito en el ejercicio 4.35: lanzar al aire una moneda de 1 centavo y tirar un dado. Sea 1 = H (cara) y 2 = T (cola) para el centavo, y 1, 2, 3, 4, 5, 6 para el dado. Presente sus resultados usando una tabla tabulada en cruz que muestre la frecuencia de cada resultado.

a. Encuentre la frecuencia relativa para caras.

b. Encuentre la frecuencia relativa para 3.

c. Encuentre la frecuencia relativa para (H, 3).

4.37 Usando una moneda, realice el experimento discutido en las páginas 213-214. Lance al aire una moneda 10 veces, observe el número de caras (o ponga 10 monedas en una taza, agítela, tire las monedas en una caja, y use cada tiro para un bloque de 10); registre los resultados. Repita hasta que haya hecho 200 tiros. Ponga en una tabla y grafique los datos como conjuntos individuales de 10 y como frecuencias relativas acumulativas. Sus datos ¿tienden a apoyar el dicho de que $P(\text{cara}) = \frac{1}{2}$? Explique.

4.38 Un merengue de chocolate se va a lanzar al aire y para que caiga en una superficie dura y lisa. (Como lanzar al aire una moneda o tirar un dado.)

a. ¿Qué proporción del tiempo piensa usted que el merengue caerá "con la punta hacia arriba" (contrario a la punta para abajo)?

b. Estimemos la probabilidad de que un merengue de chocolate caiga "con la punta hacia arriba" cuando caiga en una superficie dura y lisa después de ser lanzado. Usando un merengue de chocolate, con su envoltura puesta, realice el experimento del dado del que se habla en las páginas 213-214. Lance el merengue 10 veces, registre el número de veces que caiga "con la punta hacia arriba" (o ponga 10 merengues en una taza, agítela y descargue los merengues en una superficie dura y lisa, y usa cada tiro para un bloque de 10); registre los resultados. Repita hasta que haya hecho 200 tiros. Ponga en una tabla y grafique los datos como conjuntos individuales de 10 como frecuencias relativas acumulativas.

c. ¿Cuál es su mejor estimación para el verdadero $P($ $)$? Explique.

d. Si lanzaran al aire merengues sin envoltura, ¿cuál piensa usted que sea la probabilidad de que caigan "con la punta hacia arriba"? ¿Será diferente? Explique.

e. Desenvuelva los merengues de chocolate usados en la parte b y repita el experimento.

f. Los resultados de la parte e ¿son lo que usted anticipó? Explique.

4.39 Una caja contiene canicas de cinco colores diferentes: rojo, verde, azul, amarillo y morado. Hay un número igual de cada color. Asigne probabilidades a cada color en el espacio muestral.

4.40 Suponga que una caja de canicas contiene números iguales de canicas rojas y canicas amarillas, pero el doble de canicas verdes que las de hay de rojas. Saque una canica de la caja y observe su color. Asigne probabilidades a los elementos en el espacio muestral.

4.41 Si cuatro veces más estudiantes pasan un curso de estadística que los que reprueban y al azar se selecciona un estudiante de estadística, ¿cuál es la probabilidad de que el estudiante pase estadística?

4.42 Los eventos A, B y C se definen como espacio muestral S. Sus conjuntos correspondientes de puntos muestrales no se cruzan, y su unión es S. Además, el evento B tiene el doble de probabilidad de ocurrir que el evento A, y el evento C tiene el doble de probabilidad de ocurrir que el evento B. Determine la probabilidad de cada uno de los tres eventos.

4.43 Las posibilidades de que el equipo Patriotas gane el super tazón del año próximo son de 1 a 12.

a. ¿Cuál es la probabilidad de que los Patriotas ganen el super tazón el año próximo?

b. ¿Cuáles son las posibilidades en contra de que los Patriotas ganen el súper tazón del año próximo?

4.44 La temporada de baloncesto varonil de la NCAA se inicia con 327 equipos universitarios que sueñan todos con llegar al "gran baile" y alcanzar el campeonato nacional. Se seleccionan 64 equipos para el torneo, y sólo uno lo gana todo.

a. ¿Cuáles son las posibilidades contra un equipo para que sea seleccionado para el torneo?

b. ¿Cuáles son las posibilidades de que un equipo que está en el torneo gane el campeonato nacional?

c. ¡Espere! ¿Qué suposición hizo usted para contestar las partes a y b? ¿Le parece que esto es realista?

4.45 En el juego de bridge de naipes, las posibilidades en contra de que usted reciba una "mano" de "contrato" que tenga 13 cartas del mismo "palo" son 158 753 389 899 a 1. Las posibilidades en contra de que reciba una "mano" de "escalera" cuando juegue póker son 649 739 a 1.

a. ¿Cuál es la probabilidad de que reciba una mano de "contrato" que contenga 13 cartas todas del mismo palo?

b. ¿Cuál es la probabilidad de que reciba una mano de "escalera" en póker?

c. Exprese las respuestas a las partes a y b en notación científica (potencias de 10).

4.46 Al nivel mundial, el porcentaje de muertes por embarazo (riesgo en mujeres de morir desde la concepción al nacimiento) es 1 a 233. Por regiones en el mundo, este porcentaje es como sigue: América del Norte, 1 en 3700; Europa del Norte, 1 en 4000; África, 1 en 16, Asia, 1 en 65; y Latinoamérica/Caribe, 1 en 130. Exprese el riesgo de muertes por embarazo como (i) posibilidades a favor de morir, (ii) posibilidades en contra de morir, y (iii) probabilidad de morir para cada uno de lo siguiente:

a. En el mundo

b. América del Norte

c. Europa del Norte

d. África

e. Asia

f. Latinoamérica/Caribe

4.47 a. Una moneda balanceada se lanza dos veces al aire. Haga una lista del espacio muestral que muestre los posibles resultados.

b. Una moneda cargada (a favor de las caras en proporción de 3 a 1) se lanza dos veces al aire. Haga una lista del espacio muestral que indique los posibles resultados.

4.48 Una caja guardada en un almacén contiene 100 unidades de una pieza específica, de las cuales 10 son defectuosas y 90 en buenas condiciones. Se seleccionan tres piezas sin sustitución. Construya un diagrama de árbol que represente el espacio muestral.

4.3 Probabilidad condicional de eventos

Muchas de las probabilidades que vemos o escuchamos que se usan a diario son el resultado de condiciones que existen en ese momento. En esta sección aprenderemos acerca de las probabilidades condicionales.

Probabilidad condicional de que un evento ocurrirá: una probabilidad condicional es la frecuencia relativa con la cual un evento puede esperarse que ocurra, bajo la condición de que se conozca información preexistente acerca de algún otro evento. $P(A|B)$ se usa para simbolizar la probabilidad de que el evento A ocurra bajo la condición de que se sepa que el evento B ya existe.

Algunas formas de decir o expresar la probabilidad condicional, $P(A|B)$, son:

La *"probabilidad de A, dada B"*.

La *"probabilidad de A, conociendo B"*.

La *"probabilidad de que A ocurra, sabiendo que B ya ha ocurrido."*

El concepto de probabilidad condicional es en realidad muy conocido y se presenta con frecuencia sin que estemos conscientes de ello. Los medios noticiosos con frecuencia informan de muchos valores de probabilidad condicional, pero nadie lo dice y pasa sólo como aritmética de todos los días, como se ilustra en el siguiente ejemplo.

EJEMPLO 4.8 Para hallar probabilidades desde una tabla de porcentajes

De un sondeo de salida para elección nacional hecho a 13,660 votantes en 250 distritos electorales en todo el país, el 2 de noviembre de 2004, tenemos lo siguiente:

Género	Porcentaje de votantes	Porcentaje para Bush	Porcentaje para Kerry	Porcentaje por otros
Hombres	46	55	44	1
Mujeres	54	48	51	1
Edad				
18 – 29	17	45	54	1
30 – 44	29	53	46	1
45 – 59	30	51	48	1
60 y más	24	54	46	0

Todos los porcentajes citados son al entero más cercano.

Una persona ha de ser seleccionada al azar de la muestra de 13,600 votantes. Con el uso de la tabla, encuentre la respuesta a las siguientes preguntas de probabilidad.

1. ¿Cuál es la probabilidad de que la persona seleccionada sea hombre? Su respuesta: 0.46.

   Expresado en forma de ecuación: P(volante seleccionado es hombre) = 0.46.
2. ¿Cuál es la probabilidad de que la persona seleccionada tenga de 18 a 29 años de edad? Su respuesta: 0.17.

   Expresado en forma de ecuación: P(votando seleccionado es tiene entre 18 y 29 años) = 0.17.
3. Sabiendo que el votante seleccionado fue mujer, ¿cuál es la probabilidad de que ella votó por Kerry? Su respuesta: 0.51.

   Expresado en forma de ecuación: $P(\text{Kerry}|\text{mujer}) = 0.51$.
4. ¿Cuál es la probabilidad de que la persona seleccionada votó por Bush si el votante tenía 60 años o más? Respuesta: 0.54.

   Expresado en forma de ecuación: $P(\text{Bush}|\ 60 \text{ o más}) = 0.54$.

Nota: las primeras dos son probabilidades sencillas, mientras que las últimas dos son probabilidades condicionales.

EJEMPLO 4.9 Para hallar probabilidades condicionales de la tabla de datos de cuenta

De un sondeo de salida para elección nacional hecho a 1000 votantes en 25 distritos electorales en todo el país, el 2 de noviembre de 2004, tenemos lo siguiente:

Educación	Número por Bush	Número por Kerry	Número por otros	Número de votantes
Sin preparatoria	19	20	1	40
Egresado de preparatoria	114	103	3	220
Universitario inconcluso	172	147	1	320
Universitario egresado	135	119	6	260
Posgraduado	70	88	2	160
Total	510	477	13	1000

Una persona ha de ser seleccionada al azar de la muestra precedente de 1000 votantes. Con el uso de la tabla, encuentre la respuesta a las siguientes preguntas de probabilidad.

1. Sabiendo que el votante seleccionado era graduado de preparatoria, ¿cuál es la probabilidad de que la persona votó por Kerry? Respuesta: 103/220 = 0.46818 = 0.47.

   Expresado en forma de ecuación:

   P(Kerry|egresado de preparatoria) = 103/220 = 0.46818 = <u>0.47</u>.
2. Sabiendo que el votante seleccionado tenía alguna educación universitaria, ¿cuál es la probabilidad de que la persona votó por Bush? Respuesta: 172/220 = 0.5375 = 0.54.

   Expresado en forma de ecuación: P(Bush|universidad inconclusa) = 172/320 = 0.5375 = <u>0.54</u>
3. Sabiendo que la persona seleccionada votó por Kerry, ¿cuál es la probabilidad de que el votante tenga educación de postgrado? Respuesta: 88/147 = 0.1844 = 0.18.

   Expresado en forma de ecuación: P(postgrado|Kerry) = 88/447 = 0.1844 = <u>0.18</u>.
4. Sabiendo que la persona seleccionada votó por Bush, ¿cuál es la probabilidad de que el votante no tenga educación de preparatoria? Respuesta: 19/510 = 0.0372 = 0.04

   Expresado en forma de ecuación:

   P(sin preparatoria|Bush) = 19/510 = 0.0372 = <u>0.04</u>.

Notas:

1. La notación de probabilidad condicional es muy informativa y útil. Cuando se expresa una probabilidad condicional en forma de ecuación, es ventajoso usar la notación más completa. En esta forma, cuando se lea la información, toda ésta está ahí.
2. Cuando se encuentre una probabilidad condicional, algunos resultados de la lista de posibles resultados serán eliminados como posibilidades tan pronto como la condición se conozca. Considere la pregunta 4 del ejemplo 4.9. Tan pronto como la condicional expresó "sabiendo que la persona seleccionada votó por Bush", los 477 que votaron por Kerry y los 13 que votaron por Otros fueron eliminados, dejando los 510 posibles resultados.

SECCIÓN 4.3 EJERCICIOS

4.49 A trescientos televidentes se les preguntó si estaban satisfechos con la cobertura de un reciente desastre por TV.

	Género	
	Femenino	Masculino
Satisfecho	80	55
No satisfecho	120	45

Un televidente se ha de seleccionar al azar de entre los encuestados.

a. Encuentre P(satisfecho)

b. Encuentre P(satisfecho|femenino)

c. Encuentre P(satisfecho|masculino)

4.50 Los sábados por la mañana son horas de gran movimiento en el centro acuático Webster. Las lecciones de natación que van del nivel 2 de Cruz Roja, Habilidad Acuática Fundamental, al nivel 6 de Cruz Roja, Suficiencia en Natación y Aptitud, se ofrecen durante dos sesiones.

Nivel	*Número de personas en clase de 10:00 A.M.*	*Número de personas en clase de 11:00 A.M.*
2	16	16
3	15	11
4	9	7
5	8	3
6	0	3

Lauren, la coordinadora del programa, va a seleccionar al azar un nadador para entrevistarlo para un "spot" de la televisora local en el centro y su programa de natación. ¿Cuál es la probabilidad de que el nadador seleccionado tenga lo siguiente:

a. Una clase de nivel 4

b. La clase de 10:00 A.M.

c. Una clase de nivel 3 dada es la sesión de las 10:00 A.M.

d. La clase de 11:00 A.M. dada es la clase de nivel 5

4.51 *The World Factbook*, 2004, informa que los aeropuertos de Estados Unidos tienen el siguiente número de metros de pistas que son pavimentadas o no pavimentadas.

	Número de aeropuertos	
Total pista (metros)	Pavimentado	No pavimentado
Más de 3047	188	1
2438–3047	221	7
1524–2437	1375	160
914–1523	2383	1718
Menos de 914	961	7843
Total	5128	9729

Fuente: *The World Factbook*, January 2004, http://www.cia.gov/cia/publications/factbook/geos/us.html#People

Si uno de estos aeropuertos se selecciona al azar para inspección, ¿cuál es la probabilidad de que tendrá lo siguiente:

a. Pistas pavimentadas

b. 914 a 2437 metros de pista

c. Menos de 1524 metros de pista y no pavimentada

d. Más de 2437 metros de pista y pavimentada

e. Pista pavimentada, dado que tiene más de 1523 metros de pista

f. No pavimentada, sabiendo que tiene menos de 1524 metros de pista

g. Menos de 1524 metros de pista, dado que no está pavimentada

4.52 Durante el mes de agosto de 2002, al profesorado y personal administrativo de la universidad estatal de Boise se les pidió participar en una encuesta, para identificar el nivel general de satisfacción con la recién modificada semana de trabajo en el verano. La siguiente tabla contiene una lista de cómo los 620 entrevistados contestaron la pregunta: "¿Qué tan satisfecho está usted con el horario de verano 2002 de la universidad estatal de Boise?".

Grupo	Muy satisfecho	Un poco satisfecho	Ni satisfecho ni insatisfecho	Un poco insatisfecho	Muy insatisfecho	Total
Profesorado	65	24	21	13	9	132
Personal clasificado	190	61	16	15	2	284
Personal profesional	139	38	7	12	8	204
Todos los entrevistados	394	123	44	40	19	620

Fuente: Boise State University, http://www2.boisestate.edu/iassess/summer_schedule_survey.htm

Encuentre la probabilidad de lo siguiente para un entrevistado seleccionado al azar.

a. Estuvo "un poco satisfecho" con el horario de verano 2002

b. Era miembro del "personal profesional"

c. Estuvo "muy satisfecho" con el horario de verano 2002 dado que el entrevistado era miembro del profesorado

d. Era un miembro del "personal clasificado" dado que el entrevistado estuvo "muy insatisfecho" con el horario de verano 2002

4.53 Un artículo de *USA Today* titulado "Yum Brands hace dinastía en China" (7 de febrero, 2005) informa sobre cómo la Yum Brands, la empresa de restaurantes más grande del mundo, está llevando la industria de comida rápida a China, India y otros países grandes. La Yum Brands, filial de PepsiCo, ha estado entregando un crecimiento de utilidades de dos dígitos en el año pasado.

Ubicación y número de tiendas de comida rápida

Tienda	USA	En otros países	Total
KFC	5 450	7 676	13 126
Pizza Hut	6 306	4 680	10 986
Taco Bell	5 030	123	5 223
Long John Silver's	485	33	1 233
A&W All-American	485	209	694
Total	18 471	12 791	31 262

Fuente: USA Today, 7 de febrero, 2005, y Yum Brands

Supongamos que cuando el director general de Yum Brands fue entrevistado para este artículo, se le hicieron las siguientes preguntas. ¿Cómo podría haber contestado con base en la tabla siguiente?

a. ¿Qué porcentaje de sus locales está en Estados Unidos?

b. ¿Qué porcentaje de sus locales está en otros países?

c. ¿Qué porcentaje de sus tiendas son Pizza Huts?

d. ¿Qué porcentaje de sus tiendas son Taco Bell dado que la ubicación es en Estados Unidos?

e. ¿Qué porcentaje de sus locales está en otros países dado que la tienda es una A&W All-American?

f. ¿Qué porcentaje de sus tiendas es KFC dado que la ubicación es en otros países?

g. ¿Qué porcentaje de sus locales en otros países son KFC?

Repase sus respuestas a las partes f y g para contestar lo siguiente:

h. ¿Qué observa usted acerca de estas dos respuestas? ¿Por qué está ocurriendo?

4.54 En un censo en 2000, la National Highway Traffic Safety Administration reportó que, al nivel nacional, 2% de todos los muertos en accidentes de tránsito fueron de ciclistas. El sistema estatal de registros de tránsito integrados, de la Patrulla de Autopistas de California, informa que los muertos en bicicletas son 4% de los muertos en accidentes de tránsito en el estado. La información de ese reporte está resumida en la tabla siguiente.

Muertos y lesionados en bicicleta en California por grupo de edad, 2000

Edad (años)	Lesiones en bicicletas	Muertos en bicicletas	Total colisiones
0 a 4	0	14	14
5 a14	21	3 210	3 231
15 a 24	9	2 945	2 954
25 a 34	9	1 907	1 916
35 a 44	23	1 904	1 927
45 a 54	22	1 212	1 234
55 a 64	8	505	513
65 a 74	10	207	217
75 a 84	8	117	125
85 o más	3	22	25
No indicado	3	102	105
Total	116	12 145	12 261

Fuente: 2000 Statewide Integrated Traffic Records System

a. ¿Qué porcentaje de colisiones fue de muertos en bicicleta en California en 2000?

b. ¿Qué porcentaje de colisiones resultó en lesiones en bicicletas en California en 2000?

c. ¿Qué porcentaje de colisiones involucró a alguien del grupo de edad de 5 a 14 años?

d. ¿Qué porcentaje de lesiones ocurrió en bicicletas dado que había interés en sólo el segmento de edades de 35 a 44 años?

e. ¿Qué porcentaje de muertos en bicicletas involucró a alguien del grupo de edad de 75 a 84 años?

f. ¿Qué porcentaje del grupo de edades entre 15 y 24 años participó en lesiones en bicicletas?

4.55 El American Housing Survey publica sus descubrimientos acerca de los principales medios de transportes al trabajo, por trabajador, en Washington, DC, durante el año de 2001.

Medios de transporte	Número (miles)
Todos los trabajadores	120 191
Automóvil	105 586
Conduce él mismo	93 942
Auto colectivo	11 644
2 personas	9 036
3 personas	1 635
4 + personas	973
Transporte público	5 627
Taxi	133
Bicicleta o motocicleta	847
Sólo camina	3 408
Otros medios	1 049
Trabaja en casa	3 401

NOTA: Principales medios de transporte se refiere al modo usado con más frecuencia por las personas.

1. Transporte público se refiere a autobús, tranvía, metro o trenes elevados.
2. Otros medios incluyen transbordadores, trenes de superficie y servicio colectivo.

Fuente: U.S.Department of Housing and Urban Development, American Housing Survey, Washington, DC,2001, http://www.infoplease.com/ipa/A0908113.html

a. La columna de total no está incluida porque serían valores que no tienen sentido. Examine la tabla y explique por qué.

Una persona ha de ser seleccionada y se le hacen más preguntas como parte de esta encuesta. Si esa persona se selecciona al azar, encuentre la probabilidad de cada uno de los siguientes eventos.

b. La persona seleccionada es miembro de un auto colectivo de dos personas.

c. La persona elegida es miembro de un auto colectivo de dos personas dado que la persona viaja en auto colectivo.

d. La persona seleccionada no llega en auto.

e. La persona elegida usa transporte público sabiendo que esa persona no usa automóvil.

4.56 Los cinco colores más importantes para autos de lujo, manufacturados durante el año 2003 en América del Norte, se presentan aquí en porcentajes.

Auto de lujo	Porcentaje	Auto de lujo	Porcentaje
1. Med./Dk. Gray	23.30	4. White	12.6
2. Silver	18.8	5. Black	10.9
3. White Met.	17.8		

Fuente: DuPont Herberts Automotive Systems, Troy, Michigan, 2003 DuPont Automotive Color Popularity Survey Results, http://www.infoplease.com/ipa/A0855652.html

a. ¿Por qué no totaliza 100% la columna de porcentajes?

b. ¿Por qué son condicionales todas las probabilidades basadas en esta tabla? ¿Cuál es la condición?

c. ¿Su color favorito aparece en la lista?

Si un auto de lujo 2003 se seleccionó al azar de todos los autos de lujo manufacturados en Estados Unidos en 2003, determine la probabilidad de que su color sea lo siguiente:

d. Negro, plateado, gris, o blanco

e. No sea blanco

f. Negro, sabiendo que el auto de lujo tiene uno de los cinco colores más populares

g. Negro, sabiendo que el auto de lujo tiene uno de los cinco colores más populares pero no es blanco.

4.4 Reglas de probabilidad

Con frecuencia, uno desea saber la probabilidad de un evento compuesto y los únicos datos de que se dispone son las probabilidades de los eventos simples relacionados. (Los eventos compuestos son combinaciones de más de un evento simple.) En los siguientes párrafos se resume la relación entre estas probabilidades.

Para hallar la probabilidad de "no A"

El concepto de eventos complementarios es fundamental para hallar la probabilidad de "no A".

Eventos complementarios: el *complemento de un evento A*, $\overline{A}$, es el conjunto de todos los puntos muestrales del espacio muestral que no pertenecen al evento A.

Nota: el complemento del evento A se denota por $\overline{A}$ (léase "A complemento").

Unos pocos ejemplos de eventos complementarios son (1) el complemento del evento "éxito" es "fracaso," (2) el complemento de "votante seleccionado es republicano" es "votante seleccionado no es republicano", y (3) el complemento de "no caras" de 10 tiros de una moneda es "al menos una cara".

Al combinar la información de la definición de complemento con la Propiedad 2 (p. 212), podemos decir que

$$P(\mathrm{A}) + P(\overline{\mathrm{A}}) = 1.0 \text{ para cualquier evento A}$$

Como resultado de esta relación, tenemos la regla de complemento:

Regla de complemento

En palabras: *probabilidad de A complemento = uno − probabilidad de A*

En álgebra: $P(\overline{\mathrm{A}}) = 1 - P(\mathrm{A})$

(4.3)

Nota: todo evento A tiene un evento complementario $\overline{A}$. Las probabilidades complementarias son muy útiles cuando la pregunta pide la probabilidad de "al menos uno". Generalmente, esto representa una combinación de varios eventos, pero el evento complementario "ninguno" es un solo resultado. Es más fácil despejar el evento complementario y obtener la respuesta con el uso de la fórmula (4.3).

EJEMPLO 4.10

Uso de complementos para hallar probabilidades

Se lanzan dos dados. ¿Cuál es la probabilidad de que la suma sea al menos 3 (es decir, 3, 4, 5,. . . ,12)?

SOLUCIÓN Supongamos que un dado es negro y el otro es blanco. (Vea la tabla del ejemplo 4, páginas 208-209; presenta los 36 pares posibles de resultados cuando se lanza un par de dados.)

Más que hallar la probabilidad para cada una de las sumas 3, 4, 5,. . . ,12 por separado y sumar, es mucho más fácil hallar la probabilidad de que la suma es 2 ("menos a 3)" y luego usar la fórmula (4.3) para hallar la probabilidad de "al menos 3," porque "menor a 3" y "al menos 3" son eventos complementarios.

$P(\text{suma de 2}) = P(\mathrm{A}) = \dfrac{1}{36}$ ("2" ocurre sólo una vez en el espacio muestral de 36 puntos)

$P(\text{la suma es al menos 3}) = P(\overline{\mathrm{A}}) = 1 - P(\mathrm{A}) = 1 - \dfrac{1}{36} = \dfrac{35}{36}$ [usando la fórmula (4.3)]

Para hallar la probabilidad de "A o B"

Un trabajador que se emplea por hora desea calcular las probabilidades de "recibir una promoción u obtener un aumento de sueldo". El trabajador estaría feliz con cualquiera de estos resultados. Existe información histórica que permitirá que el trabajador estime la probabilidad de "recibir una promoción" y "obtener un aumento de sueldo" separadamente. En esta sección aprenderemos a aplicar la regla de la adición para hallar la probabilidad compuesta de interés.

Regla general de la adición

Sean A y B dos eventos definidos en un espacio muestral, *S*.

En palabras: *probabilidad de A o B = probabilidad de A + probabilidad de B − probabilidad de A y B*

En álgebra: $P(\text{A o B}) = P(\text{A}) + P(\text{B}) - P(\text{A y B})$

(4.4)

Para ver si funciona la relación expresada por la regla general de la adición, veamos el ejemplo 4.11.

EJEMPLO 4.11

Comprensión de la regla de la adición

Se lleva a cabo una encuesta de 800 votantes registrados en 25 distritos electorales del estado de Nueva York. Cada votante fue identificado como registrado como republicano, demócrata y otro y luego se le pregunta", ¿Está usted a favor o en contra del proyecto actual de presupuesto en espera de firma del gobernador?" Los totales resultantes se muestran a continuación.

	Número a favor	Número en contra	Número de votantes
Republicano	136	88	224
Demócrata	314	212	526
Otros	14	36	50
Total	464	336	800

Suponga que un votante se selecciona al azar de los 800 votantes resumidos en la tabla precedente. Consideremos los dos eventos: "El votante seleccionado está a favor" y "El votante es republicano". Supongamos, además, que un votante se escoge al azar de estos 800 votantes; encuentre las cuatro probabilidades: *P*(a favor), *P*(republicano), *P*(a favor o republicano), y *P*(a favor y republicano). A continuación use los resultados para comprobar la verdad de la regla de la adición.

SOLUCIÓN

La probabilidad de que el votante seleccionado sea "a favor" = *P*(a favor) = 464/800 = 0.58.

La probabilidad de que el votante seleccionado sea "republicano" = *P*(republicano) = 224/800 = 0.28

La probabilidad de que el votante seleccionado sea "a favor o republicano" = *P*(a favor o republicano) = (136 + 314 + 14 + 88)/800 = 552/800 = 0.69.

La probabilidad de que el votante seleccionado sea "a favor" y "republicano" = P(a favor y republicano) = 136/800 = 0.17.

Notas sobre las probabilidades precedentes:

1. La conectiva "o" significa "una o la otra o ambas"; así, "a favor o republicano" significa todos los votantes que satisfacen cualquiera de estos eventos.
2. La conectiva "y" significa "ambos" o "en común"; así, "a favor y republicano" significa todos los votantes que satisfacen ambos eventos.

Ahora usemos las probabilidades precedentes para demostrar la verdad de la regla de la adición.

Sea A = "a favor" y B = "republicano". La regla general de la adición entonces se convierte en:

P(a favor o republicano) = P(a favor) + P(republicano) − P(a favor y republicano)

Recuerde: Previamente encontramos: P(a favor o republicano) = 0.69.
Usando las otras tres probabilidades, vemos:

P(a favor) + P(republicano) − P(a favor y republicano) = 0.58 + 0.28 − 0.17 = 0.69.

Entonces, obtenemos respuestas idénticas al aplicar la regla de la adición y consultar las celdas relevantes de la tabla. Por lo general no se tiene la opción de hallar la P(A o B) en dos formas, como lo hicimos aquí. Se nos pide hallar P(A o B) empezando con la P(A) y P(B). No obstante, se necesita una tercera pieza de información. En la situación previa, necesitamos P(A y B). Se hará necesario conocer P(A y B) o alguna información que nos permita hallarla.

Para hallar la probabilidad de "A y B"

Supongamos que un profesor de justicia criminal desea que sus alumnos determinen la semejanza del evento "a un conductor se le levanta infracción por exceso de velocidad y el conductor había asistido previamente a una clase de manejo defensivo". Los estudiantes tienen confianza en hallar las probabilidades de "a un conductor se le levanta infracción por exceso de velocidad" y "un conductor ha asistido a una clase de manejo defensivo" por separado. En esta sección aprenderemos a aplicar la regla de la multiplicación para hallar la probabilidad compuesta de interés.

Regla general de la multiplicación

Sean A y B dos eventos definidos en el espacio muestral S.

En palabras: *probabilidad de A y B = probabilidad de A × probabilidad de B, conociendo A*

En álgebra: $P(\text{A y B}) = P(\text{A}) \cdot P(\text{B}|\text{A})$

(4.5)

Nota: cuando intervienen dos eventos, cualquiera de ellos puede ser identificado como A, con el otro identificado como B. La regla general de la multiplicación también podría escribirse como $P(\text{B y A}) = P(\text{B}) \cdot P(\text{A}|\text{B})$.

EJEMPLO 4.12 Comprensión de la regla de la multiplicación

Se lleva a cabo una encuesta estatal de 800 votantes registrados en 25 distritos electorales del estado de Nueva York. Cada votante fue identificado como registrado como republicano, demócrata y otro y luego se le pregunta: "¿Está usted a favor o en contra del proyecto actual de presupuesto en espera de firma del gobernador?" Los totales resultantes se muestran a continuación.

	Número a favor	Número en contra	Número de votantes
Republicano	136	88	224
Demócrata	314	212	526
Otros	14	36	50
Total	464	336	800

Suponga que un votante se selecciona al azar de los 800 votantes resumidos en la tabla precedente. Consideremos los dos eventos: "El votante seleccionado está a favor" y "El votante es republicano". Supongamos, además, que un votante se escoge al azar de estos 800 votantes; encuentre las tres probabilidades: P(a favor), P(republicano|a favor), P(a favor y republicano). A continuación use los resultados para comprobar la verdad de la regla de la multiplicación.

SOLUCIÓN

La probabilidad de que el votante seleccionado sea "a favor" = P(a favor) =

$$464/800 = \frac{464}{800}.$$

La probabilidad de que el votante seleccionado sea "republicano|dado a favor"

$$= P(\text{republicano}|\text{a favor}) = 136/464 = \frac{136}{464}.$$

La probabilidad de que el votante seleccionado sea "a favor" y "republicano"

$$= P(\text{a favor y republicano}) = 136/800 = \frac{136}{800} = \underline{0.17}$$

Notas relacionadas con el cálculo de las probabilidades anteriores:

1. La condicional "dado" significa que hay una restricción; así, "republicano|a favor" significa que comenzamos con sólo los votantes que están "a favor." En este caso, esto significa que vemos sólo a 464 votantes cuando determinemos esta probabilidad.
2. La conectiva "y" significa "ambos" o "en común"; así, "a favor y republicano" significa todos los votantes que satisfacen ambos eventos.

Usemos ahora las probabilidades previas para demostrar la verdad de la regla de la multiplicación.

Sea A = "a favor" y B = "republicano". La regla general de la multiplicación se convierte entonces en:

$$P(\text{a favor y republicano}) = P(\text{a favor}) \cdot P(\text{republicano}|\text{a favor})$$

Previamente encontramos: $P(\text{a favor y republicano}) = \dfrac{136}{800} = \underline{0.17}$.
Usando las otras dos probabilidades, vemos:

$$P(\text{a favor})P(\text{republicano}|\text{a favor}) = \frac{464}{800} \cdot \frac{136}{464} = \frac{136}{800} = \underline{0.17}.$$

Por lo general no hay la opción de hallar la P(A y B) en dos formas, como lo hicimos aquí. Cuando se nos pide hallar P(A y B), con frecuencia se nos da P(A) y P(B). No obstante, no siempre se obtiene la respuesta correcta con sólo multiplicar esas dos probabilidades. Será necesaria una tercera pieza de información; se necesita la probabilidad condicional de uno de los dos eventos o información que permita hallarla.

EJEMPLO 4.13 Sacar sin sustitución

En un juego de carnaval, el jugador saca a ciegas una canica de color a la vez, de una caja que contiene dos canicas rojas y cuatro azules. La canica escogida no se regresa a la caja después de ser seleccionada, es decir, cada saque se hace sin sustitución. Las canicas se mezclan antes de cada saque. Cuesta \$1 jugar, y si las dos primeras canicas sacadas son rojas, el jugador recibe un premio de \$2. Si las primeras cuatro canicas sacadas son todas azules, el jugador recibe un premio de \$5. De otro modo, no hay premio. Para hallar la probabilidad de ganar un premio, veamos primero la probabilidad de sacar roja o azul en saques consecutivos y organicemos la información en un diagrama de árbol.

En el primer saque (representado por los segmentos de ramas moradas en la figura 4.5), la probabilidad de rojas es dos de seis, 2/6 o 1/3, mientras que la probabilidad de azules es 4/6, o 2/3. Debido a que no se sustituye la canica, sólo cinco canicas quedan en la caja; el número de cada color restante depende del color de la primera canica que se saque. Si la primera canica fue roja, entonces las probabilidades son 1/5 y 4/5 como se ve en el diagrama de árbol (segmentos de ramas verdes en la figura 4.5). Si la primera canica fue azul, entonces las probabilidades son 2/5 y 3/5 como se muestra en el diagrama de árbol (segmentos de ramas naranja en la figura 4.5). Las probabilidades cambian con cada saque, porque el número de canicas disponible sigue decreciendo a medida que tiene lugar cada saque. El diagrama de árbol es una excelente ayuda gráfica para dar seguimiento al avance.

FIGURA 4.5 Diagrama de árbol: primeros dos saques, juego de carnaval

La probabilidad de ganar el premio de \$2 se puede hallar ahora usando la fórmula (4.5):

$$P(\text{A y B}) = P(\text{A}) \cdot P(\text{B}|\text{A})$$

$$P(\text{ganando \$2}) = P(\text{R}_1 \text{ y R}_2) = P(\text{R}_1) \cdot P(\text{R}_2 \,|\, \text{R}_1) = \frac{2}{6} \cdot \frac{1}{5} = \frac{1}{15} = 0.067$$

(Ganar el premio de \$5 se deja como ejercicio 4.77.)

Nota: el diagrama de árbol, cuando se le aplican leyendas, tiene las probabilidades necesarias para multiplicar enumeradas en la rama que representa el esfuerzo ganador.

SECCIÓN 4.4 EJERCICIOS

4.57 a. Si la probabilidad de que ocurra el evento A durante un experimento es 0.7, ¿cuál es la probabilidad de que el evento A no ocurra durante ese experimento?

b. Si los resultados de un experimento de probabilidad pueden ser cualquier entero de 16 a 28, y la probabilidad de que el entero sea menor a 20 es 0.78, ¿cuál es la probabilidad de que el entero sea 20 o más?

4.58 a. Si la probabilidad de que usted apruebe el siguiente examen de estadística es evaluado con toda precisión en 0.75, ¿cuál es la probabilidad de que no apruebe el siguiente examen de estadística?

b. El pronóstico del clima predice que hay un "70 por ciento" de probabilidad de menos de 1 pulgada de lluvia durante el siguiente periodo de 30 días. ¿Cuál es la probabilidad de al menos 1 pulgada de lluvia en los siguientes 30 días?

4.59 Según el U.S. Pet Ownership & Demographic Sourcebook, de Alimento de Pedigree para Perros (USA Snapshot "Casi todos dan amor de cachorro a sólo uno", 22 de febrero, 2005), alrededor de 66% de todos los dueños estadounidenses de perros (casi 60 millones de personas) son dueñas de un perro. Con base en esta información, encuentre la probabilidad de que el propietario de un perro tenga más de un perro.

4.60 Según el Sleep Channel (http://www.sleepchanel.net, septiembre 2002), la apnea del sueño afecta a 12 millones de individuos en Estados Unidos. Esta afección del sueño interrumpe la respiración y puede despertar a quienes la sufren hasta cinco veces en una hora. Numerosas personas no reconocen este mal aun cuando produce fuerte ronquido. Suponiendo que haya 275 millones de habitantes en Estados Unidos, ¿cuál es la probabilidad de que un individuo escogido al azar no esté afectado por la apnea del sueño?

4.61 Si $P(\text{A}) = 0.4$, $P(\text{B}) = 0.5$, y $P(\text{A y B}) = 0.1$, hállese $P(\text{A o B})$.

4.62 Si $P(\text{A}) = 0.5$, $P(\text{B}) = 0.3$, y $P(\text{A y B}) = 0.2$, hállese $P(\text{A o B})$.

4.63 Si $P(\text{A}) = 0.4$, $P(\text{B}) = 0.5$, y $P(\text{A y B}) = 0.7$, hállese $P(\text{A o B})$.

4.64 Si $P(\text{A}) = 0.4$, $P(\text{A o B}) = 0.9$, y $P(\text{A y B}) = 0.1$, hállese $P(\text{B})$.

4.65 La industria de entretenimiento de deportes emplea atletas, entrenadores, árbitros y trabajadores del ramo. De éstos, 0.37 trabajan a tiempo parcial y 0.50 ganan más de \$20 540 al año. Si 0.32 de estos empleados trabajan a tiempo completo y ganan más de \$20 540, ¿qué proporción de los empleados de la industria trabajan a tiempo completo o ganan más de \$20 540?

4.66 Jason asiste a reuniones de su secundaria. De los que asisten, 50% son mujeres. El conocimiento común dice que 88% de las personas son derechas. Siendo zurdo, Jasón sabe que de un grupo grande de personas, sólo alrededor de 6% son zurdos, hombres. Si Jasón habla a la primera persona que se encuentre en la reunión, ¿cuál es la probabilidad de que la persona sea hombre o zurdo?

4.67 Un establecimiento de piezas de repuesto vende piezas nuevas y usadas. Sesenta por ciento de las pie-

zas en existencia son usadas. Sesenta y uno por ciento son piezas usadas o defectuosas. Si 5% de las piezas del establecimiento son defectuosas, ¿qué porcentaje es de partes usadas y defectuosas? Resuelva con el uso de fórmulas. Compare su solución con su respuesta al ejercicio 4.27.

4.68 Unos dirigentes sindicales informan que 60% de los trabajadores de una gran fábrica pertenecen al sindicato, 90% ganan más de $12 por hora, y 40% pertenecen al sindicato y ganan más de $12 por hora. ¿Cree usted estos porcentajes? Explique. Resuelva usando fórmulas. Compare su solución con su respuesta al ejercicio 4.28.

4.69 A y B son eventos definidos en un espacio muestral, con $P(\text{A}) = 0.7$ y $P(\text{B}|\text{A}) = 0.4$. Encuentre $P(\text{A y B})$.

4.70 A y B son eventos definidos en un espacio muestral, con $P(\text{A}|\text{B}) = 0.5$ y $P(\text{B}) = 0.8$. Encuentre $P(\text{A y B})$.

4.71 A y B son eventos definidos en un espacio muestral, con $P(\text{A}) = 0.6$ y $P(\text{A y B}) = 0.3$. Encuentre $P(\text{B}|\text{A})$.

4.72 A y B son eventos definidos en un espacio muestral, con $P(\text{B}) = 0.4$ y $P(\text{A y B}) = 0.5$. Encuentre $P(\text{A}|\text{B})$.

4.73 Se sabe que los esteroides dan a usuarios una ventaja en concursos atléticos, pero también se sabe que su uso está prohibido en atletas. Como consecuencia de esto, se ha instituido un programa de prueba de esteroides y a atletas se les practica una prueba al azar. Se cree que los procedimientos de prueba son igualmente eficaces en usuarios y no usuarios y se dice que son 98% precisos. Si 90% de los atletas afectados por este programa de prueba son limpios, ¿cuál es la probabilidad de que el siguiente atleta probado sea usuario y no pase la prueba?

4.74 Juan vive en una gran ciudad y viaja diariamente a su trabajo en transporte colectivo metro o en taxi. Toma el metro 80% del tiempo porque cuesta menos, y toma taxi el otro 20% del tiempo. Cuando toma el metro, llega a su trabajo a tiempo 70% de las veces, mientras que llega a tiempo 90% de las veces cuando viaja en taxi.

a. ¿Cuál es la probabilidad de que Juan tome el metro y llegue a su trabajo a tiempo en cualquier día determinado?

b. ¿Cuál es la probabilidad de que Juan tome un taxi y llegue a su trabajo a tiempo en cualquier día determinado?

4.75 A nadie le gusta pagar impuestos, pero ésta no es la forma de salirse. Se cree que 10% de contribuyentes intencionalmente solicitan algunas deducciones a las que no tienen derecho. Si 9% de todos los contribuyentes intencionalmente solicitan deducciones extra y las niegan cuando se les hacen auditorías, encuentre la probabilidad de que un contribuyente que toma intencionalmente deducciones extra las niegue.

4.76 A Casey le gusta mucho tomar café a media mañana y siempre para en una de sus cafeterías preferidas para tomarlo. Cuando toma comida para llevar, hay un 0.6 de probabilidad de que siempre coma un pastel. Toma café y pastel como comida para llevar con una probabilidad de 0.48. ¿Cuál es la probabilidad de que tome comida para llevar?

4.77 Encuentre la probabilidad de ganar $5 si juega en el juego de carnaval que se describe en el ejemplo 4.13.

a. Complete las ramas del diagrama de árbol iniciado en la figura 4.5, citando las probabilidades para todos los saques posibles.

b. ¿Cuál es la probabilidad de sacar una canica roja en el segundo saque? ¿Qué información adicional es necesaria para hallar la probabilidad? ¿Qué "condiciones" podrían existir?

c. Calcule la probabilidad de ganar el premio de $5.

d. ¿Es el premio de $2 o el de $5 más difícil de ganar? ¿Cuál es más probable? Justifique su respuesta.

4.78 Suponga que las reglas para el juego de carnaval del ejemplo 4.13 se modificaron para que la canica sacada cada vez sea devuelta a la caja antes del siguiente saque.

a. Vuelva a dibujar el diagrama de árbol trazado para el ejercicio 4.77, citando las probabilidades para el juego cuando se jugó "con restitución".

b. ¿Cuál es la probabilidad de sacar una canica roja en el segundo saque? ¿Qué información adicional es necesaria para hallar la probabilidad? ¿Qué efecto tiene esto en P(roja)?

c. Calcule la probabilidad de ganar el premio de \$2.

d. Calcule la probabilidad de ganar el premio de \$5.

e. Cuando el juego se juega sin sustitución, ¿es más difícil ganar el premio de \$2 o el de \$5? ¿Cuál es más probable? Justifique su respuesta.

4.79 Suponga que A y B son eventos definidos en un espacio muestral común y que se conocen las siguientes probabilidades: $P(\text{A}) = 0.3$, $P(\text{B}) = 0.4$, y $P(\text{A}|\text{B}) = 0.2$. Encuentre P(A o B).

4.80 Suponga que A y B son eventos definidos en un espacio muestral común y que se conocen las siguientes probabilidades: $P(\text{A o B}) = 0.7$, $P(\text{B}) = 0.5$, y $P(\text{A}|\text{B}) = 0.2$. Encuentre P(A).

4.81 Suponga que A y B son eventos definidos en un espacio muestral común y que se conocen las siguientes probabilidades: $P(\text{A}) = 0.4$, $P(\text{B}) = 0.3$, y $P(\text{A o B}) = 0.66$. Encuentre $P(\text{A}|\text{B})$.

4.82 Suponga que A y B son eventos definidos en un espacio muestral común y que se conocen las siguientes probabilidades: $P(\text{A}) = 0.5$, $P(\text{A y B}) = 0.24$, y $P(\text{A}|\text{B}) = 0.4$. Encuentre P(A o B).

4.83 Dado $P(\text{A o B}) = 1.0$, $P(\overline{\text{A y B}}) = 0.7$, y $P(\overline{\text{B}}) = 0.4$, encuentre:

a. $P(\text{B})$ b. $P(\text{A})$ c. $P(\text{A}|\text{B})$

4.84 Dado $P(\text{A o B}) = 1.0$, $P(\overline{\text{A y B}}) = 0.3$, y $P(\overline{\text{B}}) = 0.4$, encuentre:

a. $P(\text{B})$ b. $P(\text{A})$ c. $P(\text{A}|\text{B})$

4.85 La probabilidad de A es 0.5. La probabilidad condicional de que A ocurra dado que B ocurre es 0.25. La probabilidad condicional de que B ocurra dado que A ocurre es 0.2.

a. ¿Cuál es la probabilidad de que B ocurra?

b. ¿Cuál es la probabilidad condicional de que B no ocurra dado que A no ocurre?

4.86 La probabilidad de C es 0.4. La probabilidad condicional de que C ocurra dado que D ocurre es 0.5. La probabilidad condicional de que C ocurra dado que D no ocurre es 0.25.

a. ¿Cuál es la probabilidad de que D ocurra?

b. ¿Cuál es la probabilidad condicional de que D ocurra dado que C ocurre?

4.5 Eventos mutuamente excluyentes

Para favorecer nuestra discusión de eventos compuestos, debe introducirse el concepto de "mutuamente excluyente".

Eventos mutuamente excluyentes: eventos no vacíos definidos en el espacio muestral con cada evento excluyendo que ocurra el otro. En otras palabras, son eventos que no comparten elementos comunes.

En álgebra: $P(\text{A y B}) = 0$

En palabras: Hay varias formas equivalentes de expresar el concepto de mutuamente excluyente:

1. Si se sabe que cualquiera de los eventos ha ocurrido, entonces el otro evento está excluido o no puede haber ocurrido.
2. Si se busca en las listas de los elementos que conforman cada evento, ninguno de los elementos citados para cualquier evento aparecerá en la lista del otro evento; "no hay elementos compartidos".

3. Si se observa un diagrama de Venn, las áreas cerradas que representan cada evento "o se cruzan", esto es, "no hay elementos compartidos", o dicho en otras palabras, "están disjuntas".
4. La ecuación dice, "la intersección de los dos eventos tiene una probabilidad de cero", lo cual significa que "la intersección es un conjunto vacío" o "no hay intersección".

Nota: el concepto de eventos mutuamente excluyentes está basado en la relación entre los conjuntos de elementos que satisfacen los eventos. Mutuamente excluyente no es un concepto de probabilidad por definición, sino que es fácil para expresar el concepto usando un enunciado de probabilidad.

Veamos algunos ejemplos.

EJEMPLO 4.14

Comprensión de eventos mutuamente excluyentes

De un sondeo nacional de salida de 1000 votantes en 25 distritos electorales en todo el país, el 2 de noviembre, 2004, tenemos lo siguiente.

Educación	Número por Bush	Número por Kerry	Número por otros	Número de votantes
Sin preparatoria	19	20	1	40
Egresado de preparatoria	114	103	3	220
Universitario inconcluso	172	147	1	320
Egresado de universidad	135	119	6	260
Posgrado	70	88	2	160
Total	510	477	13	1000

Considere los dos eventos que el votante seleccionado "votó por Bush" y el votante seleccionado "votó por Kerry". Suponga que se selecciona un votante al azar de los 1000 votantes resumidos en la tabla. Para que ocurra el evento en el que el votante seleccionado "votó por Bush", el votante seleccionado debe ser 1 de los 510 votantes que aparecen en la columna "Número por Bush". Para que ocurra el evento en el que el votante seleccionado "votó por Kerry", el votante seleccionado debe ser 1 de los 477 votantes que aparecen en la columna "Número por Kerry". Debido a que ningún votante que aparece en la columna de Bush aparece en la columna de Kerry, y debido a que ningún votante que aparece en la columna de Kerry aparece en la columna de Bush, estos dos eventos son mutuamente excluyentes.

En forma de ecuación: P(votó por Bush y votó por Kerry) = 0.

EJEMPLO 4.15

Comprensión de eventos no mutuamente excluyentes

De un sondeo nacional de salida de 1000 votantes en 25 distritos electorales en todo el país, el 2 de noviembre, 2004, tenemos lo siguiente.

Educación	Número por Bus	Número por Kerry	Número por otros	Número de votantes
Sin preparatoria	19	20	1	40
Egresado de preparatoria	114	103	3	220
Universitario inconcluso	172	147	1	320
Egresado de universidad	135	119	6	260
Postgrado	70	88	2	160
Total	510	477	13	1000

Considere los dos eventos en que el votante seleccionado "votó por Bush" y el votante seleccionado tenía "estudios inconclusos de universidad". Suponga que un votante se selecciona al azar de los 1000 votantes resumidos en la tabla. Para que ocurra el evento en el que el votante seleccionado "votó por Bush", el votante seleccionado debe ser 1 de los 510 votantes que aparecen en la columna "Número por Bush". Para que ocurra el evento en el que el votante seleccionado tenía "estudios inconclusos de universidad", el votante seleccionado debe ser 1 de los 320 votantes que aparecen en la fila de "universitario inconcluso". Debido a que los 172 votantes que aparecen en la intersección de la columna "Número por Bush" y la fila "universitario inconcluso" pertenecen de ambos eventos (el votante seleccionado "votó por Bush" y el votante seleccionado tenía "estudios inconclusos de universidad"), estos dos eventos NO SON mutuamente excluyentes.

En forma de ecuación: P(votó por Bush y tenía estudios inconclusos de universidad) = 172/1000 = 0.172; que no es igual a cero.

EJEMPLO 4.16 Eventos mutuamente excluyentes en cartas

Considere un "monte" de barajas y los dos eventos "carta sacada es una reina" y "carta sacada es un as". El monte ha de barajarse y sacarse una carta al azar. Para que ocurra el evento "carta obtenida es una reina", la carta sacada debe ser una de las cuatro reinas: reina de corazones, reina de diamantes, reina de espadas o reina de bastos. Para que ocurra el evento "carta sacada es un as", la carta sacada debe ser uno de los cuatro ases: as de corazones, as de diamantes, as de espadas, o as de bastos. Observe que no hay carta que sea reina y as. Por tanto, estos dos eventos, "carta sacada es una reina" y "carta sacada es un as", son eventos mutuamente excluyentes.

En forma de ecuación: P(reina y as) = 0.

EJEMPLO 4.17 Eventos de cartas que no son mutuamente excluyentes

Considere un "monte" regular de barajas y los dos eventos "carta sacada es una reina" y "carta sacada es un corazón". El monte ha de barajarse y sacarse una carta al azar. Los eventos "reina" y "corazones" ¿son mutuamente excluyentes? El evento "carta sacada es una reina" consta de las cuatro reinas: reina de corazones, reina de diamantes, reina de espadas, y reina de bastos. El evento "carta sacada es un corazón" consta de los 13 corazones: as de corazones, rey de corazones, reina de corazones, comodín de corazones, y los otros nueve corazones. Nótese que la "reina de corazones" está en ambas listas, con lo que es posible que ambos eventos "carta sacada es una reina" y "carta sacada es un corazón" ocurran simultáneamente. Esto significa que, cuando ocurre uno de estos dos eventos, no excluye la posibilidad de que ocurra el otro. Estos eventos no son mutuamente excluyentes.

En forma de ecuación: P(reina y corazón) = 1/52; que no es igual a cero.

EJEMPLO 4.18 Representación visual y comprensión de eventos mutuamente excluyentes

Considere un experimento en el que se lanzan dos dados. Tres eventos se definen como sigue:

A: La suma de los números de los dos dados es 7.

B: La suma de los números en los dos dados es 10.

C: Cada uno de los dos dados muestra el mismo número.

Determinemos si estos tres eventos son mutuamente excluyentes.

Podemos exponer que tres eventos son mutuamente excluyentes al demostrar que cada par de eventos es mutuamente excluyente. ¿Los eventos A y B son mutuamente excluyentes? Sí, lo son, porque la suma de los dos dados no puede ser 7 y 10 al mismo tiempo. Si ocurre una suma de 7, es imposible que la suma sea 10.

La figura 4.6 presenta el espacio muestral para este experimento. Éste es el mismo espacio muestral del ejemplo 4.2 excepto que los pares ordenados se usan en lugar de las imágenes. Los óvalos, diamantes y rectángulos muestran los pares ordenados que están en los eventos A, B y C, respectivamente. Podemos ver que los eventos A y B no se cruzan, por lo cual son mutuamente excluyentes. El punto (5,5) de la figura 4.6 satisface ambos eventos B y C. Por tanto, B y C no son mutuamente excluyentes. Dos dados pueden mostrar un 5 cada uno, lo cual satisface a C, y el total satisface a B. Como encontramos un par de eventos que no son mutuamente excluyentes, los eventos A, B y C no son mutuamente excluyentes.

FIGURA 4.6
Usos de probabilidades empíricas

Regla especial de la adición

La regla de la adición se simplifica cuando los eventos que intervienen son mutuamente excluyentes.

Si sabemos que dos eventos son mutuamente excluyentes, entonces al aplicar $P(\text{A y B}) = 0$ a la regla adición para probabilidades, se deduce que $P(\text{A o B}) = P(\text{A}) + P(\text{B}) - P(\text{A y B})$ se convierte en $P(\text{A o B}) = P(\text{A}) + P(\text{B})$.

Regla especial de la adición

Sean A y B dos eventos mutuamente excluyentes definidos en un espacio muestral S.

En palabras: *probabilidad de A o B = probabilidad de A + probabilidad de B*

En álgebra: $P(\text{A o B}) = P(\text{A}) + P(\text{B})$

(4.6)

Esta fórmula se puede expandir para considerar más de dos eventos mutuamente excluyentes:

$$P(\text{A o B o C o} \ldots \text{E}) = P(\text{A}) + P(\text{B}) + P(\text{C}) + \ldots + P(\text{E})$$

Con frecuencia, esta ecuación es conveniente para calcular probabilidades, pero no nos ayuda a comprender la relación entre los eventos A y B. Es la *definición* la que nos dice cómo debemos pensar acerca de eventos mutuamente excluyentes. Los estudiantes que comprenden la exclusividad mutua en esta forma captan la idea de lo que significa la exclusividad mutua. Esto debe llevar al lector a pensar más claramente acerca de situaciones que se refieren a eventos mutuamente excluyentes, con lo cual es menos probable que confunda el concepto de eventos mutuamente excluyentes con eventos independientes (que se definen en la sección 4.6), o que cometa otros errores comunes respecto al concepto de mutuamente excluyente.

Notas:

1. Defina eventos mutuamente excluyentes en términos de los conjuntos de elementos que satisfacen los eventos y pruebe exclusividad mutua en esa forma.
2. No use $P(\text{A y B}) = 0$ como la definición de eventos mutuamente excluyentes. Es una propiedad que resulta de la definición. Se puede usar como prueba para eventos mutuamente excluyentes, pero, como enunciado, no muestra significado o idea del concepto de eventos mutuamente excluyentes.
3. En forma de ecuación, la *definición* de eventos mutuamente excluyentes expresa que:

$P(\text{A y b}) = 0$ (Ambos no pueden ocurrir al mismo tiempo.)

$P(\text{A}|\text{B}) = 0$ y $P(\text{B}|\text{A}) = 0$
(Si se sabe que uno ha ocurrido, entonces el otro no ha ocurrido.)

Reconsidere el ejemplo 4.16, con los dos eventos "carta sacada es una reina" y "carta sacada es un as" cuando saque exactamente una carta de un monte regular de cartas. La carta sacada es una reina, o la carta sacada es un as. La carta no puede ser reina y as al mismo tiempo, lo cual hace que estos dos eventos sean mutuamente excluyentes. La regla especial de la adición, en consecuencia, aplica a la situación de hallar $P(\text{reina o as})$.

$$P(\text{reina o as}) = P(\text{reina}) + P(\text{as}) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

SECCIÓN 4.5 EJERCICIOS

4.87 Determine si cada uno de los siguientes pares de eventos es mutuamente excluyente.

a. Se lanzan al aire cinco monedas: "se observa una cara," "se observa al menos una cara".

b. Un vendedor visita a un cliente y hace una venta: "la venta rebasa los $100", "la venta rebasa los $1000".

c. Un estudiante se selecciona al azar de un grupo estudiantil: la persona seleccionada es "hombre", la persona seleccionada tiene "más de 21 años de edad".

d. Se lanzan dos dados: el total que aparece es "menos de 7", el total que aparece es "más de 9".

4.88 Determine si cada uno de los siguientes conjuntos de eventos es mutuamente excluyente.

a. Se lanzan al aire cinco monedas: "se observa no más de una cara", "se observan dos caras," "se observan tres o más caras."

b. Un vendedor visita a un cliente y hace una venta: la cantidad de la venta es "menor a $100", es "entre $100 y $1000", es "más de $500."

c. Un estudiante se selecciona al azar del cuerpo estudiantil: la persona seleccionada es "mujer", es "hombre", es "mayor de 21."

d. Se lanzan dos dados: los números de puntos que aparecen en el dado son "ambos nones," "ambos pares", "totalizan 7", "totalizan 11".

4.89 Explique por qué $P(\text{A y B}) = 0$ cuando los eventos A y B son mutuamente excluyentes.

4.90 Explique por qué $P(\text{A ocurre cuando B ha ocurrido}) = 0$ cuando los eventos A y B son mutuamente excluyentes.

4.91 Si $P(\text{A}) = 0.3$ y $P(\text{B}) = 0.4$, y si A y B son eventos mutuamente excluyentes, encuentre:

a. $P(\overline{\text{A}})$ b. $P(\overline{\text{B}})$ c. $P(\text{A o B})$ d. $P(\text{A y B})$.

4.92 Si $P(\text{A}) = 0.4$ y $P(\text{B}) = 0.5$, y si A y B son eventos mutuamente excluyentes, encuentre $P(\text{A o B})$.

4.93 Un estudiante se selecciona del alumnado de su universidad. Defina los siguientes eventos: M, el estudiante seleccionado es masculino; F, el estudiante seleccionado es femenino; S, el estudiante seleccionado está registrado para estadística.

a. ¿Los eventos M y F son mutuamente excluyentes? Explique.

b. ¿Los eventos M y S son mutuamente excluyentes? Explique.

c. ¿Los eventos F y S son mutuamente excluyentes? Explique.

d. ¿Los eventos M y F son complementarios? Explique.

e. ¿Los eventos M y S son complementarios? Explique.

f. ¿Los eventos complementarios son también eventos mutuamente excluyentes? Explique.

g. ¿Los eventos mutuamente excluyentes también son eventos complementarios? Explique.

4.94 Un estudiante se selecciona al azar de un cuerpo estudiantil. Suponga que la probabilidad de que este estudiante sea mujer es 0.5 y que la probabilidad de que este estudiante trabaje a tiempo parcial es 0.6. ¿Son los eventos "femenino" y "trabaja" mutuamente excluyentes? Explique.

4.95 Se lanzan dos dados. Defina los eventos como sigue: A: suma de 7, C: duplica, E: suma de 8.

a. ¿Qué par de eventos, A y C, A y E, o C y E, son mutuamente excluyentes? Explique.

b. Encuentre las probabilidades $P(\text{A o C})$, $P(\text{A o E})$, y $P(\text{C o E})$.

4.96 El acuario de una tienda de venta de mascotas contiene 40 peces espada color naranja (22 hembras y 18 machos) y 28 colas de espada verdes (12 hembras y 16 machos). Al azar, una persona captura un pez con una red.

a. ¿Cuál es la probabilidad de que sea un pez espada color naranja?

b. ¿Cuál es la probabilidad de que sea un pez macho?

c. ¿Cuál es la probabilidad de que sea un pez espada hembra?

d. ¿Cuál es la probabilidad de que sea hembra o cola de espada verde?

e. Los eventos "hembra" y "macho" ¿son mutuamente excluyentes? Explique.

f. Los eventos "macho" y "pez espada" ¿son mutuamente excluyentes? Explique.

4.97 Las personas ¿toman lecciones de natación en piscinas cubiertas a mediados de un verano caluroso? Seguro que lo hacen en el Webster Aquatic Center. Sólo durante el mes de julio de 2004, 179 personas participaron en diversas formas de lecciones.

Categoría de natación	Diurna	Vespertina
Preescolar	26	29
Niveles	75	39
Adultos y buceo	4	6
Total	105	74

Si un nadador fue seleccionado al azar de los participantes de julio:

a. Los eventos donde está el participante seleccionado, "diurno" y "vespertino" ¿son mutuamente excluyentes? Explique.
b. Los eventos donde está el participante seleccionado, "preescolar" y "niveles" ¿son mutuamente excluyentes? Explique.
c. Los eventos donde está el participante seleccionado, "diurno" y "preescolar" ¿son mutuamente excluyentes? Explique.
d. Encuentre P(preescolar).
e. Encuentre P(diurno).
f. Encuentre P(no niveles).
g. Encuentre P(preescolar o vespertino).
h. Encuentre P(preescolar y diurno).
i. Encuentre P(diurno | niveles).
j. Encuentre P(adultos y buceo | vespertino).

4.98 Las lesiones son desafortunadamente parte de todo deporte. El baloncesto en preparatorias no es la excepción, como lo demuestra la tabla siguiente. Los porcentajes citados son el porcentaje de lesiones indicadas que ocurren a jugadores de baloncesto en preparatorias, hombres y mujeres, y el lugar de su cuerpo que se lesionó.

Si un jugador se selecciona al azar de los incluidos en la tabla:

Lugar de lesión	Hombres	Mujeres
Tobillo/pie	38.3%	36.0%
Cadera/muslo/pierna	14.7%	16.6%
Rodilla	10.3%	13.0%
Antebrazo/muñeca/mano	11.5%	11.2%
Cara/cuero cabelludo	12.2%	8.8%
Otros	13.0%	14.4%
Total	100.0%	100.0%

a. ¿Son mutuamente excluyentes los eventos en los que estaba el jugador seleccionado, "hombre" y "mujer"? Explique.
b. ¿Son mutuamente excluyentes los eventos en los que estaba el jugador seleccionado, "tobillo/pie" y "rodilla"? Explique.
c. ¿Son mutuamente excluyentes los eventos "mujer" y "cara/cuero cabelludo"? Explique.
d. Encuentre P(tobillo/pie|hombre).
e. Encuentre P(tobillo/pie|mujer).
f. Encuentre P(no en una pierna|mujer).
g. Encuentre P(rodilla o cara/cuero cabelludo|hombre).
h. Encuentre P(rodilla o cara/cuero cabelludo|mujer).
i. Explique por qué P(rodilla) para todos los jugadores de baloncesto de preparatoria no se puede hallar usando la información de la tabla. ¿Qué información adicional se hace necesaria?

4.99 La mayoría de estadounidenses, 70%, dicen que lavarse las manos con frecuencia es la mejor forma de prevenir la gripe. A pesar de eso, cuando usan baños públicos, las mujeres se lavan las manos sólo 62% de las veces y los hombres se las lavan sólo 43% de las veces. De los adultos que usan baños públicos en una gran cadena de abarrotes, 58% son mujeres. ¿Cuál es la probabilidad de que la siguiente persona que entre al baño en esta tienda se lave las manos?

4.100 Un oficial de tránsito es lo último que usted desea ver en su espejo retrovisor cuando acelera en una autopista, pero una investigación demuestra que una infracción de tránsito reduce la probabilidad de que un conductor se vea involucrado en un accidente de mortales consecuencias, al menos por unas pocas semanas. Por grupo de edades, 13.3% de todos los automovilistas tienen menos de 25 años, 58.6% tienen entre 25 y 54 años, y 28.1% tienen 55 o más. Las estadísticas muestran que 1.6% de los automovilistas que tienen menos de 25 años de edad, 2.2% de 25 a 54, y 0.5% de 55 o más años tendrá un accidente en el mes siguiente. ¿Cuál es la probabilidad de que un conductor identificado al azar tenga un accidente el mes próximo?

4.6 Eventos independientes

El concepto de eventos independientes es necesario para continuar nuestro análisis de eventos compuestos.

Eventos independientes: dos eventos son *independientes* si el suceso (o no suceso) de uno de ellos no nos da información acerca de la probabilidad de que ocurra el otro. En otras palabras, si la probabilidad de A permanece sin cambio después que sepamos que B ha ocurrido (o no ha ocurrido), los eventos son independientes.

En álgebra: $P(\text{A}) = P(\text{A} \mid \text{B}) = P(\text{A} \mid \text{no B})$

En palabras: Hay varias formas equivalentes de expresar el concepto de independencia:

1. La probabilidad del evento A no se afecta por saber que un segundo evento, B, ha ocurrido, saber que B no ha ocurrido, o no se sepa acerca del evento sea lo que sea.
2. La probabilidad del evento A no se afecta por saber, o no saber, que un segundo evento, B, ha ocurrido o no ha ocurrido.
3. La probabilidad de un evento A (sin saber acerca del evento B) es la misma que la probabilidad del evento A, sabiendo que B ha ocurrido, y ambas son iguales que la probabilidad del evento A, sabiendo que el evento B no ha ocurrido.

No todos los eventos son independientes.

Eventos dependientes: eventos que no son independientes. Esto es, el suceso de un evento no tiene efecto sobre la probabilidad de que ocurra el otro evento.

Veamos algunos ejemplos.

EJEMPLO 4.19 Comprensión de eventos independientes

Se lleva a cabo una encuesta de 750 votantes registrados en 25 distritos electorales del estado de Nueva York. Cada votante fue identificado como registrado como republicano, demócrata y luego se le pregunta, "¿Está usted a favor o en contra del proyecto actual de presupuesto en espera de firma del gobernador?" Los totales resultantes se muestran a continuación.

	Número a favor	Número en contra	Número de votantes
Republicano	135	90	225
Demócrata	315	210	525
Totales	450	300	750

Supongamos que un votante ha de ser seleccionado al azar de los 750 votantes que se resumen en la tabla precedente. Consideremos los dos eventos, "el votante seleccionado está a favor" y "el votante es republicano". ¿Son independientes estos dos eventos?

Para contestar esto, considere las siguientes tres probabilidades: (1) probabilidad de que el votante seleccionado esté a favor; (2) probabilidad de que el votante seleccionado esté a favor, sabiendo que el votante es republicano; y (3) probabilidad de que el votante seleccionado esté a favor, sabiendo que el votante no es republicano.

Probabilidad de que el votante seleccionado esté a favor = P(a favor) = 450/750 = 0.60.

Probabilidad de que el votante seleccionado esté a favor, sabiendo que el votante es republicano = P(a favor|republicano) = 135/225 = 0.60

Probabilidad de que el votante seleccionado esté a favor, sabiendo que el votante no es republicano = Probabilidad de que el votante seleccionado está a favor, sabiendo que el votante es demócrata = P(a favor|no republicano) = P(a favor|demócrata) = 315/525 = 0.60.

Saber que la afiliación política del votante ¿tiene efecto sobre la probabilidad de que el votante esté a favor de la propuesta de presupuesto? Sin información acerca de la afiliación política, la probabilidad de estar a favor es 0.60. La información acerca del evento "republicano" no altera la probabilidad de "a favor". Todos tienen un valor de 0.60, en consecuencia, se dice que estos dos eventos son *eventos independientes.*

Cuando se prueben las tres probabilidades, $P(A)$, $P(A|B)$ y $P(A|\text{no } B)$, necesitamos comparar sólo dos de ellas. Si cualesquiera dos de estas probabilidades son iguales, la tercera tendrá el mismo valor. Además, si cualesquiera dos de las tres probabilidades son desiguales, entonces las tres tendrán valor diferente.

Nota: determine los tres valores, usando el tercero como prueba. Todos serán iguales, o todos serán diferentes, no hay otro resultado posible.

EJEMPLO 4.20 Comprensión de eventos no independientes

De un sondeo de salida para elección nacional, de 13,660 votantes registrados en 25 distritos electorales el 2 de noviembre de 2004, tenemos lo siguiente:

	Porcentaje de votantes	Porcentaje por Bush	Porcentaje por Kerry	Porcentaje por otro
Hombres	46	55	44	1
Mujeres	54	48	51	1

Supongamos que un votante se selecciona al azar de los 13,660 votantes resumidos en la tabla precedente. Consideremos los dos eventos: "el votante es mujer" y "el votante votó por Bush". ¿Son independientes estos dos eventos? Para contestar esto, considere la pregunta "¿Saber si el votante es mujer tiene efecto sobre la probabilidad de que el votante votó por Bush? ¿Cuál es la probabilidad de votar por Bush, si el votante es mujer? El lector dice, "0.48". Ahora compare esto contra la probabilidad de votar por Bush, si el votante no es mujer. Usted dice que la probabilidad es 0.55, de modo que le pregunto, "Saber que si el votante era mujer influye en la probabilidad de votar por Bush?" Sí, influyó; es 0.48 cuando el votante es mujer y 0.55 cuando no es mujer. La información acerca del evento "mujer" altera la

probabilidad de "votó por Bush". Por tanto, estos dos eventos son *no independientes* y se dice que son *eventos dependientes.*

En forma de ecuación:

$$P(\text{votó por Bush} \mid \text{votante es mujer}) = P(\text{B}|\text{W}) = 0.48$$

$$P(\text{votó por Bush} \mid \text{votante no es mujer}) = P(\text{B}|\text{no W}) = 0.51$$

Por tanto, $P(\text{B} \mid \text{W}) \neq P(\text{B} \mid \text{no W})$, y los dos eventos son no independientes.

EJEMPLO 4.21 Eventos independientes de cartas

© Samantha Grandy / Shutterstock

Considere un "monte" regular de cartas y los dos eventos "carta sacada es una reina" y "carta sacada es un corazón". Suponga que se baraja el monte, se saca una carta al azar y, antes de ver la carta, se pregunta la probabilidad de que sea "reina". El jugador dice 4/52, o 1/13. Luego veo la carta y le digo que es un "corazón". Ahora, ¿cuál es la probabilidad de que la carta sea una "reina"? Usted dice que es 1/13, igual que antes de saber que la carta era un "corazón".

La sugerencia de que la carta era un corazón le dio información adicional, pero esa información no cambió la probabilidad de que fuera una reina. Por tanto, "reina" y "corazón" son independientes. Además, suponga que después de sacar la carta y se vio, se dijo que la carta era "no un corazón". ¿Cuál sería la probabilidad de que la carta sea una "reina"? El jugador dice 3/39, o sea 1/13. De nuevo, observe que saber que la carta era "no un corazón" proporcionó información adicional, pero esa información no cambió la probabilidad de que era una "reina". Esto es lo que significa para los dos eventos "carta es una reina" y "carta es un corazón" para ser independientes.

En forma de ecuación:

$$P(\text{reina}|\text{carta es corazón}) = P/(\text{Q}|\text{H}) = P(\text{Q})$$

$$P(\text{reina}|\text{carta es no corazón}) = P(\text{Q}|\text{no H}) = P(\text{Q})$$

Por tanto, $P(\text{Q}) = P(\text{Q}|\text{H}) = P(\text{Q}|\text{no H})$, y los dos eventos son independientes.

EJEMPLO 4.22 Eventos de cartas no independientes

Ahora, consideremos los dos eventos "carta sacada es un corazón" y "carta sacada es roja". ¿Son independientes los eventos "corazón" y "roja"? Siguiendo la misma situación que en el ejemplo 4.21, se baraja el "monte" de 52 cartas, al azar se saca una carta y, antes de verla, el jugador dice que la probabilidad de que la carta desconocida es "roja" es 26/52 = 1/2. No obstante, cuando se le da información adicional de que la carta es un "corazón", cambia la probabilidad de que la carta sea "roja"

a 13/13, o 1. Esta información adicional resulta en una probabilidad diferente de "roja".

P(roja|carta es corazón) = P(R|H) = 13/13 = 1, y P(roja) = P(roja|no tener información adicional) = 26/52 = 1/2. Por tanto, la información adicional cambió la probabilidad del evento "roja". Estos dos eventos son no independientes y por tanto se dice que son eventos dependientes.

En forma de ecuación, la definición expresa que:

$$\text{A y B son independientes si y sólo si } P(\text{A}|\text{B}) = P(\text{A})$$

Nota: defina *independencia* en términos de probabilidad condicional, y pruebe la independencia de esa forma.

Regla especial de la multiplicación

La regla de la multiplicación se simplifica cuando los eventos que intervienen son independientes.

Si sabemos que dos eventos son independientes, entonces al aplicar la definición de independencia, $P(\text{B}|\text{A}) = P(\text{B})$, a la regla de la multiplicación, se deduce que:

$$P(\text{A y B}) = P(\text{A}) \cdot P(\text{B}|\text{A}) \text{ se convierte en } P(\text{A y B}) = P(\text{A}) \cdot P(\text{B})$$

Regla especial de la multiplicación

Sean A y B dos eventos independientes definidos en un espacio muestral S.

En palabras: *probabilidad de A y B = probabilidad de A × probabilidad de B*

En álgebra: $P(\text{A y B}) = P(\text{A}) \cdot P(\text{B})$ (4.7)

Esta fórmula se puede expandir para considerar más de dos eventos independientes:

$$P(\text{A y B y C y} \ldots \text{y E}) = P(\text{A}) \cdot P(\text{B}) \cdot P(\text{C}) \cdot \ldots \cdot P(\text{E})$$

Es frecuente que esta ecuación sea conveniente para calcular probabilidades, pero no nos ayuda a comprender la relación de independencia entre los eventos A y B. Es *la definición* la que nos dice cómo debemos pensar acerca de eventos independientes. Los estudiantes que comprenden la independencia de esta forma adquieren mejor idea de lo que es la independencia. Esto debe llevar al lector a pensar más claramente acerca de situaciones que se refieren a eventos independientes, con lo que es menos probable que confunda el concepto de eventos independientes con eventos mutuamente excluyentes, o que cometa otros errores comunes respecto a la independencia.

Nota: no use $P(\text{A y B}) = P(\text{A}) \cdot P(\text{B})$ como la definición de independencia. Es una propiedad que resulta de la definición. Puede usarse como prueba para independencia, pero, como enunciado, no muestra significado ni idea del concepto de eventos independientes.

SECCIÓN 4.6 EJERCICIOS

4.101 Determine si cada uno de los siguientes pares de eventos es independiente:

a. Lanzar un par de dados y observar un "1" en el primer dado y un "1" en el segundo dado

b. Sacar una "espada" de un "monte" regular de cartas y luego sacar otra "espada" del mismo monte sin restituir la primera carta

c. Igual que la parte b, excepto que la primera carta se restituye al monte antes de sacar la segunda

d. Ser dueño de un automóvil rojo y tener cabello rubio

e. Poseer un automóvil rojo y tener hoy una llanta sin aire

f. Estudiar para un examen y aprobarlo

4.102 Determine si cada uno de los siguientes pares de eventos es independiente:

a. Lanzar un par de datos y observar un "2" en uno de los dados y tener un "total de 10"

b. Sacar una carta de un monte regular de cartas y tener una carta "roja" y tener un "as"

c. Llover hoy y pasar el examen de hoy

d. Llover hoy y jugar al golf hoy mismo

e. Completar la tarea de hoy y estar a tiempo para la clase

4.103 A y B son eventos independientes, y $P(\text{A}) = 0.7$ y $P(\text{B}) = 0.4$. Encuentre $P(\text{A y B})$.

4.104 A y B son eventos independientes, y $P(\text{A}) = 0.5$ y $P(\text{B}) = 0.8$. Encuentre $P(\text{A y B})$.

4.105 A y B son eventos independientes, y $P(\text{A}) = 0.6$ y $P(\text{A y B}) = 0.3$. Encuentre $P(\text{B})$.

4.106 A y B son eventos independientes, y $P(\text{A}) = 0.4$ y $P(\text{A y B}) = 0.5$. Encuentre $P(\text{B})$.

4.107 Si $P(\text{A}) = 0.3$ y $P(\text{B}) = 0.4$ y A y B son eventos independientes, ¿cuál es la probabilidad de cada uno de lo siguiente:

a. $P(\text{A y B})$ b. $P(\text{B}|\text{A})$ c. $P(\text{A}|\text{B})$

4.108 Suponga que $P(\text{A}) = 0.3$, $P(\text{B}) = 0.4$, y $P(\text{A y B}) = 0.12$.

a. ¿Cuál es $P(\text{A}|\text{B})$?

b. ¿Cuál es $P(\text{B}|\text{A})$?

c. ¿Son independientes A y B?

4.109 Suponga que $P(\text{A}) = 0.3$, $P(\text{B}) = 0.4$, y $P(\text{A y B}) = 0.20$.

a. ¿Cuál es $P(\text{A}|\text{B})$?

b. ¿Cuál es $P(\text{B}|\text{A})$?

c. ¿Son independientes A y B?

4.110 Un estudiante es seleccionado al azar de un grupo de 200 estudiantes que se sabe está formado por 140 estudiantes de tiempo completo (80 mujeres y 60 hombres) y 60 estudiantes de tiempo parcial (40 mujeres y 20 hombres). El evento A es "el estudiante seleccionado es de tiempo completo", y el evento C es "el estudiante seleccionado es mujer".

a. ¿Son independientes los eventos A y C? Justifique su respuesta.

b. Encuentre la probabilidad $P(\text{A y C})$.

4.111 Se toma una sola carta de un "monte" regular. Sea A el evento de que "la carta es una cara" (una sota, una reina o un rey), B es una "carta roja," y C es "la carta es un corazón". Determine si los siguientes pares de eventos son independientes o dependientes:

a. A y B b. A y C c. B y C

4.112 Una caja contiene cuatro fichas de póker rojas y tres azules. Tres fichas de póker han de ser seleccionadas, una a la vez.

a. ¿Cuál es la probabilidad de que las tres fichas sean rojas si la selección se hace con restitución?

b. ¿Cuál es la probabilidad de que las tres fichas sean rojas si la selección se hace sin restitución?

c. ¿Son independientes los saques ya sea en la parte a o en la b? Justifique su respuesta.

4.113 Con exclusión de la cobertura de prestaciones laborales, aproximadamente 49% de los adultos han comprado seguros de vida. La probabilidad de que quienes tienen de 18 a 24 años y no tengan seguro de vida compren uno en el siguiente año es 15%, y en los que tienen de 25 a 34 es de 26%. (Investigación de opiniones)

a. Encuentre la probabilidad de que un adulto seleccionado al azar no haya comprado seguro de vida.

b. ¿Cuál es la probabilidad de que un adulto entre 18 y 24 años de edad compre seguro de vida dentro del año siguiente?

c. Encuentre la probabilidad de que un adulto seleccionado al azar tenga de 25 a 34 años de edad, no tenga actualmente seguro de vida y lo comprará dentro del año siguiente.

4.114 El programa espacial de Estados Unidos tiene una historia formada por muchos éxitos y algunos fracasos. La confiabilidad de vuelos espaciales es de la mayor importancia en el lanzamiento de transbordadores espaciales. La confiabilidad de toda una misión está basada en todos sus componentes. Cada una de las seis uniones del cohete propulsor del trasbordador espacial *Challenger* tenía 0.977 de confiabilidad. Las seis uniones trabajaron de manera independiente.

a. ¿Qué significa decir que las seis uniones trabajaron de manera independiente?

b. ¿Cuál fue la confiabilidad (probabilidad) de que las seis uniones trabajaran juntas?

4.115 De las familias de Estados Unidos, 18 millones, es decir 17%, tienen tres o más vehículos, como lo publica *USA Today* (12 de junio, 2002), citando la Oficina del Censo como fuente.

a. Si se seleccionan al azar dos familias de Estados Unidos, encuentre la probabilidad de que ambas tengan tres o más vehículos.

b. Si se seleccionan al azar dos familias de Estados Unidos, encuentre la probabilidad de que ninguna de ellas tenga tres o más vehículos.

c. Si se seleccionan cuatro familias, ¿cuál es la probabilidad de que las cuatro tengan tres o más vehículos?

4.116 Un artículo de *USA Today* titulado "Encuesta: Registros corruptos —Los aficionados desean pruebas antidrogas para jugadores de béisbol" (12 de junio, 2002) cita una encuesta de *USA Today*/CNN Gallup y encuentra que 86% de los aficionados al béisbol dicen que están a favor de que se realicen pruebas antidrogas a jugadores para ver si han consumido esteroides u otras drogas que mejoran el rendimiento. Si se seleccionan al azar cinco aficionados al béisbol, ¿cuál es la probabilidad de que los cinco estén a favor de la prueba antidrogas?

4.117 El 8 de julio, 2002, la edición *Democrat & Chronicle* dio los resultados del censo de 2000 de que 42% de los abuelos son responsables de "la mayor parte de las necesidades básicas" de un nieto en la casa. Si se establece comunicación con tres abuelos estadounidenses, ¿cuál es la probabilidad de que los tres sean los satisfactores principales de sus nietos?

4.118 Usted ha solicitado dos becas: una beca al mérito (M) y una beca por atletismo (A). Suponga que la probabilidad de que usted reciba la beca por atletismo es 0.25, la probabilidad de que reciba ambas becas es 0.15, y la probabilidad de que obtenga al menos una de las becas es 0.37. Use un diagrama de Venn para contestar estas preguntas:

a. ¿Cuál es la probabilidad de que reciba la beca al mérito?

b. ¿Cuál es la probabilidad de que no reciba ninguna de las dos becas?

c. ¿Cuál es la probabilidad de que reciba la beca al mérito dado que ya se le ha otorgado la beca por atletismo?

d. ¿Cuál es la probabilidad de que reciba la beca por atletismo dado que ya se le ha otorgado la beca al mérito?

e. Los eventos de "recibir una beca por atletismo" y "recibir una beca al mérito" ¿son independientes? Explique.

4.119 Los propietarios de un negocio de dos personas toman sus decisiones en forma independiente entre ellos y luego comparan sus decisiones. Si concuerdan, la decisión se toma; si no es así, entonces es necesaria

más consideración antes de llegar a una decisión. Si cada uno tiene una historia de tomar la decisión correcta 60% de las veces, ¿cuál es la probabilidad de que juntos:

a. tomen la decisión correcta al primer intento?

b. tomen la decisión equivocada al primer intento?

c. demoren la decisión para estudiarla mejor?

4.120 Las posibilidades contra lanzar un par de dados y sacar un total de 5 son 8 a 1. Las posibilidades contra lanzar un par de dados y sacar un total de 10 son 11 a 1. ¿Cuál es la probabilidad de lanzar los dados dos veces y obtener un total de 5 en el primer tiro y 10 en el segundo?

4.121 Considere el conjunto de enteros 1, 2, 3, 4 y 5.

a. Se selecciona un entero al azar. ¿Cuál es la probabilidad de que sea non?

b. Se seleccionan dos enteros al azar (uno a la vez con restitución para que cada uno de los cinco esté disponible para una segunda selección). Encuentre la probabilidad de que cualquiera de ellos sea non; exactamente uno de ellos sea non; ambos sean nones.

4.122 Una caja contiene 25 piezas, de las cuales 3 son defectuosas y 22 no son defectuosas. Si 2 partes se seleccionan sin restitución, encuentre las siguientes probabilidades:

a. P(ambas sean defectuosas)

b. P(exactamente una sea defectuosa)

c. P(ninguna sea defectuosa)

4.123 Los porcentajes para graduación alcanzaron su récord más bajo en 2001. El porcentaje de estudiantes que egresaron antes de 5 años fue de 41.9% para universidades públicas y 55.1% para privadas. Una de las razones para esto podría ser que 42% de los estudiantes asisten sólo en forma parcial. (ACT)

a. ¿Qué información adicional es necesaria para determinar la probabilidad de que un estudiante seleccionado al azar sea de tiempo parcial y egrese antes de 5 años?

b. ¿Es probable que estos dos eventos tengan la propiedad necesaria? Explique.

c. Si es apropiado, encuentre la probabilidad de que un estudiante seleccionado al azar sea de tiempo parcial y egrese antes de 5 años.

4.124 De una encuesta de adultos, 48% planea comprar dulces este año en Pascua. Los tipos de dulces que comprarán se describen en la tabla siguiente.

Chocolate	No chocolate	Gelatinas	Relleno de crema	Malvaviscos	Malteadas	No lo sabe
30%	25%	13%	11%	8%	7%	6%

Fuente: International Mass Retail Association

a. ¿Qué información adicional es necesaria para determinar la probabilidad de que un cliente seleccionado al azar compre dulces y éste sea de chocolate?

b. ¿Es probable que estos dos eventos tengan la propiedad necesaria? Explique.

c. Si es apropiado, encuentre la probabilidad de que un cliente seleccionado al azar compre dulces y éstos sean de chocolate.

4.7 ¿Existe relación entre eventos mutuamente excluyentes y la independencia?

Los eventos mutuamente excluyentes y los eventos independientes son dos conceptos muy diferentes basados en definiciones que empiezan de orientaciones muy diferentes. Los dos conceptos pueden con facilidad hacerse confusos porque interactúan el uno con el otro y están entrelazados por los enunciados de probabilidad que usamos al describir estos conceptos.

Para describir estos dos conceptos y en última instancia comprender la distinción entre ellos, así como la relación entre ambos, necesitamos estar de acuerdo en

que los eventos que se consideran son dos eventos no vacíos, definidos en el mismo espacio muestral y por tanto cada uno tiene probabilidades diferentes de cero.

Nota: a veces los estudiantes tienen problemas al ver que cuando decimos "el evento A es un evento no vacío" y escriben "$P(\text{A}) > 0$" estamos describiendo la misma situación. Es frecuente que las palabras y el álgebra no parezcan tener el mismo significado. En este caso, las palabras y el enunciado de probabilidad nos dicen ambos que el evento A existe dentro del espacio muestral.

Mutuamente excluyentes

Los eventos mutuamente excluyentes son dos eventos no vacíos definidos en el mismo espacio muestral que no comparten elementos comunes.

Esto significa:

1. En palabras: si se ve un diagrama de Venn, las áreas cerradas representan cada evento "no se cruzan"; en otras palabras, son conjuntos disjuntos, o que no hay intersección entre sus respectivos conjuntos.
2. En álgebra: $P(\text{A y B}) = 0$, que dice "la intersección de los dos eventos es un conjunto vacío"; en otras palabras, no hay intersección entre sus respectivos conjuntos.

Observe que el concepto de mutuamente excluyente está basado en la relación de los elementos que satisfacen los eventos. Mutuamente excluyente no es un concepto de probabilidad por definición, sino que ocurre que es fácil de expresar el concepto usando un enunciado de probabilidad.

Independencia

Los eventos independientes son dos eventos no vacíos definidos en el mismo espacio muestral que están relacionados, en forma tal, que el suceso de cualquiera de estos eventos no afecta la probabilidad del otro evento.

Esto significa que:

1. En palabras: si el evento A ya ha ocurrido (o se sabe que ha ocurrido), la probabilidad del evento B no resulta afectada (es decir que la probabilidad de B, después de saber que el evento A había ocurrido, sigue igual a como era antes de saber que el evento A había ocurrido).

   Además, también es el caso cuando A y B intercambian papeles que si el evento B ha ocurrido (o se sabe que ha ocurrido), la probabilidad del evento A no resulta afectada (es decir, la probabilidad de A todavía es la misma después de saber que el evento B había ocurrido como era antes).

   Ésta es una "relación mutua"; funciona en las dos formas.

2. En álgebra: $P(B|A) = P(B|\text{no } A) = P(B)$ y $P(A|\text{no } B) = P(A)$, o bien, con unas pocas palabras para ayudar a leer el álgebra, P(B, sabiendo que A ha ocurrido) = P(B, sabiendo que A no ha ocurrido) = P(B) y P(A, sabiendo que B ha ocurrido) = P(A, sabiendo que B no ha ocurrido) = P(A).

   Observe que el concepto de independencia está basado en el efecto que un evento (en este caso, la falta de efecto) tiene sobre la probabilidad del otro evento.

Veamos las siguientes cuatro demostraciones respecto a eventos mutuamente excluyentes e independientes:

Demostración A

Dado: $P(A) = 0.4$, $P(B) = 0.5$, y A y B son mutuamente excluyentes; ¿son independientes?

Respuesta: Si A y B son eventos mutuamente excluyentes, $P(A|B) = 0.0$, y como nos dan $P(A) = 0.4$, vemos que el suceso de B tiene un efecto sobre la probabilidad de A. Por tanto, A y B son eventos no independientes.

Conclusión A: si los eventos son mutuamente excluyentes, son NO independientes.

Demostración B

Dado: $P(A) = 0.4$, $P(B) = 0.5$, y A y B son independientes; ¿los eventos A y B son mutuamente excluyentes?

Respuesta: si A y B son eventos independientes, entonces la $P($ y B$) = P(A) \cdot P(B) = 0.4 \cdot 0.5 = 0.20$, y como la P(A y B) es mayor a cero, los eventos A y B deben intersecarse, lo que significa que los eventos no son mutuamente excluyentes.

Conclusión B: si los eventos son independientes, son NO mutuamente excluyentes.

Demostración C

Dado: $P(A) = 0.4$, $P(B) = 0.5$, y A y B no son mutuamente excluyentes; ¿los eventos A y B son independientes?

Respuesta: Como A y B no son eventos mutuamente excluyentes, debe ser que P(A y B) sea mayor a cero. Ahora, si sucede que la P(A y B) es exactamente 0.20, entonces los eventos A y B son independientes [$P(A) \cdot P(B) = 0.4 \cdot 0.5 = 0.20$], pero si la P(A y B) es cualquier otro valor positivo, por ejemplo 0.1, entonces los eventos A y B no son independientes. Por tanto, los eventos A y B podrían ser independientes o dependientes; se hace necesaria alguna otra información para hacer esa determinación.

Conclusión C: si los eventos no son mutuamente excluyentes, PUEDEN ser independientes o dependientes; se hace necesaria información adicional para determinar cuál.

Demostración D

Dada: $P(A) = 0.4$, $P(B) = 0.5$, y A y B no son independientes; ¿los eventos A y B son mutuamente excluyentes?

Respuesta: como A y B son eventos no independientes, debe ser que P(A y B) sea diferente de 0.20, el valor que sería si fueran independientes [$P(\text{A}) \cdot P(\text{B}) = 0.4 \cdot 0.5 = 0.30$]. Ahora, si sucede que P(A y B) es exactamente 0.00, entonces los eventos A y B son mutuamente excluyentes, pero si P(A y B) es cualquier otro valor positivo, por ejemplo 0.1, entonces los eventos A y B son no mutuamente excluyentes. Por tanto, los eventos A y B podrían ser mutuamente excluyentes o no; se hace necesaria información adicional para hacer esa determinación.

Conclusión: si los eventos no son independientes, PUEDEN ser ya sea mutuamente excluyentes o no ser mutuamente excluyentes, se necesita información adicional para determinar de cuál se trata.

Consejo

Trabaje con todo cuidado, comenzando con la información que se le dé y las definiciones de los conceptos involucrados.

Qué no hacer:

No confíe en el primer ejemplo "extraño" que pueda pensar que lo llevará a la respuesta correcta. ¡Esto no es así!

Los siguientes ejemplos da más práctica con estos conceptos de probabilidad.

EJEMPLO 4.23 Cálculo de probabilidades y regla de la adición

Se lanza un par de dados. El evento T se define como el suceso de un "total de 10 u 11", y el evento D es el suceso de "dobles". Encuentre la probabilidad P(T o D).

SOLUCIÓN Vea en la figura 4.6 (p. 239) el espacio muestral de 36 pares ordenados para el tiro de dos dados. El evento T ocurre si ocurre cualquiera de los 5 pares ordenados: (4, 6), (5, 5), (6, 4), (5, 6), (6, 5). Por tanto, $P(\text{T}) = \frac{5}{36}$. El evento D ocurre si ocurre cualquiera de los 6 pares ordenados: (1,1), (2,2), (3,3), (4,4), (5,5), (6,6). Por tanto, $P(\text{D}) = \frac{6}{36}$. Observe, sin embargo, que estos dos eventos no son mutuamente excluyentes.

Los dos eventos "comparten" el punto (5,5). Así, la probabilidad $P(\text{T y D}) = \frac{1}{36}$. En consecuencia, la probabilidad P(T y D) se encontrará con el uso de la fórmula (4.4).

$$P(\text{T o D}) = P(\text{T}) + P(\text{D}) - P(\text{T y D})$$
$$= \frac{5}{36} + \frac{6}{36} - \frac{1}{36} = \frac{10}{36} = \mathbf{\frac{5}{18}}$$

(Vea el espacio muestral de la figura 4.6 y compruebe $P(\text{T o D}) = \frac{5}{18}$.)

EJEMPLO 4.24 Uso de probabilidades condicionales para determinar independencia

En una muestra de 150 residentes, a cada persona se le pregunta si estuvo a favor del concepto de tener una sola agencia de policía para todo el condado. Éste está compuesto de una ciudad grande y muchas poblaciones suburbanas. La residencia (en la ciudad o fuera de ella) y las respuestas de los residentes se resumen en la tabla 4.4. Si uno de estos residentes había de ser seleccionado al azar, ¿cuál es la probabilidad de que la persona esté (a) a favor del concepto? (b) a favor del concepto si la persona seleccionada vive en una ciudad? (c) a favor del concepto si la persona seleccionada reside fuera de la ciudad? (d) ¿Son independientes los eventos F (a favor del concepto) y C (reside en la ciudad)?

TABLA 4.4

Uso de probabilidades condicionales para determinar independencia

Residencia	Favor (F)	Se opone ($\overline{F}$)	Total
En la ciudad (C)	80	40	120
Fuera de la ciudad ($\overline{C}$)	20	10	30
Total	100	50	150

SOLUCIÓN

(a) $P(F)$ es la proporción de la muestra total a favor del concepto. Por tanto,

$$P(\mathrm{F}) = \frac{n(\mathrm{F})}{n(\mathrm{S})} = \frac{100}{150} = \mathbf{\frac{2}{3}}$$

(b) $P(\mathrm{F}|\mathrm{C})$ es la probabilidad de que la persona seleccionada esté a favor del concepto dado que vive en la ciudad. La condición, que reside en la ciudad, reduce el espacio muestral a los 120 residentes citadinos de la muestra. De éstos, 80 están a favor del concepto; por tanto,

$$P(\mathrm{F}\,|\,\mathrm{C}) = \frac{n(\mathrm{F\ y\ C})}{n(\mathrm{C})} = \frac{80}{120} = \mathbf{\frac{2}{3}}$$

(c) $P(\mathrm{F}\,|\,\overline{\mathrm{C}})$ es la probabilidad de que la persona seleccionada esté a favor del concepto, sabiendo que la persona vive fuera de la ciudad. La condición, que vive fuera de la ciudad, reduce el espacio muestral a los 30 no residentes en la ciudad; por tanto,

$$P(\mathrm{F}\,|\,\overline{\mathrm{C}}) = \frac{n(\mathrm{F\ y}\ \overline{\mathrm{C}})}{n(\overline{\mathrm{C}})} = \frac{80}{120} = \mathbf{\frac{2}{3}}$$

(d) Las tres probabilidades tienen el mismo valor, $\frac{2}{3}$. Por tanto, podemos decir que los eventos F (a favor) y C (reside en ciudad) son independientes. La ubicación de la residencia no afectó a $P(\mathrm{F})$.

EJEMPLO 4.25 Determinación de independencia y uso de la regla de la multiplicación

Un estudiante es seleccionado al azar de un grupo de 200 que se sabe está formado de 140 estudiantes de tiempo completo (80 mujeres y 60 hombres) y 60 estudiantes de tiempo parcial (40 mujeres y 20 hombres). El evento A es "el estudiante seleccionado es de tiempo completo", y el evento C es "el estudiando seleccionado es de tiempo parcial".

(a) ¿Los eventos A y C son independientes?

(b) Encuentre la probabilidad $P(\text{A y C})$ usando la regla de la multiplicación.

SOLUCIÓN 1

(a) Encuentre primeramente las probabilidades: $P(\text{A})$, $P(\text{C})$, y $P(\text{A} \mid \text{C})$:

$$P(\text{A}) = \frac{n(\text{A})}{n(S)} = \frac{140}{200} = 0.7$$

$$P(\text{C}) = \frac{n(\text{C})}{n(S)} = \frac{120}{200} = 0.6$$

$$P(\text{A} \mid \text{C}) = \frac{n(\text{A y C})}{n(C)} = \frac{80}{120} = 0.67$$

A y C son eventos independientes porque $P(\text{A}) \neq P(\text{A} \mid \text{C})$.

(b) $P(\text{A y C}) = P(\text{C}) \cdot P(\text{A} \mid \text{C}) = \frac{120}{200} \cdot \frac{80}{120} = \frac{80}{200} = \mathbf{0.4}$

SOLUCIÓN 2

(a) Encuentre primero las probabilidades: $P(\text{A})$, $P(\text{C})$, y $P(\text{C} \mid \text{A})$:

$$P(\text{A}) = \frac{n(\text{A})}{n(S)} = \frac{140}{200} = 0.7$$

$$P(\text{C}) = \frac{n(\text{C})}{n(S)} = \frac{120}{200} = 0.6$$

$$P(\text{C} \mid \text{A}) = \frac{n(\text{C y A})}{n(\text{A})} = \frac{80}{140} = 0.57$$

A y C son eventos independientes porque $P(\text{C}) \neq P(\text{C} \mid \text{A})$.

(b) $P(\text{C y A}) = P(\text{A}) \cdot P(\text{C} \mid \text{A}) = \frac{140}{200} \cdot \frac{80}{140} = \frac{80}{200} = \mathbf{0.4}$

EJEMPLO 4.26 Uso de diversas reglas de probabilidad

PARA SU INFORMACIÓN
Una mala clasificación puede ocurrir de dos modos.

Un proceso de producción produce miles de artículos. En promedio, 20% de todos ellos son defectuosos y cada uno es inspeccionado antes de ser embarcado. El inspector clasifica mal un artículo el 10% del tiempo; es decir,

$$P(\text{clasificado como bueno} \mid \text{artículo defectuoso}) = P(\text{clasificado defectuoso} \mid \text{artículo bueno}) = 0.10$$

¿Qué proporción de artículos será "clasificado como bueno"?

SOLUCIÓN ¿Qué queremos decir con el evento "clasificado como bueno"?

G: El artículo es bueno.

D: El artículo es defectuoso.

CG: El artículo es clasificado como bueno por el inspector.

CD: El artículo es clasificado defectuoso por el inspector.

FIGURA 4.7 Tratando de vencer las posibilidades

CG está formado por dos posibilidades: "el artículo es bueno y está correctamente clasificado como bueno" y "el artículo es defectuoso y está mal clasificado como bueno". Así,

$$P(\text{CG}) = P[\text{CG y G}) \text{ o } (\text{CG y D})]$$

Como las dos posibilidades son mutuamente excluyentes, podemos empezar con usar la regla de la adición, fórmula (4.6):

$$P(\text{CG}) = P(\text{CG y G}) + P(\text{CG y D})$$

La condición de un artículo y su clasificación por el inspector no son independientes. Debe usarse la regla de la multiplicación para eventos dependientes. Por tanto,

$$P(\text{CG}) = [P(\text{G}) \cdot P(\text{CG}|\text{G})] + [P(\text{D}) \cdot P(\text{CG}|\text{D})]$$

Si sustituimos las probabilidades conocidas en la figura 4.7, obtenemos

$$\begin{aligned} P(\text{CG}) &= [(0.8)(0.9)] + [(0.2)(0.1)] \\ &= 0.72 + 0.02 \\ &= \mathbf{0.74} \end{aligned}$$

Esto es, 74% de los artículos son clasificados como buenos.

SECCIÓN 4.7 EJERCICIOS

4.125 a. Describa verbalmente lo que significa que dos eventos sean mutuamente excluyentes.

b. Describa verbalmente lo que significa que dos eventos sean independientes

c. Explique cómo mutuamente excluyente e independiente son dos propiedades muy diferentes.

4.126 a. Describa verbalmente por qué dos eventos no pueden ser independientes si ya se sabe que son mutuamente excluyentes.

b. Describa verbalmente por qué dos eventos no pueden ser mutuamente excluyentes si ya se sabe que son independientes.

4.127 $P(\text{G}) = 0.5$, $P(\text{H})0.4$, y $P\text{G y H}) = 0.1$ (ver diagrama).

a. Encuentre $P(\text{G}|\text{H})$.

b. Encuentre $P(\text{H}|\text{G})$.

c. Encuentre $P(\overline{\text{H}})$.

d. Encuentre $P(\text{G o H})$.

e. Encuentre $P(\text{G o } \overline{\text{H}})$.

f. ¿Los eventos G y H son mutuamente excluyentes? Explique.

g. ¿Los eventos G y H son independientes? Explique.

4.128 $P(\text{R}) = 0.5$, $P(\text{S}) = 0.3$, y los eventos R y S son independientes.

a. Encuentre $P(\text{R y S})$.

b. Encuentre $P(\text{R o S})$.

c. Encuentre $P(\overline{\text{S}})$

d. Encuentre $P(\text{R}|\text{S})$.

e. Encuentre $P(\text{S}|\overline{\text{R}})$.

f. ¿Los eventos R y S son mutuamente excluyentes? Explique.

4.129 $P(\text{M}) = 0.3$, $P(\text{N}) = 0.4$, y los eventos M y N son mutuamente excluyentes.

a. Encuentre $P(\text{M y N})$.

b. Encuentre $P(\text{M o N})$.

c. Encuentre $P(\text{M o } \overline{\text{N}})$.

d. Encuentre $P(\text{M}|\text{N})$.

e. Encuentre $P(\text{M}|\overline{\text{N}})$.

f. ¿Los eventos M y N son independientes? Explique.

4.130 Se seleccionan al azar dos semillas de flores de un paquete que contiene cinco semillas para flores rojas y tres semillas para flores blancas.

a. ¿Cuál es la probabilidad de que ambas semillas resulten en flores rojas?

b. ¿Cuál es la probabilidad de que se seleccione una de cada color?

c. ¿Cuál es la probabilidad de que ambas semillas sean para flores blancas?

PARA SU INFORMACIÓN

Trace un diagrama de árbol.

4.131 Fueron encuestados mil empleados en la Russell Microprocessor Company acerca de la satisfacción de trabajadores. Se selecciona un empleado al azar.

	Hombre		Mujer		
	Capacitado	No capacitado	Capacitado	No capacitado	Total
No capacitado	350	150	25	100	625
Insatisfecho	150	100	75	50	375
Total	500	250	100	150	1000

a. Encuentre la probabilidad de que un trabajador no capacitado esté satisfecho con su trabajo.

b. Encuentre la probabilidad de que una empleada capacitada esté satisfecha con su trabajo.

c. La satisfacción para empleadas ¿Es independiente de que sean capacitadas o no capacitadas?

4.132 Una empresa que manufactura zapatos tiene tres fábricas. La fábrica 1 produce 25% de los zapatos de la compañía, la fábrica 2 produce 60% y la fábrica 3 produce 15%. Uno por ciento de los zapatos producidos por la fábrica 1 está mal marcado, 0.5% de los producidos por la fábrica 2 está mal marcado, y 2% de los producidos por la fábrica 3 está mal marcado. Si una persona adquiere un par de zapatos manufacturados por esta compañía, ¿cuál es la probabilidad de que los zapatos estén mal marcados?

REPASO DEL CAPÍTULO

En retrospectiva

El lector ha estado estudiando los conceptos básicos de probabilidad, que es necesario domine antes de continuar con su estudio de estadística. La probabilidad es el vehículo de la estadística, y hemos empezado a ver cómo ocurren eventos probabilísticos. También hemos explorado probabilidades teóricas y experimentales para el mismo evento. La probabilidad experimental ¿resulta tener el mismo valor que la teórica? No exactamente, pero hemos visto que a la larga tiene casi el mismo valor.

Al completar este capítulo, el estudiante debe comprender las propiedades de exclusividad mutua e independencia, además de tener aptitud para aplicar las reglas de la multiplicación y la adición a eventos compuestos "y" y "o." También debe saber calcular probabilidades condicionales.

En los siguientes tres capítulos veremos distribuciones asociadas con eventos probabilísticos. Esto nos prepara para la estadística que sigue. Debemos tener capacidad para pronosticar la variabilidad que la muestra exhibirá respecto a la población antes de tener éxito en "estadísticas inferenciales", donde describimos la población basada en la estadística muestral disponible.

Vocabulario y conceptos clave

diagrama de árbol (p. 209)
diagrama de Venn (p. 211)
espacio muestral (p. 208)
evento (p. 208)
evento complementario (p. 229)
evento compuesto (p. 228)
eventos dependientes (p. 243)
eventos igualmente probables (p. 208)
eventos independientes (pp. 240, 243)
eventos mutuamente excluyentes (p. 236)
eventos todo incluido (p. 213)
frecuencia relativa (p. 207)
independencia (p. 243)
intersección (p. 237)
ley de números grandes (pp. 213, 215)
par ordenado (p. 209)
posibilidades (p. 215)
probabilidad condicional (p. 223)
probabilidad de un evento (p. 207)
probabilidad empírica (p. 207)
probabilidad experimental (p. 207)
probabilidad subjetiva (p. 207)
probabilidad teórica (p. 207)
promedio a largo plazo (p. 215)
punto muestral (p. 208)
regla de la adición (pp. 230, 239)
regla de la multiplicación (pp. 231, 246)
regla especial de la adición (p. 239)
regla especial de la multiplicación (p. 246)
regla general de la adición (p. 230)
regla general de la multiplicación (p. 231)
resultado (p. 208)

Objetivos de aprendizaje

✓ Comprender y ser capaz de describir el concepto básico de probabilidad	pp. 205-207
✓ Comprender y describir un evento sencillo	EJ. 4.1
✓ Comprender y ser capaz de describir las diferencias entre probabilidades empíricas, teóricas y subjetivas	pp. 208-209, 212
✓ Calcular e interpretar frecuencias relativas.	Ejer. 4.7, 4.10, 4.11, 4.133
✓ Identificar y describir un espacio muestral para un experimento.	pp. 208-209, Ejer. 4.15, 4.21, 4.22
✓ Construir tablas, diagramas de árbol, y/o diagramas de Venn para ayudar en calcular e interpretar probabilidades.	Ejer. 4.2, 4.3, 4.4, Ejer. 4.23, 4.25

✓ Entender las propiedades de números de probabilidad: 1. $0 \leq 0$ cada $P(A) \leq 1$ 2. $\sum_{\text{todos los resultados}} P(A) = 1$	pp. 212-213, Ejer. 4.29, 4.41
✓ Entender, describir y usar la ley de números grandes para determinar probabilidades	EJ. 4.5, p. 215, Ejer.4.32, 4.171
✓ Entender, calcular e interpretar posibilidades de un evento.	EJ. 4.6, Ejer. 4.43, 4.46, 4.120
✓ Entender que los eventos compuestos comprenden el suceso de más de un evento.	Ejer. 4.35, 4.51
✓ Construir, describir, calcular e interpretar una probabilidad condicional.	EJ. 4.9, Ejer. 4.49, 4.53, 4.141
✓ Entender y ser capaz de utilizar la regla del complemento.	EJ. 4.10, Ejer. 4.59, 4.60
✓ Calcular probabilidades de eventos compuestos usando la regla de la adición.	EJ. 4.11, Ejer. 4.65, Ejer. 4.23
✓ Calcular probabilidades de eventos compuestos usando la regla de la multiplicación.	EJ. 4.12, Ejer. 4.74
✓ Entender, describir y determinar eventos mutuamente excluyentes.	p. 236, EJ. 4.14, 4.15, Ejer. 4.87, 4.93
✓ Calcular probabilidades de eventos compuestos usando la regla de la adición para eventos mutuamente excluyentes.	EJ. 4.18, Ejer. 4.97
✓ Entender, describir y determinar eventos independientes.	p. 243, EJ. 4.19, 4.20, Ejer. 4.101
✓ Calcular probabilidades de eventos compuestos usando la regla de la multiplicación para eventos independientes.	Ejer. 4.11, 4.115
✓ Reconocer y comparar las diferencias entre eventos mutuamente excluyentes y eventos independientes.	pp. 250-252, Ejer. 4.127, 4.147, 4.155

Ejercicios del capítulo

4.133 La Administración Federal de Ferrocarriles proporcionó las cinco categorías principales de violaciones para el ferrocarril CSX para los años 1999-2003 en la tabla siguiente. Hubo un total de 1897 violaciones. La información estuvo contenida en el artículo *Democrat and Chronicle*, 29 de diciembre, 2004, titulado "Rail cop lacks a 'big stick'. (El uniformado no lleva 'garrote'".

Categoría	Número
Seguridad en vías	485
Equipo de seguridad en trenes	324
Horas de trabajo de empleados	323
Seguridad en furgones	289
Locomotoras	248
Todos los otros	228
Total	1897

Si una violación se selecciona al azar para repaso, ¿cuál es la probabilidad de que la violación para el CSX se deba a lo siguiente?:

a. Equipo de seguridad en trenes

b. Horas de trabajo de empleados

c. Seguridad en furgones o seguridad en vía

¿Qué pasa si se seleccionan dos violaciones?

d. ¿Sería esto un ejemplo de muestreo con o sin restitución? Explique por qué.

4.134 El número de personas que vivían en los 50 estados de Estados Unidos y el Distrito de Columbia, en septiembre de 2004,se publicó por grupos de edades en la siguiente tabla.

Grupo de edad	Porcentaje	Número (en miles)
0–17	25%	73,447.7
18–24	10%	28,855.7
25–34	13%	39,892.5
35–49	23%	66,620.3
50+	29%	84,119.8

Fuente: Sales & Marketing Management Survey of Buying Power, septiembre, 2004, para los 50 estados y el Distrito de Columbia

a. Verifique los porcentajes presentados en la tabla.

Si una persona es seleccionada al azar de todas las representadas en la tabla, ¿cuál es la probabilidad de los siguientes eventos?:

b. "Entre 18 y 24." ¿Cómo está esto relacionado al 10% de la tabla?

c. "Mayor de 17"

d. "Entre 18 y 24" y "mayor de 17"

e. "Entre 18 y 24" o "mayor de 17"

f. "No más de 24"

4.135 Mil personas seleccionadas de cierta enfermedad reciben un examen clínico. Como consecuencia del examen, la muestra de 1000 personas se clasifica de acuerdo con su estatura y situación de su enfermedad.

	Situación de enfermedad				
Estatura	Ninguno	Benigno	Moderado	Grave	Total
Alta	122	78	139	61	400
Media	74	51	90	35	250
Corta	104	71	121	54	350
Total	300	200	350	150	1000

Use la información de la tabla para estimar la probabilidad de ser de estatura media o corta y tener situación de enfermedad moderada o grave.

4.136 La Administración Federal de Autopistas periódicamente rastrea el número de conductores de vehículos con licencia, por género y edad. La tabla siguiente muestra los resultados de lo que halló la administración en 2002.

Grupo de edad (años)	Hombre	Mujer
19 o menos	4 772 152	4 526 106
20–24	8 424 540	8 115 247
25–29	8 727 305	8 372 379
30–34	9 737 052	9 378 312
35–39	10 189 184	9 936 933
40–44	10 614 344	10 584 498
45–49	9 941 582	9 997 864
50–54	8 735 627	8 788 501
55–59	7 148 429	7 141 534
60–64	5 371 340	5 377 859
65–69	4 253 857	4 284 304
70–74	3 647 137	3 788 721
75–79	2 936 969	3 173 171
80–84	1 849 298	2 079 929
85 o mayor	1 112 647	1 288 812
Total	97 461 463	96 834 170

Fuente: U.S.Department of Transportation, Federal Highway Administration, *Highway Statistics 2002*:

Suponga que al azar usted se topa con un conductor de un vehículo. Encuentre las probabilidades de los siguientes eventos:

a. El conductor es hombre y de más de 59 años.

b. El conductor es mujer o menor de 30.

c. El conductor es menor de 25 años.

d. El conductor es mujer.

e. El conductor es un hombre entre 35 y 49 años.

f. El conductor es mayor de 69 años.

g. El conductor es mujer, entre 25 y 44 años.

h. El conductor tiene entre 25 y 44 años, es mujer.

4.137 Supongamos que hay tres semáforos entre su casa y la casa de un amigo. Cuando usted llega a cada semáforo, puede estar en roja (R) o verde (G).

a. Haga una lista del espacio muestral que indique todas las posibles secuencias de semáforos en rojo y verde que podría haber en un viaje de su casa a la casa de su amigo. (RGG representa rojo en el primer semáforo y verde en los otros dos.)

Suponga que es igualmente probable que ocurra cada elemento del espacio muestral.

b. ¿Cuál es la probabilidad de que, en el siguiente viaje a la casa de su amigo, tenga que detenerse en exactamente un semáforo en rojo?

c. ¿Cuál es la probabilidad de que tenga que detenerse en al menos un semáforo en rojo?

4.138 Suponiendo que es igualmente probable que una mujer tenga un niño o una niña, use un diagrama de árbol para calcular la probabilidad de que una familia de cuatro hijos esté formada de un niño y tres niñas.

4.139 Realice la siguiente simulación para generar una familia. La "familia" dejará de tener hijos cuando tengan un niño o tres niñas, lo que ocurra primero. Suponiendo que es igualmente probable que una mujer tenga un niño o una niña, realice la simulación 24 veces. ¿Cuál es la probabilidad de que la familia tenga un niño?

4.140 Una moneda se lanza al aire tres veces.

a. Dibuje un diagrama de árbol que represente todos los posibles resultados.

b. Identifique todas las ramas que representen el evento de "ocurrió exactamente una cabeza".

c. Encuentre la probabilidad de "ocurrió exactamente una cabeza".

4.141 Una encuesta reciente de familias del estado de Nueva York les preguntó por sus hábitos en vacaciones. La siguiente tabla de dos formas muestra el número de familias según donde viven (rural, suburbana, urbana) y duración de sus últimas vacaciones (1-7 días, 8 días o más).

	Rural	Suburbana	Urbana	Total
1-7 días	90	57	52	199
8 días o más	74	38	21	123
Total	164	95	73	332

Si una familia se selecciona al azar de estas 332 familias, ¿cuál es la probabilidad de lo siguiente?:

a. Pasaron 8 días o más de vacaciones.

b. Eran una familia rural.

c. Eran una familia urbana y pasaron 8 días o más de vacaciones.

d. Eran una familia rural o pasaron de 1 a 7 días de vacaciones.

e. Pasaron 8 días o más de vacaciones, dado que eran una familia suburbana.

f. Eran una familia rural, dado que pasaron de 1 a 7 días de vacaciones.

4.142 La demografía de edad y género para los estudiantes de la universidad de la comunidad de Monroe en 2004 se indican en la tabla siguiente.

	19 años o menos	20-24	25-29	30 años o mayor
Mujer	3136	2736	1067	2648
Hombre	2877	2757	779	1502
Total	6013	5493	1846	4150

Si uno de estos estudiantes se selecciona al azar, ¿cuál es la probabilidad de que el estudiante sea lo siguiente?

a. Hombre

b. Entre 20 y 24 años de edad

c. Mujer y de 30 años o mayor

d. Hombre o de 19 años o menor

e. Entre 25 y 29 años de edad, dado que ella era una estudiante mujer

f. Hombre, dado que el estudiante tenía 20 años o más

4.143 Esta gráfica de barras muestra el número de automóviles registrados en cada uno de varios países.

a. Mencione al menos dos países no incluidos en la información.

b. ¿Por qué son probabilidades condicionales todas las probabilidades que resultan de esta información?

Con base en la información de la gráfica adjunta:

c. ¿Qué porcentaje de todos los autos de estos países está registrado en Estados Unidos?

d. Si un auto registrado se seleccionó al azar de todos éstos, ¿cuál es la probabilidad de que esté registrado en Estados Unidos?

e. Explique la relación entre sus respuestas a las partes c y d.

4.144 Las probabilidades para los eventos A, B y C están distribuidas como se muestra en la figura. Encuentre:

a. $P(\text{A y B})$ b. $P(\text{A o C})$ c. $P(\text{A}|\text{C})$

4.145 Muestre que si el evento A es un subconjunto del evento B, entonces $P(\text{A o B}) = P(\text{B})$.

4.146 Explique por qué estas probabilidades no pueden ser legítimas: $P(\text{A}) = 0.6$, $P(\text{B}) = 0.4$, $P(\text{A y B}) = 0.7$.

4.147 Un embarque de toronjas llegó conteniendo las siguientes proporciones de tipos: 10% rosas sin semilla, 20% blancas sin semilla, 30% rosas con semillas, y 40% blancas con semillas. Una toronja se selecciona al azar del embarque. Encuentre la probabilidad de estos eventos:

a. Es sin semilla.

b. Es blanca.

c. Es rosa y sin semilla.

d. Es rosa o sin semilla.

e. Es rosa, dado que es sin semilla.

f. Es sin semilla, dado que es rosa.

4.148 Un análisis de tránsito en una glorieta de mucho tráfico en Washington, DC, mostró que 0.8 de los autos que usaban la glorieta entraban de la avenida Connecticut. De los que entraban a la glorieta desde la avenida Connecticut, 0.7 continuaban sobre la avenida Connecticut en el lado opuesto de la glorieta. ¿Cuál es la probabilidad de que un auto seleccionado al azar y observado en la glorieta entrara desde Connecticut y continuara en la Connecticut?

4.149 Suponga que cuando un candidato a un trabajo llega a entrevista para un trabajo en la RJB Enterprises, la probabilidad de que desee el trabajo (A) después de la entrevista es 0.68. Del mismo modo, la probabilidad que RJB desee al candidato (B) es 0.36. La probabilidad $P(\text{A}|\text{B})$ es 0.88.

a. Encuentre $P(\text{A y B})$.

b. Encuentre $P(\text{B}|\text{A})$.

c. ¿Son independientes los eventos A y B?

d. ¿Son mutuamente excluyentes los eventos A y B? Explique.

e. ¿Qué significaría decir que A y B son eventos mutuamente excluyentes en este ejercicio?

4.150 La probabilidad de que haya tormentas eléctricas en las cercanías de un aeropuerto particular en la región del medio oeste, en un día de agosto, es 0.70. Cuando hay tormentas en las cercanías, la probabilidad de que un avión aterrice a tiempo es 0.80. Encuentre la probabilidad de que haya tormentas en las cercanías y el avión aterrice a tiempo.

4.151 Los neumáticos recuperados de un accidente de trenes están a la venta en la Getrich Tire Company. De las 15 llantas ofrecidas a la venta, 5 han sufrido averías internas y las restantes 10 no tienen averías. Se puede seleccionar al azar y comprar dos de estas llantas.

a. ¿Cuál es la probabilidad de que las llantas que compró usted sean ambas sin averías?

b. ¿Cuál es la probabilidad de que exactamente una de las llantas que compró sea sin averías?

c. ¿Cuál es la probabilidad de que al menos una de las llantas que compró sea sin averías?

4.152 Según estadísticas de accidentes automovilísticos, uno de cada seis accidentes resulta en una reclamación de seguro de \$100 o menos en daños en propiedad. Tres autos asegurados por una compañía de seguros participan en diferentes accidentes. Considere estos dos eventos:

A: La mayor parte de reclamaciones exceden de \$100.

B: Exactamente dos reclamaciones son por \$100 o menos.

a. Haga una lista de puntos muestrales para este experimento.

b. ¿Son igualmente probables los puntos muestrales?

c. Encuentre $P(A)$ y $P(B)$.

d. ¿Son independientes los eventos A y B? Justifique su respuesta.

4.153 Una organización que realiza pruebas desea clasificar una marca particular de televisor. Se seleccionan seis televisores al azar de la existencia. Si no se encuentra algo mal en cualquiera de los seis aparatos, la marca se juzga como satisfactoria.

a. ¿Cuál es la probabilidad de que la marca se clasifique como satisfactoria si 10% de los aparatos en realidad son defectuosos?

b. ¿Cuál es la probabilidad de que la marca sea clasificada como satisfactoria si 20% de los aparatos en realidad son defectuosos?

c. ¿Cuál es la probabilidad de que la marca sea clasificada como satisfactoria si 40% de los aparatos en realidad son defectuosos?

4.154 Suponga que cierto rasgo oftálmico está asociado con el color de los ojos. Se estudian 300 personas seleccionadas al azar, con los resultados dados en la tabla siguiente.

	Color del ojo			
Rasgo	Azul	Café	Otros	Total
Sí	70	30	20	120
No	20	110	50	180
Total	90	140	70	300

a. ¿Cuál es la probabilidad de que una persona seleccionada al azar tenga ojos azules?

b. ¿Cuál es la probabilidad de que una persona seleccionada al azar tenga el rasgo?

c. ¿Son independientes los eventos A (tiene ojos azules) y B (tiene el rasgo)? Justifique su respuesta.

d. ¿Cómo están relacionados los dos eventos A (tiene ojos azules) y C (tiene ojos cafés), independiente, mutuamente excluyente, complementario o todo incluido? Explique por qué o por qué no aplica cada término.

4.155 Como lo indica *The World Factbook* 2004, la estructura de edades de la población de Estados Unidos es como sigue.

	Hombre	Mujer
0-14 años	31 122 974	29 713 748
15-64 años	97 756 380	98 183 309
65 años o más	15 078 204	21 172 956

Si un ciudadano de Estados Unidos fuera a ser seleccionado al azar, ¿cuál es la probabilidad de que la persona seleccionada de esta población sea lo siguiente?:

a. Mujer

b. 0 a 14 años de edad

c. Hombre y 15 a 64 años de edad

d. Mujer o 65 años o más

e. Menor de 15 años, sabiendo que la persona es mujer

f. Hombre, dado que la persona tiene de 15 a 64 años

Los eventos "persona seleccionada es hombre" y "persona seleccionada es mujer" no son eventos independientes.

g. ¿Es correcto este enunciado? Justifique su respuesta. ¿Cuál es la relación entre mujer y hombre en esta situación?

4.156 La tabla siguiente muestra las impresiones de 2500 empleados asalariados de la compañía Spruce, en una propuesta para destacar las prestaciones en lugar de aumento de sueldo durante sus inminentes discusiones de contrato.

	Opinión			
Empleado	Favor	Neutral	En contra	Total
Hombre	800	200	500	1500
Mujer	400	100	500	1000
Total	1200	300	1000	2500

a. Calcule la probabilidad de que un empleado seleccionado al azar de este grupo sea en contra.

b. Calcule la probabilidad de que un empleado seleccionado al azar de este grupo sea mujer.

c. Calcule la probabilidad de que un empleado seleccionado al azar de este grupo sea en contra, dado que la persona es hombre.

d. ¿Son independientes los eventos "en contra" y "mujer"? Explique.

4.157 Los eventos R y S se definen en un espacio muestral. Si $P(\text{R}) = 0.2$ y $P(\text{S}) = 0.5$, explique por qué cada uno de los siguientes enunciados es verdadero o falso:

a. Si R y S son mutuamente excluyentes, entonces $P(\text{R o S}) = 0.10$.

b. Si R y S son independientes, entonces $P(\text{R o S}) = 0.06$.

c. Si R y S son mutuamente excluyentes, entonces $P(\text{R y S}) = 0.7$.

d. Si R y S son mutuamente excluyentes, entonces $P(\text{R o S}) = 0.6$.

4.158 Se cree que 3% de los pacientes de una clínica tiene cáncer. Una prueba particular de sangre indica un resultado positivo para 98% de pacientes con cáncer, pero también indica positivo para 4% de los pacientes que no tienen cáncer. Un paciente se selecciona al azar de la lista de pacientes de la clínica y se somete a prueba. ¿Cuál es la probabilidad de que si el resultado de la prueba es positivo, la persona en realidad tenga cáncer?

4.159 La caja 1 contiene dos pelotas rojas y tres pelotas verdes, y la caja 2 contiene cuatro pelotas rojas y una pelota verde. Una pelota se selecciona al azar de la caja 1 y se coloca en la caja 2. A continuación, una pelota se selecciona al azar de la caja 2. ¿Cuál es la probabilidad de que la pelota seleccionada de la caja 2 sea verde?

4.160 Los vendedores Adams y Jones visitan tres y cuatro clientes, respectivamente, en un día determinado. Adams pudo hacer 0, 1, 2 o 3 ventas, en tanto que Jones pudo hacer 0, 1, 2, 3 o 4 ventas. El espacio muestral que contiene el número de posibles ventas para cada persona en un día determinado se muestra en la tabla. (3,1 representa 3 ventas por Jones y 1 venta por Adams.)

	Jones				
Adams	0	1	2	3	4
0	0, 0	1, 0	2, 0	3, 0	4, 0
1	0, 1	1, 1	2, 1	3, 1	4, 1
2	0, 2	1, 2	2, 2	3, 2	4, 2
3	0, 3	1, 3	2, 3	3, 3	4, 3

Suponga que cada punto muestral es igualmente probable. Considere tres eventos:

A: Al menos uno de los vendedores no hizo ventas.

B: Juntos hicieron exactamente tres ventas.

C: Cada uno hizo el mismo número de ventas.

D: Adams hizo exactamente una venta.

Encuentre las probabilidades al contar puntos muestrales:

a. $P(\text{A})$ b. $P(\text{B})$ c. $P(\text{C})$

d. $P(\text{D})$ e. $P(\text{A y B})$ f. $P(\text{B y C})$

g. $P(\text{A o B})$ h. $P(\text{B o C})$ i. $P(\text{A} \mid \text{B})$

j. $P(\text{B} \mid \text{D})$ k. $P(\text{C} \mid \text{B})$ l. $P(\text{B} \mid \overline{\text{A}})$

m. $P(\text{C} \mid \overline{\text{A}})$ n. $P(\text{A o B o C})$

¿Son mutuamente excluyentes los siguientes pares de eventos? Explique.

o. A y B p. B y C q. B y D

¿Son independientes los siguientes pares de eventos? Explique.

r. A y B s. B y C t. B y D

4.161 Alex, Bill y Chen, uno por uno, lanzan al aire una moneda balanceada. El primero en lanzar una cara gana.

a. ¿Cuáles son sus respectivas probabilidades de gana si cada uno lanza sólo una vez?

b. ¿Cuáles son sus respectivas probabilidades de ganar si continúan, dándoles un máximo de dos tiros a cada uno?

PARA SU INFORMACIÓN

Elabore un diagrama de árbol.

4.162 La moneda A está cargada en forma tal que P (caras) es 0.6. La moneda B es una moneda balanceada. Se lanzan al aire ambas monedas. Encuentre:

a. El espacio muestral que representa este experimento; asigne una medida de probabilidad a cada resultado

b. P(ambas muestran caras)

c. P(exactamente sale una cara)

d. P(en ninguna moneda sale una cara)

e. P(ambas muestran caras | la moneda A muestra una cara)

f. P(ambas muestran caras | la moneda B muestra una cara)

g. P(caras en la moneda A | exactamente muestra una cara)

4.163 El profesor de francés olvida poner la alarma con una probabilidad de 0.3. Si pone la alarma, suena con una probabilidad de 0.8. Si la alarma suena, le despertará a tiempo para dar su primera clase con una probabilidad de 0.9. Si la alarma no suena, él despierta a tiempo para su primera clase con una probabilidad de 0.2. ¿Cuál es la probabilidad de que el profesor de francés despierte a tiempo para dar su primera clase mañana?

4.164 La probabilidad de que cierta puerta esté cerrada con llave es 0.6, pero la llave de la puerta es una de cinco no identificadas que cuelgan de un llavero. Una persona, al azar, selecciona dos llaves antes de aproximarse a la puerta. ¿Cuál es la probabilidad de que pueda abrir la puerta sin regresar por otra llave?

4.165 El museo local de arte ha planeado el calendario de 52 semanas del año próximo al programar una combinación de muestras de 1 semana y 2 semanas que presentan obras de 22 pintores y 20 escultores. Hay una muestra programada para cada semana del año, y sólo un artista se presenta a la vez. Hay 42 muestras diferentes programadas para el año próximo. Una persona selecciona al azar una semana para asistir y se le ha indicado que la probabilidad de que sea una muestra de 2 semanas de escultura es 3/13.

a. ¿Cuál es la probabilidad de que la muestra seleccionada sea de un pintor?

b. ¿Cuál es la probabilidad de que la muestra seleccionada es la de un escultor?

c. ¿Cuál es la probabilidad de que la muestra seleccionada sea una muestra de 1 semana?

d. ¿Cuál es la probabilidad de que la muestra seleccionada sea una muestra de 2 semanas?

4.166 Un informe de dos páginas mecanografiadas contiene un error en una de las páginas. Dos lectores de pruebas revisan la copia. Cada uno tiene 80% de probabilidad de captar el error. ¿Cuál es la probabilidad de que el error sea identificado en los siguientes casos?:

a. Cada uno lee una página diferente.

b. Cada uno lee ambas páginas.

c. El primero selecciona al azar una página para leerla y luego el segundo selecciona al azar una página sin saber cuál página fue seleccionada primero.

4.167 Es frecuente que, en deportes, los campeonatos sean decididos por dos equipos que juegan en una serie de campeonato. Los aficionados del equipo perdedor dicen que no tuvieron suerte y que su equipo en realidad es el mejor. Supongamos que el Equipo A es el mejor, y la probabilidad de que derrote al Equipo B en cualquier juego es 0.6.

a. ¿Cuál es la probabilidad de que el mejor equipo, el Equipo A, gane la serie si es una serie de un solo juego?

b. ¿Cuál es la probabilidad de que el mejor equipo, el Equipo A, gane la serie si es el mejor en una serie de tres juegos?

c. ¿Cuál es la probabilidad de que el mejor equipo, el Equipo A, gane la serie si es el mejor en una serie de siete juegos?

d. Suponga que la probabilidad de que A venza a B en cualquier juego determinado fuera en realidad 0.7. Calcule de nuevo los incisos a-c.

e. Suponga que la probabilidad de que A venza a B en cualquier juego determinado fuera en realidad 0.9. Calcule de nuevo los incisos a-c.

f. ¿Cuál es la relación entre el "mejor" equipo ganador y el número de juegos jugados? ¿Y el mejor equipo ganador y las probabilidades de que gane cada uno?

4.168 Una mujer y un hombre (no emparentados) tienen cada uno dos hijos. Al menos uno de los hijos de la mujer es niño, y el hijo mayor del hombre es un niño. La probabilidad de que la mujer tenga dos hijos ¿es mayor, igual o menor que la probabilidad de que el hombre tenga dos niños?

a. Demuestre la verdad de su respuesta usando una muestra simple para representar cada familia.

b. Demuestre la verdad de su respuesta al tomar dos muestras, una de hombres con familias de dos hijos y una de mujeres con familias de dos hijos.

c. Demuestre la verdad de su respuesta usando una simulación de computadora. Usando la función de probabilidad de Bernoulli con $p = 0.5$ (sea 0 = niña y 1 = niño), genere 500 "familias de dos hi-

jos" para el hombre y la mujer. Determine cuál de las 500 satisfacen la condición para cada una y determine la proporción observada con dos niños.

d. Demuestre la verdad de su respuesta al repetir varias veces la simulación de computadora. Repita varias veces la simulación de la parte c.

e. ¿Le parece que los procedimientos precedentes dan los mismos resultados? Explique.

4.169 Tres monedas cargadas se lanzan al aire simultáneamente. Encuentre la probabilidad de obtener tres caras, dado que al menos una de las monedas muestra caras.

a. Resuelva usando un espacio muestral igualmente probable.

b. Resuelva usando la fórmula para probabilidad condicional.

Proyecto del capítulo

Estadística y los dulces

El proyecto del capítulo nos lleva a la sección 4.1, "Estadística y los dulces" (p. 205), como una forma de evaluar lo que hemos aprendido en este capítulo. Y qué mejor forma de hacerlo con algún dulce. Podemos explorar las diferencias entre probabilidades teóricas y experimentales y ver la ley de números grandes en acción, todo con M&M's. Ahora eso es "Estadística y los dulces". Empecemos.

Trabajando en el contenido del capítulo 4

4.170 Veamos ahora el aspecto teórico de lo esperado. Mars, Inc., actualmente usa los siguientes porcentajes para mezclar los colores para dulces de chocolate con leche M&M's: 13% cafés, 13% rojos, 14% amarillos, 16% verdes, 20% naranjas, 24% azules.

a. Construya una gráfica de barras que muestre la proporción esperada (teórica) de M&M's para cada color.

b. Teóricamente, ¿qué porcentaje de M&M's rojos debe esperar en una bolsa de M&M's?

c. Si usted abre una bolsa de chocolates M&M's ahora mismo, ¿le sorprendería hallar porcentajes de colores diferentes de los dados por Mars? Explique.

Una mirada empírica (experimental) a lo que ocurrió.

d. Hágase de un paquete de chocolates M&M's (de al menos 1.69 onzas, a un costo aproximado de $0.50)

e. Registre el número de cada color en una distribución de frecuencia con encabezados "Color" y "Frecuencia".

f. Verifique el número total de los M&M's con la suma de la columna de Frecuencia.

g. Ya puede tomar un bocadillo.

h. Presente la distribución de frecuencia como distribución de frecuencia relativa, usando el encabezado "Probabilidad Empírica".

i. Verifique que la suma de la columna de Probabilidad Empírica sea igual a 1. Explique el significado de esta suma.

j. Construya una gráfica de barras que muestre la frecuencia relativa para cada color. Use el mismo orden de color que en la parte a.

k. Empíricamente, ¿qué porcentaje de chocolates M&M's rojos debe esperar en una bolsa de M&M's?

l. ¿Qué otros aspectos estadísticos podría usar para presentar los datos de la bolsa de los M&M's? Preséntelos.

m. Compare sus hallazgos empíricos (experimentales) contra las expectativas expresadas (teóricas) en la parte a.

Para su estudio

4.171 a. Use una computadora (o tabla de números aleatorios) para generar una muestra aleatoria de 56 M&M's, usando las correspondientes probabilidades teóricas para cada color.

b. Forme una distribución de frecuencia de los datos aleatorios.

c. Construya una gráfica de barras que muestre las frecuencias relativas para cada color. Use el mismo orden de color que en la parte a del ejercicio 4.170.

d. Compara sus hallazgos experimentales contra las expectativas teóricas.

e. Repita las partes a-d tres veces más.

f. Describa la variabilidad que observe entre las muestras.

g. Consolide sus cuatro distribuciones de frecuencia en una distribución de frecuencia que tenga un total de frecuencia de 224 M&M's.

h. Construya una barra de gráficas de la consolidación, que muestre frecuencias relativas para cada color. Use el mismo orden de color que en la parte a del ejercicio 4.170.

i. Compare estos hallazgos experimentales contra las expectativas teóricas.

j. Compare los hallazgos consolidados contra los cuatro hallazgos individuales previos.

k. ¿En qué forma impacta la ley de números grandes en este mini estudio?

MINITAB y Excel sólo pueden generar números aleatorios. Por tanto, es práctica común usar números en lugar de los colores (palabras). Use los números 1, 2, 3, 4, 5, 6 que correspondan a café, rojo, . . . , azul, respectivamente.

MINITAB Versión 14

a. Introduzca los números 1-6 en C1 y sus correspondientes probabilidades en C2; luego continúe con:

```
Choose:  Calc > Random Data > Discrete
Enter:   Generate: 56 (# of M&M's® in a pack)
         Store in column(s): C3
         Values in: C1 (color numbers)
         Probabilities in: C2 > OK
```

b. Para obtener la distribución de frecuencia, continúe con:

```
Choose:  Stat > Tables > Cross Tabulation & Chi Square
Enter:   Categorical variables:For rows:C3
Select:  Display: Counts and Column percents > OK
```

c. Para construir una barra de gráficas introduzca en C4 los colores reales y en C5 las correspondientes probabilidades (%) halladas en el paso b:

```
Choose:  Graph > Bar Chart > Bar represent: Values from a
         table > One Column of values: Simple > OK
Enter:   Graph variables: C5  Categorical variables: C4
Select:  Labels > Data Labels > Label Type: Use y-value
         labels > OK
Select:  Data View > Data Display: Bars > OK > OK
```

Excel

a. Introduzca los números 1-6 en la columna A y sus correspondientes probabilidades en la columna B; luego continúe con:

```
Choose:  Tools > Data Analysis > Random Number Generation
         > OK
Enter:   Number of Variables: 1
         Number of Random Numbers: 56 (# of M&M's® in a
         pack)
         Distribution: Discrete
         Value & Prob. Input Range: (A1:B7 select data
         cells)
Select:  Output range
Enter:   (C1 or select cell) > OK
```

b. La distribución de frecuencia está dada con el histograma de los datos generados. Use los comandos del histograma Excel de la página 61 usando los datos en la columna C y el rango de depósito en la columna A.

c. Divida las frecuencias entre 56 para obtener las correspondientes probabilidades. Introduzca los colores reales en la columna D (ej. D13:D18) y las correspondientes probabilidades en la columna E (ej. E13:E18). Para construir una gráfica de barras, continúe con:

```
Choose:  Chart Wizard > Column > 1st picture(usually)
         > Next
Enter:   Data range: (D13:E18 or select cells) > Next
Enter:   Chart and axes titles > Finish (Edit as
         needed)
```

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

4.1 La probabilidad de un evento es un **número entero.**

4.2 Los conceptos de probabilidad y frecuencia relativa en relación con un evento son **muy semejantes.**

4.3 El **espacio muestral** es la población teórica para problemas de probabilidad.

4.4 Los puntos muestrales de un espacio muestral son eventos **igualmente probables.**

4.5 El valor hallado para probabilidad experimental **siempre será** exactamente igual a la probabilidad teórica asignada al mismo evento.

4.6 Las probabilidades de eventos complementarios siempre **son iguales.**

4.7 Si dos eventos son mutuamente excluyentes, también son **independientes.**

4.8 Si los eventos A y B son **mutuamente excluyentes,** la suma de sus probabilidades debe ser exactamente 1.

4.9 Si los conjuntos de puntos muestrales que pertenecen a dos eventos diferentes no se intersecan, los eventos son **independientes.**

4.10 Un evento compuesto formado con la palabra "y" requiere el uso de la **regla de la adición.**

SEGUNDA PARTE: Aplicación de los conceptos

4.11 Una computadora está programada para generar los ocho enteros de un dígito 1, 2, 3, 4, 5, 6, 7 y 8 con la misma frecuencia. Considere el experimento "el siguiente entero generado" y estos eventos:

A: número non, {1, 3, 5, 7}

B: número mayor a 4, {5, 6, 7, 8}

C: 1 o 2, {1, 2}

a. Encuentre $P(\text{A})$.

b. Encuentre $P(\text{B})$.

c. Encuentre $P(\text{C})$.

d. Encuentre $P(\overline{\text{C}})$.

e. Encuentre $P(\text{A y B})$.

f. Encuentre $P(\text{A o B})$.

g. Encuentre $P(\text{B y C})$.

h. Encuentre $P(\text{B o C})$.

i. Encuentre $P(\text{A y C})$.

j. Encuentre $P(\text{A o C})$.

k. Encuentre $P(\text{A}|\text{B})$.

l. Encuentre $P(\text{B}|\text{C})$.

m. Encuentre $P(\text{A}|\text{C})$.

n. ¿Son mutuamente excluyentes los eventos A y B?

o. ¿Son mutuamente excluyentes los eventos B y C?

p. ¿Son mutuamente excluyentes los eventos A y C?

q. ¿Son independientes los eventos A y B? Explique.

r. ¿Son independientes los eventos B y C? Explique.

s. ¿Son independientes los eventos A y C? Explique.

4.12 Los eventos A y B son mutuamente excluyentes y $P(\text{A}) = 0.4$ y $P(\text{B}) = 0.3$.

a. Encuentre $P(\text{A y B})$.

b. Encuentre $P(\text{A o B})$.

c. Encuentre $P(\text{A}|\text{B})$.

d. ¿Son independientes los eventos A y B? Explique.

4.13 Los eventos E y F tienen probabilidades $P(\text{E}) = 0.5$, $P(\text{F}) = 0.4$, y $P(\text{E y F}) = 0.2$.

a. Encuentre $P(\text{E o F})$.

b. Encuentre $P(\text{E}|\text{F})$.

c. ¿Son mutuamente excluyentes E y F? Explique.

d. ¿Son independientes E y F? Explique.

e. ¿Son independientes G y H? Explique

4.14 Janice desea ser oficial de policía. Ella debe pasar un examen físico y luego uno escrito. Los registros indican que la probabilidad de pasar un examen físico es 0.85 y que una vez aprobado éste, la probabilidad de pasar el examen escrito es 0.60. ¿Cuál es la probabilidad de que Janice pase ambos exámenes?

TERCERA PARTE: Comprensión de los conceptos

4.15 El estudiante A dice que independencia y mutuamente excluyente son básicamente lo mismo; es decir, ambos significan que ningún evento tiene nada qué ver con el otro. El estudiante B dice que aun cuando lo dicho por el Estudiante A tiene algo de verdad, el Estudiante A ha equivocado el punto de estas dos propiedades. El Estudiante B tiene razón. Con todo cuidado explique por qué.

4.16 Usando oraciones completas, describa lo siguiente con sus propias palabras:

a. Eventos mutuamente excluyentes

b. Eventos independientes

c. La probabilidad de un evento

d. Una probabilidad condicional

CAPÍTULO

5 Distribuciones de probabilidad (variables discretas)

5.1 Bebidas con cafeína

5.2 Variables aleatorias

5.3 Distribuciones de probabilidad de una variable aleatoria discreta

5.4 Media y varianza de una distribución de probabilidad discreta

5.5 Distribución de probabilidad binomial

5.6 Media y desviación estándar de la distribución binomial

© Photodisc Red/Getty/Images

5.1 Bebidas con cafeína

¿Están adueñándose del país Starbucks y otros proveedores de café? Así parece. Una de las escenas más comunes es la de una persona con teléfono celular y tomando una taza de café. Veamos: ¿cuántas personas en esta situación ha visto usted hoy? ¡Quizás hasta usted sea una de ellas!

A LOS ESTADOUNIDENSES LES GUSTA TOMAR CAFÉ

Datos de Shannon Reilly y Alejandro Gonzalez. © 2005 USA Today.

Considere la gráfica "A los estadounidenses les gusta tomar café", que presenta el número de tazas o latas de bebidas con cafeína que adultos estadounidenses dicen que toman a diario. El número de tazas va de cero a cuatro tazas o más. ¿Se puede usted encontrar en la gráfica?

¿Quién más podría estar interesado en esta información además de Starbucks? Parece que la National Sleep Foundation, cuya frase de misión y metas, como lo dice su página web, es:

> La National Sleep Foundation (NSF) es una organización independiente, sin fines de lucro, dedicada a mejorar la salud y seguridad públicas al hacer que se comprenda el sueño y sus enfermedades, así como a apoyar la educación, investigación acerca del sueño, y su recomendación.
>
> *Fuente:* http://www.sleepfoundation.org

Con base en el sondeo hecho a 1506 adultos y la creencia generalizada de que el consumo de cafeína afecta el sueño, ¿le parece a usted que la cafeína debe ser un problema de la NSF? Es probable que no.

A medida que el estudiante avance en el capítulo 5, estará combinando los fundamentos de distribuciones de frecuencia del capítulo 2 con los fundamentos de probabilidad del capítulo 4. Esta combinación se denomina *distribuciones de probabilidad* que en realidad son muy semejantes a distribuciones de frecuencia. La diferencia básica entre distribuciones de probabilidad y distribuciones de frecuencia relativa es que las distribuciones de probabilidad son probabilidades teóricas (poblaciones), en tanto que las distribuciones de frecuencia son probabilidades empíricas (muestras). También podrá investigar más a fondo "A los estadounidenses les gusta tomar café" en la sección de Proyecto del capítulo con los ejercicios 5.128 y 5.129 (p.310).

SECCIÓN 5.1 EJERCICIOS

5.1 Consulte la gráfica "A los estadounidenses les gusta tomar café" para contestar las preguntas siguientes:

a. ¿Qué porcentaje de adultos no toman ninguna bebidas con cafeína?

b. ¿Qué número de tazas o latas de bebidas con cafeína tiene la probabilidad más alta?

c. ¿Qué variable podría usarse para describir los cinco eventos que se muestran en la gráfica?

d. Los eventos ¿son mutuamente excluyentes? Explique.

5.2 Consulte la gráfica "A los estadounidenses les gusta tomar café" para contestar las siguientes preguntas:

a. ¿Qué otra gráfica estadística podría usarse para presentar esta información? Dibújela.

b. ¿Qué otros métodos estadísticos podrían usarse para describir esta información?

5.2 Variables aleatorias

Si a cada uno de los resultados de un **experimento** de probabilidad se asigna un valor numérico, entonces cuando veamos los resultados del experimento estaremos observando los valores de una variable aleatoria. Este valor numérico es el *valor de la variable aleatoria.*

Variable aleatoria: es una variable que toma un valor numérico único para cada uno de los resultados del espacio muestral de un experimento de probabilidad.

En otras palabras, se utiliza una variable aleatoria para denotar los resultados de un experimento de probabilidad. La variable aleatoria puede tomar cualquier valor numérico que pertenezca al conjunto de todos los posibles resultados del experimento. (Se denomina "aleatoria" porque el valor que toma es el resultado de un evento de probabilidad, o aleatorio.) Cada uno de los eventos en un experimento de probabilidad también debe definirse en forma tal que sólo se le asigne un valor de la variable aleatoria **(eventos mutuamente excluyentes)**, y todo evento debe tener asignado un valor **(eventos todo incluido)**.

El siguiente ejemplo demuestra las variables aleatorias.

EJEMPLO 5.1 Variables aleatorias

a. Lanzamos al aire cinco monedas y observamos el "número de caras" visible. La variable aleatoria x es el número de caras observadas y puede tomar valores enteros de 0 a 5.
b. Sea el "número de llamadas telefónicas recibidas" por día por una compañía la variable aleatoria. Son posibles valores los valores enteros que van de cero a algún número muy grande.
c. Sea la "longitud del cable", de un aparato eléctrico, una variable aleatoria. La variable aleatoria es un valor numérico entre 12 y 72 pulgadas para casi todos los aparatos.
d. Sea la "velocidad para calificar" una variable aleatoria en autos de carrera que tratan de calificar para la Indianápolis 500. Dependiendo de qué tan rápido corra el piloto, las velocidades son aproximadamente 220 y más y se miden en millas por hora (al milésimo más cercano).

Las variables numéricas aleatorias se pueden subdividir en dos clasificaciones: *variables aleatorias discretas* y *variables aleatorias continuas.*

PARA SU INFORMACIÓN
Las variables discretas y continuas se definen en la página 11.

Variable discreta aleatoria: es una variable cuantitativa aleatoria que puede tomar un número contable de valores.

Variable continua aleatoria: es una variable cuantitativa aleatoria que puede tomar un número incontable de valores.

Las variables aleatorias "número de caras" y "número de llamadas telefónicas recibidas" del ejemplo 5.1, partes a y b, son discretas. Cada una de ellas representa una cuenta y, por tanto, hay un número contable de posibles valores. Las variables aleatorias "longitud de cable" y "velocidad para calificar", partes c y d, son continuas. Cada una representa mediciones que pueden tomar cualquier valor en todo un intervalo y, por tanto, hay un número infinito de valores posibles.

SECCIÓN 5.2 EJERCICIOS

5.3 Haga una encuesta entre sus compañeros de clase acerca del número de hermanos que tienen y la duración de la última conversación que tuvieron con su madre. Identifique las dos variables aleatorias de interés y haga una lista de sus posibles valores.

5.4 a. Explique por qué la variable "cantidad de números telefónicos guardados en el teléfono celular de una persona" es discreta.
b. Explique por qué la variable "peso de un libro de texto de estadística" es continuo.

5.5 a. Las variables del ejercicio 5.3 pueden ser discretas o continuas. ¿Cuáles son y por qué?
b. Explique por qué la variable "número de invitados para la cena de día de gracias" es discreta.
c. Explique por qué la variable "número de millas a la casa de su abuela" es continua.

5.6 Una trabajadora social interviene en un estudio acerca de la estructura familiar. De los datos del censo, ella obtiene información relacionada con el número

de hijos por familiar para cierta comunidad. Identifique la variable aleatoria de interés, determine si es discreta o continua y haga una lista de sus posibles valores.

5.7 El personal de *Fortune* recientemente aisló a las que consideraron eran las mejores 100 empresas de Estados Unidos para trabajar. Muchas de las de esa lista estuvieron contratando personal el año pasado. Las que aumentaron su personal aparecen a continuación.

Compañía	Nuevos empleos	Compañía	Nuevos empleos
Marriott International	3679	Booz Allen Hamilton	2463
Whole Foods Market	3569		

Fuente: Fortune, "The 100 Best Companies to Work for 2005"

a. ¿Cuál es la variable aleatoria que interviene en este estudio?

b. ¿Es discreta o continua la variable aleatoria? Explique.

5.8 El calor por encima del promedio se extendió hacia el este y el sudeste el 13 de enero de 2005. Las temperaturas altas pronosticadas para el día en cuatro ciudades de la zona afectada fueron como sigue:

Ciudad	Temperatura	Ciudad	Temperatura
Burlington, VT	55°F	Durham, NC	74°F
Williamsburg, VA	74°F	Augusta, GA	75°F

a. ¿Cuál es la variable aleatoria que interviene en este estudio?

b. ¿Es discreta o continua la variable aleatoria? Explique.

5.9 Un arquero dispara flechas a una "diana" de un blanco y mide la distancia desde el centro del blanco a la flecha. Identifique la variable aleatoria de interés, determine si es discreta o continua y haga una lista de sus posibles valores.

5.10 Un ejemplar de USA Snapshot titulado "Are we getting a summer job?" (¿Tendremos trabajo para el verano?) reportó que 49% de estudiantes de preparatoria dijeron "¿Tener? Ya tengo uno"; 26% dijeron, "Quizá. Depende de mi situación económica"; y 25% dijeron, "No, nada interfiere con mi tiempo en la playa".

a. Cuál es la variable que interviene, y cuáles son los posibles valores?

b. ¿Por qué no es aleatoria esta variable?

5.11 Un artículo de *USA Today* titulado "Un estudio descubre que el mundo electrónico se traga el tiempo de los niños" (10 de marzo, 2005) presentó la tabla siguiente, que describe el tiempo promedio de niños de 8 a 18 años que pasan a diario en varias actividades. La fundación de la familia Kaiser había realizado el estudio de 2000 niños de grado escolar del 3 al 12.

Actividad	Tiempo promedio
Ver TV	3 horas, 51 minutos
Escuchar música	1 hora, 44 minutos
Usar computadora	1 hora, 2 minutos
Juegos de video	49 minutos
Leer	43 minutos
Ver cine	25 minutos

a. ¿Cuál es la variable aleatoria que interviene en este estudio?

b. ¿Es discreta o continua la variable aleatoria? Explique.

5.12 Si se pudiera detener el tiempo y vivir para siempre en buena salud, ¿qué edad escogería usted? Las respuestas a esta pregunta se reportaron en un artículo de USA Snapshot. La edad promedio ideal para cada grupo de edad aparece en la tabla siguiente; se encontró que 41 es el promedio de edad ideal para todos los adultos. Es curioso que los menores de 30 años desean ser mayores, en tanto que los de más de 30 años desean sea más jóvenes.

Grupo de edad	18–24	25–29	30–39	40–49	50–64	65+
Edad ideal	27	31	37	40	44	59

Se usa la edad dos veces como variable en esta aplicación.

a. La edad de una persona entrevistada no es la variable aleatoria en esta situación. Explique por qué y describa cómo se usa "edad" respecto al grupo de edad.

b. ¿Cuál es la variable aleatoria que interviene en este estudio? Describa su papel en esta situación.

c. ¿Es discreta o continua la variable aleatoria? Explique.

5.3 Distribuciones de probabilidad de una variable aleatoria discreta

Considere un experimento de lanzar al aire dos monedas, donde no salen "caras", sale una y salen dos caras. Si definimos la variable aleatoria x como el número de "caras" observado cuando se lanzan dos monedas, x puede tomar el valor de 0, 1 y 2. La probabilidad de cada uno de estos tres eventos se puede calcular usando técnicas del capítulo 4:

$$P(x = 0) = P(0\text{H}) = P(\text{TT}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

$$P(x = 1) = P(1\text{H}) = P(\text{HT or TH}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} = 0.50$$

$$P(x = 2) = P(2\text{H}) = P(\text{HH}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

TABLA 5.1

Distribución de probabilidad: lanzar dos monedas

x	$P(x)$
0	0.25
1	0.50
2	0.25

Estas probabilidades se pueden poner en lista en varias formas. Una de las más convenientes es un formato de tabla conocido como *distribución de probabilidad* (vea la tabla 5.1).

PARA SU INFORMACIÓN

¿Puede usted ver por qué se usa el nombre de "distribución de probabilidad"?

Distribución de probabilidad: es una distribución de probabilidades asociada con cada uno de los valores de una variable aleatoria. La distribución de probabilidad es una distribución teórica; se usa para representar poblaciones.

En un experimento donde un solo dado se lanza y se observa el número de puntos de su cara superior, la variable aleatoria es el número observado. La distribución de probabilidad para esta variable aleatoria se muestra en la tabla 5.2.

TABLA 5.2

Distribución de probabilidad: lanzar un dado

x	1	2	3	4	5	6
$P(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

A veces es conveniente escribir una regla que algebraicamente expresa la probabilidad de un evento en términos del valor de la variable aleatoria. Esta expresión típicamente se escribe como fórmula y se denomina *función de probabilidad.*

Función de probabilidad: es una regla que asigna probabilidades a los valores de las variables aleatorias.

Una función de probabilidad se puede ver tan sencilla como una lista que parea los valores de una variable aleatoria con sus probabilidades\. Las tablas 5.1 y 5.2 muestran estas dos listas; no obstante, una función de probabilidad se expresa con más frecuencia como fórmula.

Considere un dado que ha sido modificado de modo que tiene una cara con un punto, dos caras con dos puntos, y tres caras con tres puntos. Sea x el número de puntos observados cuando se tira el dado. La distribución de probabilidad para este experimento se presenta en la tabla 5.3.

TABLA 5.3

Distribución de probabilidad: tiro del disco modificado

x	$P(x)$
1	$\frac{1}{6}$
2	$\frac{2}{6}$
3	$\frac{3}{6}$

Cada una de las probabilidades pueden ser representadas por el valor de x dividido entre 6; es decir, cada $P(x)$ es igual al valor de x dividido entre 6, donde x = 1,2 o 3. Así.

$$P(x) = \frac{x}{6} \qquad \text{para} \qquad x = 1, 2, 3$$

es la fórmula para la función de probabilidad de este experimento.

La función de probabilidad para el experimento de tirar un dado ordinario es

$$P(x) = \frac{1}{6} \qquad \text{para} \qquad x = 1, 2, 3, 4, 5, 6$$

Esta función particular se denomina **función constante** porque el valor de $P(x)$ no cambia cuando x cambia.

Toda función de probabilidad debe presentar las dos propiedades básicas de probabilidad (vea p. 212). Estas dos propiedades son (1) la probabilidad asignada a cada valor de la variable aleatoria debe ser cero y uno, inclusivo y (2) la suma de las probabilidades asignadas a todos los valores de la variable aleatoria debe ser igual a 1, es decir,

PARA SU INFORMACIÓN
Estas propiedades fueron presentadas en el capítulo 4.

Propiedad 1 $0 \leq \text{cada } P(x) \leq 1$

Propiedad 2 $\sum_{\text{toda } x} P(x) = 1$

EJEMPLO 5.2 Determinación de una función de probabilidad

¿Es $P(x) = \frac{x}{10}$ para x = 1, 2, 3, 4 una función de probabilidad?

SOLUCIÓN Para contestar esta pregunta sólo necesitamos probar la función en términos de las dos propiedades básicas. La distribución de probabilidad se muestra en la tabla 5.4.

TABLA 5.4

Distribución de probabilidad para $P(x) = \frac{x}{10}$ para x = 1, 2, 3, 4

x	$P(x)$	
1	$\frac{1}{10} = 0.1$	✓
2	$\frac{2}{10} = 0.2$	✓
3	$\frac{3}{10} = 0.3$	✓
4	$\frac{4}{10} = 0.4$	✓
	$\frac{10}{10} = 1.0$	ck

La propiedad 1 se satisface porque 0.1, 0.2, 0.3 y 0.4 son todos ellos valores numéricos entre 0 y 1. (Vea el símbolo ✓ que indica que cada uno de los valores ha sido comprobado.) La propiedad 2 también queda satisfecha porque la suma de las cuatro probabilidades es exactamente uno. (Vea el ck que indica que la suma ha sido comprobada.) Como ambas propiedades han sido satisfechas, podemos concluir que $P(x) = \frac{x}{10}$ para x = 1, 2, 3, 4 es una función de probabilidad.

¿Qué hay acerca de $P(x = 5)$ (o cualquier valor que no sea x = 1, 2, 3 o 4) para la función $P(x) = \frac{x}{10}$ para x = 1, 2, 3, 4? $P(x = 5)$ se considera como cero. Esto es, la función de probabilidad produce una probabilidad de cero para todos los valores de x que no sean los valores especificados como parte del dominio.

Las distribuciones de probabilidad se pueden presentar gráficamente. Cualquiera que sea la representación gráfica específica que se utilice, los valores de la variable aleatoria se grafican en la escala horizontal, y la probabilidad asociada con cada valor de la variable aleatoria se grafican en la escala vertical. La distribución de probabilidad de una variable aleatoria discreta podría ser presentada por un conjunto de segmentos de recta trazados en los valores de x con longitudes que representan

la probabilidad de cada x. La figura 5.1 muestra la distribución de probabilidad de $P(x) = \dfrac{x}{10}$ para $x = 1, 2, 3, 4$.

FIGURA 5.1 Representación de recta: distribución de probabilidad para $P(x) = \dfrac{x}{10}$ para $x = 1, 2, 3, 4$

FIGURA 5.2 Histograma: distribución de probabilidad para $P(x) = \dfrac{x}{10}$ para $x = 1, 2, 3, 4$

PARA SU INFORMACIÓN
La gráfica de la figura 5.1 a veces se llama gráfica de aguja.

Un histograma regular se emplea con más frecuencia para presentar distribuciones de probabilidad. La figura 5.2 presenta la distribución de probabilidad de la figura 5.1 como un **histograma de probabilidad.** El histograma de una distribución de probabilidad usa el área física de cada barra para representar su probabilidad asignada. La barra para $x = 2$ es de 1 unidad de ancho (de 1.5 a 2.5) y de 0.2 unidades de alto. Por tanto, su área (longitud ancho) es $(1)(0.2) = 0.2$, la probabilidad asignada a $x = 2$. Las áreas de las otras barras se pueden determinar de modo semejante. Esta representación de área será un concepto importante en el capítulo 6 cuando empecemos a trabajar con variables aleatorias continuas.

INSTRUCCIONES DE TECNOLOGÍA: GENERAR DATOS ALEATORIOS

MINITAB (Versión 14) Introduzca los posibles valores de la variable aleatoria en C1 y las probabilidades correspondientes en C2; luego continúe con:

```
Choose:     Calc > Random Data > Discrete
Enter:      Generate: 25 (number wanted)
            Store in column(s): C3
            Values (of x) in: C1
            Probabilities in: C2 > OK
```

Excel Introduzca los posibles valores de la variable aleatoria en la columna A y las probabilidades correspondientes en la columna B; luego continúe con:

```
Choose:     Tools > Data Analysis >Random Number Generation > OK
Enter:      Number of Variables: 1
            Number of Random Numbers: 25 (# wanted)
            Distribution: Discrete
            Value & Prob. Input Range: (A2:B5 select data cells, not labels)
Select:     Output Range
Enter       (C1 or select cell)
```

CASO PRÁCTICO 5.3 Solicitud de admisión

Estudiantes compensan sus apuestas

La mayoría de estudiantes hacen solicitud en más de una escuela, lo cual hace difícil que los colegios pronostiquen a cuántos inscribirán en realidad. Al grupo de primer año del semestre pasado se le preguntó:

¿En cuántos colegios solicitó usted admisión este año, además de aquel en el que estuvo inscrito?

Fuente: The Aerican Freshman: National Norms for Fall 2001; encuesta de 281, 064 alumnos de primer año que entraron a 421 colegios y universidades de 4 años.

Datos de Julie Snider, 2002 USA Today

LOS COLEGIOS LUCHAN POR LLENAR DORMITORIOS

Por Mary Beth Marklein, *USA Today*

Colegios y universidades enviarán por correo su último lote de ofertas de admisión en los próximos días, pero el proceso está lejos de terminar.

Ahora, los estudiantes tienen hasta el 1 de mayo para decidir a dónde irán este verano. Con preocupaciones persistentes acerca de la economía y temores residuales en viajes y seguridad desde el 11 de septiembre, muchos oficiales de admisiones podrán este año pronosticar cómo responderán los estudiantes.

Nótese la distribución descrita en la gráfica de barras. Tiene las cualidades de una distribución discreta de probabilidad. La variable aleatoria, "número de colegios a los que se hace solicitud", es una variable aleatoria discreta con valores de 0 a 11 o más. Cada uno de los valores tiene una probabilidad correspondiente, y la suma de las probabilidades es igual a 1.

SECCIÓN 5.3 EJERCICIOS

5.13 Exprese el tiro al aire de una moneda como distribución de probabilidad de x, el número de caras que salen (es decir, $x = 1$ si sale una cara y $x = 0$ si sale cruz).

5.14 a. Exprese $P(x) = \frac{1}{6}$; para x = 1, 2, 3, 4, 5, 6, en forma de distribución.

b. Construya un histograma de la distribución de probabilidad $P(x) = \frac{1}{6}$; para $x = 1, 2, 3, 4, 5, 6$.

c. Describa la forma del histograma de la parte b.

5.15 a. Explique el modo en que los diversos valores de x de una distribución de probabilidad forman un conjunto de eventos mutuamente exclusivos.

b. Explique el modo en que los diversos valores de x de una distribución de probabilidad forman un conjunto de eventos "todo incluido".

5.16 Compruebe la siguiente función para determinar si es una función de probabilidad. Si no lo es, trate de hacer que sea una función de probabilidad.

$$R(x) = 0.2 \text{ para } x = 1, 2, 3, 4.$$

a. Haga una lista de la distribución de probabilidades.

b. Dibuje un histograma.

5.17 Compruebe la siguiente función para determinar si es una función de probabilidad.

$$P(x) = \frac{x^2 + 5}{50}, \text{ para } x = 1, 2, 3, 4.$$

a. Haga una lista de la distribución de probabilidades.

b. Dibuje un histograma.

5.18 Compruebe la siguiente función para determinar si es una función de probabilidad. Si no lo es, trate de hacer que sea una función de probabilidad.

$$S(x) = \frac{6 - |x - 7|}{36}, \text{ para } x = 2, 3, 4, 5, 6, 7, \ldots, 11, 12$$

a. Haga una lista de la distribución de probabilidades.

b. ¿Reconoce a $S(x)$? Si es así, identifíquela.

5.19 Con frecuencia se usan los datos del censo para obtener distribuciones de probabilidad para diversas variables aleatorias. Los datos del censo para familiar en un estado en particular con un ingreso combinado de $50 000 o más muestran que 20% de estas familias no tienen hijos, 30% tienen un hijo, 40% tienen dos hijos, y 10% tienen tres hijos. De esta información, construya la distribución de probabilidad para x, donde x representa el número de hijos por familia para este grupo de ingresos.

5.20 "¿El perro, es el mejor amigo del hombre"? Lo pensaríamos así por los 60 millones de perros mascota que hay en todo el país. Pero, ¿cuántos amigos se necesitan? En un artículo de USA Snapshot (22 de febrero, 2005), se publicaron las siguientes estadísticas.

Número de perros mascotas	Porcentaje	Número de perros mascotas	Porcentaje
Uno	6	Cuatro	3
Dos	24	Cinco o más	2
Tres	5		

Fuente: U.S.Pet Ownership & Demographics Sourcebook, Pedigree Food for Dogs

a. ¿Es ésta una distribución de probabilidad? Explique.

b. Trace un histograma de frecuencia relativa para describir los resultados mostrados en la tabla.

5.21 ¿En cuántos colegios hizo usted solicitud, además de aquel en el que está inscrito? Ésta fue exactamente la pregunta hecha e ilustrada en el Caso práctico 5.3, "Solicitud de admisión," en la página 276.

a. Usando la variable x, enumere las solicitudes adicionales de admisión con sus datos, exprese la información en la gráfica de barras "Estudiantes compensan sus apuestas" como una distribución de probabilidad discreta.

b. Explique la forma en que la distribución apoya la frase inicial del artículo, "pero el proceso está lejos de terminar".

5.22 En febrero de 2004, el Oregon's Medically Needy Program Survey publicó las siguientes estadísticas respecto de la población médicamente necesitada en sus estados. Debido a recortes presupuestales, fue eliminado el programa para los médicamente necesitados que había proporcionado asistencia médica a ciertos grupos, que no tenían derecho a atención médica, pero tenían importantes necesidades de salud.

Número de enfermedades crónicas	Población de Oregon médicamente necesitada	Número de enfermedades crónicas	Población de Oregon médicamente necesitada
0	2%	3	21%
1	12%	4–5	31%
2	23%		

Fuente: http://www.ohpr.state.or.us/UHRECwelcome2_files/ReportsandBriefs/MedicallyNeedyFINAL.pdf

a. ¿Es ésta una distribución de probabilidad? Explique.

b. ¿Qué información podría usted agregar para que sea una distribución de probabilidad?

c. Trace un histograma de frecuencia para describir los resultados mostrados en la tabla más la parte b.

5.23 Como parte de un informe de preferencias del consumidor del año 2003, se reunió la siguiente información. Indica el porcentaje, como frecuencia relativa, de clientes con casa nueva que deseaban cada uno de ellos tener elementos exteriores como parte de su nueva casa. ¿Es ésta una distribución de frecuencia? Explique.

Elemento exterior	Porcentaje (Frec. Rel.)	Elemento exterior	Porcentaje (Frec. Rel.)
Pórtico de entrada	0.56	Fencing	0.23
Piso	0.49	Landscape wall	0.14
Deck	0.35		

Fuente: NAHB Research Center

5.24 Un artículo de USA Snapshot (10 de marzo, 2005) presentó una gráfica de barras que describía la impresión, de quienes viajaban por trabajo, respecto a los tiempos de espera en las líneas de seguridad de aeropuertos en los últimos 12 meses. Las estadísticas se obtuvieron de 2034 personas que respondieron a una encuesta de la Asociación de la Industria de Viajes de la American Business Traveler. ¿Es ésta una distribución de probabilidad? Explique.

Impresión	Porcentaje	Impresión	Porcentaje	Impresión	Porcentaje
Peor	49	Igual	40	Mejor	11

5.25 a. Use computadora (tabla de números aleatorios) para generar una muestra aleatoria de 25 observaciones extraídas de la distribución de probabilidad discreta.

x	1	2	3	4	5
$P(x)$	0.2	0.3	0.3	0.1	0.1

Compare los datos resultantes contra sus expectativas.

b. Forme una distribución de frecuencia relativa de los datos aleatorios.

c. Construya un histograma de probabilidad de la distribución dada y un histograma de frecuencia relativa de los datos observados usando los puntos medios de clase de 1, 2, 3, 4 y 5.

d. Compare los datos observados con la distribución teórica. Describa sus conclusiones.

e. Repita las partes a-d varias veces con $n = 25$. Describa la variabilidad que usted observe entre muestras.

f. Repita las partes a-d varias veces con $n = 250$. Describa la variabilidad que vea entre muestras de este tamaño mucho mayor.

MINITAB (Versión 14)

a. Introduzca los valores x de la variable aleatoria en C1 y sus correspondientes probabilidades, $P(x)$, en C2; luego continúe con los comandos MINITAB que generan datos aleatorios de la página 275.

b. Para obtener la distribución de frecuencia, continúe con:

```
Choose:   Stat > Tables > Cross Tabulation
Enter:    Categorical variables: For rows: C3
Select:   Display: Total percents > OK
```

c. Para construir el histograma de los datos generados en C3, continúe con los comandos MINITAB del histograma de la página 61, seleccionando escala>Y-Scale Type>Percent. (Use Binning seguido por punto medio y posiciones de punto medio 1:5/1 si es necesario.)
Para construir una gráfica de barras de la distribución dada, continúe con los comandos MINITAB de gráfica de barras de la página 266, usando C2 como la variable de gráfica y C1 como la variable categórica.

Excel

a. Introduzca los valores x de la variable aleatoria en la columna A y sus correspondientes probabilidades, $P(x)$, en la columna B; luego continúe con los comandos Excel de generación de datos aleatorios de la página 275 para $n = 25$.

b. y c. La distribución de frecuencia está dada con el histograma de los datos generados. Use los comandos Excel del histograma de las páginas 61-62 usando los datos en la columna C y el rango del depósito en la columna A.
Para construir un histograma de la distribución dada, continúe con:

```
Choose:   Chart Wizard > Column > 1st picture(usually) >
          Next
Enter:    Data range: (A1:B6 or select cells)
Choose:   Series > Remove (Series 1: x column) > Next >
          Titles
Enter:    Chart and axes titles > Finish  (Edit as needed)
```

5.26 a. Use una computadora (o tabla de números aleatorios) y genere una muestra aleatoria de 100 observaciones extraídas de la población de probabilidad discreta $P(x) = \dfrac{5 - x}{10}$ para $x = 1, 2, 3, 4$. Haga una lista de la muestra resultante. (Use los comandos de computadora del ejercicio 5.25; simplemente cambie los argumentos.)

b. Forme una distribución de frecuencia relativa de los datos aleatorios.

c. Forme una distribución de probabilidad de la distribución de probabilidad esperada. Compare los datos resultantes con sus expectativas.

d. Construya un histograma de probabilidad de la distribución dada y un histograma de frecuencia relativa de los datos observados, usando puntos medios de clase de 1, 2, 3 y 4.

e. Compare los datos observados con la distribución teórica. Describa sus conclusiones.

f. Repita las partes a-d varias veces con $n = 100$. Describa la variabilidad que observe entre muestras.

5.4 Media y varianza de una distribución de probabilidad discreta

Recuerde que en el capítulo 2 calculamos varias estadísticas muestrales numéricas (media, varianza, desviación estándar y otras) para describir conjuntos empíricos de datos. Las distribuciones de probabilidad se pueden usar para representar pobla-

ciones teóricas, las similares de las muestras. Usamos **parámetros de población** (media, varianza y desviación estándar) para describir estas distribuciones de probabilidad igual que usamos **estadísticas muestrales** para describir muestras.

Notas:

1. $\bar{x}$ es la media de la muestra.
2. s^2 y s son la varianza y la desviación estándar de la muestra, respectivamente.
3. $\bar{x}$, s^2 y s se llaman *estadísticas muestrales*.
4. μ es la media de la población.
5. σ^2 es la varianza de la población.
6. σ es la desviación estándar de la población.
7. μ, σ^2 y σ son llaman *parámetros de población.* (Un parámetro es una constante; μ, σ^2 y σ son por lo general valores desconocidos en problemas de estadística reales. Casi siempre, la única vez que son conocidos es en un problema de un libro de texto con la finalidad de aprender y entender.)

La *media de la distribución de probabilidad* de una variable aleatoria discreta, o la *media de una variable aleatoria discreta,* se encuentra de un modo un tanto semejante al que se emplea para hallar la media de una distribución de frecuencia. La media de una variable aleatoria discreta frecuentemente se conoce como su *valor esperado.*

Media de una variable aleatoria discreta (valor esperado): la media, μ, de una variable aleatoria discreta x se encuentra al multiplicar cada posible valor de x por su propia probabilidad y luego sumar todos los productos:

media de x: mu = suma de (cada x multiplicada por su propia probabilidad)

$$\mu = \sum[xP(x)] \tag{5.1}$$

La varianza de una variable aleatoria discreta se define en forma muy semejante que la varianza de datos muestrales, la media de las desviaciones cuadradas desde la media.

Varianza de una variable aleatoria discreta: la varianza, μ^2, de una variable aleatoria discreta x se encuentra multiplicando cada posible valor del cuadrado de la desviación desde la media, $(x - \mu)^2$, por su propia probabilidad y luego sumando todos los productos:

varianza: sigma cuadrada = suma de (cuadrado de la desviación por la probabilidad)

D

$$\sigma^2 = \sum[(x - \mu)^2 P(x)] \tag{5.2}$$

Con frecuencia no es conveniente usar la fórmula (5.2); puede trabajarse otra vez en la(s) forma(s) siguiente(s):

varianza: sigma cuadrada = suma de (x^2 por probabilidad)−[suma de (x por probabilidad)]^2

$$\sigma^2 = \sum[x^2P(x)] - \{\sum[xP(x)]\}^2 \tag{5.3a}$$

o

$$\sigma^2 = \sum[x^2P(x)] - \mu^2 \tag{5.3b}$$

Del mismo modo, la desviación estándar de una variable aleatoria se calcula en la misma forma que la desviación estándar de datos muestrales.

Desviación estándar de una variable aleatoria discreta: es la raíz cuadrada positiva de varianza.

$$\text{desviación estándar: } = \sigma = \sqrt{\sigma^2} \tag{5.4}$$

EJEMPLO 5.4 Estadística para una función (distribución) de probabilidad

Encuentre la media, varianza y desviación estándar de la función de probabilidad

$$P(x) = \frac{x}{10} \quad \text{para } x = 1, 2, 3, 4$$

SOLUCIÓN Encontraremos la media con la fórmula (5.1), la varianza con la fórmula (5.3a) y la desviación estándar con la fórmula (5.4). La forma más conveniente que necesitamos para organizar los productos y hallar los totales es expandir la distribución de probabilidad en una tabla de extensiones (vea tabla 5.5).

TABLA 5.5

Tabla de extensiones: distribución de probabilidad, $P(x) = \frac{x}{10}$ para $x = 1, 2, 3, 4$

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
1	$\frac{1}{10} = 0.1$ ✓	0.1	1	0.1
2	$\frac{2}{10} = 0.2$ ✓	0.4	4	0.8
3	$\frac{3}{10} = 0.3$ ✓	0.9	9	2.7
4	$\frac{4}{10} = 0.4$ ✓	1.6	16	6.4
	$\frac{10}{10} = 1.0$ ck	$\sum[xP(x)] = 3.0$		$\sum[x^2P(x)] = 10.0$

Encuentre la media de x: la columna $xP(x)$ contiene cada uno de los valores de x multiplicado por su probabilidad correspondiente, y la suma que aparece abajo es el valor necesario en la fórmula (5.1):

$$\mu = \sum[xP(x)] = \mathbf{3.0}$$

Encuentre la varianza de x: los totales que aparecen debajo de las columnas $xP(x)$ y $x^2P(x)$ se sustituyen en la fórmula (5.3a):

$$\begin{aligned}\sigma^2 &= \sum[x^2P(x)] - \{\sum[xP(x)]\}^2 \\ &= 10.0 - \{3.0\}^2 = \mathbf{1.0}\end{aligned}$$

Encuentre la desviación estándar de x: use la fórmula (5.4):

$$\sigma = \sqrt{\sigma^2} = \sqrt{1.0} = \mathbf{1.0}$$

Notas:

1. El propósito de la tabla de extensiones es organizar el proceso de hallar los tres totales de columna: $\Sigma\,[P(x)]$, $xP(x)]$, y $\Sigma\,[x^2P(x)]$.
2. Las otras columnas, x y x^2, no deben sumarse; no se usan.
3. $\Sigma[P(x)]$ siempre será 1.0; use esto sólo como prueba.
4. $\Sigma[xP(x)]$ y $\Sigma\,[x^2P(x)]$ se usan para hallar la media y varianza de x.

EJEMPLO 5.5 Media, varianza y desviación estándar de una variable aleatoria discreta

Una moneda se lanza al aire tres veces. Sea el "número de caras", que sale en esos tres lanzamientos, la variable aleatoria, x. Encuentre la media, varianza y desviación estándar de x.

SOLUCIÓN Hay ocho posibles resultados (todos igualmente probables) a este experimento: {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}. Un resultado se manifiesta en $x = 0$, tres en $x = 1$, tres en $x = 2$, y uno en $x = 3$. Por tanto, las probabilidades para esta variable aleatoria son $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, y $\frac{1}{8}$. La distribución de probabilidad asociada con este experimento se muestra en la figura 5.3 y en la tabla 5.6. Las extensiones y sumas necesarias para el cálculo de la media, varianza y desviación estándar también se muestran en la tabla 5.6.

FIGURA 5.3
Distribución de probabilidad: número de caras y tres tiros de moneda

TABLA 5.6

Tabla de extensiones de distribución de probabilidad del número de caras en tres tiros de moneda

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	$\frac{1}{8}$ ✓	$\frac{0}{8}$	0	$\frac{0}{8}$
1	$\frac{3}{8}$ ✓	$\frac{3}{8}$	1	$\frac{3}{8}$
2	$\frac{3}{8}$ ✓	$\frac{6}{8}$	4	$\frac{12}{8}$
3	$\frac{1}{8}$ ✓	$\frac{3}{8}$	9	$\frac{9}{8}$

$\sum[P(x)] = \frac{8}{8} = 1.0$ ck $\quad \sum[xP(x)] = \frac{12}{8} = 1.5 \quad \sum[x^2P(x)] = \frac{24}{8} = 3.0$

La media se encuentra con la fórmula (5.1):

$$\mu = \sum [xP(x)] = \mathbf{1.5}$$

Este resultado, 1.5, es la media de la distribución teórica para la variable aleatoria "número de caras" observado por conjunto de tres tiros de moneda. Se espera que la media para muchos valores observados de la variable aleatoria también sea aproximadamente igual a este valor.

La varianza se encuentra usando la fórmula (5.3a):

$$\begin{aligned}\sigma^2 &= \sum [x^2P(x)] - \left\{\sum [xP(x)]\right\}^2 \\ &= 3.0 - \{1.5\}^2 = 3.0 - 2.25 = \mathbf{0.75}\end{aligned}$$

La desviación estándar se encuentra usando la fórmula (5.4):

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.75} = 0.866 = \mathbf{0.87}$$

Esto es, 0.87 es la desviación estándar de la distribución teórica para la variable aleatoria "número de caras" observado por conjunto de tres tiros de moneda. Se espera que la desviación estándar para muchos valores observados de la variable aleatoria sea aproximadamente igual a este valor.

SECCIÓN 5.4 EJERCICIOS

5.27 Verifique que las fórmulas (5.3a) y (5.3b) sean equivalentes a la fórmula (5.2).

5.28 a. Forme la tabla de distribución de probabilidad para $P(x) = \frac{x}{6}$, para $x = 1, 2, 3$.

b. Encuentre las extensiones $xP(x)$ y $x^2P(x)$ para cada x.

c. Encuentre $\Sigma[xP(x)]$ y $\Sigma[x^2P(x)]$.

d. Encuentre la media para $P(x) = \frac{x}{6}$, para $x = 1, 2, 3$.

e. Encuentre la varianza para $P(x) = \frac{x}{6}$, para $x = 1, 2, 3$.

f. Encuentre la desviación estándar para $P(x) = \frac{x}{6}$, para $x = 1,2,3$.

5.29 Si usted encuentra la suma de las columnas de x y de x^2 en la tabla de extensiones, ¿exactamente qué ha encontrado?

5.30 Dada la función de probabilidad $P(x) = \frac{5 - x}{10}$ para $x = 1, 2, 3, 4$, encuentre la media y desviación estándar.

5.31 Dada la función de probabilidad $R(x) = 0.2$ para $x = 1, 2, 3, 4$, encuentre la media y desviación estándar.

5.32 a. Dibuje un histograma de la distribución de probabilidad para los números aleatorios de un solo dígito 0, 1, 2, . . . ,9.

b. Calcule la media y desviación estándar asociadas con la población de números aleatorios de un solo dígito.

c. Represente (1) la ubicación de la media en el histograma con una recta vertical y (2) la magnitud de la desviación estándar con un segmento de recta.

d. ¿Cuánto de esta distribución de probabilidad está a no más de 2 desviaciones estándar de la media?

5.33 El pronóstico de huracanes se ha convertido en una de las bellas artes en Florida. Se hace necesaria una combinación de meteorología y estadística para construir modelos de pronósticos. La siguiente distribución de probabilidad se publicó en "Lo que significan los pronósticos de huracanes estacionales para los residentes de Florida" en abril de 2003.

Número de huracanes en Florida	Probabilidad anual	Número de huracanes en Florida	Probabilidad anual
0	0.60	3	0.02
1	0.30	4	0.01
2	0.07		

Fuente: http://garnet.acns.fsu.edu/~jelsner/PDF/Research/Floridafcsts.pdf

a. Construya una tabla de extensiones de la distribución de probabilidad y úsela para hallar la media y desviación estándar del número de huracanes que se presentan anualmente en Florida.

b. Dibuje el histograma de las frecuencias relativas.

5.34 En un artículo de USA Snapshot (12 de junio, 2002), la oficina del censo de Estados Unidos describe el número de vehículos por familia en Estados Unidos como sigue:

Número	Porcentaje	Número (millones)	Número	Porcentaje	Número (millones)
0	10.3%	10.9	2	38.4%	40.5
1	34.2%	36.1	3 o más	17.1%	18.0

a. Sustituyendo la categoría "3 o más" con exactamente "3", encuentre la media y desviación estándar del número de vehículos por familia en Estados Unidos.

b. Explique el efecto que tuvo sustituir la categoría "3 o más" con "3" en la media y desviación estándar.

5.35 El número de barcos que llegan a un puerto en cualquier día dado es una variable aleatoria representada por x. La distribución de probabilidad para x es como sigue:

x	10	11	12	13	14
$P(x)$	0.4	0.2	0.2	0.1	0.1

Encuentre la media y desviación estándar del número de barcos que llegan a puerto en un día determinado.

5.36 En un artículo de USA Today (22 de febrero, 2005), se publicaron las siguientes estadísticas sobre el número de perros por familia.

Número de perros	Porcentaje	Número de perros	Porcentaje
Uno	66	Cuatro	3
Dos	24	Cinco o más	2
Tres	5		

Fuente: U.S. Pet Ownership & Demographics Sourcebook, Pedigree Food for Dogs

a. Sustituyendo la categoría "cinco o más" con exactamente "cinco", encuentre la media y desviación estándar del número de perros mascotas por familia.

b. ¿Cómo interpreta la media?

c. Explique el efecto que sustituir la categoría "cinco o más" con "cinco" tuvo en la media y desviación estándar.

5.37 La variable aleatoria A tiene la siguiente distribución de probabilidad:

A	1	2	3	4	5
$P(A)$	0.6	0.1	0.1	0.1	0.1

a. Encuentre la media y desviación estándar de A.

b. ¿Cuánto de la distribución de probabilidad está a no más de 2 desviaciones estándar de la media?

c. ¿Cuál es la probabilidad de que A se encuentre entre $\mu - 2\sigma$ y $\mu + \sigma 2$?

5.38 La variable aleatoria $\bar{x}$ tiene la siguiente distribución de probabilidad:

$\bar{X}$	1	2	3	4	5
$P(\bar{X})$	0.6	0.1	0.1	0.1	0.1

a. Encuentre la media y desviación estándar de $\bar{x}$.

b. ¿Cuál es la probabilidad de que $\bar{x}$ se encuentre entre $\mu + \sigma$?

5.39 Elabore un ejercicio donde se simula que participa en un juego donde un jugador tiene 0.2 de probabilidad de ganar \$3 y una probabilidad de 0.8 de perder \$1. Realice simulaciones para varios conjuntos de 10 jugadas.

a. ¿Qué estimaría usted para su valor esperado (promedio de ganancia o pérdida) a partir de los resultados?

b. Usando la siguiente distribución de probabilidad, calcule la media.

x	$P(x)$
\$3	0.2
−\$1	0.8

c. ¿Cómo se comparan sus respuestas a las partes a y b? ¿Consideraría que éste es un juego limpio? ¿Por qué?

5.40 Todos los martes, en el negocio de video de Jason hay un día de "tirar los dados". Un cliente puede tirar dos dados no "cargados" y rentar una segunda película por una cantidad (en centavos) determinada por los números que salgan en los dados, primero el mayor. Por ejemplo, si el cliente tira un uno y un cinco, la segunda película puede rentarla en \$0.51. Sea x la cantidad pagada por la segunda película al tirar el dado un martes.

a. Use un espacio muestral para el tiro de un par de dados y exprese el costo de renta de la segunda película, x, como distribución de probabilidad.

b. ¿Cuál es el costo de renta medio esperado (media de x) de la segunda película el martes de tirar los dados.

c. ¿Cuál es la desviación estándar de x?

d. Usando una computadora y la distribución de probabilidad hallada en la parte a, genere una muestra aleatoria de 30 valores para x y determine el costo total de rentar la segunda película para 30 rentas.

e. Usando una computadora, obtenga un estimado para la probabilidad de que la cantidad total pagada por 30 segundas películas será más de \$15.00 repitiendo la parte d 500 veces y usando los 500 resultados.

5.5 Distribución de probabilidad binomial

Considere el siguiente experimento de probabilidad. El maestro hace al grupo una prueba sorpresa de cuatro preguntas de opción múltiple. Usted no ha estudiado el material y, por tanto, decide contestar las cuatro preguntas adivinando al azar las respuestas sin leer las preguntas ni las respuestas.

Página de respuesta a la prueba

Instrucciones: circule la mejor respuesta a cada una de las preguntas siguientes.

1. a b c
2. a b c
3. a b c
4. a b c

PARA SU INFORMACIÓN

Eso está bien, ¡adivine!

Circule sus respuestas antes de continuar.

Antes de ver las respuestas correctas a la prueba y averiguar qué tan bien lo hizo, pensemos en algunas de las cosas que podrían suceder si contesta de esta manera a una prueba.

1. De las cuatro preguntas, ¿cuántas de ellas es probable que haya contestado correctamente?
2. ¿Cuál es la probabilidad de que haya contestado correctamente más de la mitad de ellas?

3. ¿Cuál es la probabilidad de que usted haya seleccionado las respuestas correctas a las cuatro preguntas?
4. ¿Cuál es la probabilidad de que usted haya seleccionado las respuestas equivocadas a las cuatro preguntas?
5. Si todo el grupo contesta la prueba por tanteo, ¿qué piensa usted que será el número "promedio" de respuestas correctas del grupo?

Para hallar las respuestas a estas preguntas, empecemos con un diagrama de árbol del espacio austral, mostrando las 16 posibles formas de contestar la prueba de cuatro preguntas. Cada una de las cuatro preguntas se contesta con una respuesta correcta (C) o con una incorrecta (W). Vea la figura 5.4.

FIGURA 5.4
Diagrama de árbol: posibles respuestas a una prueba de cuatro preguntas

PARA SU INFORMACIÓN
WWWW representa incorrectas en 1, incorrecta en 2, incorrecta en 3 e incorrecta en 4; por tanto, su probabilidad se encuentra usando la regla de la multiplicación, fórmula (4.7)

Podemos convertir la información del diagrama de árbol en una distribución de probabilidad. Sea x el "número de respuestas correctas" en la prueba de una persona cuando la prueba se hizo por tanteo aleatorio. La variable aleatoria x puede tomar cualquiera de los valores 0, 1, 2, 3 o 4 para cada cuestionario. La figura 5.4 muestra 16 ramas que representan cinco valores diferentes de x. Nótese que el evento $x = 4$, "cuatro respuestas correctas", está representado por la rama superior del diagrama de árbol, y el evento $x = 0$ "cero respuestas correctas", se muestra en la rama de hasta abajo. Los otros eventos, "una respuesta correcta", "dos respuestas correctas" y "tres respuestas correctas", están representados cada uno de ellos por varias ramas del árbol. Encontramos que el evento $x = 1$ ocurre en cuatro ramas diferentes, el evento $x = 2$ en seis ramas, y el evento $x = 3$ en cuatro ramas.

Cada pregunta individual tiene sólo una respuesta correcta entre las tres posibles respuestas, de modo que la probabilidad de seleccionar la respuesta correcta a una pregunta individual es $\frac{1}{3}$. La probabilidad de que se seleccione una respuesta equivocada a una pregunta individual es $\frac{2}{3}$. La probabilidad de que cada uno de los valores de x puede hallarse al calcular las probabilidades de todas las ramas y luego combinar las probabilidades para las ramas que tienen los mismos valores

TABLA 5.7

Distribución de probabilidad para la prueba de cuatro preguntas

x	$P(x)$
0	0.198
1	0.395
2	0.296
3	0.099
4	0.012
	1.000 ck

de x. Los cálculos siguen, y la distribución de probabilidad resultante aparece en la tabla 5.7.

$P(x = 0)$ es la probabilidad de que las respuestas correctas se den para cero preguntas y las respuestas equivocadas se den para cuatro preguntas (hay sólo una rama en la figura 5.4 donde las cuatro están equivocadas: WWWW):

$$P(x = 0) = \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = \left(\frac{2}{3}\right)^4 = \frac{16}{81} = \mathbf{0.198}$$

Nota: contestar cada pregunta individual es un evento separado e independiente, lo cual nos permite usar la fórmula (4.7) que indica que debemos multiplicar las probabilidades.

$P(x = 1)$ es la probabilidad de que se den respuestas correctas para exactamente una pregunta y se den respuestas equivocadas para las otras tres (hay cuatro ramas en la figura 5.4 donde ocurre, es decir, CWWW, WCWW, WWCW, WWWC, y cada una tiene la misma probabilidad):

$$P(x = 1) = (4) \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} \times \frac{2}{3} = (4)\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^3 = \mathbf{0.395}$$

$P(x = 2)$ es la probabilidad de que se den respuestas correctas a exactamente dos preguntas y se den respuestas equivocadas para las otras dos (hay seis ramas en la Figura 5.4 donde esto ocurre —CCWW, CWCW, CWWC, WCCW, WCWC, WWCC— y cada una tiene la misma probabilidad):

$$P(x = 2) = (6) \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{2}{3} = (6)\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2 = \mathbf{0.296}$$

$P(x = 3)$ es la probabilidad de que se den respuestas correctas para exactamente tres preguntas y se dé una respuesta equivocada para la otra (hay cuatro ramas en la Figura 5.4 donde esto ocurre —CCCW, CCWC, CWCC, WCCC— y cada una tiene la misma probabilidad):

$$P(x = 3) = (4) \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = (4)\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1 = \mathbf{0.099}$$

$P(x = 4)$ es la probabilidad de que se den respuestas correctas para las cuatro preguntas (hay sólo una rama en la figura 5.4 donde las cuatro son correctasCCCC):

$$P(x = 4) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \left(\frac{1}{3}\right)^4 = \frac{1}{81} = \mathbf{0.012}$$

Ahora podemos contestar las cinco preguntas formuladas acerca de la prueba de cuatro preguntas (pp. 284-285).

Respuesta 1: el suceso más probable sería obtener una respuesta correcta; tiene una probabilidad de 0.395. Se espera que resulten cero, una o dos respuestas correctas aproximadamente 89% del tiempo (0.198 + 0.395 + 0.296 = 0.889).

Respuesta 2: tener más de la mitad correcta está representada por $x = 3$ o 4; su probabilidad total es 0.099 + 0.012 = 0.111. (Pasará este cuestionario sólo 11% del tiempo por tanteo aleatorio.)

Respuesta 3: P(las cuatro correctas) = $P(x = 4)$ = 0.012. (Todas correctas suceden sólo 1% del tiempo.)

Respuesta 4: P(todas equivocadas) $= P(x = 0) = 0.198$. (Es casi 20% del tiempo.)

Respuesta 5: Se espera que el promedio del grupo sea ⅓ de 4, o 1.33 de respuestas correctas.

Las respuestas correctas al cuestionario son b, c, b, a. ¿Cuántas respuestas correctas tuvo? ¿Cuál rama del árbol de la figura 5.4 representa sus resultados de la prueba? Podría pedir a varias personas que contesten este mismo cuestionario por tanteo de respuestas. Luego construya una distribución de frecuencia relativa y compárela con la distribución que se muestra en la tabla 5.7.

Numerosos experimentos están compuestos de intentos repetidos cuyos resultados se pueden clasificar en una de dos categorías: **éxito** o **fracaso.** Ejemplos de estos experimentos son el tiro de monedas al aire, respuestas correctas/(equivocadas) de pruebas y otros experimentos más prácticos como por ejemplo determinar si un producto ejecutó o no ejecutó su trabajo prescrito y si un candidato resulta o no resulta electo. Hay experimentos en los que los intentos tienen muchos resultados que, bajo las condiciones adecuadas, pueden caber en esta descripción general de ser clasificados en una de dos categorías. Por ejemplo, cuando tiramos un solo dado, por lo general consideramos seis posibles resultados. No obstante, si estamos interesados sólo en saber si aparece un "uno" o no aparece, hay realmente sólo dos resultados: aparece el "uno" o aparece "algo más". Los experimentos que acabamos de describir se denominan *experimentos de probabilidad binomial*.

Experimento de probabilidad binomial: experimento formado por intentos repetidos que posee las siguientes propiedades:

1. Hay n intentos independientes idénticos repetidos.
2. Cada intento tiene dos posibles resultados (éxito o fracaso).
3. P(éxito) $= p$, P(fracaso) $= q$, y $p + q = 1$.
4. La variable aleatoria binomial x es la cuenta del número de intentos con éxito que sucedieron; x puede tomar cualquier valor entero de cero a n.

Notas:

1. Las propiedades 1 y 2 describen las dos características básicas de cualquier experimento binomial.
2. Por *intentos independientes* queremos decir que el resultado de un intento no afecta la probabilidad de éxito en cualquier otro intento del experimento. En otras palabras, la probabilidad de éxito permanece constante en todo el experimento.
3. La propiedad 3 da la notación algebraica para cada intento.
4. La propiedad 4 se refiere a la notación algebraica para el experimento completo.
5. Es de la mayor importancia que x y p sean asociados con "éxito".

La prueba de cuatro preguntas se clasifica como experimento binomial que consta de cuatro intentos cuando las cuatro respuestas se obtienen por tanteo aleatorio.

Propiedad 1: Un <u>intento</u> es la <u>respuesta de una pregunta</u>, y se repite $\underline{n = 4}$ veces. Los intentos son <u>independientes</u> porque la probabilidad de una respuesta correcta en cualquier pregunta no es afectada por las respuestas a otras preguntas.

Propiedad 2: Los dos posibles resultados en cada intento son éxito = C, respuesta correcta, y fracaso = W, respuesta equivocada.

Propiedad 3: Por cada intento (cada pregunta): $p = P(\text{correcta}) = \frac{1}{3}$ y $q = P(\text{equivocada}) = \frac{2}{3}$. [$p + q = 1$ ck]

Propiedad 4: Para el experimento total (la prueba): x = número de respuestas correctas y puede ser cualquier valor entero de cero a $n = 4$.

EJEMPLO 5.6 Demostración de las propiedades de un experimento de probabilidad binomial

Considere el experimento de lanzar 12 veces un dado y observar un "uno" o "algo más". Al final de los 12 tiros, se indica el número de "unos". La variable aleatoria x es el número de veces que se observa un "uno" en los $n = 12$ intentos. Como "uno" es el resultado de interés, se considera "éxito"; por lo tanto, $p = P(\text{uno}) = \frac{1}{6}$ y $q = P(\text{no uno}) = \frac{5}{6}$. Este experimento es binomial.

EJEMPLO 5.7 Demostración de las propiedades de un experimento de probabilidad binomial

Si el lector fuera inspector en una línea de montaje de una planta donde se manufacturan televisores, su interés sería identificar el número de receptores defectuosos. Es probable que defina "éxito" como el suceso de un televisor defectuoso. Esto no es lo que normalmente consideramos como un éxito, pero, si contamos aparatos "defectuosos" en un experimento binomial, debemos definir "éxito" como "defectuoso". La variable aleatoria x indica el número de aparatos defectuosos hallados por lote de n aparatos; $p = P(\text{televisor es defectuoso})$ y $q = P(\text{televisor está bien})$.

La clave para trabajar con cualquier experimento de probabilidad es su distribución de probabilidad. Todos los experimentos de probabilidad binomiales tienen las mismas propiedades, y por tanto se puede usar el mismo esquema de organización para representarlos todos. La *función de probabilidad binomial* nos permite hallar la probabilidad para cada posible valor de x.

Función de probabilidad binomial: para un experimento binomial, representemos con p la probabilidad de un "éxito" y q representa la probabilidad de un "fracaso" en un solo intento. Entonces $P(x)$, la probabilidad de que habrá exactamente x éxitos en n intentos, es

$$P(x) = \binom{n}{x}(p^x)(q^{n-x}) \qquad \text{para } x = 0, 1, 2, \ldots, n \tag{5.5}$$

Cuando vea la función de probabilidad, observe que es el producto de tres factores básicos:

1. El número de formas en que pueden suceder exactamente x éxitos en n intentos, $\binom{n}{x}$
2. La probabilidad de que haya exactamente x éxitos, p^x
3. La probabilidad de que suceda un fracaso en los restantes $(n\text{-}x)$ intentos, q^{n-x}

El número de formas en que pueden suceder exactamente x éxitos en un conjunto de n está representado por el símbolo $\binom{n}{x}$, que siempre debe ser un entero positivo. Este término se denomina **coeficiente binomial** y se encuentra usando la fórmula

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5.6}$$

Notas:

1. $n!$ (*"n factorial"*) es una abreviatura para el producto de la sucesión de enteros que se inician con n y terminan en uno. Por ejemplo, $3! = 3 \cdot 2 \cdot 1 = 6$ y $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. Hay un caso especial, $0!$, que se define como 1.
2. Los valores para $n!$ y $\binom{n}{x}$ se pueden hallar fácilmente si se usa una calculadora científica.
3. El coeficiente binomial $\binom{n}{x}$ es equivalente al número de combinaciones ${}_nC_x$, el símbolo más probable en su calculadora.

Consideremos el ejemplo 5.5 (pp. 281-282): una moneda se lanza tres veces al aire y observamos el número de caras que suceden en los tres tiros. Éste es un experimento binomial porque presenta todas las propiedades de un experimento binomial:

1. Hay $\underline{n = 3}$ intentos repetidos <u>independientes</u> (cada tiro de la moneda es un intento separado, y el resultado de cualquier intento no tiene efecto en la probabilidad de otro).
2. Cada uno de los intentos (cada tiro de la moneda) resulta en uno de dos posibles resultados: éxito = <u>caras</u> (lo que estamos contando) y fracaso = <u>cruces.</u>
3. La probabilidad de éxito es $p = P(\text{H}) = \underline{0.5}$, y la probabilidad de fracaso es $q = P(\text{T}) = \underline{0.5}$. [$p + q = 0.5 + 0.5 = 1$ ck]
4. La variable aleatoria x es el <u>número de caras</u> que suceden en los tres intentos. x tomará exactamente uno de los valores <u>0,1,2 o 3</u> cuando se complete el experimento.

La función de probabilidad binomial para el tiro de tres monedas es

$$P(x) = \binom{n}{x}(p^x)\ (q^{n-x}) = \binom{3}{x}(0.5)^x(0.5)^{n-x} \qquad \text{para } x = 0,\ 1,\ 2,\ 3$$

Encontremos la probabilidad de $x = 1$ usando la función de probabilidad binomial precedente:

$$P(x = 1) = \binom{3}{1}(0.5)^1(0.5)^2 = 3(0.5)(0.25) = \mathbf{0.375}$$

PARA SU INFORMACIÓN
En la tabla 5.6 (p. 281), $P(1) = \frac{3}{8}$.
Aquí, $P(1) = 0.375$ y $\frac{3}{8} = 0.375$.

Nótese que éste es el mismo valor hallado en el ejemplo 5.5 (p. 281).

EJEMPLO 5.8 Determinación de un experimento binominal y sus probabilidades

Considere un experimento que requiere sacar cinco cartas, una a la vez con restitución, de un "monte" de cartas bien barajado. La carta sacada es identificada como de espadas o no de espadas, es devuelta al monte, éste se vuelve a barajar, y así sucesivamente. La variable aleatoria x es el número de espadas observado en el conjunto de cinco saques de barajas. ¿Es éste un experimento binomial? Identifiquemos las cuatro propiedades.

1. Hay cinco saques repetidos; $n = 5$. Estos intentos individuales son independientes porque la carta sacada es devuelta al monte y éste se vuelve a barajar antes de sacar la siguiente carta.
2. Cada saque es un intento y tiene dos resultados: espadas o no de espadas.
3. $p = P(\text{espadas}) = \dfrac{13}{52}$ y $q = P(\text{no de espadas}) = \dfrac{39}{52}$. [$p + q = 1$ ck]
4. x es el número de espadas registradas al terminar los cinco intentos; los posibles valores son 0, 1, 2, . . . ,5.

La función de probabilidad binomial es

$$P(x) = \binom{5}{x}\left(\frac{13}{52}\right)^x\left(\frac{39}{52}\right)^{5-x} = \binom{5}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{5-x} = \binom{5}{x}(0.25)^x(0.75)^{5-x}$$

para $x = 0, 1, \ldots, 5$

$$P(0) = \binom{5}{0}(0.25)^0(0.75)^5 = (1)(1)(0.2373) = \mathbf{0.2373}$$

$$P(1) = \binom{5}{1}(0.25)^1(0.75)^4 = (5)(0.25)(0.3164) = \mathbf{0.3955}$$

$$P(2) = \binom{5}{2}(0.25)^2(0.75)^3 = (10)(0.0625)(0.421875) = \mathbf{0.2637}$$

$$P(3) = \binom{5}{3}(0.25)^3(0.75)^2 = (10)(0.015625)(0.5625) = \mathbf{0.0879}$$

Las dos probabilidades restantes se dejan al estudiante para que las calcule en el ejercicio 5.54.

PARA SU INFORMACIÓN
Respuesta: cinco

La distribución precedente de probabilidades indica que el valor individual más probable de x es uno, el evento de observar exactamente una espada en una mano de cinco cartas. ¿Cuál es el número menos probable de espadas que se observaría?

EJEMPLO 5.9 Probabilidad binomial de "huevos podridos"

El gerente del Mercado de Alimentos de Steve garantiza que ninguna de sus cajas de una docena de huevos contendrá más de un huevo podrido. Si una caja contiene más de un huevo podrido, le daremos toda la caja y dejaremos que el cliente conserve la caja de huevos original. Si la probabilidad de que un huevo individual sea malo es 0.05, ¿cuál es la probabilidad de que el gerente tenga que cambiar la caja de huevos dada?

SOLUCIÓN A primera vista, la situación del gerente parece adaptarse a las propiedades de un experimento binomial si x es el número de huevos podridos encontrados en una caja de una docena de huevos, $p = P(\text{malos}) = 0.05$, y si la inspección de cada huevo es un intento que resulta en hallar un huevo "podrido" o "no

podrido". Habrá $n = 12$ intentos a considerar para los 12 huevos de una caja. No obstante, los intentos de un experimento binomial deben ser independientes; por tanto, supondremos que la calidad de un huevo de una caja es independiente de la calidad de cualquiera otro de los huevos. (Esto puede ser una suposición muy grande, pero con ella podremos usar la distribución de probabilidad binomial como modelo.) Ahora, con base en esta suposición, podremos hallar/estimar la probabilidad de que el gerente tenga que hacer buena garantía. La función de probabilidad asociada con este experimento será:

$$P(x) = \binom{12}{x}(0.05)^x(0.95)^{12-x} \qquad \text{para } x = 0, 1, 2, \ldots, 12$$

La probabilidad de que el gerente cambie una docena de huevos es la probabilidad de que $x = 2, 3, 4, \ldots, 12$. Recuerde que $\sum P(x) = 1$; es decir,

$$\boldsymbol{P(0) + P(1)} + P(2) + \cdots + P(12) = 1$$
$$P(\text{restitución}) = P(2) + P(3) + \cdots + P(12) = 1 - \boldsymbol{[P(0) + P(1)]}$$

Es más fácil hallar la probabilidad de restitución si hallamos $P(x = 0)$ y $P(x = 1)$ y de 1 restamos su total al hallar todas las otras probabilidades. Tenemos

$$P(x) = \binom{12}{x}(0.05)^x(0.95)^{12-x}$$
$$P(0) = \binom{12}{0}(0.05)^0(0.95)^{12} = \mathbf{0.540}$$
$$P(1) = \binom{12}{1}(0.05)^1(0.95)^{11} = \mathbf{0.341}$$
$$P(\text{restitución}) = 1 - (0.540 + 0.341) = \mathbf{0.119}$$

Si $p = 0.05$ es correcto, entonces el gerente estará ocupado haciendo cambio de cajas de huevos. Si él cambia 11.9% de todas las cajas de huevos que venda, ciertamente estará regalando una parte importante de los huevos que venda. Esto sugiere que debe ajustar su garantía (o vender huevos mejores). Por ejemplo, si tuviera que cambiar una caja de huevos sólo cuando se encontraran cuatro o más huevos podridos, él esperaría cambiar sólo 3 de cada 1000 cajas [1.0 − (0.540 + 0.341 + 0.099 + 0.017)], o sea 0.3% de las cajas vendidas. Nótese que el gerente podrá controlar su "riesgo" (probabilidad de cambio) si ajusta el valor de la variable aleatoria indicada en su garantía.

Nota: el valor de muchas probabilidades binomiales para valores de $n \leq 15$ y valores comunes de p se encuentran en la tabla 2 del apéndice B. En este ejemplo, tenemos $n = 12$ y $p = 0.05$, y buscamos las probabilidades para $x = 0$ y 1. Necesitamos ubicar la sección de la tabla 2 donde $n = 12$, encontrar la columna con encabezado $p = 0.05$, y leer los números en sentido horizontal de $x = 0$ a $x = 1$. Encontramos .540 y .341, como se ve en la tabla 5.8. (Busque estos valores en la tabla 2 del apéndice B.)

TABLA 5.8

Extracto de la tabla 2 del apéndice B, probabilidades binomiales

		p													
n	*x*	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	*x*
	:														
	:														
12	0	.886	.540	.282	.069	.014	.002	0+	0+	0+	0+	0+	0+	0+	0
	1	.107	.341	.377	.206	.071	.017	.003	0+	0+	0+	0+	0+	0+	1
	2	.006	.099	.230	.283	.168	.064	.016	.002	0+	0+	0+	0+	0+	2
	3	0+	.017	.085	.236	.240	.142	.054	.012	.001	0+	0+	0+	0+	3
	4	0+	.002	.021	.133	.231	.213	.121	.042	.008	.001	0+	0+	0+	4
	:														
	:														

Nota: una notación conveniente para identificar la distribución de probabilidad binomial para un experimento binomial con $n = 12$ y $p = 0.05$ es $B(12, 0.05)$. $B(12,0.05)$, léase *"distribución binomial para $n = 12$ y $p = 0.05$"*, representa toda la distribución o "bloque" de probabilidades mostradas en color violeta en la tabla 5.8. Cuando se usa en combinación con la notación $P(x)$, $P(x = 1|B(12, 0.05))$ indica la probabilidad de $x = 1$ de esta distribución, o 0.341 como se ve en la tabla 5.8.

INSTRUCCIONES DE TECNOLOGÍA: PROBABILIDADES BINOMIALES Y BINOMIALES ACUMULATIVAS

MINITAB (Versión 14) Para probabilidades binomiales, introduzca valores de x en C1; luego continúe con:

```
Choose:      Calc > Probability Distributions > Binomial
Select:      Probability *
Enter:       Number of trials: n
             Probability of success: p
Select:      Input column
Enter:       C1
             Optional Storage:  C2  (not necessary) > OK
Or
Select:      Input constant
Enter:       One single x value > OK
```

*Para probabilidades binomiales acumulativas, repita los comandos precedentes pero cambie la selección de probabilidad con:

```
Select:      Cumulative Probability
```

Excel Para probabilidades binomiales, introduzca valores de x en la columna A y active la celda de la columna B desde el primer valor de x; luego continúe con:

```
Choose:      Insert function, fx > Statistical > BINOMDIST > OK
Enter:       Number_s: (A1:A4 or select 'x value' cells)
             Trials: n
             Probability_s: p
             Cumulative: false* (gives individual probabilities) > OK
```

```
Drag:        Bottom right corner of probability value cell in column B down
             to give other probabilities
```

*Para probabilidades binomiales acumulativas, repita los comandos precedentes pero cambie la acumulativa falsa con:

```
Cumulative:  true (gives cumulative probabilities) > OK
```

TI-83/84 Plus

Para obtener una lista completa de probabilidades para una n y p particulares, continúe con:

```
Choose:      2nd > DISTR > 0:binompdf(
Enter:       n, p)
```

Use la tecla de flecha a la derecha para desplazarse por las probabilidades.
Para desplazarse por una lista vertical en L1:

```
Choose:      STO0  > L1 > ENTER
             STAT > EDIT > 1:Edit
```

Para obtener probabilidades individuales para n, p y x particulares, continúe con:

```
Choose:      2nd > DISTR > 0:binompdf(
Enter:       n, p, x)
```

Para obtener probabilidades acumulativas para $x = 0$ a $x = n$ para n y p particulares, continúe con:

```
Choose:      2nd > DISTR > A:binomcdf(
Enter:       n, p)*     (see previous for scrolling through probabilities)
```

*Para obtener probabilidades acumulativas individuales para n, p y x particulares, repita los comandos precedentes pero cambie la entrada con:

```
Enter:       n, p, x)
```

CASO PRÁCTICO 5.10 Viviendo con la ley

¿QUÉ ES UN PROGRAMA DE ACCIÓN AFIRMATIVA (AAP)?

Como condición para hacer negocio con el gobierno federal, los contratistas federales que satisfacen ciertos niveles de contrato y población de empleados convienen en elaborar, de acuerdo con reglamentos federales en 41 CFR 60-1,60-2, etc., un programa de acción afirmativa (AAP). El AAP de un contratista es una combinación de informes numéricos, compromisos de acción y descripción de políticas. Un rápido repaso de un AAP con base en los reglamentos federales (41 CFR 60-2.10), es como sigue:
60-2.10):

Los AAP deben ser desarrollados para

- Minorías étnicas y mujeres (41 CFR 60-1 y 60-2)
- Veteranos discapacitados especiales, veteranos de Vietnam, y otros veteranos amparados (41 VFR 60-250)
- Personas con discapacidades (41 CFR 60-741)

Fuente: http://eeosource.peopleclick.com/maintopic/default.asp?MainTopicID=1

Los reglamentos del AAP no apoyan el uso de un examen específico para determinar si el porcentaje de minorías étnicas o mujeres es menos de lo que se esperaría

razonablemente. No obstante, por lo general se usan varias pruebas, una de las cuales se denomina *prueba binomial exacta* como se define a continuación.

EXAMEN BINOMIAL EXACTO

Las variables empleadas son:

T = Número total de empleados en el grupo de trabajo

M = Número de mujeres o minorías en el grupo de trabajo

A = Porcentaje de disponibilidad de mujeres o minorías para el grupo de trabajo

Este examen comprende el cálculo de una probabilidad, denotado como P, y la comparación de esa probabilidad contra 0.05. Si P es menor o igual a 0.05, el porcentaje de minorías o mujeres como "menos de lo que se esperaría razonable mente". La fórmula para calcular p es como sigue:

1. Calcule la probabilidad, Q, que es la probabilidad binomial acumulativa para la distribución de probabilidad binomial con $n = T$, $x = M$, y $p = A/100$.
2. Si Q es menor o igual a 0.5, entonces $P = 2Q$; de otro modo, $P = Q$.

Por ejemplo, si T = 50 empleados y M = 2 mujeres, A = 6% de disponibilidad de mujeres.

Usando una computadora, se encuentra el valor Q: $Q = 0.41625$. Como Q es menor a 0.5, $P = 2Q = 0.8325$. P, 0.8325, es mayor a 0.05, de modo que se encuentra que el porcentaje de mujeres es "**no** menos de lo que se esperaría razonablemente".

SECCIÓN 5.5 EJERCICIOS

5.41 Considere la prueba de cuatro preguntas de opción múltiple presentada al principio de esta sección (pp. 284-287).

a. Explique por qué las cuatro preguntas representan cuatro intentos independientes.

b. Explique por qué el número 4 se multiplica en la $P(x = 1)$.

c. En la respuesta 5 de la página 287, ¿de dónde provienen $\frac{1}{3}$ y 4? ¿Por qué multiplicarlos para hallar el promedio esperado?

5.42 Identifique las propiedades por las que lanzar al aire una moneda 50 veces y dar seguimiento a "caras" es un experimento binomial.

5.43 Indique una razón práctica por la que un artículo defectuoso en una situación industrial podría ser definido como "éxito" en un experimento binomial.

5.44 ¿Qué significa que los intentos sean independientes en un experimento binomial?

5.45 Evalúe cada uno de lo siguiente.

a. $4!$ b. $7!$ c. $0!$ d. $\dfrac{6!}{2!}$

e. $\dfrac{5!}{2!3!}$ f. $\dfrac{6!}{4!(6-4)!}$ g. $(0.3)^4$ h. $\binom{7}{3}$

i. $\binom{5}{2}$ j. $\binom{3}{0}$ k. $\binom{4}{1}(0.2)^1(0.8)^3$

l. $\binom{5}{0}(0.3)^0(0.7)^5$

5.46 Demuestre que cada uno de lo siguiente es verdadero para cualesquier valores de n y k. Use dos conjuntos específicos de valores para n y k para demostrar que cada uno es verdadero.

a. $\binom{n}{0} = 1$ y $\binom{n}{n} = 1$

b. $\binom{n}{1} = n$ y $\binom{n}{n-1} = n$ c. $\binom{n}{k} = \binom{n}{n-k}$

5.47 Se somete a inspección una caja que contiene 100 camisetas. Cada camiseta se clasifica como "de primera calidad" o "irregular". Después de haber inspeccionado las 100 camisetas, se informa el número de irregulares como una variable aleatoria. Explique por qué x es una variable aleatoria binomial.

5.48 Un dado se tira 20 veces, y el número de "cincos" que sucede se reportan como la variable aleatoria. Explique por qué x es una variable aleatoria binomial.

5.49 Se seleccionan cuatro cartas, una a la vez, de un "monte" estándar de 52 barajas de juego. Represente por x el número de ases sacado del conjunto de cuatro cartas.

a. Si este experimento se completa sin restitución, explique por qué x no es una variable aleatoria binomial.

b. Si este experimento se completa con restitución, explique por qué x es una variable aleatoria binomial.

5.50 Los empleados de una planta de ensamble de General Motors son encuestados a la salida de su trabajo. A cada uno se le pregunta, "¿En un auto de qué marca se dirige usted a casa?" La variable aleatoria a ser reportada es el número de cada marca mencionada. ¿Es x una variable aleatoria binomial? Justifique su respuesta.

5.51 Considere un experimento binomial formado por tres intentos con resultados de éxito, S, y fracaso, F, donde $P(S) = p$ y $P(F) = q$.

a. Complete el siguiente diagrama de árbol. Aplique leyenda a todas las ramas sin faltar ninguna.

b. En la columna (b) del diagrama de árbol, exprese la probabilidad de cada resultado representado por las ramas como producto de potencias de p y q.

c. Sea x la variable aleatoria, el número de éxitos observado. En la columna (c), identifique el valor de x para cada rama del diagrama de árbol.

d. Observe que todos los productos de la columna (b) están formados por tres factores y que el valor de la variable aleatoria es la misma que el exponente del número p.

e. Escriba la ecuación para la función de probabilidad binomial para esta situación.

5.52 Trace un diagrama de árbol que describa un experimento binomial de cuatro intentos.

5.53 Use la función de probabilidad para tres tiros de moneda como se demuestra en la página 289 y verifique las probabilidades para $x = 0$, 2 y 3.

5.54 a. Calcule $P(4)$ y $P(5)$ para el ejemplo 5.8 de la página 290.

b. Verifique que las seis posibilidades $P(0)$, $P(1)$, $P(2)$, . . . ,$P(5)$ formen una distribución de probabilidad.

5.55 Realice un ejercicio donde demuestre el cálculo de una probabilidad binomial junto con una interpretación visual. Suponga que usted compra 20 plantas de un invernadero y que este negocio dice que 95% de sus plantas sobreviven cuando son plantadas. Utilice $n = 20$ y $p = 0.95$, calcule lo siguiente:

a. La probabilidad de que sobrevivan las 20 plantas

b. La probabilidad de que a lo sumo sobrevivan 16 plantas

c. La probabilidad de que al menos sobrevivan 18 plantas

5.56 Realice un ejercicio donde demuestre el cálculo de una probabilidad binomial junto con una interpretación visual. Suponga que usted está en un grupo

de 30 estudiantes y se supone que aproximadamente 11% de la población es de zurdos. Introduciendo $n = 30$ y $p = 11$, calcule lo siguiente:

a. La probabilidad de que exactamente cinco estudiantes sean zurdos.

b. La probabilidad de que a lo sumo cuatro estudiantes sean zurdos.

c. La probabilidad de que al menos seis estudiantes sean zurdos.

5.57 Si x es una variable aleatoria binomial, calcule la probabilidad de x para cada caso.

a. $n = 4, x = 1, p = 0.3$ b. $n = 3, x = 2, p = 0.8$

c. $n = 2, x = 0, p = \frac{1}{4}$ d. $n = 5, x = 2, p = \frac{1}{3}$

e. $n = 4, x = 2, p = 0.5$ f. $n = 3, x = 3, p = \frac{1}{6}$

5.58 Si x es una variable aleatoria binomial, use la tabla 2 del apéndice B para determinar la probabilidad de x para cada uno de lo siguiente:

a. $n = 10, x = 8, p = 0.3$ b. $n = 8, x = 7, p = 0.95$

c. $n = 15, x = 3, p = 0.05$ d. $n = 12, x = 12, p = 0.99$

e. $n = 9, x = 0, p = 0.5$ f. $n = 6, x = 1, p = 0.01$

g. Explique el significado del símbolo 0+ que aparece en la tabla 2.

5.59 Pruebe la siguiente función para determinar si es o no es una función de probabilidad binomial. Haga una lista de la distribución de probabilidades y trace

$$T(x) = \binom{5}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{5-x} \qquad \text{para} \quad x = 0, 1, 2, 3, 4, 5$$

5.60 Sea x una variable aleatoria con la siguiente distribución de probabilidad:

x	0	1	2	3
$P(x)$	0.4	0.3	0.2	0.1

¿Tiene x una distribución binomial? Justifique su respuesta.

5.61 De los tres árboles plantados por una empresa de paisajes, 90% sobreviven. ¿Cuál es la probabilidad de que sobrevivan 8 o más de los 10 árboles que acaban de plantar? (Encuentre la respuesta usando una tabla.)

5.62 Según la Federal Trade Commission, más de la mitad de las 358,603 quejas de consumidores por fraude en 2004 estuvieron relacionadas con el Internet —53% para ser más exactos—. En un grupo de 20 personas que han presentado una queja por fraude, ¿cuál es la probabilidad de que exactamente la mitad están relacionados con el Internet?

5.63 Es muy probable que los alumnos de primer grado que tienen mala conducta en la escuela sean fumadores como adultos jóvenes, según un nuevo estudio presentado en la edición de julio de 2004 de la *American Journal of Epidemiology.* Después de hacer seguimiento de un grupo de alumnos de primer año durante 15 años, se encontró que entre esos chicos que habían fumado y tenido mala conducta, 66% eran fumadores consuetudinarios.

Fuente: http://preventdisease.com/news/articles/troubled_kids_more_likely_smokers.shtml

a. ¿Cuál es la probabilidad de que exactamente dos de los siguientes tres jóvenes adultos seleccionados al azar, que tuvieron mala conducta en sus primeros años y han fumado, sean fumadores consuetudinarios?

b. ¿Cuál es la probabilidad de que exactamente ocho de los siguientes 12 jóvenes adultos seleccionados al azar, que tuvieron mala conducta en sus primeros años y han fumado, sean fumadores consuetudinarios?

c. ¿Cuál es la probabilidad de que exactamente 20 de los siguientes 30 jóvenes adultos seleccionados al azar, que tuvieron mala conducta en sus primeros años y han fumado, sean fumadores consuetudinarios?

5.64 El proyecto de vida Pew Internet & American elabora reportes que exploran el impacto de la Internet en innumerables facetas de nuestra vida diaria, ya sea en casa, escuela u oficina. En su encuesta de seguimiento de enero de 2005, encontró que más de 80% de personas de 18 a 29 años de edad usan la Internet. Considere un grupo de personas de 18 a 29 años de edad, seleccionado al azar.

Fuente: http://www.pewinternet.org/trends/User_Demo_03.07.05.htm

a. ¿Cuál es la probabilidad de que exactamente ocho usen la Internet?

b. ¿Cuál es la probabilidad de que al menos cinco usen la Internet?

5.65 En el evento de biatlón de los Juegos Olímpicos, un participante realiza una prueba de esquí a campo traviesa y en cuatro ocasiones intermitentes se detiene en un campo de tiro con rifle y dispara una serie de cinco tiros. Si acierta en el centro del blanco, no se le aplican puntos de castigo. Si un hombre en particular tiene un registro de acertar al centro del blanco en 90% de sus tiros, ¿cuál es la probabilidad de lo siguiente?:

a. Acertar en el centro del blanco con los cinco de su siguiente serie de cinco tiros.

b. Acertar en el centro del blanco con al menos cuatro de su siguiente serie de cinco tiros. (Suponga independencia.)

5.66 El porcentaje de supervivencia durante una operación riesgosa para pacientes sin ninguna otra esperanza de supervivencia es 80%. ¿Cuál es la probabilidad de que exactamente cuatro de los siguientes cinco pacientes salga con vida de esta operación?

5.67 De las piezas producidas por una máquina en particular, 0.5% son defectuosas. Si una muestra aleatoria de 10 piezas producidas por esta máquina contiene 2 o más piezas defectuosas, la máquina es aparada para hacerle reparaciones. Encuentre la probabilidad de que la máquina será apagada para repararla con base en este plan de muestreo.

5.68 Una encuesta hecha a motociclistas y realizada en enero de 2005, encargada por el Grupo de Progreso de Compañías de Seguros, mostró que 40% de los motociclistas tienen figuras artísticas en su cuerpo, por ejemplo tatuajes y perforaciones. Un grupo de 10 moto ciclistas están en el proceso de comprar seguro de motocicletas.

Fuente: http://www.syracuse.com/business/poststandard/index.ssf?/base/business-1/

a. ¿Cuál es la probabilidad de que ninguno de los 10 tenga ninguna figura artística en su cuerpo?

b. ¿Cuál es la probabilidad de que exactamente 3 tengan alguna figura artística en su cuerpo?

c. ¿Cuál es la probabilidad de que exactamente 4 tengan alguna figura artística en su cuerpo?

d. ¿Cuál es la probabilidad de que no más de 2 tengan alguna figura artística en su cuerpo?

5.69 Si hay igual probabilidad de que nazcan niños y niñas, ¿cuál es la probabilidad de que en una familia de seis hijos, seleccionada al azar, haya al menos un niño? (Encuentre la respuesta usando una fórmula.)

5.70 La cuarta parte de cierta variedad de conejos hacen con pelo largo. ¿Cuál es la probabilidad de que en una camada de seis conejos exactamente tres tengan pelo largo? (Encuentre la respuesta usando una fórmula.)

5.71 El jugador de béisbol Cardenales de San Luis, Albert Pujols, tiene un promedio de bateo en 3 años (razón entre *hits* y veces al bat) de 0.334 para las temporadas 2002-2004. Supongamos que Pujols tiene cinco veces oficiales al bat durante su siguiente juego. Suponiendo que no haya circunstancias extenuantes y que el modelo binomial producirá aproximaciones razonables, ¿cuál es la probabilidad de lo siguiente?:

a. Pujols conecta menos de dos hits.

b. Pujols conecta más de tres hits.

c. Pujols conecta cinco *hits* en cinco veces al bat.

5.72 Como inspector de control de calidad para camiones de juguete, usted ha observado que 3% del tiempo a las ruedas de madera se les hace un agujero fuera del centro. Si se usan seis ruedas de madera en cada uno de los camiones de juguete producidos, ¿cuál es la probabilidad de que un camión de juguete seleccionado al azar no tenga ruedas con agujero fuera del centro?

5.73 Considere al gerente del mercado de alimentos Steve's, como se ilustró en el ejemplo 5.9. ¿Cuál sería el "riesgo" del gerente si compró huevos "mejores", digamos con $P(\text{malos}) = 0.01$ usando la garantía de "más de uno"?

5.74 Según el artículo de USA Snapshot "Conociendo a drogadictos", 45% de estadounidenses conocen a alguien que se hizo adicto a una droga diferente del alcohol. Suponiendo que esto es verdadero, ¿cuál es la probabilidad de lo siguiente?:

a. Exactamente tres personas de una muestra aleatoria de 5 conocen a alguien que se hizo adicto. Calcule el valor.

b. Exactamente siete personas de una muestra aleatoria de 15 conocen a alguien que se hizo adicto. Haga su estimación con el uso de la tabla 2 del apéndice B.

c. Al menos siete personas de una muestra aleatoria de 15 conocen a alguien que se hizo adicto. Haga su estimación usando la tabla 2.

d. No más de siete personas de una muestra aleatoria de 15 conocen a alguien que se hizo adicto. Haga su estimación usando la tabla 2.

5.75 De todas las ejecuciones de hipoteca en Estados Unidos, 48% son causadas por discapacidad. Las personas lesionadas o enfermas no pueden trabajar, con lo cual pierden su trabajo y sus ingresos. Sin ingresos, no pueden hacer pagos de sus hipotecas y el banco realiza una ejecución.

Fuente: http://www.ricedelman.com06.11.02

Dado que 20 ejecuciones de hipoteca son auditadas por una gran institución de préstamos, encuentre la probabilidad de lo siguiente:

a. Cinco o menos de las ejecuciones se deben a una discapacidad.

b. Al menos tres ejecuciones se deben a una discapacidad.

5.76 a. Use calculadora o computadora para hallar la probabilidad de que $x = 3$ en un experimento binomial, donde $n = 12$ y $p = 0.30$: $P(x = 3|B(12, 0.30))$. (Vea Nota acerca de esta notación en la p. 292.)

b. Use la tabla 8 para verificar la respuesta de la parte a.

5.77 Use computadora para hallar las probabilidades para todos los posibles valores x para un experimento binomial donde $n = 30$ y $p = 0.35$.

MINITAB (Versión 14)

```
Choose:  Calc > Make Patterned Data > Simple Set of
         Numbers
Enter:   Store patterned data in: C1
         From ?rst value: 0
         To last value: 30
         In steps of: 1 > OK
```

Continúe con los comandos MINITAB de probabilidad binomial de la página 292, usando $n = 30$, $p = 0.35$ y C2 para memoria adicional.

Excel

```
Enter:   0,1,2, . . . , 30 into column A
```

Continúe con los comandos Excel de probabilidad binomial de las páginas 292-293, usando $n = 30$ y $p = 0.35$.

TI-83/84 Plus

Use the binomial probability TI-83/84 commands on pages 293, using $n = 30$ and $p = 0.35$.

5.78 Use computadora para hallar las probabilidades acumulativas para todos los posibles valores x para un experimento binomial donde $n = 45$ y $p = 0.125$.

a. Explique por qué hay tantos miles en la lista.

b. Explique qué está representado por cada número de la lista.

MINITAB (Versión 14)

```
Choose:  Calc > Make Patterned Data > Simple Set of
         Numbers . . .
Enter:   Store patterned data in: C1
         From ?rst value: 0
         To last value: 45
         In steps of: 1 > OK
```

Continúe con los comandos MINITAB de probabilidad binomial <u>acumulativa</u> de la página 292, usando $n = 45$, $p = 0.125$ y C2 como memoria opcional.

Excel

```
Enter:   0,1,2, . . . , 45 into column A
```

Continúe con los comandos Excel de probabilidad binomial <u>acumulativa</u> de las páginas 292-293, usando $n = 45$ y $p = 0.125$.

TI-83/84 Plus

Use los comandos TI-83/84 de probabilidad binomial <u>acumulativa</u> de la página 293, usando $n = 45$ y $p = 0.125$.

5.79 El aumento en el uso de Internet de los últimos años ha sido fenomenal, como lo demuestra el informe de febrero de 2004 de Pew Internet & American Life Project. La encuesta a estadounidenses de 65 años de edad o más (unos 8 millones de adultos) indicó que 22% tienen acceso a Internet. En contraste, 58% de personas entre 50 y 64 años de edad, 75% de entre 30 y 49, y 77% de entre 18 y 29 años actualmente están en línea.

Fuente: http://www.suddenlysenior.com/maturemarketstatsmore.html

Suponga que 50 adultos de cada grupo han de ser entrevistados.

a. ¿Cuál es la probabilidad de que "tengo acceso a Internet" sea la respuesta de 10 a 20 adultos del grupo de 65 años de edad o más?

b. ¿Cuál es la probabilidad de que "tengo acceso a Internet" sea la respuesta de 30 a 40 adultos del grupo de 50 a 64 años de edad?

c. ¿Cuál es la probabilidad de que "tengo acceso a Internet" sea la respuesta de 30 a 40 adultos del grupo de 30 a 49 años de edad?

d. ¿Cuál es la probabilidad de que "tengo acceso a Internet" sea la respuesta de 30 a 40 adultos del grupo de 18 a 29 años de edad?

e. ¿Por qué las respuestas para las partes a y d son casi iguales? Explique.

f. ¿Qué efecto tuvieron los diversos valores de p en las probabilidades? Explique.

5.80 ¿A dónde se van todos esos dulces de Halloween? La edición de octubre de 2004 de *Reader's Digest* citó que "90% de padres reconocen tomar dulces de las bolsas en que sus hijos recibieron dulces de Halloween".

La fuente de información fue el National Confectioners Association. Suponga que se realizan entrevistas a 25 padres, ¿cuál es la probabilidad de que 20 o más padres de familia tomen los dulces de Haloweeen de las bolsas de dulces de sus hijos?

5.81 Harris Interactive llevó a cabo una encuesta para el Tylenol PM, donde preguntó a automovilistas de Estados Unidos qué hacen si conducen un vehículo cuando están cansados. Los resultados se publicaron en la USA Snapshot del 18 de enero, 2005, con 40% de quienes respondieron diciendo que "abren las ventanas" para combatir el sueño. Supongamos que se entrevista a 35 conductores. ¿Cuál es la probabilidad de que entre 10 y 20 de ellos digan que "abren las ventanas" para combatir el sueño?

5.82 a. Cuando se use la prueba binomial exacta (Caso práctico 5.10, pp. 293-294), ¿cuál es la interpretación de la situación cuando el valor calculado de P sea menor o igual a 0.05?

b. Cuando se use la prueba binomial exacta, ¿cuál es la interpretación de la situación cuando el valor calculado de P es mayor a 0.05?

c. Una empresa tiene 15 empleados en un grupo de trabajo muy especializado, de los cuales 2 son de minorías. Con base en la información del censo de 2000, la proporción de minorías disponibles para este tipo de trabajo es 5%. Con el uso de la prueba binomial, ¿sería razonable esperar ese porcentaje de minorías?

d. Para esta misma empresa y el mismo grupo de trabajo, hay tres empleadas. El porcentaje de disponibilidad de mujeres para este trabajo es 50%. ¿Le parece que es razonable esperar este porcentaje de mujeres?

5.83 Prolongado a tiempos extras en el juego 7 de gira de las eliminatorias de la NBA de 2002, el dos veces campeón defensor Los Ángeles Lakers hicieron lo que mejor hacen: se crecen cuando la presión es máxima. Los dos jugadores estrellas de los Lakers tuvieron su oportunidad en la línea de tiros de castigo al final del tiempo extra.

a. Restando 1:27 minutos del tiempo extra y el juego empatado a 106-106, Shaquille (Shaq) O'Neal estaba en la línea para ejecutar dos tiros libres. Él tiene un historial de anotar en 0.555 de sus tiros libres y durante este juego, antes de estos dos tiros, había anotado en 9 de sus 13 oportunidades. Justifique la frase de "La ley de los promedios está trabajando en contra de él".

b. Con 0:06 segundos restantes del tiempo extra y el marcador en 110-106, Kobe Bryant estaba en la línea para realizar dos tiros libres. Él tiene un historial de anotar en 0.829 de sus tiros libres y durante este juego, antes de estos dos tiros, había anotado en 6 de sus 8 oportunidades. Justifique la frase de "La ley de promedios está trabajando contra él".

Ambos jugadores anotaron en sus dos tiros, terminando así la serie con los Reyes de Sacramento.

5.84 Si el binomio $(q + p)$ se eleva al cuadrado, el resultado es $(q + p)^2 = q^2 + 2qp + p^2$. Para el experimento binomial con $n = 2$, la probabilidad de que no haya éxito en dos intentos es q^2 (el primer término de la expansión), la probabilidad de un éxito en dos intentos es $2qp$ (el segundo término de la expansión), y la probabilidad de dos éxitos en dos intentos es p^2 (el tercer término de la expansión). Encuentre $(q + P)^3$ y compare sus términos contra las probabilidades binomiales para $n = 3$ intentos.

5.6 Media y desviación estándar de la distribución binomial

¿SABÍA USTED...?

Huellas dactilares

A Sir Francis Galton se le acredita el "descubrimiento" de las huellas dactilares (que las huellas dactilares son únicas en cada persona), y fue él quien perfeccionó los medios para identificarlas. Es la presencia de marcas irregulares y cortaduras en las figuras del dedo que hacen que cada huella sea única. Estas marcas se conocen como Marcas de Galton. El sistema Galton-Henry de clasificación de huellas dactilares fue publicado en junio de 1900, se empezó a usar en la Scotland Yard en 1901, y pronto se usó en todo el mundo como identificador en investigaciones criminales.

La media y desviación estándar de una distribución teórica de probabilidad binomial se puede hallar con estas dos fórmulas:

Media de distribución binomial

$$\mu = np \tag{5.7}$$

y

Desviación estándar de distribución binomial

$$\sigma = \sqrt{npq} \tag{5.8}$$

La fórmula para la media, μ, parece apropiada: el número de intentos multiplicado por la probabilidad de "éxito". [Recuerde que el número medio de respuestas correctas en el cuestionario binomial (Respuesta 5, p. 287) se esperaba ser de $\frac{1}{3}$ de 4, $4(\frac{1}{3})$, o np.] La fórmula para la desviación estándar, , no se entiende fácilmente. Así, en este punto es apropiado ver un ejemplo, que demuestra que las fórmulas (5.7) y (5.8) dan los mismos resultados que las fórmulas (5.1), (5.3a), y (5.4).

En el ejemplo 5.5 (pp. 281-282), x es el número de caras en tres tiros de moneda, $n = 3$, y $p = \frac{1}{2} = 0.5$. Usando la fórmula (5.7), encontramos que la media de x es

$$\mu = np = (3)(0.5) = \mathbf{1.5}$$

Usando la fórmula (5.8), encontramos que la desviación estándar de x es

$$\sigma = \sqrt{npq} = \sqrt{(3)(0.5)(0.5)} = \sqrt{0.75} = 0.866 = \mathbf{0.87}$$

Ahora volvamos a la solución del ejemplo 5.5 (p. 282). Nótese que los resultados son los mismos, cualquiera que sea la fórmula que se use. No obstante, las fórmulas (5.7) y (5.8) son mucho más fáciles de usar cuando x es una variable aleatoria binomial.

EJEMPLO 5.11

Cálculo de la media y desviación estándar de una distribución binomial

Encuentre la media y desviación estándar de la distribución binomial cuando $n = 20$ y $p = \frac{1}{5}$ (o 0.2, en forma decimal). Recuerde que la "distribución binomial donde $n = 20$ y $p = 0.2$" tiene la función de probabilidad

$$P(x) = \binom{20}{x}(0.2)^x(0.8)^{20-x} \qquad \text{para } x = 0, 1, 2, \ldots, 20$$

y una distribución correspondiente con 21 valores x y 21 probabilidades, como se muestra en la tabla de distribución, tabla 5.9, y en el histograma de la figura 5.5.

TABLA 5.9

Distribución binomial
$n = 20$, $p = 0.2$

x	$P(x)$
0	0.012
1	0.058
2	0.137
3	0.205
4	0.218
5	0.175
6	0.109
7	0.055
8	0.022
9	0.007
10	0.002
11	0+
12	0+
13	0+
.	.
:	:
20	0+

FIGURA 5.5 Histograma de distribución binomial $B(20, 0.2)$

Encontremos la media y desviación estándar de esta distribución de x usando las fórmulas (5.7) y (5.8):

$$\mu = np = (20)(0.2) = \mathbf{4.0}$$
$$\sigma = \sqrt{npq} = \sqrt{(20)(0.2)(0.8)} = \sqrt{3.2} = \mathbf{1.79}$$

FIGURA 5.6 Histograma de distribución binomial $B(20, 0.2)$

La figura 5.6 muestra la media, $\mu = 4$, (indicada por la ubicación de la recta vertical azul por el eje x) relativa a la variable x. Este 4.0 es el valor medio esperado para x, el número de éxitos en cada muestra aleatoria de tamaño 20 extraída de una población con $p = 0.2$. La figura 5.6 también indica el tamaño de la desviación estándar, $\sigma = 1.79$ (como lo indica la longitud del segmento de recta roja horizontal). Es la desviación estándar esperada para los valores de la variable aleatoria x que se presenta en muestras de tamaño 20 extraídas de esta misma población.

SECCIÓN 5.6 EJERCICIOS

5.85 Encuentre la media y desviación estándar para la variable aleatoria binomial x con $n = 30$ y $p = 0.6$, usando las fórmulas (5.7) y (5.8).

5.86 Considere la distribución binomial donde $n = 11$ y $p = 0.05$.

a. Encuentre la media y desviación estándar usando las fórmulas (5.7) y (5.8).

b. Usando la tabla 2 del apéndice B, haga una lista de la distribución de probabilidad y trace un histograma.

c. Localice μ y σ en el histograma.

5.87 Considere la distribución binomial donde $n = 11$ y $p = 0.05$ (vea el ejercicio 5.86).

a. Use la distribución [ejercicio 5.86(b) o la tabla 2] y encuentre la media y desviación estándar usando las fórmulas (5.1), (5.3a) y (5.4).

b. Compare los resultados de la parte a con las respuestas halladas en el ejercicio 5.86(a).

5.88 Dada la función de probabilidad binomial

$$P(x) = \binom{5}{x} \cdot \left(\frac{1}{2}\right)^x \cdot \left(\frac{1}{2}\right)^{5-x} \quad \text{para } x = 0, 1, 2, 3, 4, 5$$

a. Calcule la media y desviación estándar de la variable aleatoria usando las fórmulas (5.1), (5.3a) y (5.4).

b. Calcule la media y desviación estándar usando las fórmulas (5.7) y (5.8).

c. Compare los resultados de las partes a y b.

5.89 Encuentre la media y desviación estándar de x para cada una de las siguientes variables aleatorias binomiales:

a. El número de cruces visto en 50 tiros al aire de una moneda

b. El número de estudiantes zurdos en un grupo de 40 estudiantes (suponga que 11% de la población es de zurdos)

c. El número de autos que se encuentran con llantas inseguras entre 400 autos detenidos en un retén de carretera para inspecciones (suponga que 6% de todos los autos tienen una o más llantas inseguras)

d. El número de semillas de sandía que germinan cuando se planta un paquete de 50 semillas (el paquete indica que la probabilidad de germinación es 0.88)

5.90 Encuentre la media y desviación estándar para cada una de las siguientes variables aleatorias binomiales en las partes a-c:

a. El número de seises visto en 50 tiros de un dado

b. El número de televisores defectuoso en un embarque de 125 (El fabricante dice que 98% de los aparatos están en buenas condiciones.)

c. El número de televisores en buenas condiciones en un embarque de 125 (el fabricante dice que 98% de los aparatos están en buenas condiciones)

d. ¿Cómo están relacionadas las partes b y c? Explique.

5.91 Según United Mileage Plus Visa (22 de noviembre, 2004), 41% de los pasajeros dicen que se ponen "audífonos" para evitar ser molestados por sus vecinos de asiento durante un vuelo. Para mostrar la importancia, o no importancia, de audífonos para personas, considere que la variable x es el número de personas en una muestra de 12 que dicen que "se ponen audífonos" para evitar a sus vecinos de asiento. Suponga que 41% es verdadero para toda la población de viajeros de aerolíneas y que se selecciona una muestra al azar.

a. ¿Es x una variable aleatoria binomial? Justifique su respuesta.

b. Encuentre la probabilidad de que $x = 4$ o 5.

c. Encuentre la media y desviación estándar de x.

d. Trace un histograma de la distribución de x: aplique leyendas en su totalidad, resalte el área que represente $x = 4$ y $x = 5$, trace una línea vertical en el valor de la media, y marque la ubicación de x que sea una desviación estándar mayor que la media.

5.92 Según la American Payroll Association (10 de septiembre, 2004), cuando se hace la pregunta "¿Apoya usted el uso de tecnología biomédica (huellas dactilares o reconocimiento de mano) para registrar el tiempo y asistencia de empleados?" la mayoría de trabajadores dicen que no les importa. Sin embargo, 28% pusieron objeciones, pensando que es una invasión a su privacidad personal. Para entender mejor el alcance de esta objeción, considere que la variable x es el número de personas de la muestra de 25 que dicen que "se oponen". Suponga que el 28% es verdadero para toda la población de trabajadores y que se selecciona una muestra aleatoria.

a. ¿Es x una variable aleatoria binomial? Justifique su respuesta.

b. Encuentre la probabilidad de que x tiene un valor de 4 a 10.

c. Encuentre la media y desviación estándar de x.

d. Trace un histograma de la distribución de x: aplique leyendas en su totalidad, resalte el área que represente $x = 4$ a $x = 10$, trace una línea vertical en el

valor de la media, y marque la ubicación de x que sea 1 desviación estándar mayor que la media.

5.93 Una variable aleatoria binomial tiene una media igual a 200 y una desviación estándar de 10. Encuentre los valores de n y p.

5.94 Se sabe que la probabilidad de éxito en un solo intento de un experimento binomial es $\frac{1}{4}$. La variable aleatoria x, número de éxitos, tiene un valor medio de 80. Encuentre el número de intentos involucrados en este experimento y la desviación estándar de x.

5.95 Una variable aleatoria x está basada en 15 intentos con la probabilidad de éxito igual a 0.4. Encuentre la probabilidad de que esta variable tome un valor que sea más de 2 desviaciones estándar de la media.

5.96 Una variable aleatoria x está basada en 15 intentos con la probabilidad de éxito igual a 0.2. Encuentre la probabilidad de que esta variable tome un valor que sea más de 2 desviaciones estándar de la media.

5.97 Imprints Galore compra camisetas (para imprimirles una frase elegida por el comprador) de un fabricante que garantiza que las camisetas han sido inspeccionadas y que no más de 1% tienen alguna imperfección. Las camisetas llegan en cajas de 12. Sea x el número de camisetas con imperfecciones halladas en una caja cualquiera.

a. Haga una lista de distribuciones de probabilidad y trace el histograma de x.

b. ¿Cuál es la probabilidad de que una caja cualquiera no tenga camisetas imperfectas?

c. ¿Cuál es la probabilidad de que una caja cualquiera no tenga más de una camiseta imperfecta?

d. Encuentre la media y desviación estándar de x.

e. ¿Qué proporción de la distribución está entre $\mu - \sigma$ y $\mu + \sigma$?

f. ¿Qué proporción de la distribución está entre $\mu - 2\sigma$ y $\mu + 2\sigma$?

g. ¿Cómo se relaciona esta información con la regla empírica y el teorema de Chebyshev? Explique.

h. Use computadora para simular la compra que Imprints Galore's hace de 200 cajas de camisetas y observe x, el número de camisetas imperfectas por caja de 12. Describa cómo se compara la información de la simulación contra lo que se esperaba (las respuestas a las partes a-g describen los resultados esperados).

i. Repita la parte h varias veces. Describa cómo se comparan estos resultados con los de las partes a-g con la parte h.

MINITAB (Versión 14)

a.

```
Choose:  Calc > Make Patterned Data > Simple Set of
         Numbers . . .
Enter:   Store patterned data in: C1
         From ?rst value: −1 (see note)
         To last value: 12
         In steps of: 1   > OK
```

c. Continúe con los comandos MINITAB de probabilidad binomial de la página 292, usando $n = 12$, $p = 0.01$ y C2 para memoria opcional.

```
Choose:  Graph > Scatterplot > Simple > OK
Enter:   Y variables: C2 X variables: C1
Select:  Data view: Data Display: Area > OK
The graph is not a histogram, but can be converted to a
histogram by double clicking on 'area' of graph.
Select:  Options  Select: Step > OK > OK
```

g. Continúe con los comandos MINITAB de probabilidad binomial <u>acumulativa</u> de la página 292, usando $n = 12$, $p = 0.01$ y C3 para memoria opcional.

```
Choose:  Calc > Random Data > Binomial
Enter:   Generate: 200 rows of data
         Store in column C4
         Number of trials: 12
         Probability: .01   > OK
Choose:  Stat > Tables > Cross Tabulation
Enter:   Categorical variables: For rows: C4
Select:  Display: Total percents   >OK
Choose:  Calc > Column Statistics
Select:  Statistic: Mean
Enter:   Input variable: C4   > OK
Choose:  Calc > Column Statistics
Select:  Statistic: Standard deviation
Enter:   Input variable: C4   > OK
```

Continúe con los comandos MINITAB del histograma de la página 61, usando los datos de C4 y seleccionando las opciones: porcentaje y punto medio con intervalos 0:12/1.

Nota: la variable binomial x no puede tomar el valor −1. El uso de −1 (el supuesto punto medio de clase siguiente a la izquierda de 0) permite a MINITAB trazar el histograma de una distribución de probabilidad. Sin −1, PLOT trazará sólo la mitad de la barra que representa a $x = 0$.

Excel

a.

```
Enter:   0,1,2, . . . ,12 into column A
```

Continúe con los comandos Excel de probabilidad binomial de las páginas 292-293, usando $n = 12$ y $p = 0.01$. Active las columnas A y B; luego continúe con:

```
Choose:  Chart Wizard > Column > 1st picture(usually) >
         Next > Series
Choose:  Series 1 > Remove
Enter:   Category (x)axis labels: (A1:A13 or select 'x
         value' cells)
Choose:  Next > Finish
Click on:Anywhere clear on the chart
         —use handles to size so x values fall under
         corresponding bars
```

Continúe los comandos Excel de probabilidad binomial acumulativa en las páginas 292-293, usando $n = 12$ y $p = 0.01$, y la columna C para la celda activada.

h.

```
Choose:  Tools > Data analysis > Random Number Genera-
         tion > OK
Enter:   Number of Variables: 1
         Number of Random Numbers: 200
         Distribution: Binomial
         p Value = 0.01
         Number of Trials = 12
Select:  Output Options: Output Range
Enter    (D1 or select cell) > OK
         Activate the E1 cell, then:
Choose:  Insert function, fx > Statistical > AVERAGE > OK
Enter:   Number 1: D1:D200 > OK
Activate the E2 cell, then:
Choose:  Insert function, fx > Statistical > STDEV > OK
Enter:   Number 1: D1:D200 > OK
```

Continúe con los comandos Excel del histograma de las páginas 61-62, usando los datos de la columna D y el rango de depósito de la columna A.

TI-83/84 Plus

a.

```
Choose:  STAT > EDIT > 1:Edit
Enter:   L1: 0,1,2,3,4,5,6,7,8,9,10,11,12
Choose:  2nd QUIT > 2nd DISTR > 0:binompdf(
Enter:   12, 0.01) > ENTER
Choose:  STO0  > L2 > ENTER
Choose:  2nd > STAT PLOT > 1:Plot1
Screen capture 5.5A
Choose:  WINDOW
Enter:   0, 13, 1, −.1, .9, .1, 1
Choose:  TRACE > > >
```

c.

```
Choose:  2nd > DISTR > A:binomcdf(
Enter:   12, 0.01)
Choose:  STO0  > L3 > ENTER
         STAT > EDIT > 1:Edit
```

h.

```
Choose:  MATH > PRB > 7:randBin(
Enter:   12, .01, 200) (takes a while to process)
Choose:  STO0  > L4 > ENTER
Choose:  2nd LIST > Math > 3:mean(
Enter:   L4
Choose:  2nd LIST > Math > 7:StdDev(
Enter:   L4
```

Continúe con los comandos TI-83/84 del histograma de las páginas 62-63, usando los datos de la columna L4 y ajustando la ventana después de la vista inicial usando ZoomStat.

5.98 ¿Alguna vez compró usted una bombilla eléctrica que fallara (se quemara o no funcionara) la primera vez que la encendió? Cuando usted pone una bombilla en una lámpara, se espera que encienda, y casi siempre así es. Considere unos paquetes de 8 bombillas de 60 watts y sea x el número de bombillas del paquete que "fallan" la primera vez que se usan. Si 0.02 de todas las bombillas de este tipo fallan en su primer uso y cada paquete de 8 bombillas se considera una muestra aleatoria,

a. Haga una lista de la distribución de probabilidad y trace el histograma de x.

b. ¿Cuál es la probabilidad de que cualquier paquete de 8 bombillas no tenga alguna que falle en su primer uso?

c. ¿Cuál es la probabilidad de que cualquier paquete de 8 bombillas no tenga más de una bombilla que falle en su primer uso?

d. Encuentre la media y desviación estándar de x.

e. ¿Qué proporción de la distribución está entre $\mu - \sigma$ y $\mu + \sigma$?

f. ¿Qué proporción de la distribución está entre $\mu - 2\sigma$ y $\mu + 2\sigma$?

g. ¿Cómo se relaciona esta información con la regla empírica y el teorema de Chebyshev? Explique.

h. Use computadora para simular probar 100 paquetes de 8 bombillas y observe x, el número de fallas por paquete de 8 piezas. Describa cómo se compara la información de la simulación con lo que se esperaba (las respuestas a las partes a-g describen los resultados esperados).

i. Repite la parte h varias veces. Describa cómo se comparan estos resultados con los de las partes a-g y con la parte h.

REPASO DEL CAPÍTULO

En retrospectiva

En este capítulo combinamos conceptos de probabilidad con algunas de las ideas presentadas en el capítulo 2. Ahora estamos en aptitud de manejar distribuciones de valores de probabilidad y encontrar medias, desviaciones estándar y otras estadísticas.

En el capítulo 4 exploramos los conceptos de eventos mutuamente excluyentes y eventos independientes. Empleamos las reglas de la adición y la multiplicación en varias ocasiones en este capítulo, pero dijimos muy poco acerca de exclusividad o independencia mutuas. Recuerde que cada vez que sumamos probabilidades, como lo hicimos en cada una de las distribuciones de probabilidad, necesitamos saber que los eventos asociados son mutuamente excluyentes. Si lee de nuevo el capítulo, notará que la variable aleatoria en realidad requiere que los eventos sean mutuamente excluyentes; por tanto, no hicimos hincapié en este concepto. El mismo comentario básico puede hacerse con respecto a la multiplicación de probabilidades y el concepto de eventos independientes. En todo este capítulo, multiplicamos probabilidades y ocasionalmente mencionamos la independencia. Ésta, por supuesto, es necesaria para estar en aptitud de multiplicar probabilidades.

Ahora, después de completar el capítulo 5 si fuéramos a ver más de cerca algunos de los conjuntos de datos del capítulo 2, veríamos que los diversos problemas podrían reconocerse para formar distribuciones de probabilidad. Veamos a continuación algunos ejemplos: (1) sea x el número de horas de crédito para las que un estudiante está registrado este semestre, pareado con el porcentaje de todo el alumnado que se indica por cada valor de x. (2) Sea x el número de pasillos correctos por los que un animal de laboratorio experimental pasa antes de tomar uno equivocado, pareado con la probabilidad de cada valor de x. (3) Sea x el número de solicitudes a universidad hechas y que no sean para aquella en la que está inscrito (Caso práctico 5.3), pareado con la probabilidad de cada valor de x. La lista de ejemplos es interminable.

Estamos listos para ampliar estos conceptos a las variables aleatorias continuas del capítulo 6.

Vocabulario y conceptos clave

coeficiente binomial (p. 289)
distribución de probabilidad (p. 273)
estadística muestral (p. 278)
eventos mutuamente excluyentes (p. 270)
éxito (p. 287)
experimento (p. 270)
experimento binomial (p. 287)
falla (p. 287)
función constante (p. 274)
función de probabilidad (p. 273)
función de probabilidad binomial (p. 288)
histograma de probabilidad (p. 275)
intento (p. 287)
intentos independientes (p. 287)
media de variable aleatoria discreta (p. 279)
parámetro de población (p. 278)
variable aleatoria (p. 280)
variable aleatoria (p. 270)
variable aleatoria binomial (p. 287)
variable aleatoria continua (p. 271)
variable aleatoria discreta (p. 271)
varianza de variable aleatoria discreta (p. 279)

Objetivos de aprendizaje

- ✓ Entender que una variable aleatoria es una cantidad numérica cuyo valor depende de las condiciones y probabilidades asociadas con un experimento. — pp. 270-271, EJ. 5.1
- ✓ Entender la diferencia entre una variable aleatoria discreta y una continua. — Ejer. 5.4, 5.5, 5.9

✓ Ser capaz de construir una distribución de probabilidad discreta basada en un experimento o función determinada.	pp. 273-274, Ejer.5.13 5.19
✓ Comprender los términos *mutuamente excluyente* y *todo incluido* cuando se aplican a las variables para distribuciones de probabilidad.	p. 270, Ejer. 5.15
✓ Entender las similitudes y diferencias entre distribuciones de frecuencia y distribuciones de probabilidad.	p. 270 Ejer. 5.100
✓ Comprender y ser capaz de utilizar las dos principales propiedades de distribuciones de probabilidad para verificar su conformidad.	p. 274, EJ. 5.2, Ejer. 5.17, 5.99, 5.101
✓ Entender que una distribución de probabilidad es una distribución de probabilidad teórica y que la media y desviación estándar (μ y σ, respectivamente) son parámetros	pp. 278-280, Ejer. 5.100
✓ Calcular, describir e interpretar la media y desviación estándar de una distribución de probabilidad.	Ejer. 5.5, Ejer. 5.33, 5.35
✓ Entender los elementos clave de un experimento binomial y ser capaces de definir x, n, p y q.	p. 287, EJ. 5.57, 5.63
✓ Saber y ser capaz de calcular probabilidades binomiales usando la función de probabilidad binomial	EJ. 5.8, Ejer. 5.57,5.63
✓ Entender y ser capaz de usar la tabla 2 del apéndice B, Probabilidades binomiales, para determinar probabilidades binomiales.	p. 292, Ejer. 5.58,5.111
✓ Calcular, describir e interpretar la media y desviación estándar de una distribución de probabilidad binomial.	EJ.5.11, Ej.589, 5.91

Ejercicios del capítulo

5.99 ¿Cuáles son las dos propiedades básicas de toda distribución de probabilidad?

5.100 A. Explique la diferencia y la relación entre una distribución.

b. Explique la diferencia y la relación entre una distribución de probabilidad y una distribución de frecuencia, y explique cómo se relacionan con una población y una muestra.

5.101 Verifique si cada uno de lo siguiente es una función de probabilidad. Exprese su conclusión y explique.

a. $f(x) = \dfrac{\frac{3}{4}}{x!(3-x)!}$ para $x = 0, 1, 2, 3$

b. $f(x) = 0.25$ para $x = 9, 10, 11, 12$

c. $f(x) = (3 - x)/2$ para $x = 1, 2, 3, 4$

d. $f(x) = (x^2 + x + 1)/25$ para $x = 0, 1, 2, 3$

5.102 Verifique si cada uno de lo siguiente es una función de probabilidad. Exprese su conclusión y explique.

a. $f(x) = \dfrac{3x}{8x!}$ para $x = 1, 2, 3, 4$

b. $f(x) = 0.125$ para $x = 0, 1, 2, 3$ y $f(x) = 0.25$ para $x = 4, 5$

c. $f(x) = (7 - x)/28$ para $x = 0, 1, 2, 3, 4, 5, 6,$

d. $f(x) = (x^2 + 1)/60$ para $x = 0, 1, 2, 3, 4, 5$

5.103 El número de barcos que llegan a puerto en cualquier día determinado es una variable aleatoria representada por x. La distribución de probabilidad para x es como sigue:

x	10	11	12	13	14
$P(x)$	0.4	0.2	0.2	0.1	0.1

Encuentre la probabilidad de lo siguiente para cualquier día determinado:

a. Llegan exactamente 14 barcos.

b. Llegan al menos 12 barcos.

c. Llegan a lo sumo 11 barcos.

5.104 "¿Cuántos televisores hay en su casa?" fue una de las preguntas de un cuestionario enviado a 5000 personas en Japón. Los datos recolectados resultaron en la siguiente distribución:

Número de TV por familia	0	1	2	3	4	5 o más
Porcentaje	1.9	31.4	23.0	24.4	13.0	6.3

Fuente: http://www.japan-guide.com/topic/0107.html

Una de estas familias se selecciona al azar.

a. ¿Qué porcentaje de familias tiene al menos una TV?

b. ¿Qué porcentaje de familias tiene a lo sumo tres TV?

c. ¿Qué porcentaje de familias tiene tres o más TV?

d. ¿Es éste un experimento de probabilidad binomial? Justifique su respuesta.

e. Sea x el número de televisores por familia. ¿Es ésta una distribución de probabilidad? Explique.

f. Asigne $x = 5$ para "5 o más" y encuentre la media y desviación estándar de x.

5.105 Los pacientes a quienes se ha hecho implante de cadera artificial experimentan dolor el primer día después de la cirugía. Por lo general, el dolor se mide en una escala subjetiva de 1 a 5. Represente con x la variable aleatoria, la intensidad del dolor como lo determine el paciente. La distribución de probabilidad para x se cree que es:

x	1	2	3	4	5
$P(x)$	0.10	0.15	0.25	0.35	0.15

a. Encuentre la media de x.

b. Encuentre la desviación estándar de x.

5.106 El censo del año 2000 produjo las siguientes cifras para la ciudad de Loveland, Colorado, respecto al número de vehículos disponibles por familia:

x	0	1	2	3 o más
Porcentaje	4.6	30.0	43.3	22.1

Fuente:
http://www.co.larimer.co.us/compass/vehicleperhousehold_cd_trans.htm#chart2

Sustituyendo la categoría de "3 o más" con "3":

a. Encuentre la media de x.

b. Encuentre la desviación estándar de x.

5.107 Una doctora sabe por experiencia que 10% de los pacientes a quienes prescribe cierto medicamento tendrán efectos colaterales indeseables. Encuentre las probabilidades de que entre los 10 pacientes a quienes les da el medicamento:

a. A lo sumo dos tendrán efectos colaterales indeseables.

b. Al menos dos tendrán efectos colaterales indeseables.

5.108 En una encuesta reciente de mujeres, 90% reconocieron que nunca habían visto un ejemplar de la revista *Vogue*. Suponiendo que ésta sea una información precisa, ¿cuál es la probabilidad de que una muestra aleatoria de tres mujeres muestre que menos de dos hayan leído la revista?

5.109 De quienes buscan obtener licencia de manejo, 70% reconocieron que nunca informarían de alguien que les copiara algunas respuestas durante el examen escrito. Usted acaba de entrar al salón y ve 10 personas a la espera de pasar el examen escrito. ¿Cuál es la probabilidad de que, si el incidente ocurrió, 5 de las 10 no reportarían lo que vieron?

5.110 Los motores de un avión operan de manera independiente. La probabilidad de que un motor opere para un viaje determinado es 0.95. Un avión podrá terminar con éxito un viaje si al menos la mitad de sus motores operan durante todo el trayecto. Determine si un avión de cuatro motores o de dos motores tiene la probabilidad más alta de un viaje exitoso.

5.111 El proyecto Pew Internet & American Life encontró que casi 70% de adultos "conectados" están en línea todos los días. En un grupo seleccionado al azar de 15 ciudadanos "conectados":

a. ¿Cuál es la probabilidad de que más de cuatro digan que están en línea todos los días?

b. ¿Cuál es la probabilidad de que exactamente 10 digan que están en línea todos los días?

c. ¿Cuál es la probabilidad de que menos de 10 digan que están en línea todos los días?

5.112 Las melodías R&B/hip-hop constituyeron más de 60% de los 100 más grandes éxitos de la radio en 2004, según datos de Nielsen BDS y Arbitron, como se informa en *USA Today* el 5 de enero, 2005. Una nueva estación de radio, adecuadamente llamada Fickle (veleidosa), transmite todo tipo de música de la lista de los 100 más grandes éxitos incluyendo R&B/hip-hop, rock y de *country*. En el siguiente grupo seleccionado al azar de 14 melodías que se transmite en Fickle, y usando el 60% para melodías R&B/hip-hop:

a. ¿Cuál es la probabilidad de que más de siete melodías sean de R&B/hip-hop?

b. ¿Cuál es la probabilidad de que exactamente 10 melodías sean de R&B/hip-hop?

c. ¿Cuál es la probabilidad de que menos de cinco melodías sean de R&B/hip-hop?

5.113 Imagine que está usted en la compra de un billete de lotería y la persona que está tras el mostrador imprime demasiados billetes con los números de usted. ¿Qué haría usted? Los resultados de una encuesta en línea fueron como sigue:

¿Dejar que conserve los boletos?	30.77%
¿Confiar en que la persona los borre?	15.38%
¿Comprar los billetes extra y esperar que ganen?	30.77%
Otro	23.08%

¿Es ésta una distribución de probabilidad? Explique.

5.114 El aprendizaje es una actividad que dura toda la vida. Para algunos, significa aprender de las experiencias diarias; para otros, es tomar clases en una atmósfera más tradicional. El porcentaje de personas que participan en situaciones organizadas de aprendizaje durante 2002 para cada grupo de edad se informa aquí por la NIACE.

Grupo de edad	17–19	20–24	25–35	35–44	45–54	55–64	65–74	75 +	All
Porcentaje	78	72	51	49	44	30	20	10	42

Fuente: NIACE Adult Participation in Learning Surveys

¿Es ésta una distribución de probabilidad? Explique.

5.115 El concejo de la ciudad tiene nueve miembros. Se ha presentado una proposición para establecer una nueva industria en esta ciudad, y todas las proposiciones deben tener al menos dos tercios de los votos para ser aceptadas. Si sabemos que dos miembros del concejo se oponen y que los otros votan al azar "a favor" y "en contra", ¿cuál es la probabilidad de que la proposición sea aceptada?

5.116 Hay 750 jugadores en las listas de activos de los 30 equipos de béisbol de ligas mayores. Se va a seleccionar y probar una muestra aleatoria de 15 jugadores en busca de uso de drogas.

a. Si 5% de todos los jugadores están usando drogas ilegales en el momento de la prueba, ¿cuál es la probabilidad de que 1 o más jugadores den positivo y no pasen la prueba?

b. Si 10% de todos los jugadores están usando drogas ilegales en el momento de la prueba, ¿cuál es la probabilidad de que 1 o más jugadores den positivo y no pasen la prueba?

c. Si 20% de todos los jugadores están usando drogas ilegales en el momento de la prueba, ¿cuál es la probabilidad de que 1 o más jugadores den positivo y no pasen la prueba?

5.117 Una caja contiene 10 artículos, 3 de los cuales son defectuosos y 7 no son defectuosos. Dos de los artículos se seleccionan sin reposición, y x es el número de artículos defectuosos en la muestra de dos. Explique por qué x no es una variable aleatoria binomial.

5.118 Una caja contiene 10 artículos, 3 de los cuales son defectuosos y 7 no son defectuosos. Dos de los artículos se seleccionan al azar, uno a la vez, con reposición, y x es el número de artículos defectuosos en la muestra de dos. Explique por qué x no es una variable aleatoria binomial.

5.119 Un embarque importante de radios es aceptado al recibirlo si una inspección de 10 radios seleccionados al azar da no más de 1 radio defectuoso.

a. Encuentre la probabilidad de que este embarque sea aceptado si 5% del embarque total es defectuoso.

b. Encuentre la probabilidad de que este embarque sea aceptado si 20% del embarque total es defectuoso.

c. La distribución de probabilidad binomial se emplea con frecuencia en situaciones semejantes a ésta, es decir, se muestren grandes poblaciones sin reposición. Explique por qué el binomio da una buena estimación.

5.120 El ingeniero de diseño de puentes estatales ha ideado un plan para reparar los 4706 puentes de Carolina del Norte que actualmente están considerados como en malas o regulares condiciones. El estado tiene un total de 13,268 puentes. Antes que el gobernador incluya el costo de este plan en su presupuesto, él ha decidido visitar e inspeccionar personalmente cinco puentes que han de seleccionarse al azar. ¿Cuál es la probabilidad de que, en la muestra de cinco puentes, el gobernador visite lo siguiente?:

a. Ningún puente sea clasificado como malo o regular
b. Uno de dos puentes sea clasificado como malo o regular
c. Cinco puentes sean clasificados como malos o regulares

5.121 Una variable aleatoria discreta tiene una desviación estándar igual a 10 y una media igual a 50. Encuentre $\sum x^2 P(x)$.

5.122 Una variable aleatoria discreta está basada en $n = 20$ y $p = 0.4$. Encuentre $\sum x^2 P(x)$.

5.123 En una prueba de germinación, se plantaron 50 semillas en cada uno de 40 surcos. El número de semillas que germinaron en cada surco se registró como aparece en la tabla siguiente:

Número germinado	Número de surcos	Número germinado	Número de surcos
39	1	45	8
40	2	46	4
41	3	47	3
42	4	48	1
43	6	49	1
44	7		

a. Use la tabla precedente de distribución de frecuencia para determinar el porcentaje observado de germinación de estas semillas.
b. El experimento de probabilidad binomial con su correspondiente distribución de probabilidad se puede usar con la variable "número de semillas que germinan por surco" cuando se plantan 50 semillas en cada surco. Identifique la función binomial específica y haga una lista de su distribución, usando el porcentaje de germinación hallado en la parte a. Justifique su respuesta.
c. Suponga que usted está planeando repetir este experimento al plantar 40 surcos de estas semillas, con 50 semillas en cada surco. Use su modelo de probabilidad de la parte b para hallar la distribución de frecuencia para x que usted esperaría resulte de su experimento planeado.
d. Compare su respuesta en la parte c con los resultados que se dieron en la tabla precedente. Describa cualesquiera similitudes y diferencias.

5.124 En otro experimento de germinación donde intervienen semillas viejas, se plantaron 50 surcos de semillas. En la siguiente tabla se registró el número de semillas que germinaron en cada surco (cada surco contenía el mismo número de semillas).

Número germinado	Número de surcos	Número germinado	Número de surcos
0	17	3	2
1	20	4	1
2	10	5 o más	0

a. ¿Qué distribución de probabilidad (o función) sería útil para modelar la variable "número de semillas que germinan por surco"? Justifique su respuesta.
b. ¿Qué información falta para aplicar la distribución de probabilidad que escogió en la parte a?
c. Con base en la información que tenga, ¿cuál es el porcentaje más alto o más bajo de germinación que usted puede estimar para estas semillas? Explique.

5.125 Una empresa financiera está considerando dos inversiones. Seleccionará aquella que prometa el rendimiento más alto. ¿Cuál de las inversiones debe aceptar? (La utilidad media mide el rendimiento.)

Invertir en taller de herramientas		Invertir en librería	
Utilidad	Probabilidad	Utilidad	Probabilidad
$100 000	0.10	$400 000	0.20
50 000	0.30	90 000	0.10
20 000	0.30	−20 000	0.40
−80 000	0.30	−250 000	0.30
Total	1.00	Total	1.00

5.126 Bill ha completado un examen de 10 preguntas con respuestas opcionales, 7 de las cuales contestó correctamente. Cada pregunta tenía una respuesta correcta para ser seleccionada de cinco alternativas. Hill dice que él contestó el examen al elegir al azar las respuestas sin leer las preguntas o respuestas.

a. Defina la variable aleatoria x como el número de respuestas correctas en este examen, y construya la distribución de probabilidad si las respuestas se obtuvieron por selección aleatoria.

b. ¿Cuál es la probabilidad de que Hill seleccione correctamente 7 de las 10 respuestas?

c. ¿Cuál es la probabilidad de que alguien puede seleccionar correctamente seis o más respuestas?

d. ¿Piensa usted que Hill en realidad seleccionó al azar como lo dice? Explique.

5.127 Se dice que una variable aleatoria que pueda tomar cualquiera de los valores enteros 1,2, . . . ,n con iguales probabilidades de $\frac{1}{n}$ tiene una distribución uniforme.

La función de probabilidad se escribe $P(x) = \frac{1}{n}$, para $x = 1, 2, 3, \ldots, n$. Demuestre que $\mu = \frac{(n+1)}{2}$. (*Sugerencia:* $1+2+3+ \ldots +n = [n(n+1)]/2$.)

Proyecto del capítulo

Bebidas de cafeína

Veamos de nuevo la sección 5.1 "Bebidas de cafeína" (p. 269), y pongamos a prueba nuestro conocimiento del material presentado en este capítulo. Con base en la USA Snapshot, tenemos el número de tazas o latas de bebidas de cafeína que adultos estadounidenses dicen que toman a diario, y sus correspondientes probabilidades. Considere dónde podría usted caber en esta situación.

Trabajando en el contenido del capítulo 5

5.128 a. ¿Qué variable podría usarse para describir los cinco eventos mostrados en la gráfica "A los estadounidenses les gusta tomar café" (p. 269)?

b. ¿Es discreta o continua la variable de la parte a? ¿Por qué?

c. ¿Son mutuamente excluyentes los eventos $x = 1,2$? Explique por qué sí o por qué no.

d. ¿Qué características de una gráfica de pastel la hacen apropiada para usarla con una distribución de probabilidad? Sea específico.

e. Construya una gráfica de pastel que describa la información descrita en la gráfica.

f. Exprese la información de la gráfica de pastel como una distribución de probabilidad.

g. Suponiendo que la información en la gráfica de pastel represente la población, encuentre la media y desviación estándar de la variable descrita en la parte a.

h. Trace un histograma para mostrar la información en la gráfica. Describa el histograma. ¿Es una distribución normal? Explique.

i. Localice la media y desviación estándar hallada en la parte g en el histograma trazado en la parte h.

j. ¿Aplican las reglas empírica y de Chebyshev? Justifique su respuesta.

Para su estudio

5.129 Diseñe su propio estudio de tomar bebidas de cafeína.

a. Defina una población específica que usted va a muestrear, describa su plan de muestreo, y recolecte sus datos.

b. Exprese su muestra como una distribución de frecuencia relativa y trace un histograma.

c. Exprese su muestra como una distribución de frecuencia y encuentre la media muestral y desviación estándar muestral.

d. Discuta las diferencias y similitudes entre su muestra y la distribución mostrada en la gráfica "A los estadounidenses les gusta tomar café".

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

5.1 El número de horas que estuvo usted en una fila de espera para registrarse este semestre es un ejemplo de una variable aleatoria **discreta.**

5.2 El número de accidentes automovilísticos en los que usted participó como conductor el año pasado es un ejemplo de una variable aleatoria **discreta.**

5.3 La suma de todas las probabilidades en cualquier distribución de probabilidad es siempre exactamente **dos.**

5.4 Los diversos valores de una variable aleatoria forman una lista de **eventos mutuamente excluyentes.**

5.5 Un experimento binomial siempre tiene **tres o más** posibles resultados en cada intento.

5.6 La fórmula $\mu = np$ puede usarse para calcular la media de una población **discreta.**

5.7 El parámetro binomial p es la probabilidad de **un éxito que sucede en *n* intentos** cuando se realiza un experimento binomial.

5.8 Un parámetro es una medida estadística de algún aspecto de una **muestra.**

5.9 Las **estadísticas muestrales** se representan por letras del alfabeto griego.

5.10 La probabilidad del evento A o B es igual a la suma de la probabilidad del evento A y la probabilidad del evento B, cuando A y B sean **eventos mutuamente excluyentes.**

SEGUNDA PARTE: Aplicando los conceptos

5.11 a. Demuestre que lo siguiente es una distribución de probabilidad:

x	1	3	4	5
$P(x)$	0.2	0.3	0.4	0.1

b. Encuentre $P(x = 1)$.

c. Encuentre $P(x = 2)$.

d. Encuentre $P(x > 2)$.

e. Encuentre la media de x.

f. Encuentre la desviación estándar de x.

5.12 Una empresa fabricante de camisetas anuncia que la probabilidad de que una pieza irregular sea irregular es de 0.1. Una caja de 12 de estas camisetas se selecciona e inspecciona al azar.

a. ¿Cuál es la probabilidad de que exactamente 2 de las 12 camisetas sean irregulares?

b. ¿Cuál es la probabilidad de que exactamente 9 de las 12 camisetas no sean irregulares?

Sea x el número de camisetas que son irregulares en todas esas cajas de 12 camisetas.

c. Encuentre la media de x.

d. Encuentre la desviación estándar de x.

TERCERA PARTE: Entendiendo los conceptos

5.13 ¿Qué propiedades debe poseer un experimento para que sea experimento de probabilidad binomial?

5.14 La estudiante A emplea una distribución de frecuencia relativa para un conjunto de datos muestrales y calcula la media y desviación estándar usando fórmulas del capítulo 5. La estudiante A justifica su de fórmulas diciendo que, como las frecuencias relativas son probabilidades empíricas, su muestra está representada por una distribución de probabilidad y por tanto su selección de fórmulas fue correcta. La Estudiante B dice que como la distribución representaba una muestra, la media y desviación estándar comprendidas se conocen como $\bar{x}$ y s y deben ser calculadas usando la correspondiente distribución de frecuencia y fórmulas del capítulo 2. ¿Quién tiene razón, A o B? Justifique su respuesta.

5.15 La Estudiante A y la Estudiante B estuvieron analizando una entrada de una tabla de distribución de probabilidad:

x	$P(x)$
−2	0.1

La Estudiante B pensó que esta entrada estaba bien porque $P(x)$ era un valor entre 0.0 y 1.0. La Estudiante A dice que esta entrada era imposible para una distribución de probabilidad porque x era −2 y los negativos no son posibles. ¿Quién tiene razón, A o B? Justifique su elección.

CAPÍTULO

6 Distribuciones de probabilidad normal

6.1 Medición de la inteligencia

6.2 Distribuciones de probabilidad normal

6.3 La distribución normal estándar

6.4 Aplicaciones de la distribución normal

6.5 Notación

6.6 Aproximación normal de la binomial

6.1 Medición de la inteligencia

MEDIDAS DE INTELIGENCIA

Pruebas de aptitud y su interpretación

Hay numerosas clases de pruebas de aptitud. Algunas de ellas son para fines específicos, por ejemplo medir la destreza con los dedos, algo que podría ser importante en un trabajo en particular. Otras son de aptitudes más generales. Las llamadas pruebas de inteligencia son ejemplos de pruebas de aptitud.

La escala Binet de inteligencia. Alfred Binet, quien ideó la primera prueba general de aptitud a principios del siglo XX, definió la inteligencia como *la capacidad de hacer adaptaciones*. El propósito general de la prueba fue determinar cuáles niños en París podrían beneficiarse por la escuela. La prueba de Binet, al igual que sus modificaciones subsiguientes, consta de una serie de tareas cada vez más difíciles que niños de edades diferentes pueden completar con éxito. Se dice que un niño que puede resolver problemas que por lo general son resueltos por niños de un nivel particular de edad tiene esa edad mental. Por ejemplo, si un niño puede con éxito hacer las mismas tareas que un niño de ocho años, tiene una edad mental de ocho. El *cociente de inteligencia*, o IQ, está definido por la fórmula:

Cociente de inteligencia =
100 × (edad mental/edad cronológica)

Ha habido una gran controversia en años recientes sobre lo que miden las pruebas de inteligencia. Muchos de los elementos de la prueba dependen del idioma o de otras experiencias culturales específicas para ser resueltos. No obstante, con estas pruebas se puede pronosticar de manera efectiva el éxito escolar. Si la escuela requiere capacidad de lenguaje en un punto particular de la vida de un niño, entonces la prueba es un pronosticador mejor que la probabilidad de rendimiento escolar.

Mediciones de desviación de IQ. Las pruebas de inteligencia u otras capacidades de hoy día emplean *mediciones de desviación*. Estas mediciones representan la desviación de una persona particular respecto del promedio de medición para personas semejantes. Supongamos que el lector toma un "examen general de aptitud" y obtiene una puntuación de 115. Esto no significa que su edad mental sea mayor que su edad cronológica; significa que está "arriba del promedio" en algún grado. Como nos hemos acostumbrado a pensar en que una medición de IQ de 100 como promedio, casi todas las pruebas de aptitud general se miden en forma tal que 100 es el promedio. Una persona que obtiene 115 por lo general tendría una puntuación más alta que la de alrededor de 85 por ciento de las personas que toman la prueba; una puntuación de 84 sería mejor que la de 16 por ciento. La interpretación exacta de una puntuación de prueba depende de la prueba en particular, pero la figura 2.2 (p. 314) muestra la forma en que las puntuaciones de varias pruebas de aptitud comúnmente empleadas se interpretan en términos de cómo se compara una persona contra un grupo.

La figura 2.2 ilustra la comparación de varias medidas de desviación y la distribución normal: las medidas estándar tienen una media de cero y una desviación estándar de 1.0. Las medidas de la Prueba de Aptitud Escolar (SAT) tienen una media de 500 y una desviación estándar de 100. Las medidas de la Prueba de Inteligencia de Binet tienen una media de 100 y una desviación estándar de 16. En cada caso hay 34 por ciento de medidas entre la media y una desviación estándar, 14 por ciento entre una y dos desviaciones estándar, y 2 por ciento de más de dos desviaciones estándar.

FIGURA 2.2

Fuente: Robert C. Beck, *Applying Psychology, Critical and Creative Thinking,* 3rd ed. (Englewood Cliffs, NJ: Prentice Hall, 1992)

Después de completar el capítulo 6, investigue más a fondo las mediciones de inteligencia en el proyecto del capítulo con los ejercicios 6.137 y 6.138 (p. 356).

SECCIÓN 6.1 EJERCICIOS

6.1 a. Explique por qué la medición del IQ es una variable continua.

b. ¿Cuáles son la media y la desviación estándar para la distribución de mediciones del IQ? ¿Y de las mediciones de SAT? ¿Y de las mediciones estándar?

c. Exprese, algebraicamente o como una ecuación, la relación entre mediciones estándar y mediciones del IQ y entre mediciones estándar y mediciones de la SAT.

d. ¿Qué medición estándar está a 2 desviaciones estándar arriba de la media? ¿Cuál medición del IQ está 2 desviaciones estándar arriba de la media? ¿Cuál medición de la SAT está 2 desviaciones estándar arriba de la media?

e. Compare la información acerca del porcentaje de distribución mostrado en la figura 2.2 con la regla empírica estudiada en el capítulo 2. Explique las similitudes.

6.2 Examine el cociente de inteligencia, o IQ, como está definido por la fórmula:

$$\text{cociente de inteligencia} = 100 \times (\text{edad mental/edad cronológica})$$

Justifique por qué es razonable que la media sea 100.

6.2 Distribuciones de probabilidad normal

La **distribución de probabilidad normal** es considerada como la más importante distribución de probabilidad. Un número ilimitado de **variables aleatorias continuas** tienen una distribución ya sea normal o una aproximadamente normal. Varias otras probabilidades de distribución de variables aleatorias discretas y continuas también son aproximadamente normales bajo ciertas condiciones.

Recuerde que en el capítulo 5 aprendimos a usar una función de probabilidad para calcular las probabilidades asociadas con **variables aleatorias discretas.** La distribución de probabilidad normal tiene una variable aleatoria continua y emplea dos funciones: una función para determinar las ordenadas (valores y) de la gráfica que representa la distribución y una segunda para determinar las probabilidades. La fórmula (6.1) expresa la ordenada (valor y) que corresponde a cada abscisa (valor x).

Función de distribución de probabilidad normal

$$y = f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad \text{para todo real } x \tag{6.1}$$

Cuando se traza una gráfica de todos estos puntos, la **curva normal (en forma de campana)** aparecerá como se ve en la figura 6.1.

FIGURA 6.1
Distribución de probabilidad normal

Nota: cada par diferente de valores para la media, μ, y desviación estándar, σ, resultarán en una función diferente de distribución de probabilidad normal.

La fórmula (6.2) da la probabilidad asociada con el intervalo de $x = a$ a $x = b$:

$$P(a \le x \le b) = \int_a^b f(x)\ dx \tag{6.2}$$

FIGURA 6.2
Área sombreada: $P(a \le x \le b)$

La probabilidad de que x se encuentre dentro del intervalo de $x = a$ a $x = b$ se muestra como el área sombreada en la figura 6.2.

No estaremos usando las fórmulas precedentes para calcular probabilidades para distribuciones normales. La integral definida de la fórmula (6.2) es un tema de cálculo y está matemáticamente fuera de lo que se espera en estadística elemental. (Estas fórmulas aparecen con frecuencia en la parte superior de tablas de probabilidad normal como identificación.) En lugar de usar las fórmulas (6.1) y (6.2), usaremos una tabla para hallar probabilidades para distribuciones normales. Las fórmulas 6.1 y 6.2 se usaron para generar esa tabla. Antes que aprendamos a usar la tabla, sin embargo, debe señalarse que la tabla está expresada en forma "estandarizada" de modo que esta tabla pueda usarse para hallar probabilidades para todas las combinaciones de valores de la media, μ, y desviación estándar, σ. Esto es, la distribución de probabilidad normal con media de 38 y desviación estándar de 7 es semejante a la distribución de probabilidad normal con media de 123 y desviación estándar de 32. Recuerde la regla empírica y los porcentajes de la distribución que caen dentro de ciertos intervalos de la media (p. 116). Los mismos tres porcentajes se cumplen para todas las distribuciones normales.

Nota: porcentaje, proporción y **probabilidad** son básicamente los mismos conceptos. Se usa el porcentaje (25%) o proporción ($\frac{1}{4}$) cuando se habla de parte de una población, siendo el porcentaje lo más común. La probabilidad suele usarse cuando se habla de la posibilidad de que el siguiente elemento individual posea cierta propiedad. El área es la representación gráfica de las tres cuando trazamos una figura para ilustrar la situación. La regla empírica es un medio de medición más bien burdo; con ella podemos hallar probabilidades asociadas sólo con múltiplos enteros de la desviación estándar (no más de 1, 2 o 3 desviaciones estándar de la media). Con frecuencia estaremos interesados en las probabilidades asociadas con partes fraccionales de la desviación estándar. Por ejemplo, podríamos desear conocer la probabilidad de que x se encuentre a no más de 1.37 desviaciones estándar de la media. Por tanto, debemos refinar la regla empírica para que podamos manejar mediciones más precisas. Este refinamiento se estudia en la siguiente sección.

SECCIÓN 6.2 EJERCICIOS

6.3 Porcentaje, proporción o probabilidad: identifique cuál está ilustrado por cada una de las siguientes frases.

a. Un tercio de la multitud tuvo una vista clara del evento.

b. Quince por ciento de los votantes fueron encuestados cuando salieron del distrito electoral de votación.

c. La posibilidad de lluvia durante el día de mañana es 0.2.

6.4 Porcentaje, proporción o probabilidad: en sus propias palabras, usando entre 25 y 50 palabras para cada una, describa lo siguiente:

a. Modo en que el porcentaje es diferente de los otros dos

b. Modo en que la proporción es diferente de los otros dos

c. Modo en que la probabilidad es diferente de los otros dos

d. Modo en que los tres son básicamente lo mismo

6.3 La distribución normal estándar

Hay un número ilimitado de distribuciones de probabilidad normal, pero por fortuna todas están relacionadas con una distribución: la **distribución normal estándar.** La distribución normal estándar es la distribución normal de la variable estándar z (llamada **"puntaje estándar"** o **"puntaje** z**"**).

Propiedades de la distribución normal estándar:

1. El área total bajo la curva normal es igual a 1.
2. La distribución tiene forma de campana y es simétrica; se extiende indefinidamente en ambas direcciones, aproximándose pero sin tocar el eje horizontal.
3. La distribución tiene una media de 0 y una desviación estándar de 1.
4. La media divide el área en dos: 0.50 a cada lado.
5. Casi toda el área está entre $z = -3.00$ y $z = 3.00$.

La tabla 3 del apéndice B es una lista de las probabilidades asociadas con los intervalos desde la media (ubicada en $z = 0.00$) hasta un valor específico de z. Las probabilidades de otros intervalos pueden hallarse usando las entradas de tabla y las operaciones de adición y sustracción, de acuerdo con las propiedades precedentes. Veamos varias ilustraciones que demuestran la forma de usar la tabla 3 para hallar probabilidades del puntaje normal estándar, z.

EJEMPLO 6.1 Para hallar el área a la derecha de $z = 0$

Encuentre el área bajo la curva normal estándar entre $z = 0$ y $z = 1.52$ (vea figura 6.3).

FIGURA 6.3
Área de $z = 0$ a $z = 1.52$

SOLUCIÓN La tabla 3 está diseñada para dar el área entre $z = 0$ y $z = 1.52$ directamente. El puntaje z está ubicado en los márgenes, con las unidades y décimas de dígito por todo el lado izquierdo y centésimas de dígito en la parte superior. Para $z = 1.52$, localice la fila marcada 1.5 y la columna marcada 0.02; en su intersección encontrará 0.4357, la medida del área o la probabilidad para el intervalo $z = 0.00$ a $z = 1.52$ (vea la tabla 6.1). Expresado como una probabilidad: $P(0.00 < z < 1.52) = \mathbf{0.4357}$.

TABLA 6.1

Una parte de la tabla 3

z	*0.00*	*0.01*	*0.02*	*. . .*
⋮				
1.5			0.4357	. . .
⋮				

Recuerde que una de las propiedades básicas de probabilidad es que la suma de todas las probabilidades es exactamente 1.0. Como el área bajo la curva normal representa la medida de probabilidad, el área total bajo la curva en forma de campana es exactamente 1. Esta distribución también es simétrica respecto a la recta vertical trazada por $z = 0$, que corta el área en dos en la media. ¿Puede usted verificar este dato al inspeccionar la fórmula (6.1)? Esto es, el área bajo la curva a la derecha de la media es exactamente la mitad, 0.5, y el área a la izquierda también es la mitad, 0.5. Las áreas (probabilidades) que no se dan directamente en la tabla se pueden hallar si confiamos en estos datos.

A continuación veamos algunos ejemplos.

¿SABÍA USTED...?

La curva en forma de campana

En los siglos XVIII y XIX, astrónomos y físicos describían sus observaciones usando fórmulas matemáticas precisas. A continuación explicaban que la diferencia, entre los valores observados y los pronosticados, era resultado de la falta de precisión en sus instrumentos y por lo tanto no era de importancia. A medida que los instrumentos se hicieron más precisos, se hizo evidente que este error era la aleatoriedad inherente de las observaciones. En 1820, Laplace describió esto con su función de error. Esta distribución de error ganó popularidad y ahora se conoce como la distribución de probabilidad normal y con frecuencia recibe el nombre de *curva en forma de campana*.

EJEMPLO 6.2 Para hallar el área en la cola derecha de una curva normal

Encuentre el área bajo la curva normal a la derecha de $z = 1.52$: $P(z > 1.52)$.

SOLUCIÓN El área a la derecha de la media (toda el área sombreada de la figura) es exactamente 0.5000. El problema pide el área sombreada que no está incluida en 0.4357. Por tanto, restamos 0.4357 de 0.5000:

$$P(z > 1.52) = 0.5000 - 0.4357 = \mathbf{0.0643}$$

Notas: 1. Como lo hemos hecho aquí, siempre trace y aplique leyendas a un dibujo; es muy útil. 2. Fórmese el hábito de escribir z con dos lugares decimales y áreas y probabilidades con cuatro lugares decimales, como en la tabla 3.

EJEMPLO 6.3 Para hallar el área a la izquierda de un valor positivo de *z*

Encuentre el área a la izquierda de $z = 1.52$: $P(z < 1.52)$.

SOLUCIÓN El total del área sombreada está formado por 0.4357 hallado en la tabla y el 0.5000 que está a la izquierda de la media. Por tanto, sumamos 0.4357 a 0.5000:

$$P(z \leq 1.52) = P(z < 0) + P(0 < z < 1.52)$$
$$= 0.5000 + 0.4357 = \mathbf{0.9357}$$

Nota: la suma y resta hechas en los ejemplos 6.2 y 6.3 son correctas porque las "áreas" representan eventos mutuamente excluyentes (estudiados en la sección 4.5).

La simetría de la distribución normal es un factor clave para determinar probabilidades asociadas con valores debajo de la media (a la izquierda). El área entre la media y $z = -1.52$ es exactamente la misma que el área entre la media y $z = +1.52$. Este dato nos permite hallar valores relacionados con el lado izquierdo de la distribución, como se ilustra en los siguientes dos ejemplos.

EJEMPLO 6.4 Para hallar el área de una *z* negativa a *z* = 0

El área entre la media ($z = 0$) y $z = -2.1$ es igual que el área entre $z = 0$ y $z = +2.1$; esto es,

$$P(-2.1 < z < 0) = P(0 < z < 2.1)$$

Así, tenemos

$$P(-2.1 < z < 0) = P(0 < z < 2.1) = \mathbf{0.4821}$$

EJEMPLO 6.5 Para hallar el área en la cola izquierda de una curva normal

El área a la izquierda de $z = -1.35$ se encuentra al restar 0.4115 de 0.5000.

Por tanto, obtenemos

$$P(z < -1.35) = P(z < 0) - P(-1.35 < z < 0)$$
$$= 0.5000 - 0.4115 = \mathbf{0.0885}$$

EJEMPLO 6.6 Para hallar el área de una *z* negativa a una *z* positiva

El área entre $z = -1.5$ y $z = 2.1$, $P(-1.5 < z < 2.1)$, se encuentra al sumar dos áreas. Ambas probabilidades pedidas se leen directamente de la tabla 3.

Por tanto, obtenemos

$$P(-1.5 < z < 2.1) = P(-1.5 < z < 0) + P(0 < z < 2.1)$$
$$= 0.4332 + 0.4821 = \mathbf{0.9153}$$

EJEMPLO 6.7 Para hallar el área entre dos valores *z* del mismo signo

El área entre $z = 0.7$ y $z = 2.1$, $P(0.7 < z < 2.1)$, se encuentra por resta. El área entre $z = 0$ y $z = 2.1$ incluye toda el área entre $z = 0$ y $z = 0.7$. Por tanto, restamos el área entre $z = 0$ y $z = 0.7$ del área entre $z = 0$ y $z = 2.1$.

Así, tenemos

$$P(0.7 < z < 2.1) = P(0 < z < 2.1) - P(0 < z < 0.7)$$
$$= 0.4821 - 0.2580 = \mathbf{0.2241}$$

La tabla de distribución normal estándar también se puede usar para hallar un puntaje z cuando nos dan un área. El siguiente ejemplo considera esta idea.

EJEMPLO 6.8 Para hallar puntajes *z* asociados con un percentil

¿Cuál es el puntaje z asociado con el 75avo percentil de una distribución normal? Vea la figura 6.4.

SOLUCIÓN

FIGURA 6.4
P_{75} y su puntaje *z* asociado

Para hallar este puntaje z, vea la tabla 3 del apéndice B y encuentre a la entrada de "área" más cercana a 0.2500; esta entrada de área es 0.2486. Ahora lea el puntaje z que corresponde a esta área.

z	...	**0.07**		0.08	...
⋮					
0.6		0.2486	*0.2500*	0.2517	...
⋮					

De la tabla, el puntaje z se encuentra que es z = **0.67**. Esto dice que el 75avo percentil en una distribución normal es 0.67 (aproximadamente ($\frac{2}{3}$) de la desviación estándar arriba de la media.

EJEMPLO 6.9 Para hallar los puntajes *z* que limitan un área

¿Qué puntajes z limitan el 95% central de una distribución normal?

SOLUCIÓN El 95% está dividido en dos partes iguales por la media, de modo que 0.4750 es el área (porcentaje) entre $z = 0$, la media, y el puntaje z en la frontera derecha. Vea la figura 6.5.

FIGURA 6.5 Mitad del 95% de la distribución y su puntaje z asociado

Como tenemos el área, buscamos en la tabla 3 la entrada más cercana a 0.4750 (que es exactamente 0.4750) y leemos el puntaje z. Obtenemos $z = 1.96$.

z	...	**0.06**	...
⋮			
1.9		0.4750	...
⋮			

Por tanto, $z = -$**1.96** y z = **1.96** limitan el 95% central de una distribución normal.

SECCIÓN 6.3 EJERCICIOS

6.5 a. Describa la distribución de puntaje normal estándar z.

b. ¿Por qué esta distribución se denomina normal estándar?

6.6 Encuentre el área bajo la curva normal estándar entre $z = 0$ y $z = 1.37$.

6.7 Encuentre el área bajo la curva normal que está entre los siguientes pares de valores z:

a. $z = 0$ a $z = 1.30$ b. $z = 0$ a $z = 1.28$

c. $z = 0$ a $z = -3.20$ d. $z = 0$ a $z = -1.98$

6.8 Encuentre la probabilidad de que un valor de datos tomado al azar de una población normal tendrá un puntaje estándar (z) que se encuentre entre los siguientes pares de valores.

a. $z = 0$ a $z = 2.10$ b. $z = 0$ a $z = 2.57$

c. $z = 0$ a $z = -1.20$ d. $z = 0$ a $z = -1.57$

6.9 Encuentre el área bajo la curva normal estándar a la derecha de $z = 2.03$, $P(z > 2.03)$.

6.10 Encuentre el área bajo la curva normal estándar a la izquierda de $z = 1.73$, $P(z < 1.73)$.

6.11 Encuentre el área bajo la curva normal estándar entre −1.39 y la media, $P(-1.39 < z < 0.00)$.

6.12 Encuentre el área bajo la curva normal estándar a la izquierda de $z = -1.53$, $P(z < -1.53)$.

6.13 Encuentre el área bajo la curva normal estándar entre $z = -1.83$ y $z = 1.23$, $P(-1.83 < z < 1.23)$.

6.14 Encuentre el área bajo la curva normal estándar entre $z = -2.46$ y $z = 1.46$, $P(-2.46 < z < 1.46)$.

6.15 Encuentre el área bajo la curva normal estándar que corresponde a los siguientes valores z:

a. Entre 0 y 1.55 b. A la derecha de 1.55

c. A la izquierda de 1.55 d. Entre −1.55 y 1.55

6.16 Encuentre la probabilidad de que un valor de datos tomado al azar de una población normalmente distribuida tenga un puntaje estándar (z) que corresponde a lo siguiente:

a. Entre 0 y 0.84 b. A la derecha de 0.84

c. A la izquierda de 0.84 d. Entre −0.84 y 0.84

6.17 Encuentre las siguientes áreas bajo la curva normal.

a. A la derecha de $z = 0.00$

b. A la derecha de $z = 1.05$

c. A la derecha de $z = -2.3$

d. A la izquierda de $z = 1.60$

e. A la izquierda de $z = -1.60$

6.18 Encuentre la probabilidad de que un valor de datos tomado al azar de una población normalmente distribuida tenga un puntaje estándar que corresponda a lo siguiente.

a. Menor de 3.00 b. Mayor de −1.55

c. Menor de −0.75 d. Menor de 1.25

e. Mayor de −1.25

6.19 Encuentre lo siguiente:

a. $P(0.00 < z < 2.35)$ b. $P(-2.10 < z < 2.34)$

c. $P(z > 0.13)$ d. $P(z < 1.48)$

6.20 Encuentre lo siguiente:

a. $P(-2.05 < z < 0.00)$ b. $P(-1.83 < z < 2.07)$

c. $P(z < 1.52)$ d. $P(z < -0.43)$

6.21 Encuentre lo siguiente:

a. $P(0.00 < z < 0.74)$

b. $P(-1.17 < z < 1.94)$

c. $P(z < 1.25)$

d. $P(z < 1.75)$

6.22 Encuentre lo siguiente:

a. $P(3.05 < z < 0.00)$

b. $P(-2.43 < z < 1.37)$

c. $P(z < -2.17)$

d. $P(z > 2.43)$

6.23 Encuentre el área bajo la curva normal estándar entre $z = 0.75$ y $z = 2.25$, $P(0.75 < z < 2.25)$.

6.24 Encuentre el área bajo la curva normal estándar entre $z = -2.75$ y $z = -1.28$, $P(-2.75 < z < -1.28)$.

6.25 Encuentre el área bajo la curva normal estándar que se encuentra entre los siguientes pares de valores z:

a. $z = -1.20$ a $z = 1.22$ b. $z = -1.75$ a $z = 1.54$

c. $z = -1.30$ a $z = 2.58$ d. $z = -3.5$ a $z = -0.35$

6.26 Encuentre la probabilidad de que un valor de datos tomado al azar de una población normalmente distribuida tenga un puntaje estándar (z) que se encuentre entre los siguientes pares de valores z:

a. $z = -2.75$ a $z = 1.38$ b. $z = 0.67$ a $z = 2.95$

c. $z = -2.95$ a $z = -1.18$

6.27 Encuentre el puntaje z para la distribución normal estándar que se muestra en cada uno de los diagramas siguientes.

6.28 Encuentre el puntaje z para la distribución normal estándar que se muestra en cada uno de los diagramas siguientes.

6.29 Encuentre el puntaje (z) estándar que se muestra en cada uno de los diagramas siguientes.

6.30 Encuentre el puntaje (z) estándar que se muestra en cada uno de los diagramas siguientes.

6.31 Encuentre un valor de z tal que 40% de la distribución se encuentre entre él y la media. (Hay dos posibles respuestas.)

6.32 Encuentre el puntaje z estándar que corresponda a lo siguiente:

a. Ochenta por ciento de la distribución está debajo de este valor (a la izquierda).

b. El área a la derecha de este valor es 0.15.

6.33 Encuentre los dos puntajes z que limitan el 50% central de una distribución normal.

6.34 Encuentre los dos puntajes estándar (z) que correspondan a lo siguiente:

a. El 90% central de una distribución normal está limitado por ellos.

b. El 98% central de una distribución normal está limitado por ellos.

6.35 a. Encuentre el puntaje z para el 80avo percentil de la distribución normal estándar.

b. Encuentre los puntajes z que limitan el 75% central de la distribución normal estándar.

6.36 a. Encuentre el puntaje z para el 33avo percentil de la distribución normal estándar.

b. Encuentre los puntajes z que limitan el 40% central de la distribución normal estándar.

6.37 Suponiendo una distribución normal, encuentre el puntaje z asociado con lo siguiente?:

a. El 90avo percentil

b. El 95avo percentil

c. El 99avo percentil

6.38 Suponiendo una distribución normal, ¿cuál es el puntaje z asociado con lo siguiente:

a. Primer cuartil

b. Segundo cuartil

c. Tercer cuartil

6.4 Aplicaciones de la distribución normal

En la sección 6.3 aprendimos a usar la tabla 3 del apéndice B para convertir información acerca de la variable normal estándar z en probabilidad y viceversa a convertir información de probabilidad acerca de la distribución normal estándar en puntajes z. Ahora estamos listos para aplicar esta metodología a todas las distribuciones normales. La clave es el puntaje estándar, z. La información asociada con una distribución normal será en términos de valores x o probabilidades. Usaremos el puntaje z y la tabla 3 como las herramientas para "pasar entre" la información dada y la respuesta deseada.

Recuerde que el puntaje estándar, z, se definió en el capítulo 2.

Puntaje estándar

En palabras: $$z = \frac{x - (\text{media de } x)}{\text{desviación estándar de } x}$$

En álgebra: $$z = \frac{x - \mu}{\sigma} \tag{6.3}$$

(Observe que cuando $x = \mu$, el puntaje estándar es $z = 0$.)

EJEMPLO 6.10 Conversión a una curva normal estándar para hallar probabilidades

Considere las puntuaciones de IQ. Éstas están normalmente distribuidas con una media de 100 y una desviación estándar de 16. Si al azar se selecciona una persona, ¿cuál es la probabilidad de que su IQ sea entre 100 y 115, es decir, cuál es $P(100 < x < 115)$?

SOLUCIÓN $P(100 < x < 115)$ está representada por el área sombreada de la figura.

La variable x debe ser estandarizada con la fórmula (6.3). Los valores z se muestran en la siguiente figura.

$$z = \frac{x - \mu}{\sigma}$$

$$\text{cuando } x = 100: \quad z = \frac{100 - 100}{16} = \mathbf{0.00}$$

$$\text{cuando } x = 115: \quad z = \frac{115 - 100}{16} = \mathbf{0.94}$$

Por tanto,

$$P(100 < x < 115) = P(0.00 < z < 0.94) = \mathbf{0.3264}$$

Así, la probabilidad es 0.3264 de que una persona seleccionada al azar tenga un IQ entre 100 y 115.

PARA SU INFORMACIÓN
El valor 0.3264 se halla usando la tabla 3 del apéndice B

EJEMPLO 6.11 Cálculo de la probabilidad bajo "cualquier" curva normal

Encuentre la probabilidad de que una persona seleccionada al azar tenga un IQ mayor a 90.

SOLUCIÓN

$$z = \frac{x - \mu}{\sigma} = \frac{90 - 100}{16} = \frac{-10}{16} = -0.625 = -0.63$$

$$P(x > 90) = P(z > -0.63)$$

$$= 0.2357 + 0.5000 = \mathbf{0.7357}$$

Así, la probabilidad es 0.7357 de que una persona seleccionada al azar tenga un IQ mayor a 90.

Es posible usar la tabla normal para contestar numerosas clases de preguntas que comprenden una distribución normal. Muchas veces un problema pedirá la ubicación de un "punto de corte", es decir, un valor particular de x tal que exactamente cierto porcentaje se encuentre en un área especificada. Los siguientes ejemplos exponen algunos de estos problemas.

EJEMPLO 6.12

Uso de la curva normal y *z* para determinar valores de datos

En un grupo de alumnos grande, supongamos que la profesora dice que es necesario que usted obtenga una puntuación en el 10% superior de su grupo, para obtener una A en un examen en particular. Por la experiencia de ella, puede estimar que la media y desviación estándar en este examen serán 72 y 13, respectivamente. ¿Cuál será la puntuación mínima necesaria para obtener una A? (Suponga que las puntuaciones estarán normalmente distribuidas en forma aproximada.)

SOLUCIÓN Empiece por convertir el 10% a información que sea compatible con la tabla 3 con la siguiente resta:

PARA SU INFORMACIÓN
¿Por qué se usa 0.5000?

$$10\% = 0.1000; \quad 0.5000 - 0.1000 = 0.4000$$

Busque en la tabla 3 para hallar el valor de z asociado con la entrada de área más cercana a 0.4000; es $z = 1.28$. Así,

$$P(z > 1.28) = 0.10$$

Ahora encuentre el valor de x que corresponda a $z = 1.28$ usando para ello la fórmula (6.3):

$$z = \frac{x - \mu}{\sigma}: \quad 1.28 = \frac{x - 72}{13}$$

$$x - 72 = (13)(1.28)$$

$$x = 72 + (13)(1.28) = 72 + 16.64 = 88.64, \text{ o } \mathbf{89}$$

Entonces, si usted recibe un 89 o mayor, puede esperar estar en el 10% superior (lo que significa una A).

EJEMPLO 6.13 Uso de la curva normal y z para determinar percentiles

Encuentre el 33avo percentil para puntuaciones de IQ ($\mu = 100$ y $\sigma = 16$ del ejemplo 6.10, p. 324).

SOLUCIÓN

z	...	0.04	...
⋮			
0.4	...	0.1700	...

$$P(0 < z < a) = 0.17$$

$$a = \mathbf{0.44} \quad \text{(valor de corte de } z \text{ de la tabla 3)}$$

33avo percentil de $z = -0.44$ (debajo de la media)

Ahora convertimos el 33avo percentil de los puntajes z, -0.44, a un puntaje x usando la fórmula (6.3):

$$z = \frac{x - \mu}{\sigma}: \quad -0.44 = \frac{x - 100}{16}$$

$$x - 100 = 16(-0.44)$$

$$x = 100 - 7.04 = \mathbf{92.96}$$

Entonces, 92.96 es el 33avo percentil para puntuaciones de IQ.

El ejemplo 6.14 expone una situación en la que se pide al lector hallar la media, μ, cuando se le da información relacionada.

EJEMPLO 6.14 Uso de la curva normal y *z* para determinar parámetros poblacionales

Los ingresos de ejecutivos subalternos en una gran empresa están normalmente distribuidos con una desviación estándar de \$1200. Un recorte está pendiente, en cuyo momento quienes ganan menos de \$28 000 serán despedidos. Si ese recorte representa 10% de los ejecutivos subalternos, ¿cuál es el salario medio actual del grupo de ejecutivos subalternos?

SOLUCIÓN Si 10% de los salarios es menor a \$28 000, entonces 40% (o 0.4000) es entre \$28,000 y la media, μ. La tabla 3 indica que $z = -1.28$ es la puntuación estándar que ocurre en $x = \$28\,000$.

Usando la fórmula (6.3) podemos hallar el valor de μ:

$$z = \frac{x - \mu}{\sigma}: \quad -1.28 = \frac{28{,}000 - \mu}{1{,}200}$$

$$-1{,}536 = 28{,}000 - \mu$$

$$\mu = 28{,}000 + 1{,}536 = \mathbf{\$29{,}536}$$

Esto es, el salario medio actual de ejecutivos subalternos es $29,536.

Con referencia a las puntuaciones de IQ nuevamente, ¿cuál es la probabilidad de que una persona seleccionada al azar tenga un IQ de 125: $P(x = 125)$? (Las puntuaciones de IQ están normalmente distribuidas con una media de 100 y una desviación estándar de 16.) Esta situación tiene dos interpretaciones: teórica y práctica. Veamos primero la teórica. Recuerde que la probabilidad asociada con un intervalo para una variable aleatoria continua está representada por el área bajo la curva; esto es, $P(a \leq x \leq b)$ es igual al área entre a y b bajo la curva. $P(x = 125)$ (es decir, x es exactamente 125) es entonces $P(125 \leq x \leq 125)$, o el área del segmento de recta vertical en $x = 125$. Esta área es cero. No obstante, éste no es el significado práctico de $x = 125$. En general quiere decir 125 al valor entero más cercano. Entonces, es muy probable que $P(x = 125)$ se interprete como

$$P(124.5 < x < 125.5)$$

El intervalo de 124.5 a 125.5 bajo la curva tiene un área mensurable y es diferente de cero. En situaciones de esta naturaleza, es necesario estar seguro de cuál significado está en uso.

Nota: una notación estándar que se emplea para abreviar "distribución normal con media μ y desviación estándar σ" es $N(\mu, \sigma)$. Esto es, $N(58, 7)$ representa "una distribución normal con media = 58 y desviación estándar = 7".

INSTRUCCIONES DE TECNOLOGÍA: GENERAR DATOS ALEATORIOS A PARTIR DE UNA DISTRIBUCIÓN NORMAL

MINITAB (Versión 14)

```
Choose:     Calc > Random Data > Normal
Enter:      Generate:  n  rows of data
            Store in column(s):  C1
            Mean:  μ
            Stand. dev.:  σ > OK
```

Si se desean muestras múltiples (12, por ejemplo), todas del mismo tamaño, modifique los comandos precedentes: guarde en columna(s): C1–C12.

Nota: para hallar estadísticas descriptivas para cada una de estas muestras, use los comandos: Stat > Basic Statistics > Display Descriptive Statistics for C1–C12.

Excel

```
Choose:     Tools > Data Analysis > Random Number Generation > OK
Enter:      Number of Variables: 1
            Number of Random Numbers: n
            Distribution: Normal
            Mean = : μ
            Standard Deviation = : σ
Select:     Output Options: Output Range
Enter:      (A1 or select cell) > OK
```

Si se desean muestras múltiples (12, por ejemplo), todas del mismo tamaño, modifique los comandos precedentes: Número de variables: 12.

Nota: para hallar estadísticas descriptivas para cada una de estas muestras, use los comandos: Tools > Data Analysis > Descriptive Statistics for columns A through L.

TI-83/84 Plus

```
Choose:     MATH > PRB > 6:randNorm(
Enter:      μ, σ, # of trials)
Choose:     STOO  > L1 > ENTER
```

Si se desean muestras múltiples (seis, por ejemplo), todas del mismo tamaño, repita seis veces los comandos precedentes y guarde en L1–L6.

Nota: para hallar estadísticas descriptivas para cada una de estas muestras, use los comandos: STAT > CALC > 1:1–VarStats for L1–L6.

INSTRUCCIONES DE TECNOLOGÍA: CÁLCULO DE VALORES DE LA ORDENADA (y) PARA UNA CURVA DE DISTRIBUCIÓN NORMAL

MINITAB (Versión 14) Introduzca las abscisas (valores de x) deseadas en C1; luego continúe con:

```
Choose:     Calc > Probability Distributions > Normal
Select:     Probability Density
Enter:      Mean: μ
            Stand. dev.: σ
            Input column: C1
            Optional Storage: C2 > OK
```

Para trazar la gráfica de una curva de probabilidad normal con los valores de x en C1 y los valores de y en C2, continúe con:

```
Choose:     Graph > Scatterplot
Select:     With Connect Line > OK
Enter:      Y variables: C2 X variables: C1 > OK
```

Excel Introduzca las abscisas (valores de x) en la columna A y active B1; luego continúe con:

```
Choose:     Insert function fx > Statistical > NORMDIST > OK
Enter:      X: (A1:A100 or select 'x value' cells)
```

```
            Mean: μ
            Standard dev.: σ
            Cumulative: False > OK
Drag:       Bottom right corner of the ordinate value box down to give
            other ordinates
```

Para trazar la gráfica de una curva de probabilidad normal con los valores de x en la columna A y los valores de y en la columna B, continúe con:

```
Choose:     Chart Wizard > XY(Scatter) > 1st picture > Next > Data Range
Enter:      Data range: (A1:B100 or select x & y cells)
Choose:     Next > Finish
```

TI-83/84 Plus Los valores de la ordenada se pueden calcular para valores individuales de la abscisa, x:

```
Choose:     2nd > DISTR > 1:normalpdf(
Enter:      x, μ, σ)
```

Para trazar la gráfica de una curva de probabilidad normal para una μ y σ, particulares, continúe con:

```
Choose:     WINDOW
Enter:      μ − 3σ, μ + 3σ, σ, − .05, 1, .1, 0)
Choose:     Y = > 2nd > DISTR > 1:normalpdf(
Enter:      x, μ, σ)
```

Después de una gráfica inicial, ajuste con 0:ZoomFit del menú ZOOM.

INSTRUCCIONES DE TECNOLOGÍA: PROBABILIDAD ACUMULATIVA PARA DISTRIBUCIONES NORMALES

MINITAB (Versión 14) Introduzca las abscisas deseadas (valores de x) en C1; luego continúe con:

```
Choose:     Calc > Probability Distributions > Normal
Select:     Cumulative probability
Enter:      Mean: μ
            Stand. dev.: σ
            Input column: C1
            Optional Storage: C3 > OK
```

Notas:

1. Para hallar la probabilidad entre dos valores de x, introduzca los dos valores en C1, use los comandos precedentes y reste usando los números de C3.
2. Para trazar una gráfica de la distribución de probabilidad acumulativa (ojiva), use los comandos Scatterplot de la página 328 con C3 como la variable y.

Excel Introduzca las abscisas deseadas (valores de x) en la columna A y active C1; luego continúe con:

```
Choose:     Insert function fx > Statistical > NORMDIST > OK
Enter:      X: (A1:A100 or select 'x value' cells)
            Mean: μ
            Standard dev.: σ
            Cumulative: True > OK
Drag:       Bottom right corner of the cumulative probability box down to
            give other cumulative probabilities
```

Notas:

1. Para hallar la probabilidad entre dos valores de x, introduzca los dos valores en la columna A, use los comandos precedentes y reste usando los números de la columna C.
2. Para trazar una gráfica de la distribución de probabilidad acumulativa (ojiva), use los comandos Chart Wizard de la página 329, eligiendo el subcomando Series con la columna C como los valores y y la columna A como los valores x.

TI-83/84 Plus Las propiedades acumulativas se pueden calcular para valores individuales de abscisa, x:

```
Choose:     2nd > DISTR > 2:normalcdf(
Enter:      −1 EE 99, x, μ, σ)
```

Notas:

1. Para hallar la probabilidad entre dos valores de x, introduzca los dos valores en lugar de –1 EE 99 y la x.
2. Para trazar una gráfica de la distribución de probabilidad acumulativa (ojiva), use el comando Scatter bajo STATPLOTS, con los valores de x y sus probabilidades acumulativas en un par de listas, o bien normalcdf(−IEE99, x, X, X) en el Y = editor.

CASO PRÁCTICO 6.15 Tapones de corcho

© Brand X Pictures/Getty Images

Es probable que usted conozca ese pequeño cilindro aparentemente insignificante de material compresible que es el tapón de botellas, pero, ¿sabe que el proceso por el cual el corcho sin elaborar se convierte en tapón no es nada sencillo? La industria del corcho tiene estándares muy altos, y hay leyes internacionales muy estrictas que lo abarcan todo desde la cosecha del corcho hasta la entrega de los corchos al usuario.

El corcho empieza como la corteza del árbol *Quercus suber*, que después de cortarse del árbol pasa por una serie de procesos de almacenamiento y cocción para estabilizar, limpiar y aumentar la elasticidad del corcho; a continuación se corta en tiras de las que se troquelan los corchos. A esto sigue una serie de operaciones de lavado, blanqueo, desinfección y aplicación de color, con inspección y clasificación permanentes. Los procesos de terminado incluyen inspecciones, recubrimientos, impresión, eliminación de humedad, tratamiento de superficie, esterilización, empaque y certificación de control de calidad.

Los corchos de tamaño 9 estándar miden 24 mm de diámetro por 1.75 pulgadas (45 mm) de largo. Algunas de sus características (y especificaciones empleadas) que deben pasar inspección son las siguientes:

- Defectos/fallas (por ejemplo, agujeros hechos por gusanos, grietas, poros, madera verde)
- Longitud (45.0 + 1.00 mm/−0.5 mm)
- Diámetro promedio (24 mm + 0.6 mm/−0.4 mm)
- Ovalidad (fuera de redondez, < 1.00 mm)
- Peso (gramos)
- Peso específico (g/cc)
- Humedad (requisito del cliente ± 1.5%)

- Peróxido residual (< 0.2 ppm)
- Fuerza de extracción (300 N + 100 N/−150 N)

La longitud es la variable que no es tan importante en la evaluación de corchos porque tiene poco qué ver con la efectividad de un corcho para preservar el vino. Se prefieren corchos largos sobre los cortos principalmente debido a su aspecto estético el fuerte sonido que se oye cuando se descorcha una botella es atractivo.

Algunas de las variables citadas líneas antes tienen distribuciones normales; otras no las tienen. Dos de ellas con distribuciones normales son el diámetro promedio del corcho y la fuerza de extracción. El diámetro de cada corcho se mide en varios lugares, y se informa de un diámetro promedio para el corcho. Tiene una distribución normal con media de 24.0 mm y desviación estándar de 0.13 mm. Una muestra de 250 corchos produjo el siguiente resumen.

Para obtener la fuerza de extracción, cada botella es llenada, se le pone el corcho y se deja reposar durante 24 horas. A continuación se coloca en una máquina que retira el corcho y registra la fuerza requerida para extraerlo de la botella. Esta fuerza tiene una distribución normal con una media de 310 newtons. (Un newton es una unidad de fuerza; 1 N = 1 kilogramo metro/s^2). Una muestra de 400 corchos produjo este resumen.

La ovalidad (la medida fuera de redondez) es la diferencia entre el diámetro máximo de un corcho y el diámetro mínimo. Como es de esperarse, la ovalidad no tiene una distribución normal; su valor más bajo posible es 0 y aumenta a partir de ese valor. Tiene una distribución agrupada pero sesgada a la derecha.

¿Qué clase de distribución anticipa usted para las variables de longitud, peso y peso específico?

Fuente: Cortesía de Gültig GmbH

SECCIÓN 6.4 EJERCICIOS

6.39 El siguiente ejercicio demuestra que la probabilidad es igual al área bajo la curva. Dado que estudiantes universitarios duermen un promedio de 7 horas por noche, con una desviación estándar de 1.7 horas, realice una simulación para determinar lo siguiente:

a. P(un estudiante duerme entre 5 y 9 horas)

b. P(un estudiante duerme menos de 4 horas)

c. P(un estudiante duerme entre 8 y 11 horas)

6.40 El siguiente ejercicio demuestra los efectos que la media y desviación estándar tienen en una curva normal.

a. Dejando la desviación estándar en 1, aumente la media a 3. ¿Qué le pasa a la curva?

b. Restablezca la media a 0 y aumente la desviación estándar a 2. ¿Qué le pasa a la curva?

c. Si pudiera usted reducir la desviación estándar a 0.5, ¿qué piensa que podría ocurrir a la curva normal?

6.41 Dada $x = 58$, $\mu = 43$, y $\sigma = 5.2$, encuentre z.

6.42 Dada $x = 237$, $\mu = 220$, y $\sigma = 12.3$, encuentre z.

6.43 Dado que x es una variable aleatoria normalmente distribuida con una media de 60 y una desviación estándar de 10, encuentre las siguientes probabilidades:

a. $P(x > 60)$ b. $P(60 < x < 72)$ c. $P(57 < x < 83)$

d. $P(65 < x < 82)$ e. $P(38 < x < 78)$ e. $P(x < 38)$

6.44 Dado que x es una variable aleatoria normalmente distribuida con una media de 28 y una desviación estándar de 7, encuentre las siguientes probabilidades:

a. $P(x > 28)$ b. $P(28 < x < 38)$ c. $P(24 < x < 40)$

d. $P(30 < x < 45)$ e. $P(19 < x < 35)$ e. $P(x < 48)$

6.45 Usando la información dada en el ejemplo 6.10 (p. 324):

a. Encuentre la probabilidad de que una persona seleccionada al azar tenga una puntuación de IQ entre 100 y 120.

b. Encuentre la probabilidad de que una persona seleccionada al azar tenga una puntuación de IQ arriba de 80.

6.46 Usando la información dada en el ejemplo 6.14 (p. 326):

a. Encuentre la probabilidad de que un ejecutivo subalterno seleccionado al azar tenga un salario entre \$27 000 y \$31 000.

b. Encuentre la probabilidad de que un ejecutivo subalterno seleccionado al azar tenga un salario mayor a \$33 000.

6.47 Dependiendo de dónde viva usted y de la calidad de la asistencia diaria, los costos de esta asistencia pueden variar de $3000 a $15 000 al año (o $250 a $1250 al mes) para un niño, según el Baby Center (Centro de Atención a Bebés). Los centros de asistencia en ciudades grandes como Nueva York y San Francisco son notablemente costosos.

Fuente: http://www.babycenter.com/refcap/baby/babychildcare/6056.html#04/16/2005

Suponga que los costos anuales de asistencia diaria están normalmente distribuidos con una media igual a $9000 y una desviación estándar igual a $1800.

a. ¿Qué porcentaje de centros de asistencia diaria costarán entre $7200 y $10 800 al año?
b. ¿Qué porcentaje de centros de asistencia diaria costarán entre $5400 y $12 600 al año?
c. ¿Qué porcentaje de centros de asistencia diaria costarán entre $3600 y $14 400 al año?
d. Compare los resultados de las partes a —c con la regla empírica. Explique la relación.

6.48 Según Wageweb.com (http://www.wageweb.com/hr1.htm), el salario promedio nacional a octubre de 2003 para un empleado de recursos humanos fue $29,932. Si suponemos que los salarios anuales para empleados están normalmente distribuidos con una desviación estándar de $1850, encuentre lo siguiente:

a. El porcentaje que ganan menos de $27 000
b. El porcentaje que ganan más de $32 000

6.49 Según las estadísticas de carreteras para el año 2003 de la Federal Highway Administration (http://www.fhwa.dot.gov), la distribución de edades para conductores con licencia tiene una media de 44.5 años y una desviación estándar de 17.1 años. Suponiendo que la distribución de edades está normalmente distribuida, ¿qué porcentaje de los conductores están

a. entre las edades de 17 y 22?
b. menores de 25 años?
c. mayores de 21 años?
d. entre 45 y 65 años?
e. mayores de 75 años?

6.50 Hay una nueva clase trabajadora con dinero para gastar según un artículo de *USA Today* titulado "Nuevos trabajadores jóvenes de 'cuello de oro' ganan influencia" (1 de marzo, 2005). "Cuello de oro" es un subconjunto de trabajadores de cuello azul (obreros) definido por investigadores como aquellos que trabajan en restaurantes de comida rápida y comercios de venta al menudeo o como guardias de seguridad, oficinistas o salones de belleza. Estos trabajadores de "cuello de oro" de entre 18 y 25 años tienen un promedio de gastos personales de $729 al mes (contra $267 para estudiantes universitarios y $609 para obreros). Suponiendo que este gasto está normalmente distribuido con una desviación estándar de $92.00, ¿qué porcentaje de trabajadores de cuello de oro tienen gastos personales de:

a. entre $600 y $900 al mes?
b. entre $400 y $1000 al mes?
c. más de $1050 al mes?
d. menos de $500 al mes?

6.51 La International Bottled Water Association dice que los estadounidenses toman en promedio 4.6 (8 onzas) porciones de agua al día (http://www.bottledwater.org/public/summary.htm). Suponiendo que el número de porciones de 8 onzas de agua está aproximada y normalmente distribuida con una desviación estándar de 1.4 porciones, ¿qué proporción de estadounidenses toman:

a. más de las 8 porciones recomendadas?
b. menos de la mitad de las 8 porciones recomendadas?

6.52 Según el American College Test (ACT), los resultados del examen ACT de 2004 encontraron que los estudiantes tenían una calificación media de lectura de 21.3, con una desviación estándar de 6.0. Suponiendo que las calificaciones están normalmente distribuidas:

a. Encuentre la probabilidad de que un estudiante seleccionado al azar tenga una calificación ACT de lectura menor a 20.
b. Encuentre la probabilidad de que un estudiante seleccionado al azar tenga una calificación ACT entre 18 y 24.
c. Encuentre la probabilidad de que un estudiante seleccionado al azar tenga una calificación ACT mayor a 30.
d. Encuentre el valor del 75avo percentil para calificaciones del ACT.

6.53 La máquina llenadora de una cervecería está ajustada para llenar botellas de cuarto de galón, con una media de 32.0 onzas de cerveza y una varianza de 0.003. Periódicamente, se comprueba una botella y se toma nota de la cantidad de cerveza.

a. Suponiendo que la cantidad de llenado está normalmente distribuida, ¿cuál es la probabilidad de que la siguiente botella verificada al azar contenga más de 32.02 onzas?

b. Digamos que usted compra 100 botellas de cuarto de galón de esta cerveza para una fiesta. ¿Cuántas botellas esperaría hallar que contengan más de 32.02 onzas de cerveza?

6.54 La fuerza de extracción, para quitar el tapón de corcho de una botella de vino, tiene una distribución normal con una media de 310 newtons y una desviación estándar de 36 newtons.

a. Las especificaciones para esta variable, dada en el caso práctico 6.15, fueron "300 N + 100 N/−150 N". Exprese estas especificaciones como intervalo.

b. ¿Qué porcentaje de los tapones se espera que caiga dentro de especificaciones?

c. ¿Qué porcentaje de los corchos probados tendrá una fuerza de extracción de más de 250 newtons?

d. ¿Qué porcentaje de los corchos probados tendrá una fuerza de extracción con variación no mayor a 50 newtons de 310?

6.55 El diámetro de cada corcho, como se describe en el caso práctico 6.15, se mide en varios lugares y se informa de un diámetro promedio para el corcho. El diámetro promedio tiene una distribución normal con una media de 24.0 mm y una desviación estándar de 0.13 mm.

a. Las especificaciones para esta variable, dadas en el Caso práctico 6.15, fueron "24 mm + 0.6 mm/−0.4 mm". Exprese estas especificaciones como intervalo.

b. ¿Qué porcentaje de los corchos se espera que caiga dentro de las especificaciones?

c. ¿Qué porcentaje de los corchos probados tendrá un diámetro promedio de más de 24.5 mm?

d. ¿Qué porcentaje de los corchos probados tendrá un diámetro promedio con variación no mayor a 0.35 mm de 24?

6.56 Usando la curva normal estándar y z:

a. Encuentre la calificación mínima necesaria para recibir una A si la profesora del ejemplo 6.12 (p. 325) dijo que el 15% de calificaciones más altas recibiría una A.

b. Encuentre el 25avo percentil para puntuaciones de IQ en el ejemplo 6.10 (p. 324).

c. Si 20% de los salarios del ejemplo 6.14 (p. 326) están debajo de $28 000, encuentre el salario medio actual.

6.57 Por lo general, y en forma aproximada, los promedios finales están normalmente distribuidos con una media de 72 y una desviación estándar de 12.5. Su profesora dice que el 8% más alto de la clase recibirá una A; el siguiente 20% una B, el siguiente 42% una C, el siguiente 18% una D y el 12% final una F.

a. ¿Qué promedio debe usted rebasar para obtener una A?

b. ¿Qué promedio debe usted rebasar para recibir una calificación mejor a una C?

c. ¿Qué promedio debe usted obtener para pasar el curso? (Necesitará una D o mejor.)

6.58 Se utiliza un radar para medir la velocidad de automóviles en una autopista durante el tránsito en horas de mayor movimiento. Las velocidades de automóviles individuales están normalmente distribuidas con una media de 62 millas por hora (mph).

a. Encuentre la desviación estándar para todas las velocidades, si 3% de los automóviles corren a más de 72 mph.

b. Usando la desviación estándar hallada en la parte a, encuentre el porcentaje de estos autos que corren a menos de 55 mph.

c. Usando la desviación estándar hallada en la parte a, encuentre el 95avo percentil para la variable "velocidad".

6.59 Los pesos de sandías maduras producidas en la granja de Mr. Smith están normalmente distribuidos con una desviación estándar de 2.8 libras. Encuentre el peso medio de las sandías maduras de Mr. Smith si sólo 3% pesan menos de 15 libras.

6.60 Una máquina llena recipientes con un peso medio de 16.0 onzas por recipiente. Si no más de 5% de los recipientes deben pesar menos de 15.8 onzas, ¿a qué debe ser igual la desviación estándar de los pesos? (Suponga normalidad.)

6.61 Se sabe que los tiempos "en espera", para quien hace llamadas a una empresa local de televisión por cable, están normalmente distribuidos con una desviación estándar de 1.3 minutos. Encuentre el tiempo promedio "en espera" de quien llama, si la compañía afirma que no más de 10% de quienes llaman esperan más de 6 minutos.

6.62 En un día determinado, el número de pies cuadrados de espacio de oficina disponible para renta en una pequeña ciudad es una variable aleatoria normalmente distribuida con una media de 750 000 pies cuadrados y una desviación estándar de 60 000 pies cuadrados. El número de pies cuadrados disponible en otra pequeña ciudad está normalmente distribuido con una media de 800 000 pies cuadrados y una desviación estándar de 60 000 pies cuadrados.

a. Trace, en la misma gráfica, la distribución de espacio de oficina que se pueda rentar para ambas ciudades.

b. ¿Cuál es la probabilidad de que el número de pies cuadrados disponible en la primera ciudad sea menor a 800 000?

c. ¿Cuál es la probabilidad de que el número de pies cuadrados disponible en la segunda ciudad sea menor a 750 000?

6.63 Los datos son los pesos netos (en gramos) para una muestra de 30 bolsas de M&M's. El peso neto anunciado es 47.9 gramos por bolsa.

46.22	46.72	46.94	47.61	47.67	47.70
47.98	48.28	48.33	48.45	48.49	48.72
48.74	48.95	48.98	49.16	49.40	49.69
49.79	49.80	49.80	50.01	50.23	50.40
50.43	50.97	51.53	51.68	51.71	52.06

Fuente: http://www.math.uah.edu/stat/, Christine Nickel and Jason York, ST 687 project, fall 1998

La FDA exige que (casi) toda bolsa contenga el peso anunciado; de otro modo, las violaciones (menos de 47.9 gramos por bolsa) causarán multas por mandato. (M&M's son elaborados y distribuidos por Mars Inc.)

a. ¿Qué porcentaje de las bolsas de la muestra están en violación?

b. Si el peso de todas las bolsas llenas está normalmente distribuido con un peso medio de 47.9 gramos, ¿qué porcentaje de las bolsas estará en violación?

c. Suponiendo que los pesos de las bolsas están normalmente distribuidos con una desviación estándar de 1.5 gramos, ¿qué valor medio dejaría 5% de los pesos debajo de 47.9 gramos?

d. Suponiendo que los pesos de las bolsas están normalmente distribuidos con una desviación estándar de 1.0 gramos, ¿qué valor medio dejaría 5% de los pesos debajo de 47.9 gramos?

e. Suponiendo que los pesos de las bolsas están normalmente distribuidos con una desviación estándar de 1.5 gramos, ¿qué valor medio dejaría 1% de los pesos debajo de 47.9 gramos?

f. ¿Por qué es importante para Mars mantener bajo el porcentaje de violaciones?

g. Es importante para Mars mantener el estándar de desviación tan pequeño como sea posible de modo que, a su vez, la media pueda ser tan pequeña como sea posible para mantener el peso neto. Explique la relación entre la desviación estándar y la media. Explique por qué esto es importante para Mars.

6.64 a. Genere una muestra aleatoria de 100 valores simulados desde una distribución normal con una media de 50 y una desviación estándar de 12.

b. Usando la muestra aleatoria de 100 valores simulados hallados en la parte a y los comandos de tecnología para calcular valores ordinarios de la página 328, encuentre los correspondientes 100 valores y para la curva de distribución normal, con una media de 50 y desviación estándar de 12.

c. Use los 100 pares ordenados hallados en la parte b y trace la curva para la distribución normal con una media de 50 y desviación estándar de 12. (Los comandos de tecnología están incluidos con los comandos de la parte b en las páginas 328-329.)

d. Usando los comandos de tecnología para probabilidad acumulativa de la página 329, encuentre la probabilidad de que un valor seleccionado al azar de una distribución normal, con una media de 50 y desviación estándar de 12, estará entre 55 y 65. Verifique sus resultados mediante el uso de la tabla 3 del apéndice B.

6.65 Use computadora o calculadora para hallar la probabilidad de que un valor de x seleccionado al azar de una distribución normal (media de 584.2 y desviación estándar de 37.3) tenga un valor que corresponda a lo siguiente:

a. Menos de 525

b. Entre 525 y 590

c. Al menos 590

d. Verifique los resultados de las partes a-c usando la tabla 3.

e. Explique cualesquiera diferencias que pueda hallar entre las respuestas de la parte d y las de las partes a-c.

MINITAB

Input 525 and 590 into C1; then continue with the cumulative probability commands on page 329, using 584.2 as μ, 37.3 as σ, and C2 as optional storage.

Excel

Input 525 and 590 into column A and activate the B1 cell; then continue with the cumulative probability commands on page 329, using 584.2 as μ and 37.3 as σ.

TI-83/84

Input 525 and 590 into L1; then continue with the cumulative probability commands on page 330 in L2, using 584.2 as μ and 37.3 as σ.

6.66 a. Use una computadora para generar su propia tabla abreviada de probabilidad normal estándar (una versión breve de la tabla 3). Use valores z de 0.0 a 5.0 en intervalos de 0.1.

b. ¿Cómo están relacionados los valores obtenidos con las entradas de la tabla 3? Haga el ajuste necesario y guarde los resultados en una columna.

c. Compare sus resultados de la parte b con la primera columna de la tabla 3. Comente sobre algunas diferencias que vea.

MINITAB (Versión 14)

a.

```
Choose:  Calc > Make Patterned Data > Simple Set of
         Numbers
Enter:   Store patterned data in: C1
From ?rst value: 0
To last value: 5
In steps of:  0.1 > OK
```

Continúe con los comandos de probabilidad acumulativa de la página 329, usando 0 como μ, 1 como σ, y C2 como memoria opcional.

b.

```
Choose:  Calc > Calculator
Enter:   Store result in variable: C3
         Expression: C2 - 0.5 > OK
Choose:  Data > Display Data
Enter:   Columns to display: C1 C3 > OK
```

Excel

a.

```
Choose:  Tools > Data Analysis > Random Number Genera-
         tion > OK
Enter:   Number of variables: 1
         Distribution: Patterned
         From: 0 to 5.0 in steps of 0.1
         Repeat each number: 1 times
Select:  Output Range
Enter:   (A1 or select cell)
```

Continúe con los comandos de probabilidad acumulativa de la página 329, activando la celda B1 y usando 0 como μ y 1 como σ.

b. Active la celda C1; luego continúe con:

```
Enter:   = B1 - 0.5 > Enter
Drag:    Bottom right corner of the C1 box down to give
         probabilities for the x values
```

6.67 Use una computadora para comparar una muestra aleatoria contra la población de la cual se tomó la muestra. Considere la población normal con media de 100 y desviación estándar de 16.

a. Haga una lista de valores de x de $\mu - 4\sigma$ a $\mu + 4\sigma$ en incrementos de medias desviaciones estándar y guárdelas en una columna.

b. Encuentre la ordenada (valor y) correspondiente a cada abscisa (valor x) para la curva de distribución normal para $N(100, 16)$ y guárdelas en una columna.

c. Grafique la curva de distribución de probabilidad normal para $N(100, 16)$.

d. Genere una muestra aleatoria de 100 valores simulados de la distribución $N(100, 16)$ y guárdelas en una columna.

e. Grafique el histograma de los 100 valores obtenidos en la parte d usando los números que aparecen en la lista de la parte a como fronteras de clase.

f. Calcule otras estadísticas descriptivas útiles de los 100 valores y compare los datos con la distribución esperada. Comente sobre las similitudes y las diferencias que vea.

MINITAB (Versión 14)

a. Use los comandos Make Patterned Data del ejercicio 6.66, sustituyendo el primer valor con 36, el último valor con 164, y los intervalos con 8.

b.

```
Choose:  Calc > Prob. Dist. > Normal
Select:  Probability density
Enter:   Mean:    100
         Stand. dev.: 16
         Input column: C1
         Optional Storage: C2 > OK
```

c. Use los comandos Scatterplot de la página 328 para los datos en C1 y C2.

d. Use los comandos Calculate RANDOM DATA de la página 327, sustituyendo *n* con 100, guarde con C3, media con 100 y desviación estándar con 16.

e. Use los comandos HISTOGRAM with Fits de la página 61 para los datos de C3. Para ajustar el histograma, seleccione las posiciones 36:148/8 Binning with cutpoint y cutpoint.

f. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de C3.

Excel

a. Use los comandos RANDOM NUMBER GENERATION Patterned Distribution del ejercicio 6.66, sustituyendo el primer valor con 36, el último valor con 172, y los intervalos con 8.

b. Active B1; luego continúe con:

```
Choose:  Insert function fx > Statistical > NORMDIST >
         OK
Enter:   X: (A1:A? or select 'x value' cells)
         Mean: 100
         Standard dev.: 16
         Cumulative: False > OK
Drag:    Bottom right corner of the ordinate value box
         down to give other ordinates
```

c. Use los comandos CHART WIZARD XY(Scatter) de la página 329 para los datos de las columnas A y B.

d. Active la celda C1; luego use los comandos Normal RANDOM NUMBER GENERATION de la página 328, sustituyendo con 100 el número de números aleatorios, media con 100, y desviación estándar con 16.

e. Use los comandos HISTOGRAM de la página 61 con la columna C como el rango de entrada y la columna A como el rango de depósito.

f. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de la columna C.

6.68 Use una computadora para comparar una muestra aleatoria contra la población de la cual se tomó la muestra. Considere la población normal con media de 75 y desviación estándar de 14. Conteste las preguntas a-f del ejercicio 6.67 usando $N(75, 14)$.

6.69 Suponga que se deben generar varias muestras aleatorias, todas del mismo tamaño, todas de la misma distribución de probabilidad normal. ¿Todas serán iguales? ¿Cómo difieren? ¿En cuánto difieren?

a. Use una computadora o calculadora para generar 10 muestras diferentes, todas de tamaño 100, todas de la distribución de probabilidad normal de media 200 y desviación estándar 25.

b. Trace histogramas de las 10 muestras usando las mismas fronteras de clase.

c. Calcule varias estadísticas descriptivas para las 10 muestras, separadamente.

d. Comente sobre las similitudes y las diferencias que vea.

MINITAB (Versión 14)

a. Use los comandos generate RANDOM DATA de la página 327, sustituyendo *n* con 100, guarde con C1–C10, media con 200, y desviación estándar con 25.

b. Use los comandos HISTOGRAM de la página 61 para los datos en C1–C10. Para ajustar el histograma, seleccione las posiciones 36:148/8 de Binning with cutpoint y cutpoint.

c. Use el comando DISPLAY DESCRIPTIVE STATISTICS de la página 98 para los datos en C1–C10.

Excel

a. Use los comandos Normal RANDOM NUMBER GENERATION de la página 328, sustituyendo el número de variables con 10, el número de números aleatorios con 100, media con 200, y desviación estándar con 25.

b. Use los comandos RANDOM NUMBER GENERATION Patterned Distribution del ejercicio 6.66, sustituyendo el primer valor con 100, el último valor con 300, los intervalos con 25, y el rango de salida con K1. Use los comandos HISTOGRAM de la página 61 para cada una de las columnas de la A a la J (rango de entrada) con la columna K como el rango de depósito.

c. Use los comandos DESCRIPTIVE STATISTICS de la página 98 para los datos de las columnas A a J.

TI–83/84 Plus

a. Use los comandos 6:randNorm de la página 328, sustituyendo la media con 200, la desviación estándar con 25, y el número de intentos con 100. Repita seis veces, usando L1–L6 para guardar.

b. Use los comandos HISTOGRAM de la página 62 para los datos en L1–L6, introduciendo valores WINDOW de 100, 300, 25, −10, 60, 10 y 1. Ajuste con ZoomStat.

c. Use el comando 1–Var Stats de la página 98 para los datos de L1–L6.

6.70 Genere 10 muestras aleatorias, cada una de tamaño 25, de una distribución normal con media 75 y desviación estándar 14. Conteste las preguntas de las partes b-d del ejercicio 6.69.

6.5 Notación

El puntaje z se utiliza en estadísticas en varias formas; no obstante, la relación entre el valor numérico de z y el área bajo la curva **distribución normal estándar** no cambia. Como z se usará con gran frecuencia, deseamos una notación conveniente para identificar la información necesaria. La convención de que usaremos un "nombre algebraico" para un puntaje z específico es $z(\alpha)$, donde representa el "área a la derecha" de la z que se menciona.

EJEMPLO 6.16 Interpretación visual de $z(\alpha)$

a. $z(0.05)$ (léase "z de 0.05") es el nombre algebraico para z tal que el área a la derecha y bajo la curva normal estándar es exactamente 0.05, como se ve en la figura 6.6.

FIGURA 6.6 Área asociada con $z(0.05)$

b. $z(0.60)$ (léase "z de 0.60") es el valor de z tal que 0.60 del área está a su derecha, como se ve en la figura 6.7.

FIGURA 6.7 Área asociada con $z(0.60)$

Ahora hallemos los valores numéricos de $z(0.05)$, $z(0.60)$ y $z(0.95)$.

EJEMPLO 6.17 Determinación que corresponde a valores z para $z(\alpha)$

a. Encuentre el valor numérico de $z(0.05)$.

SOLUCIÓN

Debemos convertir la información de área de la notación en información que podamos usar con la tabla 3 del apéndice B. Vea las áreas que se muestran en la figura 6.8.

FIGURA 6.8 Encuentre el valor de $z(0.05)$

Cuando vemos la tabla 3, buscamos un área tan cercana como sea posible a 0.4500.

z	...	0.04			0.05	...
⋮			↑			
1.6	...	0.4495	0.4500		0.4505	...
⋮						

Por tanto, $z(0.05) =$ **1.65.**

Nota: usaremos la z correspondiente al área más cercana en valor. Si el valor está exactamente a la mitad entre las entradas de la tabla, siempre use el valor más grande de z.

b. Encuentre el valor numérico de $z(0.60)$.

SOLUCIÓN El valor 0.60 está relacionado a la tabla 3 por el uso del área 0.1000, como se ve en el diagrama.

Los valores más cercanos en la tabla 3 son 0.0987 y 0.1026.

z	...	0.05			0.06	...
⋮			↑			
0.2	...	0.0987	0.1000		0.1026	...
⋮						

Por tanto, $z(0.60)$ está relacionado con 0.25. Como $z(0.60)$ está debajo de la media, concluimos que $z(0.60) =$ **−0.25.**

c. Encuentre $z(0.95)$.

SOLUCIÓN $z(0.95)$ está ubicada en el lado izquierdo de la distribución normal porque el área a la derecha es 0.95. El área en la cola a la izquierda contiene entonces el otro 0.05, como se ve en la figura 6.9.

FIGURA 6.9 Área asociada con $z(0.95)$

Debido a la naturaleza simétrica de la distribución normal, $z(0.95)$ es $-z(0.05)$, es decir, $z(0.05)$ con su signo cambiado. Entonces, $z(0.95) = \mathbf{-1.65}$.

En los siguientes capítulos usaremos con regularidad esta notación. Los valores de z que se usarán normalmente provienen de una de las siguientes situaciones: (1) el puntaje z tal que hay un área especificada en una cola de la distribución normal o (2) los puntajes z que limitan una proporción central especificada de la distribución normal. Cuando la proporción central de una distribución normal está especificada, podemos todavía usar la notación "área a la derecha" para identificar el puntaje z específico de que se trate.

EJEMPLO 6.18 Determinación de puntajes *z* para áreas acotadas

Encuentre los puntajes z que limitan el 0.95 central de la distribución normal.

SOLUCIÓN Dado 0.95 como el área central (vea la figura 6.10), las dos colas deben contener un total de 0.05. Por tanto, cada cola contiene $\frac{1}{2}$ de 0.05, o sea 0.025, como se ve en la figura 6.11.

FIGURA 6.10 Área asociada con el 0.95 central

FIGURA 6.11 Hallar puntajes *z* para 0.95 central

Para hallar $z(0.025)$ en la tabla 3, debemos determinar el área entre la media y $z(0.025)$. Es $-0.5000 = 0.250 = 0.4750$, como se muestra en la figura 6.12.

FIGURA 6.12 Hallar el valor de $z(0.025)$

La tabla 3 nos muestra:

z	...	0.06	...
⋮			
1.9		0.4750	...
⋮			

Por tanto, $z(0.025) = 1.96$ y $z(0.975) = -z(0.025) = 1.96$. El 0.95 central de la distribución normal está limitado por **−1.96** y **1.96.**

SECCIÓN 6.5 EJERCICIOS

6.71 Usando la notación $z(\alpha)$ (identifica el valor de α empleado dentro del paréntesis), mencione cada una de las variables z normales estándar que se muestran en los diagramas siguientes.

a.

b.

c.

d.

e.

f.

6.72 Usando la notación $z(\alpha)$ (identifica el valor de α empleado dentro del paréntesis), mencione cada una de las variables z normales estándar que se muestran en los diagramas siguientes.

a.

b.

c.

d.

e.

f.

6.73 Usando la notación $z(\alpha)$ (identifica el valor de empleado dentro del paréntesis), mencione cada una de las variables z normales estándar que se muestran en los diagramas siguientes.

6.74 Usando la notación $z(\alpha)$ (identifica el valor de empleado dentro del paréntesis), mencione cada una de las variables z normales estándar que se muestran en los diagramas siguientes.

6.75 Trace una figura de la curva normal estándar que muestre:

a. $z(0.15)$ b. $z(0.82)$

6.76 Trace una figura de la curva normal estándar que muestre:

a. $z(0.04)$ b. $z(0.94)$

6.77 Con frecuencia estamos interesados en hallar el valor de z que limita un área determinada en la cola derecha de la distribución normal, como se ve en la figura siguiente. La notación $z(\alpha)$ representa el valor de z tal que $P(z > z(\alpha)) = \alpha$.

Encuentre lo siguiente:

a. $z(0.025)$ b. $z(0.05)$ c. $z(0.01)$

6.78 Encuentre el valor de lo siguiente:

a. $z(0.15)$ b. $z(0.82)$

6.79 Encuentre el valor de lo siguiente:

a. $z(0.08)$ b. $z(0.92)$

6.80 Use la tabla 3 del apéndice B para hallar los siguientes valores de z.

a. $z(0.05)$ b. $z(0.01)$ c. $z(0.025)$

d. $z(0.975)$ e. $z(0.98)$

6.81 Complete las tablas siguientes de puntajes z. El área A dada en las tablas es el área a la derecha bajo la distribución normal de las figuras.

a. Puntajes z asociados con la cola derecha: dada el área A, encuentre $z(A)$.

A	0.10	0.05	0.025	0.02	0.01	0.005
z(A)						

b. Puntajes z asociados con la cola izquierda: dada el área B, encuentre $z(B)$:

B	0.995	0.99	0.98	0.975	0.95	0.90
z(B)						

6.82 a. Encuentre el área bajo la curva normal para z entre $z(0.95)$ y $z(0.025)$.

b. Encuentre $z(0.025) - z(0.95)$.

6.83 La notación z, $z(\alpha)$, combina dos conceptos relacionados entre sí —el puntaje z y el área a la derecha— en un símbolo matemático. Identifique la letra en cada uno de lo siguiente como puntaje z o área; luego, con ayuda de un diagrama, explique lo que el número dado y la letra dada representan en la curva normal estándar.

a. $z(A) = 0.10$ b. $z(0.10) = B$

c. $z(\mathrm{C}) = -0.05$ d. $-z(0.05) = D$

6.84 Entender la notación z, $z(\alpha)$, nos exige saber si tenemos un puntaje z o un área. Cada una de las expresiones siguientes usa la notación z en varias formas, algunas típicas y otras no típicas. Encuentre el valor pedido en cada uno de lo siguiente; luego, con ayuda de un diagrama, explique lo que representan sus respuestas.

a. $z(0.08)$

b. El área entre $z(0.98)$ y $z(0.02)$

c. $z(1.00 - 0.01)$

d. $z(0.025) - z(0.975)$

6.6 Aproximación normal de la binomial

En el capítulo 5 introdujimos la **distribución binomial**. Recuerde que la distribución binomial es una distribución de probabilidad de la variable aleatoria discreta x, el número de éxitos observados en n intentos independientes repetidos. A continuación veremos la forma en que las **probabilidades binomiales**, es decir, probabilidades asociadas con una distribución binomial, pueden ser razonablemente aproximadas mediante el uso de la distribución de probabilidad normal.

Veamos primero unas pocas distribuciones binomiales específicas. La figura 6.13 muestra las probabilidades de x para 0 a n para tres situaciones: $n = 4$, $n = 8$ y $n = 24$. Para cada una de estas distribuciones, la probabilidad de éxito para un intento es 0.5. Observe que cuando n se hace más grande, la distribución aparece más y más como la distribución normal.

FIGURA 6.13 Distribuciones binomiales

Para hacer la aproximación deseada, necesitamos tomar en cuenta una diferencia importante entre la distribución de probabilidad binomial y la normal. La variable aleatoria binomial es **discreta**, mientras que la variable aleatoria normal es **continua**. Recuerde que el capítulo 5 demostró que la probabilidad asignada a un valor particular de x debe demostrarse en un diagrama, por medio de un segmento de recta cuya longitud representa la probabilidad (como en la figura 6.13). El capítulo 5 sugería, no obstante, que también podemos usar un histograma en el que el área de cada barra es igual a la probabilidad de x.

Veamos la distribución de la variable binomial x, cuando $n = 14$ y $p = 0.5$. Las probabilidades para cada uno de los valores de x se pueden obtener de la tabla 2 del apéndice B. Esta distribución de x se muestra en la figura 6.14. Vemos la misma distribución en la figura 6.15 en forma de histograma.

FIGURA 6.14 La distribución de x cuando $n = 14$, $p = 0.5$

FIGURA 6.15 Histograma de la distribución de x cuando $n = 14$, $p = 0.5$

Examinemos $P(x = 4)$ para $n = 14$ y $p = 0.5$ para estudiar la técnica de aproximación. $P(x = 4)$ es igual a 0.061 (vea la tabla 2 del apéndice B), el área de la barra (rectángulo) arriba de $x = 4$ en la figura 6.16.

FIGURA 6.16 El área de la barra arriba de $x = 4$ es 0.061, para $B(n = 14, p = 0.5)$

El área del rectángulo es el producto de su ancho y altura. En este caso la altura es 0.061 y el ancho es 1.0, de modo que el área es 0.061. Veamos más de cerca el ancho. Para $x = 4$, la barra empieza en 3.5 y termina en 4.5, por lo cual estamos viendo un área limitada por $x = 3.5$ y $x = 4.5$. La adición y sustracción de 0.5 al valor de x por lo general se denomina **factor de corrección de continuidad**. Es nuestro método de convertir una variable discreta en una variable continua.

Ahora veamos la distribución normal relacionada con esta situación. Primero necesitaremos una distribución normal con una media y una desviación estándar iguales a las de la distribución binomial que estamos estudiando. Las fórmulas (5.7) y (5.8) nos dan estos valores:

$$\mu = np = (14)(0.5) = \mathbf{7.0}$$

$$\sigma = \sqrt{npq} = \sqrt{(14)(0.5)(0.5)} = \sqrt{3.5} = \mathbf{1.87}$$

La probabilidad de que $x = 4$ es aproximada por el área bajo la curva normal entre $x = 3.5$ y $x = 4.5$, como se ve en la figura 6.17. La figura 6.18 muestra toda la distribución de la variable binomial x con una distribución normal de la misma media y desviación estándar sobrepuestas. Nótese que las barras y las áreas de intervalo bajo la curva cubren casi la misma área.

FIGURA 6.17 Probabilidad de que $x = 4$ sea aproximada por el área sombreada

FIGURA 6.18 Distribución normal sobrepuesta a la distribución para la variable binomial x

La probabilidad de que x entre 3.5 y 4.5 bajo esta curva normal se encuentra usando la fórmula (6.3), tabla 3 y los métodos indicados en la sección 6.4:

$$z = \frac{x - \mu}{\sigma}: \qquad P(3.5 < x < 4.5) = P\left(\frac{3.5 - 7.0}{1.87} < z < \frac{4.5 - 7.0}{1.87}\right)$$

$$= P(-1.87 < z < -1.34)$$

$$= 0.4693 - 0.4099 = \mathbf{0.0594}$$

Como la probabilidad binomial de 0.061 y la probabilidad normal de 0.0594 están razonablemente cercanas, la distribución de probabilidad normal parece ser una aproximación razonable de la distribución binomial.

La aproximación normal de la distribución binomial también es útil para valores de p que no sean cercanos a 0.5. Las distribuciones de probabilidad binomial que se muestran en las figuras 6.19 y 6.20 sugieren que las probabilidades binomiales se puedan aproximar usando la distribución normal. Nótese que cuando n aumenta, la distribución binomial empieza a verse como la distribución normal. A medida que el valor de p se aleja de 0.5, se hace necesaria una n más grande para que la aproxi-

FIGURA 6.19 Distribuciones binomiales

FIGURA 6.20 Distribuciones binomiales

mación normal sea razonable. La siguiente *regla práctica* se usa generalmente como guía:

> **Regla:** la distribución normal proporciona una aproximación razonable a una distribución de probabilidad binomial donde los valores de np y $n(1 - p)$ sean iguales o mayores a 5.

Por ahora el lector puede estar pensando, "¿Y esto? Con sólo usar la tabla binomial hallaré las probabilidades directamente y me evitaré todo el trabajo extra". No obstante, considere por un momento la situación presentada en el ejemplo 6.19.

EJEMPLO 6.19 Resolución de un problema de probabilidad binomial con la distribución normal

Una falla mecánica no detectada ha causado que $\frac{1}{3}$ de la producción de 5000 percutores de un taller mecánico sea defectuoso. ¿Cuál es la probabilidad de que un inspector encuentre no más de 3 percutores defectuosos en una muestra aleatoria de 25?

SOLUCIÓN En este ejemplo de un experimento binomial, x es el número de piezas defectuosas halladas en la muestra, $n = 25$, y $p = P(\text{defectuosos}) = \frac{1}{3}$. Para contestar la pregunta usando la distribución binomial, necesitaremos usar la función de probabilidad binomial, fórmula (5.5):

$$P(x) = \binom{25}{x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{25-x} \qquad \text{para } x = 0, 1, 2, \ldots, 25$$

Debemos calcular los valores para $P(0)$, $P(1)$, $P(2)$ y $P(3)$, porque no aparecen en la tabla 2. Éste es un trabajo tedioso debido al tamaño del exponente. En situaciones como ésta, podemos usar el método de aproximación normal.

Busquemos ahora $P(x \leq 3)$ con el uso del método de aproximación normal. Primero necesitamos hallar la media y desviación estándar de x, fórmulas (5.7) y (5.8):

$$\mu = np = (25)\left(\frac{1}{3}\right) = \mathbf{8.333}$$

$$\sigma = \sqrt{npq} = \sqrt{(25)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)} = \sqrt{5.55556} = \mathbf{2.357}$$

Estos valores se muestran en la figura. El área de la región sombreada ($x < 35$) representa la probabilidad de que $x = 0, 1, 2$ o 3. Recuerde que $x = 3$, la variable binomial discreta, cubre el intervalo continuo de 2.5 a 3.5.

$$P(x \text{ no es mayor a } 3) = P(x \leq 3)(\text{para una variable discreta } x)$$
$$= P(x < 3.5) \quad (\text{para una variable continua } x)$$

$$z = \frac{x - \mu}{\sigma}: \qquad P(X < 3.5) = P\left(z < \frac{3.5 - 8.333}{2.357}\right) = P(z < -2.05)$$
$$= 0.5000 - 0.4798 = \mathbf{0.0202}$$

Por tanto, P(no más de tres piezas defectuosas) es aproximadamente 0.02.

SECCIÓN 6.6 EJERCICIOS

6.85 Encuentre los valores np y nq (recuerde: $q = 1 - p$) para un experimento binomial con $n = 100$ y $p = 0.02$. ¿Esta distribución binomial satisface la regla para aproximación normal? Explique.

6.86 ¿En cuál de las siguientes distribuciones binomiales proporciona la distribución normal una aproximación razonable? Use comandos de computadora para generar una gráfica de la distribución y comparar los resultados contra la "regla práctica". Exprese sus conclusiones.

a. $n = 10, p = 0.3$ b. $n = 100, p = 0.005$

c. $n = 500, p = 0.1$ d. $n = 50, p = 0.2$

MINITAB (Versión 14)

Inserte las n y p específicas según sea necesario en el siguiente procedimiento.

Use los comandos Make Patterned Data del ejercicio 6.66, sustituyendo el primer valor con 0, el último valor con n , y los intervalos con 1.

Use los comandos Binomial Probability Distribution de la página 292, usando C2 como memoria opcional.

Use los comandos Scatterplot Simple para los datos en C1 y C2. Seleccione Data View, Data Display, Project Lines para completar la gráfica.

Excel

Inserte las n y p específicas según sea necesario en el siguiente procedimiento.

Use los comandos RANDOM NUMBER GENERATION Patterned Distribution en el ejercicio 6.66, sustituyendo el primer valor con 0, el último valor con n, los intervalos con 1, y el rango de salida con A1.

Active la celda B1; luego use los comandos Binomial Probability Distribution de la página 292.

Use los comandos Chart Wizard Column para los datos en las columnas A y B. Seleccionando el subcomando Series, introduzca la columna B para los valores y y la columna A para las leyendas del eje (x) de categoría.

6.87 Para ver lo que ocurre cuando la aproximación normal se usa incorrectamente, considere la distribución binomial con $n = 15$ y $p = 0.05$. Como $np = 0.75$, la regla práctica ($np > 5$ y $nq > 5$) no se satisfa-

ce. Usando las tablas binomiales, encuentre la probabilidad de uno o menos éxitos y compare esto con la aproximación normal.

6.88 Encuentre la aproximación normal para la probabilidad binomial $P(x = 6)$, donde $n = 12$ y $p = 0.6$. Compare esto con el valor de $P(x = 6)$ obtenido de la tabla 2 en el apéndice B.

6.89 Encuentre la aproximación normal para la probabilidad binomial $P(x = 4, 5)$, donde $n = 14$ y $p = 0.5$. Compare esto con el valor de $P(x = 4, 5)$ obtenido de la tabla 2 en el apéndice B.

6.90 Encuentre la aproximación normal para la probabilidad binomial $P(x \leq 8)$, donde $n = 14$ y $p = 0.4$. Compare esto con el valor de $P(x \leq 8)$ obtenido de la tabla 2 en el apéndice B.

6.91 Encuentre la aproximación normal para la probabilidad binomial $P(x \geq 9)$, donde $n = 13$ y $p = 0.7$. Compare esto con el valor de $P(x \geq 9)$ obtenido de la tabla 2 en el apéndice B.

6.92 Con referencia al ejemplo 6.19 (p. 346):

a. Calcule $P(x \leq 3 \mid B(25, \frac{1}{3}))$. (*Sugerencia:* si usa computadora o calculadora, use los comandos de la página 292.)

b. ¿Qué tan buena fue la aproximación normal? Explique.

6.93 Un melanoma es la forma más grave de cáncer de la piel y está aumentando a un ritmo más alto que cualquier otro tipo de cáncer en Estados Unidos. Si se detecta en una etapa temprana, el porcentaje de supervivencia para pacientes es casi de 90% en Estados Unidos. ¿Cuál es la probabilidad de que 200 o más de algún grupo de 250 pacientes de etapa temprana sobrevivan al melanoma?

Fuente: http://www.channelonenews.com/articles/2002/05/31/skin.cancer/

6.94 Si 30% de todos los estudiantes que entran a cierta universidad abandonan sus estudios durante su primer año, o al final de éste, ¿cuál es la probabilidad de que más de 600 del alumnado de 1800 que entran este año abandonen sus estudios durante su primer año, o al final de éste?

6.95 Según la Federal Highway Administration, casi 50% de automovilistas en Estados Unidos son mujeres. Suponga que una muestra aleatoria de 50 automovilistas ha de seleccionarse para una encuesta.

a. ¿Cuál es la probabilidad de que no más de la mitad (25) de automovilistas sean mujeres?

b. ¿Cuál es la probabilidad de que al menos ¾ (38) de automovilistas sean mujeres?

6.96 De acuerdo con un estudio de febrero de 2005, terminado por Pew Internet & American Life Project, alrededor de 36 millones de estadounidenses, o 27% de usuarios de Internet, dicen que descargan archivos de música o de video. Suponiendo que el porcentaje es correcto, use la aproximación normal a la binomial para hallar la probabilidad de lo siguiente en un estudio de 2000 usuarios estadounidenses de Internet.

Fuente: http://www.pewinternet.org/PPF/r/153/report_display.asp

a. Al menos 500 descargan archivos de música o de video

b. Al menos 575 descargan archivos de música o de video

c. A lo sumo 520 descargan archivos de música o de video

d. A lo sumo 555 descargan archivos de música o de video

6.97 No todos los directores técnicos de la NBA que disfrutaron de largas carreras ganaron consistentemente campeonatos con los equipos que dirigieron. Por ejemplo, Bill Fitch, que dirigió durante 25 temporadas de baloncesto profesional después de iniciar su carrera como director en la Universidad de Minnesota, ganó 944 juegos pero perdió 1106 cuando trabajó con los Cavaliers, Celtics, Rockets, Nets y Clippers. Si usted fuera a seleccionar al azar los resultados finales de 60 partidos de récord históricos en los que Bill Fitch dirigió uno de los equipos, ¿cuál es la probabilidad de que menos de la mitad de ellos muestren que el equipo de él ganó? Para obtener su respuesta, use la aproximación normal a la distribución binomial.

Fuente: http://www.basketball–reference.com

6.98 Un sondeo halló que más de 6 de 10 votantes dicen que creen que Estados Unidos está listo para una presidente mujer. El sondeo fue efectuado en febrero de 2005 por el Siena College Research Institute y patrocinado por Hearst Newspapers. Suponiendo que la proporción es 0.6, ¿cuál es la probabilidad de que otro sondeo de 1125 votantes registrados efectuado al azar resulte en lo siguiente?:

a. Más de dos tercios piensan que Estados Unidos está listo para una mujer presidente.

b. Menos de 58% piensan que Estados Unidos está listo para una mujer presidente.

6.99 Según un informe de abril de 2005 de la Substance Abuse and Mental Health Services Administration (SAMHSA), 35% de las personas que reciben tratamiento empezaron a ingerir bebidas alcohólicas en gran cantidad entre 15 y 17 años de edad (http://www.jointogether.org). Use la aproximación normal a la distribución binomial para hallar la probabilidad de que en un sondeo de 1200 personas bajo tratamiento, entre 450 y 500 inclusive habrán empezado a ingerir bebidas alcohólicas en gran cantidad entre 15 y 17 años de edad.

a. Resuelva usando aproximación normal y la tabla 3 del apéndice B.

b. Resuelva usando computadora o calculadora y el método de aproximación normal.

c. Resuelva usando computadora o calculadora y la función de probabilidad binomial.

6.100 En 2003, de los casi 105 millones de asalariados y trabajadores en Estados Unidos, alrededor de 15.4 millones estaban afiliados a un sindicato. Use la aproximación normal a la distribución binomial para hallar la probabilidad de que en un estudio nacional de 2500 trabajadores, al menos 400 serán miembros de un sindicato.

a. Resuelva usando aproximación normal y la tabla 3 del apéndice B.

b. Resuelva usando computadora o calculadora y el método de aproximación normal.

REPASO DEL CAPÍTULO

En retrospectiva

Hemos estudiado la distribución de probabilidad normal estándar, la familia más importante de variables aleatorias continuas. Hemos aprendido a aplicarla a todas las otras distribuciones de probabilidad normal y cómo usarla para estimar probabilidades de distribuciones binomiales. Hemos visto una amplia diversidad de variables que tienen esta distribución normal o que son aproximados razonablemente bien por ella.

En el siguiente capítulo examinaremos distribuciones muestrales y aprenderemos a usar la probabilidad normal estándar para resolver aplicaciones adicionales.

Vocabulario y conceptos clave

aproximación normal de la binomial (p. 343)

continuidad (p. 344)

curva en forma de campana (p. 315)

curva normal (p. 316)

distribución binomial (p. 343)

distribución normal (p. 315)

distribución normal estándar (pp. 316, 323, 338)

factor de corrección de

porcentaje (p. 316)

probabilidad (p. 316)

probabilidad (p. 316)

proporción (p. 316)

puntaje estándar (pp. 316, 323)

puntaje z (pp. 316, 323)

representación de área para

variable aleatoria (p. 315)

variable aleatoria continua (pp. 315, 344)

variable aleatoria discreta (pp. 315, 344)

Objetivos de aprendizaje

✓ Entender la diferencia entre una variable discreta y una continua.	p. 315
✓ Entender la relación entre la regla empírica y la curva normal.	p. 313-314, Ejer. 6.1
✓ Entender que una curva normal es una curva en forma de campana, con área total bajo la curva igual a 1.	pp. 315-316, EJ. 6.1, Ejer. 6.40
✓ Entender que la curva normal es simétrica alrededor de la media, con un área de 0.5000 en cada lado de la media.	pp. 315-317, EJ. 6-2
✓ Ser capaz de trazar una curva normal, aplicando leyenda a la media y diversos puntajes z.	p. 314
✓ Entender y ser capaz de usar la tabla 3, áreas de la distribución normal estándar, en el apéndice B.	EJ. 6.1-6.7
✓ Calcular probabilidades para intervalos definidos en la distribución normal estándar.	Ejer. 6.7, 6.15, 6.23
✓ Determinar valores z para intervalos correspondientes en la distribución normal estándar.	EJ. 6.8, 6.9, Ejer. 6.27 6.31, 6.35, 6.103
✓ Calcular, describir e interpretar un valor z para un valor de datos de una distribución normal.	EJ. 6.10, 6.11, Ejer. 6.43
✓ Calcular puntajes z y probabilidades para aplicaciones de la distribución normal.	Ejer. 6.47, 6.49, 6.59
✓ Trazar, calcular e interpretar z de notación alfa, $z(\alpha)$.	EJ. 6.16, 6.17, Ejer. 6.75, 6.79, 6.81
✓ Entender los elementos clave de un experimento binomial: x, n, p, q. Saber su media y fórmulas de desviación estándar.	pp. 343-344
✓ Entender que la distribución normal se puede usar para calcular probabilidades binomiales siempre que se satisfagan ciertas condiciones.	pp. 344-345, Ejer. 6.85
✓ Entender y ser capaz de usar el factor de corrección de continuidad cuando se calculen puntajes z.	p. 344, Ejer. 6.89, 6.91
✓ Calcular puntajes z y probabilidades para aproximaciones normales a la binomial.	EJ. 6.19, Ejer. 6.93, 6.129

Ejercicios del capítulo

6.101 Según el teorema de Chebyshev, ¿al menos cuánta área hay bajo la distribución normal estándar entre $z = -2$ y $z = +2$? ¿Cuál es el área real bajo la distribución normal estándar entre $z = -2$ y $z = +2$?

6.102 ¿El 60% central de la población normalmente distribuida está entre cuáles dos puntajes estándar?

6.103 Encuentre el puntaje estándar (z) tal que el área arriba de la media y debajo de z bajo la curva normal es:

a. 0.3962 b. 0.4846 c. 0.3712

6.104 Encuentre el puntaje estándar (z) tal que el área bajo la media y arriba de z bajo la curva normal es:

a. 0.3212 b. 0.4788 c. 0.2700

6.105 Dado que z es la variable normal estándar, encuentre el valor de k tal que:

a. $P(|z| > 1.68) = k$ b. $P(|z| > 2.15) = k$

6.106 Dado que z es la variable normal estándar, encuentre el valor de c tal que:

a. $P(|z| > c) = 0.0384$ b. $P(|z| < c) = 0.8740$

6.107 Encuentre los siguientes valores de z

a. $z(0.12)$ b. $z(0.28)$ c. $z(0.85)$ d. $z(0.99)$

6.108 Encuentre el área bajo la curva normal que se encuentra entre los siguientes pares de valores z:

a. $z = -3.00$ y $z = 3.00$

b. $z(0.975)$ y $z(0.025)$

c. $z(0.10)$ y $z(0.01)$

6.109 Con base en los datos del American College Test (ACT) en 2004, el promedio de calificaciones del examen de razonamiento en ciencias fue 20.9, con una desviación estándar de 4.6. Suponiendo que las calificaciones están normalmente distribuidas:

a. Encuentre la probabilidad de que un estudiante seleccionado al azar tenga una calificación ACT de al menos 25 en razonamiento de ciencias.

b. Encuentre la probabilidad de que un estudiante seleccionado al azar tenga una calificación ACT entre 20 y 26 en razonamiento de ciencias.

c. Encuentre la probabilidad de que un estudiante seleccionado al azar tenga una calificación ACT de menos de 16 en razonamiento de ciencias.

6.110 El récord de 70 años de largo plazo del clima muestra que, para el estado de Nueva York, la precipitación anual tiene una media de 39.67 pulgadas y una desviación estándar de 4.38 pulgadas.

Fuente: Department of Commerce; State, Regional and National Monthly Precipitation Report

Si la cantidad de precipitación anual tiene una distribución normal, cuál es la probabilidad de que el año próximo la precipitación total para el estado de Nueva York sea:

a. Más de 50.0 pulgadas

b. Entre 42.0 y 48.0 pulgadas

c. Entre 30.0 y 37.5 pulgadas

d. Más de 35.0 pulgadas

e. Menos de 45.0 pulgadas

f. Menos de 32.0 pulgadas

6.111 American Express cobra a comerciantes tarifas más altas que otras tarjetas de crédito o de débito, según el artículo de *USA Today* "American Express fees take flak" (Las tarifas de American Express reciben duras críticas) (23 de diciembre, 2004). La compañía piensa que pueden hacer esto porque dicen que los clientes que usan la tarjeta American Express gastan más. El promedio de cargos anuales por tarjeta en 2003 fue \$9600 dólares de acuerdo con datos de American Express y *The Neilson Report.* Suponiendo que los cargos anuales por tarjeta están normalmente distribuidos en forma aproximada con una desviación estándar de \$2100 dólares, ¿cuál es la probabilidad de que los cargos anuales a un cliente de American Express sean:

a. Menos de \$4000

b. Entre \$5000 y \$10 000

c. Mayores a \$16 000

6.112 Una compañía que produce remaches usados por fabricantes de aviones comerciales sabe que la resistencia al corte (fuerza requerida para romper) de sus remaches es de la mayor preocupación. La compañía piensa que la resistencia al corte de sus remaches está normalmente distribuida con una media de 925 libras y una desviación estándar de 18 libras.

a. Si la compañía tiene razón, ¿qué porcentaje de sus remaches tienen una resistencia al corte mayor a 900 libras?

b. ¿Cuál es el límite superior para la resistencia al corte del 1% más débil de los remaches?

c. Si un remache se selecciona al azar de todos los remaches, ¿cuál es la probabilidad de que requiera una fuerza de al menos 920 libras para romperlo?

d. Usando la probabilidad hallada en la parte c, redondeada al décimo más cercano, ¿cuál es la probabilidad de que 3 remaches de una muestra aleatoria de 10 se rompan a una fuerza menor a 920 libras?

6.113 En un estudio de la duración de juegos de béisbol de las ligas mayores a principios de la temporada de 2005, la variable "tiempo de juego" parece estar normalmente distribuida con una media de 2 horas, 50.1 minutos y una desviación estándar de 20.99 minutos.

Fuente: MLB.com

a. Algunos aficionados describen un juego como "inimaginablemente largo" si tarda más de 3 horas. ¿Cuál es la probabilidad de que un juego identificado al azar fuera inimaginablemente largo?

b. Numerosos aficionados describen un juego que dura menos de 2 horas, 30 minutos como "rápido". ¿Cuál es la probabilidad de que un juego seleccionado al azar sea rápido?

c. ¿Cuáles son los límites del rango intercuartil para la variable "tiempo de juego"?

d. ¿Cuáles son los límites del 90% central de la variable "tiempo de juego"?

6.114 Cierto tipo de refrigerador tiene una vida útil de servicio que está distribuida normalmente en forma aproximada, con una media de 4.8 años y una desviación estándar de 1.3 años.

a. Si este aparato electrodoméstico está garantizado por dos años, ¿cuál es la probabilidad de que el aparato que compró usted requiera reemplazarlo bajo la garantía?

b. ¿Cuál es el tiempo que debe dar el fabricante como garantía si está dispuesto a reemplazar sólo 0.5% de los refrigeradores?

6.115 Una máquina está programada para llenar recipientes de 10 onzas de un limpiador. No obstante, la variabilidad inherente en cualquier máquina hace que varíen las cantidades reales de llenado. La distribución es normal con una desviación estándar de 0.02 onzas. ¿Cuál debe ser la cantidad media para que sólo 5% de los recipientes reciba menos de 10 onzas?

6.116 En un gran complejo industrial, el departamento de mantenimiento ha recibido orden de reemplazar bombillas eléctricas antes que se fundan. Se sabe que la duración de las bombillas está normalmente distribuida con una vida media de 900 horas de uso y una desviación estándar de 75 horas. ¿Cuándo deben cambiarse las bombillas para que no más de 10% de ellas se queme cuando está en uso?

6.117 Las calificaciones en un examen cuya media es 525 y desviación estándar de 80 están normalmente distribuidas.

a. Cualquiera que obtenga calificación debajo de 350 tendrá que pasar de nuevo el examen. ¿Qué porcentaje representa esto?

b. El 12% más alto en calificaciones recibirá una mención especial. ¿Qué calificación debe rebasarse para recibir esta mención especial?

c. El rango de intercuartil de una distribución es la diferencia entre Q_1 y Q_3 (esto es, $Q_3 - Q_1$). Encuentre el rango de intercuartil para las calificaciones en este examen.

d. Encuentre la calificación tal que sólo 1 de 500 obtendrá una más alta.

6.118 Una máquina expendedora de bebidas gaseosas puede ser controlada para asegurar que dosifique un promedio de onzas de líquido por vaso.

a. Si la cantidad de onzas dosificadas por vaso están normalmente distribuidas con una desviación estándar de 0.2 onzas, encuentre el ajuste para μ que permitirá que un vaso de 6 onzas contenga (sin rebosarse) la cantidad dosificada el 99% del tiempo.

b. Use computadora o calculadora para simular sacar una muestra de 40 vasos de bebida gaseosa de la máquina (haga el ajuste usando su respuesta a la parte a).

MINITAB (Versión 14)

Use los comandos Calculate RANDOM DATA de la página 327, sustituyendo n con 40, guarde con C1, media con el valor calculado en la parte a, y desviación estándar con 0.2.

Use los comandos HISTOGRAM de la página 61 para los datos en C1. Para ajustar el histograma, seleccione las posiciones 5:6.2/0.05 Binning with cutpoint y cutpoint.

Excel

Use los comandos Normal RANDOM NUMBER GENERATION de la página 328, sustituyendo n con 40, la media con el valor calculado en la parte a, la desviación estándar con 0.2, y el rango de salida con A1.

Use los comandos RANDOM NUMBER GENERATION Patterned Distribution de la página 336, sustituyendo el primer valor con 5, el último valor con 6.2, los intervalos con 0.05, y el rango de salida con B1.

Use los comandos HISTOGRAM de la página 61 con la columna A como el rango de entrada y la columna B como el rango de depósito.

TI–83/84 Plus

Use los comandos 6:randNorm de la página 328, sustituyendo la media con el valor calculado en la parte a, la desviación estándar con 0.2, y el número de intentos con 40. Guarde con L1.

Use los comandos HISTOGRAM de la página 62 para los datos en L1, introduciendo los siguientes WINDOW VALUES: 5, 6.2, 0.05, −1, 10, 1, 1.

c. ¿Qué porcentaje de su muestra hubiera rebosado la taza?

d. La muestra de usted, ¿parece indicar que el ajuste para μ está funcionando bien? Explique.

PARA SU INFORMACIÓN Repita la parte b unas cuantas veces. Intente con un valor diferente para la cantidad media dosificada y repita la parte b. Observe cuántas se hubieran rebosado en cada ajuste de 40.

6.119 Suponga que x tiene una distribución binomial con $n = 25$ y $p = 0.3$.

a. Explique por qué la aproximación normal es razonable.

b. Encuentre la media y desviación estándar de la distribución normal que se use en la aproximación.

6.120 Sea x una variable aleatoria binomial para $n = 30$ y $p = 0.1$.

a. Explique por qué la aproximación normal no es razonable.

b. Encuentre la función empleada para calcular la probabilidad de cualquier x desde $x = 0$ hasta $x = 30$.

c. Use computadora o calculadora para hacer una lista de la distribución de probabilidad.

6.121 a. Use computadora o calculadora para hacer una lista de probabilidades binomiales para la distribución donde $n = 50$ y $p = 0.1$.

b. Use los resultados de la parte a y encuentre $P(x \leq 6)$.

c. Encuentre la aproximación normal para $P(x \leq 6)$, y compare los resultados con los de la parte b.

6.122 a. Use computadora o calculadora para hacer una lista de la distribución de probabilidad y la distribución de probabilidad acumulativa, para el experimento de probabilidad binomial con $n = 40$ y $p = 0.4$.

b. Explique la relación entre las dos distribuciones halladas en la parte a.

c. Si usted pudiera usar sólo una de estas listas cuando resuelva problemas, ¿cuál preferiría y por qué?

6.123 Considere el experimento binomial con $n = 300$ y $p = 0.2$.

a. Establezca, pero no evalúe, la expresión de probabilidad para 75 o menos éxitos en los 300 intentos.

b. Use computadora o calculadora para hallar $P(x \leq 75)$ usando la función de probabilidad binomial.

c. Use computadora o calculadora para hallar $P(x \leq 75)$ usando la aproximación normal.

d. Compare las respuestas de las partes b y c.

PARA SU INFORMACIÓN Use los comandos de probabilidad acumulativa.

6.124 Se sabe que una máquina para calificar exámenes registra una calificación incorrecta en 5% de los exámenes que califica. Use el método apropiado para hallar la probabilidad de que la máquina registre lo siguiente:

a. Exactamente 3 calificaciones incorrectas en un conjunto de 5 exámenes.

b. No más de 3 calificaciones incorrectas en un conjunto de 5 exámenes.

c. No más de 3 calificaciones incorrectas en un conjunto de 15 exámenes.

d. No más de 3 calificaciones incorrectas en un conjunto de 150 exámenes.

6.125 Una compañía dice que 80% de los clientes que compran su podadora especial no solicitarán repararla durante los primeros 2 años de haberla comprado. El estudio personal de usted ha demostrado que sólo 70 de las 100 en su muestra duraron los 2 años sin incurrir en gastos de reparaciones. ¿Cuál es la probabilidad de que el resultado de su muestra sea menor si el porcentaje real de reparaciones sin costo es 80%?

6.126 Se cree que 58% de parejas con hijos están de acuerdo en los métodos para disciplinar a sus hijos. Suponiendo que éste sea el caso, ¿cuál es la probabilidad de que en una encuesta efectuada a 200 parejas casadas seleccionadas al azar encontremos esto?

a. Exactamente 110 parejas que están de acuerdo

b. Menos de 110 parejas que están de acuerdo

c. Más de 110 parejas que están de acuerdo

6.127 En una encuesta realizada en febrero de 2005, dirigida por Salary.com, los bomberos arrasaron con la competencia y ganaron el título del "trabajo más sexy", con 16% de los votos. Supongamos que usted selecciona al azar 50 adultos. Use la aproximación normal a la distribución binomial para hallar la probabilidad de que de su colección:

Fuente: http://sdalary.com/careers/layoutscripts/cre1_display.asp?tab = cre&cat = nocat&ser = Ser348&part = Par516

a. Más de 12 de los adultos escogen ser bombero como el trabajo más sexy.

b. Menos de 8 de los adultos escogen ser bombero como el trabajo más sexy.

c. De 7 a 14 de los adultos escogen ser bombero como el trabajo más sexy.

6.128 La encuesta de 2004 de Pew Internet & American Life Project reveló que 4 de 10 estadounidenses que están en línea, alrededor de 53 millones de adultos, usan software de mensajes instantáneos (IM).

Fuente: http://www.pewinternet.org/PPF/r/133/report_display.asp

Use la aproximación normal a la binomial para hallar la probabilidad de que en una muestra aleatoria de 100 usuarios de Internet, no más de 50 usan programas IM.

6.129 La National Coffee Drinking Trends es "la publicación" en la industria del café. Durante más de cinco décadas, ha rastreado los patrones de consumo anual en una amplia variedad de situaciones y categorías. La edición de 2004 dice que 39% del total de quienes toman café, de 18 años de edad o más, en 2004 compraron café que creció a la sombra.

Fuente: http://fwww.ncausa.org/public/pages/index.cfm?pageid = 38

Si este porcentaje es verdadero para quienes toman café en la cafetería de Crimson Light's, ¿cuál es la probabilidad de lo siguiente para los próximos 50 clientes que compran café en Crimson Light's?:

a. Más de 20 han comprado una variedad que crece a la sombra.

b. Menos de 15 han comprado una variedad que crece a la sombra.

6.130 Aparentemente, pasar el tiempo en juegos de video, ver TV y enviar mensajes instantáneos a amigos no es suficientemente relajante. En un sondeo realizado en febrero de 2005 de Yesawich, Pepperdine, Brown and Russell encontraron que un tercio de los niños encuestados dijo que por Internet ellos ayudaron a investigar algún aspecto de las vacaciones de su familia. Si se toma un estudio de seguimiento de 100 de estos niños, ¿cuál es la probabilidad de lo siguiente?

a. Menos de 25% de la nueva muestra dirá que por Internet ellos ayudaron a investigar las vacaciones de la familia.

b. Más de 40% de la nueva muestra dirá que por Internet ellos ayudaron a investigar las vacaciones de la familia.

6.131 La fuerza laboral civil en Estados Unidos compuesta por 148 157 000 trabajadores estaba en un 94.8% empleada en marzo de 2005. Si se toma una muestra aleatoria de 2500 de la fuerza laboral civil, ¿cuál es la probabilidad de lo siguiente?

a. Más de 6% de la muestra estará desempleada.

b. Menos de 5% de la muestra estará desempleada.

6.132 Durante los primeros dos meses de 2005, hubo 1 140 256 vuelos comerciales que entraron y salieron de aeropuertos de Estados Unidos. De éstos, 74.35% fueron llegadas a tiempo y 18.96% fueron salidas con demora. Se han de identificar 300 vuelos al azar de todos estos vuelos, y se examinan cuidadosamente sus bitácoras. ¿Cuál es la probabilidad de lo siguiente?:

a. Más de 80% de la muestra será de una llegada a tiempo.

b. Menos de 15% de la muestra tendrá una salida con demora.

6.133 Es frecuente que los porcentajes de mortalidad infantil se usen para evaluar la calidad de vida y lo adecuado de la atención a la salud. El porcentaje está basado en el número de fallecimientos de niños menores de un año en un año determinado, por 1000 nacimientos de niños vivos en el mismo año. A continuación aparecen los porcentajes de mortalidad infantil, al entero más cercano, para ocho naciones en todo el mundo, como se ve en *The World Factbook*, 2004.

Nación	Mortalidad infantil (por 1000 nacimientos)	Nación	Mortalidad infantil (por 1000 nacimientos)
China	25	México	22
Alemania	4	Russia	17
India	58	S. África	62
Japón	3	Estados Unidos	7

Fuente: http://www.cia.gov/cia/publications/factbook/docs/notesanddefs.html

Suponga que se da seguimiento a los siguientes 2000 nacimientos dentro de cada nación para detectar fallecimientos de infantes.

a. Construya una tabla que muestre la media y desviación estándar de las distribuciones binomiales asociadas.

b. En la columna final de la tabla, encuentre la probabilidad de que al menos 70 infantes de las muestras, dentro de cada nación, sean fallecimientos que contribuyan al porcentaje de mortalidad de la nación. Muestre todo el trabajo.

c. Explique qué hizo variar tanto las respuestas.

6.134 Se seleccionó una gran muestra de un producto de la competencia y se evaluó en cuanto a una dimensión particular de lentes. Luego se comparó con su rango de especificación nominal (0.000) ± 0.030 de unidad. Se evaluó un total de 110 lentes. Los datos se codificaron en dos formas y aparecen a continuación:

−0.020	−0.043	−0.002	0.002	−0.018
−0.016	−0.051	0.024	−0.024	−0.032

*** Los datos para este ejercicio están en http//: latinoamerica.cengage.com/johnson

Fuente: Cortesía de Bausch & Lomb (la variable no se menciona y los datos se codifican a solicitud de B&L).

a. Calcule la media y desviación estándar de los datos.

b. Cree un histograma y la comente sobre el patrón de variabilidad de los datos.

c. Use pruebas para normalidad y/o la regla empírica como confirmación del aspecto normal. Explique lo que encuentre.

d. Determine el porcentaje observado apegado a especificación. Esto es, ¿qué porcentaje de las medidas cayeron dentro del rango de especificación de 0.000 ± 0.030 de unidad?

6.135 Suponga que la distribución de los datos del ejercicio 6.134 estaba distribuida normalmente en forma exacta con la media de 0.00 y desviación estándar de 0.020.

a. Encuentre los límites del 95% central de la distribución.

b. ¿Qué porcentaje de los datos en realidad está dentro del intervalo hallado en la parte a?

c. Usando puntajes z, determine el porcentaje de conformidad estimada a especificación. Esto es, ¿qué porcentaje de las medidas se esperaría cayeran dentro del rango de especificación de 0.000 ± 0.030 de unidad?

6.136 La siguiente distribución triangular da una aproximación a la distribución normal. El segmento de recta l_1 tiene la ecuación $y = x/9 + 1/3$, y el segmento l_2 tiene la ecuación $y = -x/9 + 1/3$.

a. Encuentre el área bajo toda la distribución triangular.

b. Encuentre el área bajo la distribución triangular entre 0 y 2.

c. Encuentre el área bajo la distribución normal estándar entre 0 y 2.

d. Discuta la efectividad de esta aproximación "triangular".

Proyecto del capítulo

Medición de la inteligencia

Todas las distribuciones de probabilidad normales tienen la misma forma y distribución respecto a la media y desviación estándar. En este capítulo aprendimos a usar la distribución de probabilidad normal estándar para contestar preguntas acerca de todas las distribuciones normales. Regresemos a la distribución de puntuaciones de IQ que estudiamos en la sección 6.1, "Medición de la inteligencia" (p. 313), y pongamos a prueba nuestro nuevo conocimiento.

Trabajando en el contenido del capítulo 6

6.137 Veamos de nuevo las puntuaciones de IQ normalmente distribuidas que se ilustran en la sección 6.1, "Medición de la inteligencia" (p. 313). Si están completas, use como base sus soluciones del ejercicio 6.1.

a. ¿Cómo se convierte una puntuación de IQ en una puntuación estándar?

b. ¿Cuál es la calificación estándar para una calificación de IQ de 90? ¿de 110? ¿de 120?

c. ¿Cuál es la calificación estándar para una calificación de la Prueba de Aptitud Escolar (SAT) de 465? ¿de 575? ¿de 650?

Usando la figura 2.2 de la página 314 con la regla empírica:

d. ¿Qué porcentaje de calificaciones de IQ es mayor a 132?

e. ¿Qué porcentaje de calificaciones del SAT es menor a 700?

Usando la tabla 3 del apéndice B:

f. ¿Cuál es la probabilidad de que una calificación de IQ sea mayor a 132?

g. ¿Cuál es la probabilidad de que una calificación del SAT sea menor a 700?

h. Compare sus respuestas a las partes f y g con sus respuestas a las partes d y e que usaron la regla empírica y la figura 2.2 de la página 314. Explique cualesquiera similitudes.

i. ¿Qué proporción de las calificaciones de IQ caen dentro del rango de 80 a 120?

j. ¿Qué proporción de las calificaciones de IQ rebasan 125?

k. ¿Qué porcentaje de las calificaciones del SAT está debajo de 450?

l. ¿Qué porcentaje de las calificaciones del SAT está arriba de 575?

m. ¿Qué calificación del SAT está en el 95avo percentil? Explique lo que esto significa.

Para su estudio

6.138 Pruebas de inteligencia

Las pruebas Wechsler, Wechsler Adult Intelligence Scale–Revised, WAIS–R, y Wechsler Intelligence Scale for Children, WISC–III, son alternativas de amplio uso de la prueba Stanford–Binet. La prueba Wechsler califica el rendimiento de inteligencia (no verbal), además de inteligencia verbal, y puede descomponerse para dejar ver puntos fuertes y débiles en varios aspectos.

Con base en puntuaciones de un gran número de personas seleccionadas al azar, los límites del IQ se han clasificado como se muestra en la tabla 10.4. Una vista a los porcentajes deja ver un patrón definido. La distribución de los IQ se aproxima a una curva normal, en la que la mayor parte de puntuaciones caen cerca del promedio, con menos en los extremos.

TABLA 10-4

Distribución de puntuaciones de IQ para adultos en la WAIS–R

IQ	Descripción	Porcentaje
Arriba de 130	Muy superior	2.2
120–129	Superior	6.7
110–119	Inteligente normal	16.1
90–109	Promedio	50.0
80–89	Torpe normal	16.1
70–79	Línea de frontera	6.7
Debajo de 70	Mentalmente retardado	2.2

Fuente: Dennis Coon, *Essentials of Psychology, Exploration and Application,* 8th ed. (Belmont, CA: Wadsworth, 1999)

a. Use la información de la tabla 10.4 de la página 356 y estime la desviación estándar para calificaciones WAIS-R para adultos. Use al menos dos piezas diferentes de información para obtener dos estimados separados. Determine su respuesta.

b. ¿La puntuación del IQ estudiada aquí parece tener una distribución normal? Dé razones para apoyar su respuesta.

c. ¿Qué porcentaje de la población adulta tiene inteligencia "superior"?

d. ¿Cuál es la probabilidad de seleccionar al azar una persona de esta población que se clasifique abajo del "promedio"?

e. ¿Qué puntuación del IQ está en el 95avo percentil? Explique lo que significa.

Examen de práctica del capítulo

PRIMERA PARTE: Conocer las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

6.1 La distribución de probabilidad normal es simétrica alrededor de **cero.**

6.2 El área total bajo la curva de cualquier distribución normal es **1.0.**

6.3 La probabilidad teórica de que ocurra un valor particular de una variable aleatoria **continua** es exactamente cero.

6.4 La unidad de medida para la calificación estándar es la **misma que la unidad de medida de los datos.**

6.5 Todas las distribuciones **normales** tienen las mismas funciones de probabilidad general y de distribución.

6.6 En la notación $z(0.05)$, el número dentro del paréntesis es la medida del área a la **izquierda** del puntaje z.

6.7 Las calificaciones normales estándar tienen una media de **uno** y una desviación estándar de **cero**.

6.8 Las distribuciones de probabilidad de **todas** las variables aleatorias continuas están normalmente distribuidas.

6.9 Podemos sumar y restar las áreas bajo la curva de una distribución continua porque estas áreas representan probabilidades de eventos **independientes**.

6.10 La distribución más común de una variable aleatoria continua es la probabilidad **binomial.**

SEGUNDA PARTE: Aplicación de conceptos

6.11 Encuentre las siguientes probabilidades para z, la calificación normal estándar:

a. $P(0 < z < 2.42)$ b. $P(z < 1.38)$

c. $P(z < -1.27)$ d. $P(-1.35 < z2.72)$

6.12 Encuentre el valor de cada puntaje z:

a. $P(z > ?) = 0.2643$ b. $P(z < ?) = 0.17$

c. $z(0.04)$

6.13 Use la notación simbólica $z()$ para dar el nombre simbólico para cada puntaje z que se muestra en la figura de esta página.

6.14 La vida útil de baterías para linternas eléctricas está normalmente distribuida alrededor de una media de 35.6 horas, con una desviación estándar de 5.4 horas. Kevin seleccionó al azar una de estas baterías y la probó. ¿Cuál es la probabilidad de que esta batería dure menos de 40.0 horas?

6.15 Se cree que los tiempos, x, que estudiantes pierden en el transporte diario para ir a la universidad, tienen una media de 22 minutos con una desviación estándar de 9 minutos. Si los tiempos empleados en transporte están normalmente distribuidos en forma aproximada, encuentre el tiempo, x, que separa el 25% de quienes pasan más tiempo en transporte que el resto de los viajeros.

Figura para el ejercicio 6.13

6.16 Miles de estudiantes de preparatoria toman el examen SAT todos los años. Las calificaciones alcanzadas por los estudiantes de cierta ciudad están normalmente distribuidas en forma aproximada, con una media de 490 y una desviación estándar de 70. Encuentre:

a. el porcentaje de estudiantes que obtienen entre 600 y 700

b. el porcentaje de estudiantes que obtienen menos de 650

c. el tercer cuartil

d. el 15avo percentil, P_{15}

e. el 95avo percentil, P_{95}

TERCERA PARTE: Comprensión de conceptos

6.17 En 50 palabras, describa la distribución normal estándar.

6.18 Describa el significado del símbolo $z(\alpha)$.

6.19 Explique por qué la distribución normal estándar, como se calcula en la tabla 3 del Apéndice B, se puede usar para hallar probabilidades para todas las distribuciones normales.

CAPÍTULO

7 Variabilidad de la muestra

7.1 275 millones de norteamericanos

7.2 Distribuciones muestrales

7.3 Distribución de medias muestrales

7.4 Aplicación de la distribución de medias muestrales

© Spencer Grant/PhotoEdit

Recuerde nuestra primera pregunta, "¿Qué se puede deducir acerca de la población estadística de la cual se toma la muestra?" El objetivo de este capítulo es estudiar las medidas y los patrones de variabilidad para la distribución formada por valores de una media muestral observados repetidamente.

7.1 275 millones de norteamericanos

El censo de Estados Unidos y su muestreo

Según el censo del año 2000, la población en Estados Unidos consta de más de 275 millones de habitantes. Con frecuencia leemos y escuchamos algo sobre esta población; los medios noticiosos informan de los resultados de muestras casi todos los días. Una de las variables de interés para muchos es la "edad" de norteamericanos.

Según el censo de 2000, los aproximadamente 275 millones de norteamericanos tienen una edad media de 36.5 años y una desviación estándar de 22.5 años. Las edades están distribuidas como se muestra en el histograma siguiente.

Un censo en Estados Unidos se lleva a cabo sólo cada 10 años. Es un trabajo enorme y abrumador, pero la información que se obtiene es vital para la organización y estructura de nuestro país. Surgen problemas y cambian los tiempos; la información se necesita y un censo es impráctico. Aquí es donde entra una muestra representativa.

EL PROBLEMA DEL MUESTREO

La meta fundamental de una encuesta es proponer los mismos resultados que se hubieran obtenido si se hubiera entrevistado a cada miembro de la población. Para los sondeos nacionales Gallup, en otras palabras, la meta es presentar las opiniones de una muestra de personas, que son exactamente las mismas opiniones que se hubieran obtenido en caso de haber sido posible entrevistar a todos los norteamericanos adultos del país.

La clave para alcanzar esta meta es un principio fundamental llamado *igual probabilidad de selección*, que expresa que si todo miembro de una población tiene igual probabilidad de ser seleccionado en una muestra, entonces esa muestra será representativa de la población. Así de sencillo.

Por tanto, la meta de Gallup es seleccionar muestras para que todo norteamericano adulto tenga igual oportunidad de caer en la muestra. La forma en que se hace esto, por supuesto, es la clave para el éxito o fracaso del proceso.

Fuente: http://www.gallup.com/help/FAQs/poll1.asp

Suponga que se tomó una muestra aleatoria de 100 edades de la distribución del censo de 2000.

45	78	55	15	47	85	93	46	13	41
87	78	7	7	94	48	11	41	81	32
59	8	15	20	49	66	11	61	16	19
39	74	34	6	46	8	46	21	44	41
52	84	27	53	33	48	80	6	62	21
47	11	17	3	31	43	46	23	52	20
35	24	30	37	54	90	26	55	89	2
58	44	30	45	15	25	47	13	28	10
80	41	30	57	63	79	75	7	26	4
2	10	21	19	5	62	32	59	40	16

¿Qué tan bien es que esta muestra representa la población? ¿Qué debemos ver? ¿Cómo haremos una comparación? Después de completar el capítulo 7, investigue con más detalle estas preguntas respecto a las edades de norteamericanos con base en el censo del año 2000 del proyecto del capítulo de la página 389.

SECCIÓN 7.1 EJERCICIOS

7.1 a. ¿En qué forma describiría usted gráficamente las 100 "edades" de la muestra aleatoria precedente tomada de la distribución del censo del año 2000? Construya la gráfica.

b. Usando la gráfica que construyó en la parte a, describa la forma de la distribución de datos muestrales.

c. ¿Qué tan bien describió la muestra a la población de edades del censo 2000? Explique usando las presentaciones gráficas.

d. Si se recolectara otra muestra, ¿esperaría usted los mismos resultados? Explique.

7.2 a. ¿Cómo describiría usted numéricamente las 100 "edades" de la muestra aleatoria precedente tomada de la distribución del censo 2000? Calcule las estadísticas.

b. ¿Qué tan bien se comparan las estadísticas calculadas en la parte a contra los parámetros del censo 2000? Sea específico.

c. Si se recolectara otra muestra, ¿esperaría usted los mismos resultados? Explique.

7.2 Distribuciones muestrales

Para hacer inferencias acerca de la población, es necesario examinar un poco más los resultados muestrales. Una media muestral, $\bar{x}$, se obtiene de una muestra. ¿Espera que este valor, $\bar{x}$, sea exactamente igual al valor de la media poblacional, μ Su respuesta debe ser negativa. No esperamos que las medias sean idénticas, pero estaremos satisfechos con nuestros resultados muestrales si la media muestral es "cercana" al valor de la media poblacional. Consideremos una segunda pregunta: Si se toma una segunda muestra, ¿ésta tendrá una media igual a la media poblacional? ¿será igual a la primera media muestral? Negativo, de nuevo; no esperamos que la media muestral sea igual a la media poblacional, ni esperamos que la segunda media muestral sea una repetición de la primera. No obstante, otra vez esperamos que los valores sean "cercanos." (Este argumento no debe cumplirse para cualquier otra estadística muestral y su correspondiente valor poblacional.)

Las siguientes preguntas ya deben estar a la vista: ¿Qué es "cercanos"? ¿Cómo determinar (y medir) esta cercanía? Exactamente, ¿cómo estarán distribuidas las **estadísticas muestrales repetidas**? Para contestar estas preguntas debemos ver la *distribución de muestreo.*

Distribución muestral de una estadística muestral: Es la distribución de valores para una estadística muestral obtenida de muestras repetidas, todas del mismo tamaño y todas extraídas de la misma población.

Empecemos por investigar dos distribuciones de muestreo teóricas, pequeñas y diferentes.

EJEMPLO 7.1

Formación de una distribución muestral de medias y rangos

PARA SU INFORMACIÓN
Las muestras se extraen con restitución.

Considere como población al conjunto de enteros pares de un solo dígito, {0,2,4,6,8}. Además, considere todas las posibles muestras de tamaño 2. Veremos dos diferentes distribuciones de muestreo que podrían formarse: la distribución de muestreo de medias muestrales y la distribución de muestreo de rangos muestrales.

Primero necesitamos hacer una lista de todas las posibles muestras de tamaño 2; hay 25 posibles muestras:

{0, 0}	{2, 0}	{4, 0}	{6, 0}	{8, 0}
{0, 2}	{2, 2}	{4, 2}	{6, 2}	{8, 2}
{0, 4}	{2, 4}	{4, 4}	{6, 4}	{8, 4}
{0, 6}	{2, 6}	{4, 6}	{6, 6}	{8, 6}
{0, 8}	{2, 8}	{4, 8}	{6, 8}	{8, 8}

Cada una de estas muestras tiene una media $\bar{x}$. Estas medias son, respectivamente:

0	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

TABLA 7.1

Distribución de probabilidad: Distribución de medias muestrales

$\bar{x}$	$P(\bar{x})$
0	0.04
1	0.08
2	0.12
3	0.16
4	0.20
5	0.16
6	0.12
7	0.08
8	0.04

Cada una de estas muestras es igualmente probable, y por tanto a cada una de las 25 medias muestrales se le puede asignar una probabilidad de $\frac{1}{25} = 0.04$. La **distribución de medias muestrales** se muestra en la tabla 7.1 como una **distribución de probabilidad** y se ve en la figura 7.1 como un histograma.

FIGURA 7.1
Histograma: Distribución de medias muestrales

Para el mismo conjunto de todas las posibles muestras de tamaño 2, encontremos la distribución muestral de rangos muestrales. Cada muestra tiene un rango R. Los rangos son:

0	2	4	6	8
2	0	2	4	6
4	2	0	2	4
6	4	2	0	2
8	6	4	2	0

De nuevo, cada uno de estos 25 rangos muestrales tiene una probabilidad de 0.04. La tabla 7.2 muestra la distribución muestral de rangos muestrales como una distribución de probabilidad, y la figura 7.2 muestra la distribución muestral como un histograma.

TABLA 7.2

Distribución de probabilidad: Distribución muestral de rangos muestrales

R	$P(R)$
0	0.20
2	0.32
4	0.24
6	0.16
8	0.08

FIGURA 7.2
Histograma: Distribución muestral de rangos muestrales

El ejemplo 7.1 es teórico en su naturaleza y por ello se expresa en probabilidades. Como esta población es pequeña, es fácil elaborar una lista de las 25 muestras posibles de tamaño 2 (un espacio muestral) y asignar probabilidades. No obstante, no siempre es posible hacer esto.

Ahora, investiguemos en forma empírica (es decir, por experimentación) otra distribución muestral.

EJEMPLO 7.2 Creación de una distribución de medias muestrales

Consideremos una población formada por cinco enteros igualmente probables: 1, 2, 3, 4 y 5. La figura 7.3 muestra una representación en histograma de la población. Podemos observar una parte de la distribución de medias muestrales cuando se seleccionen al azar 30 muestras de tamaño 5.

La tabla 7.3 presenta 30 muestras y sus medias. En la figura 7.4 aparece la distribución muestral resultante, una **distribución de frecuencia,** de medias muestrales. Nótese que esta distribución de medias muestrales no se ve como la población sino que, más bien, parece exhibir las características de una distribución normal; tiene forma de montículo y es casi simétrica alrededor de su media (aproximadamente 3.0).

FIGURA 7.3 La población: distribución teórica de probabilidad

sacan muestras

TABLA 7.3

30 muestras de tamaño 5

No.	Muestra	$\bar{x}$	No.	Muestra	$\bar{x}$
1	4,5,1,4,5	3.8	16	4,5,5,3,5	4.4
2	1,1,3,5,1	2.2	17	3,3,1,2,1	2.0
3	2,5,1,5,1	2.8	18	2,1,3,2,2	2.0
4	4,3,3,1,1	2.4	19	4,3,4,2,1	2.8
5	1,2,5,2,4	2.8	20	5,3,1,4,2	3.0
6	4,2,2,5,4	3.4	21	4,4,2,2,5	3.4
7	1,4,5,5,2	3.4	22	3,3,5,3,5	3.8
8	4,5,3,1,2	3.0	23	3,4,4,2,2	3.0
9	5,3,3,3,5	3.8	24	3,3,4,5,3	3.6
10	5,2,1,1,2	2.2	25	5,1,5,2,3	3.2
11	2,1,4,1,3	2.2	26	3,3,3,5,2	3.2
12	5,4,3,1,1	2.8	27	3,4,4,4,4	3.8
13	1,3,1,5,5	3.0	28	2,3,2,4,1	2.4
14	3,4,5,1,1	2.8	29	2,1,1,2,4	2.0
15	3,1,5,3,1	2.6	30	5,3,3,2,5	3.6

usando las 30 medias

FIGURA 7.4 Distribución empírica de medias muestrales

Nota: La variable para la distribución muestral es $\bar{x}$ 1; por tanto, la media de las $\bar{x}$'s es $\bar{\bar{x}}$ y la desviación estándar de $\bar{x}$ es $s_{\bar{x}}$.

La teoría involucrada con distribuciones muestrales que se describirán en el resto de este capítulo requiere de *muestreo aleatorio.*

Muestra aleatoria: es una muestra obtenida en forma tal que cada una de las posibles muestras de tamaño fijo *n* tiene igual probabilidad de ser seleccionada (vea p. 22).

La figura 7.5 muestra cómo está formada la distribución de medias muestrales.

FIGURA 7.5 Distribución de medias muestrales

CASO PRÁCTICO 7.3 Edad promedio de vehículos ferroviarios en tránsito urbano

Hay numerosas razones para recolectar datos repetidamente. No todas las recolecciones de datos repetidos se efectúan para formar una distribución muestral. Considere la estadística "Promedio de edad de vehículos ferroviarios en tránsito urbano (Años)" del U.S.Department of Transportation que sigue. La tabla muestra el promedio de edad para cuatro clasificaciones diferentes de vehículos ferroviarios en tránsito rastreados durante varios años. Al estudiar el patrón de cambio de la edad promedio para cada clase de vehículo, una persona puede sacar conclusiones acerca de lo que ha estado ocurriendo a la flota durante varios años. Es probable que las personas encargadas del mantenimiento de cada flota también puedan detectar cuándo es necesario un cambio en políticas respecto al cambio de vehículos viejos. A pesar de lo útil que es esta información, no hay aquí una distribución muestral.

Promedio de edad de vehículos ferroviarios en tránsito urbano (años)

	1985	1990	1995	2000	2003
Rieles					
Locomotoras en rutas suburbanas[a]	16.3	15.7	15.9	13.4	16.6
Coches de pasajeros en rutas suburbanas	19.1	17.6	21.4	16.9	20.5
Coches de pasajeros en rieles pesados	17.1	16.2	19.3	22.9	19.0
Vehículos en rieles ligeros (tranvías)	20.6	15.2	16.8	16.1	15.6

[a]No se incluyen locomotoras de Amtrak en servicio de pasajeros en rutas suburbanas.

Fuente: U.S.Department of Transportation, Federal Transit Administration

SECCIÓN 7.2 EJERCICIOS

7.3 Algunos fabricantes usan muestras aleatorias para saber si sus productos satisfacen especificaciones. Estas muestras podrían ser personas, piezas manufacturadas, o hasta muestras durante la elaboración de hojuelas de patatas.

a. ¿Piensa usted que todas las muestras aleatorias tomadas de la misma población llevarán al mismo resultado?

b. ¿Qué característica (o propiedad) de muestras aleatorias podrían observarse durante el proceso de muestreo?

7.4 Consulte la tabla 7.1 del ejemplo 7.1 (p. 363) y explique por qué las muestras son igualmente probables; esto es, ¿por qué $P(0)=0.04$, y por qué $P(2)=0.12$.

7.5 a. ¿Cuál es la distribución de medias muestrales?

b. Se toma una muestra de tamaño 3 de una población, y se encuentra la media muestral. Describa cómo está relacionada esta media muestral con la distribución de medias muestrales.

7.6 Considere el conjunto de enteros impares de un solo dígito {1, 3, 5, 7, 9}.

a. Haga una lista de todas las muestras de tamaño 2 que puedan sacarse de este conjunto de enteros. (Muestreo con restitución; es decir, se saca el primer número, se observa, y luego se restituye [se devuelve al conjunto muestral] antes de sacar de nuevo.)

b. Construya la distribución de medias muestrales para muestras de tamaño 2 seleccionadas de este conjunto.

c. Construya las distribuciones muestrales de rangos muestrales para muestras de tamaño 2.

7.7 Considere el conjunto de enteros pares de un solo dígito {0, 2, 4, 6, 8}.

a. Haga una lista de todas las posibles muestras de tamaño 3 que se puedan sacar de este conjunto de enteros. (Muestreo con restitución; es decir, se saca el primer número, se observa, y luego se restituye [se devuelve al conjunto muestral] antes de sacar de nuevo.)

b. Construya la distribución muestral de las medianas muestrales para muestras de tamaño 3.

c. Construya la distribución muestral de las medias muestrales para muestras de tamaño 3.

7.8 Usando los números telefónicos que aparecen en su guía local como su población, obtenga al azar 20 muestras de tamaño 3. Para cada número telefónico identificado como fuente, tome el cuarto, quinto y sexto dígitos. (Por ejemplo, para 245-8269, usted tomaría el 8, el 2 y el 6 como muestra de tamaño 3.)

a. Calcule la media de las 20 muestras.

b. Trace un histograma que muestre las 20 medias muestrales. (Use clase −0.5 a 0.5, 0.5 a 1.5, 1.5 a 2.5, y así sucesivamente.)

c. Describa la distribución de $\bar{x}$'s, que usted vea en la parte b (forma de distribución, centro y cantidad de dispersión).

d. Saque 20 muestras más y sume las 20 nuevas $\bar{x}$'s al histograma de la parte b. Describa la distribución que parece estar formándose.

7.9 Usando un conjunto de cinco dados, lance los dados y determine el número medio de puntos que aparecen en los cinco dados. Repita el experimento hasta que tenga 25 medias muestrales.

a. Trace una gráfica de puntos que presente la distribución de las 25 medias muestrales. (Vea ejemplo 7.2, p. 364.)

b. Describa la distribución de $\bar{x}$'s en la parte a.

c. Repita el experimento para obtener 25 medias muestrales más y sume estas 25 $\bar{x}$'s a su gráfica de puntos. Describa la distribución de 50 medias.

7.10 Considerando la población de cinco enteros igualmente probables del ejemplo 7.2:

a. Verifique μ y σ para la población del ejemplo 7.2.

b. La tabla 7.3 es una lista de 30 valores $\bar{x}$. Construya una distribución de frecuencia agrupada para verificar la distribución de frecuencia mostrada en la figura 7.4.

c. Encuentre la media y desviación estándar de los 30 valores $\bar{x}$ de la tabla 7.3 para verificar los valores para $\bar{\bar{x}}$ y $s_{\bar{x}}$. Explique el significado de los dos símbolos $\bar{\bar{x}}$ y $s_{\bar{x}}$.

7.11 Con referencia al caso práctico 7.3 de la página 366:

a. Explique por qué los valores numéricos de esta tabla no forman una distribución muestral.

b. Explique cómo difiere esta reunión repetida de datos con respecto a la idea de muestreo repetido para conjuntar información acerca de una distribución muestral.

7.12 De la tabla de números aleatorios de la tabla 1 del Apéndice B, construya otra tabla que muestre 20 conjuntos de 5 enteros de un solo dígito y seleccionados al azar. Encuentre la media de cada conjunto (la gran media) y compare este valor con la media poblacional teórica, , usando la diferencia absoluta y el error porcentual. Muestre todo su trabajo.

7.13 a. Usando computadora o una tabla de números aleatorios, simule sacar 100 muestras, cada una de tamaño 5, de la distribución de probabilidad uniforme de enteros de un solo dígito, 0 a 9.

b. Encuentre la media para cada muestra.

c. Construya un histograma de las medias muestrales. (Use valores enteros como puntos medios de clase.)

d. Describa la distribución muestral presentada en el histograma de la parte c.

MINITAB (Versión 14)

a. Use los comandos Integer RANDOM DATA de la página 101, sustituyendo generar con 100, guarda con C1-C5, valor mínimo con 0 y valor máximo con 9.

b.
```
Choose:   Calc > Row Statistics
Select:   Mean
Enter:    Input variables: C1–C5
          Store result in: C6 > OK
```

c. Use los comandos HISTOGRAM de la página 61 para los datos en C6. Para ajustar el histograma, seleccione Binning with midpoint y midpoint positions 0:9/1.

Excel

a. Introduzca de 0 a 9 en la columna A y los correspondientes 0.1 en la columna B; luego continúe con:

```
Choose:  Tools > Data Analysis > Random Number Generation > OK
Enter:   Number of Variables: 5
         Number of Random Numbers: 100
         Distribution: Discrete
         Value and Probability Input Range: (A1:B10 or select cells)
Select:  Output Range:
Enter:   (C1 or select cell) > OK
```

b. Active la celda H1.

```
Choose:  Insert function, fx > Statistical > AVERAGE > OK
Enter:   Number1: (C1:G1 or select cells)
Drag:    Bottom right corner of average value box down to give other averages
```

c. Use los comandos HISTOGRAM de las páginas 61-62 con la columna H como el rango de entrada y la columna A como rango de depósito.

TI-83/84 Plus

a. Use los comandos Integer RANDOM DATA and STO de la página 101, sustituyendo el Enter con 0,9,100). Repita cuatro veces los comandos precedentes, guardando datos en L2, L3, L4 y L5, respectivamente.

b.
```
Choose:     STAT > EDIT > 1:Edit
Highlight:  L6 (column heading)
Enter:      (L1+L2+L3+L4+L5)/5
```

c.
```
Choose:     2nd > STAT PLOT > 1:Plot1
Choose:     Window
Enter:      0, 9, 1, 0,
            30, 5, 1
Choose:     Trace > > >
```

7.14 a. Usando computadora o tabla de números aleatorios, simule sacar 250 muestras, cada una de tamaño 18, de la distribución de probabilidad uniforme de enteros de un solo dígito, 0 a 9.

b. Encuentre la media para cada muestra.

c. Construya un histograma de las medias muestrales.

d. Describa la distribución muestral presentada en el histograma de la parte c.

7.15 a. Use computadora para sacar 200 muestras aleatorias, cada una de tamaño 10, de la distribución de probabilidad normal con media 100 y desviación estándar 20.

b. Encuentre la media para cada muestra.

c. Construya un histograma de frecuencia de las 200 medias muestrales.

d. Describa la distribución muestral presentada en el histograma de la parte c.

MINITAB (Versión 14)

a. Use los comandos Normal RANDOM DATA de la página 101, sustituyendo generar con 200, guardar con C1-C10, media con 100, y desviación estándar con 20.

b.
```
Choose:     Calc > Row Statistics
Select:     Mean
Enter:      Input variables: C1–C10
            Store result in: C11 > OK
```

c. Use los comandos HISTOGRAM de la página 61 para los datos en C11. Para ajustar el histograma, seleccione Binning with midpoint and midpoint positions 74.8:125.2/6.3.

Excel

a. Use los comandos Normal RANDOM NUMBER GENERATION de la página 101, sustituyendo número de variables con 10, número de números aleatorios con 200, media con 100, y desviación estándar con 20.

b. Active la celda K1.

```
Choose:     Insert function, fx > Statistical >
            AVERAGE > OK
Enter:      Number1: (A1:J1 or select cells)
Drag:       Bottom right corner of average value
            box down to give other averages
```

c. Use los comandos RANDOM NUMBER GENERATION Patterned Distribution del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 74.8, el último valor con 125.2, los intervalos con 6.3, y el rango de salida con L1. Use los comandos HISTOGRAM de las páginas 61-62 con la columna K como rango de entrada y la columna L como rango de depósito.

7.16 a. Use computadora para sacar 500 muestras aleatorias, cada una de tamaño 20, de la distribución de probabilidad normal con media 80 y desviación estándar 15.

b. Encuentre la media para cada muestra.

c. Construya un histograma de frecuencia de las 500 medias muestrales.

d. Describa la distribución muestral presentada en el histograma en la parte c, incluyendo la media y desviación estándar.

7.3 Distribución de medias muestrales

En las páginas precedentes estudiamos las distribuciones muestrales de dos estadísticas: medias muestrales y rangos muestrales. Muchas otras podrían discutirse, pero la única distribución muestral de interés para nosotros en este momento es la distribución de medias muestrales.

PARA SU INFORMACIÓN
¡Ésta es información muy útil!

Distribución de medias muestrales (SDSM): Si todas las posibles muestras aleatorias, cada una de tamaño n, se toman de cualquier población con media μ y desviación estándar σ , entonces la distribución de medias muestrales tendrá lo siguiente:

1. Una media $\mu_{\bar{x}}$ igual a μ
2. Una desviación estándar $\sigma_{\bar{x}}$ igual a $\dfrac{\sigma}{\sqrt{n}}$.

Además, si la población muestreada tiene una distribución normal, entonces la distribución muestral de $\bar{x}$ también será normal para muestras de todos los tamaños.

Éste es un enunciado muy interesante de dos partes. La primera nos da información acerca de la relación entre la media poblacional y la desviación estándar, y la media de distribución muestral y desviación estándar para todas las distribuciones muestrales de medias muestrales. La desviación estándar de la distribución muestral se denota por $\sigma_{\bar{x}}$ y se le da un nombre específico para evitar confusión con la desviación estándar poblacional, μ.

Error estándar de la media ($\sigma_{\bar{x}}$): Es la desviación estándar de la distribución de medias muestrales.

La segunda parte indica que esta información no siempre es útil. Dicho de otro modo, dice que el valor medio de sólo unas pocas observaciones serán normalmente distribuidas cuando se sacan muestras de una población normalmente distribuida, pero no será normalmente distribuida cuando la población muestreada sea uniforme, sesgada o que de otro modo no sea normal. No obstante, el *teorema de límite central* nos da alguna información adicional muy útil acerca de la distribución de medias muestrales.

PARA SU INFORMACIÓN

Verdaderamente sorprendente: $\bar{x}$ es normalmente distribuida cuando n es grande lo suficiente, sin importar la forma de la población.

Teorema de límite central (CLT): La distribución de medias muestrales se asemejará más estrechamente a la distribución normal cuando aumente el tamaño muestral.

¿SABÍA USTED...?

Teorema de límite central

Abraham de Moivre fue un pionero en la teoría de probabilidad y publicó *La Doctrina de la Probabilidad*, primero en latín en 1711 y luego en ediciones aumentadas en 1718, 1738 y 1756. La edición de 1756 contenía su aportación más importante: la aproximación de las distribuciones binomiales para jun gran número de intentos usando la distribución normal. La definición de independencia estadística también hizo su debut junto con muchos juegos de dados y otros. de Moivre probó que el teorema de límite central se cumple para números que resultan de juegos de probabilidad. Con el uso de matemáticas, también predijo con éxito la fecha de su propia muerte.

Si la distribución muestreada es normal, entonces la distribución de medias muestrales (SDSM) es normal, como ya dijimos, y el teorema de límite central (CLT) no se hace necesario. Pero, si la población muestreada no es normal, el CLT nos dice que la distribución muestral todavía será normalmente distribuida en forma aproximada bajo condiciones apropiadas. Si la distribución de población muestreada es casi normal, la distribución $\bar{x}$ es aproximadamente normal para una n más bien pequeña (quizá de sólo 15). Cuando la distribución poblacional muestreada carece de simetría, n puede tener que ser muy grande (quizá 50 o más) antes que la distribución normal dé una aproximación satisfactoria.

Al combinar la información precedente, podemos describir por completo la distribución muestral de $\bar{x}$: (1) la ubicación del centro (media), (2) una medida de dispersión que indica con qué amplitud está dispersa la distribución (error estándar de la media) y (3) una indicación de cómo está distribuida.

1. $\mu_{\bar{x}} = \mu$; la media de la distribución muestral ($\mu_{\bar{x}}$) es igual a la media de la población (μ).
2. $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$; el error estándar de la media ($\sigma_{\bar{x}}$) es igual a la desviación estándar de la población (σ) dividida entre la raíz cuadrada del tamaño muestral, n.
3. La distribución de medias muestrales es normal cuando la población progenitora está normalmente distribuida, y el teorema de límite central (CLT) nos dice que la distribución de medias muestrales se hace aproximadamente normal (cualquiera que sea la forma de la población progenitora) cuando el tamaño muestral es grande lo suficiente.

Nota: la n a la que nos referimos es el tamaño de cada muestra de la distribución muestral. (El número de muestras repetidas empleadas en una situación empírica no tiene efecto en el error estándar.)

No mostramos la prueba para los tres datos precedentes de este texto; no obstante, la validez de éstos se demostrará al examinar dos ejemplos. Para el primer ejemplo, consideremos una población para la cual podemos construir la distribución muestral teórica de todas las muestras posibles.

EJEMPLO 7.4 Construcción de una distribución muestral de medias muestrales

Consideremos todas las muestras posibles de tamaño 2 que podrían sacarse de una población que contiene los tres números, 2,4 y 6. Veamos primero la población misma. Construya un histograma para representar su distribución, figura 7.6; calcule la media, μ, y la desviación estándar, σ, tabla 7.4. (Recuerde: debemos usar las técnicas del capítulo 5 para distribuciones de probabilidad discretas.)

FIGURA 7.6 Población

TABLA 7.4

Tabla de extensiones para x

x	$P(x)$	$xP(x)$	$x^2P(x)$
2	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{4}{3}$
4	$\frac{1}{3}$	$\frac{4}{3}$	$\frac{16}{3}$
6	$\frac{1}{3}$	$\frac{6}{3}$	$\frac{36}{3}$
Σ	$\frac{3}{3}$	$\frac{12}{3}$	$\frac{56}{3}$
	1.0	4.0	$18.6\overline{6}$

$\mu = \mathbf{4.0}$

$\sigma = \sqrt{18.6\overline{6} - (4.0)^2} = \sqrt{2.6\overline{6}} = \mathbf{1.63}$

La tabla 7.5 (vea p. 372) es una lista de todas las muestras posibles de tamaño 2 que se pueden sacar de esta población. (Se saca un número, se observa, y luego se regresa a la población antes de sacar el segundo número.) La tabla 7.5 también enumera las medias de estas muestras. Las medias muestrales se recolectan luego para formar la distribución muestral. La distribución para estas medias y las extensiones se dan en la tabla 7.6 (p. 372), junto con el cálculo de la media y el error estándar de la media para la distribución muestral. El histograma para la distribución de medias muestrales se ve en la figura 7.7 (p. 372).

Comprobemos ahora la verdad de los tres datos acerca de la distribución de medias muestrales:

1. La media $\mu_{\bar{x}}$ de la distribución muestral será igual a la media μ de la población: tanto μ como $\mu_{\bar{x}}$ tienen el mismo valor de **4.0.**
2. El error estándar de la media $\sigma_{\bar{x}}$ para la distribución muestral será igual a la desviación estándar σ de la población dividida entre la raíz cuadrada del tamaño muestral: n: $\sigma_{\bar{x}} = \mathbf{1.15}$ y $\sigma = 1.63$, $n = 2$, $\dfrac{\sigma}{\sqrt{n}} = \dfrac{1.63}{\sqrt{2}} = \mathbf{1.15;}$ son iguales: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.
3. La distribución se hará normalmente distribuida en forma aproximada: el histograma de la figura 7.7 con gran fuerza sugiere normalidad.

TABLA 7.5

Las nueve posibles muestras de tamaño 2

Muestra	$\bar{x}$	Muestra	$\bar{x}$	Muestra	$\bar{x}$
2, 2	2	4, 2	3	6, 2	4
2, 4	3	4, 4	4	6, 4	5
2, 6	4	4, 6	5	6, 6	6

TABLA 7.6

Tabla de extensiones para $\bar{x}$

$\bar{x}$	$P(\bar{x})$	$\bar{x}P(\bar{x})$	$\bar{x}^2P(\bar{x})$
2	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{4}{9}$
3	$\frac{2}{9}$	$\frac{6}{9}$	$\frac{18}{9}$
4	$\frac{3}{9}$	$\frac{12}{9}$	$\frac{48}{9}$
5	$\frac{2}{9}$	$\frac{10}{9}$	$\frac{50}{9}$
6	$\frac{1}{9}$	$\frac{6}{9}$	$\frac{36}{9}$
Σ	$\frac{9}{9}$	$\frac{36}{9}$	$\frac{156}{9}$
	1.0	4.0	$17.3\overline{3}$

$\mu_{\bar{x}} = \mathbf{4.0}$

$\sigma_{\bar{x}} = \sqrt{17.3\overline{3} - (4.0)^2} = \sqrt{1.3\overline{3}} = \mathbf{1.15}$

FIGURA 7.7 Distribución de medias muestrales

El ejemplo 7.4, una situación teórica, sugiere que los tres datos parezcan cumplirse. ¿Se cumplen estos tres datos cuando se recolectan datos reales? Vamos de nuevo al ejemplo 7.2 (p. 364) y veamos si los tres datos están apoyados por la distribución muestral empírica ahí.

Primero, veamos la poblaciónla distribución de probabilidad teórica de la cual se sacaron las muestras del ejemplo 7.2. La figura 7.3 es un histograma que presenta la distribución de probabilidad para datos seleccionados al azar de la población de enteros igualmente probables 1, 2, 3, 4, 5. La media poblacional μ es igual a 3.0. La desviación estándar poblacional σ es $\sqrt{2}$, o 1.41. La población tiene una distribución uniforme.

Ahora veamos la distribución empírica de las 30 medias muestrales halladas en el ejemplo 7.2. De los 30 valores de $\bar{x}$ en la tabla 7.3, la media observada de las $\bar{x}$'s, $\bar{\bar{x}}$, es 2.98 y el error estándar observado de la media, $s_{\bar{x}}$, es 0.638. El histograma de la distribución muestral de la figura 7.4 aparece en forma de montículo, aproximadamente simétrico, y centrado cerca del valor 3.0.

A continuación verifiquemos la verdad de las tres propiedades específicas:

1. $\mu_{\bar{x}}$ y μ serán iguales. La media de la población es 3.0, y la media de distribución muestral observada $\bar{\bar{x}}$ es 2.98; son muy cercanas en valor.
2. $\sigma_{\bar{x}}$ será igual a $\frac{\sigma}{\sqrt{n}}$. $\sigma = 1.41$ y $n = 5$; por lo tanto, $\frac{\sigma}{\sqrt{n}} = \frac{1.41}{\sqrt{5}} = \mathbf{0.632,}$ y $s_{\bar{x}} = \mathbf{0.638;}$ son muy cercanas en valor. (Recuerde que hemos tomado sólo 30 muestras, no todas las muestras posibles, de tamaño 5.)

3. La distribución muestral de $\bar{x}$ estará normalmente distribuida en forma aproximada. Aun cuando la población tiene una distribución rectangular, el histograma de la figura 7.4 sugiere que la distribución x tiene algunas de las propiedades de normalidad (en forma de montículo, simétrica).

Aun cuando los ejemplos 7.2 y 7.4 no constituyen una prueba, la evidencia parece sugerir con gran fuerza que ambos enunciados, la distribución muestral de medas muestrales y el teorema de límite central, CLT, son verdaderos.

Habiendo visto estos dos ejemplos específicos, veamos ahora cuatro ilustraciones gráficas que presentan la información de distribución muestral y el CLT en una forma ligeramente diferente. Cada una de estas ilustraciones tiene cuatro distribuciones. La primera gráfica muestra la distribución de la población progenitora, la distribución de los valores x individuales. Cada una de las otras tres gráficas muestra una distribución de medias muestrales, $\bar{x}$'s, que usa tres tamaños muestrales diferentes.

En la figura 7.8 tenemos una distribución uniforme, muy semejante a la figura 7.3 para la ilustración de enteros, y las distribuciones resultantes de medias muestrales para muestras de tamaños 2, 5 y 30.

FIGURA 7.8
Distribución uniforme

La figura 7.9 presenta una población en forma de U y las tres distribuciones muestrales.

FIGURA 7.9
Distribución en forma de U

La figura 7.10 muestra una población en forma de J y las tres distribuciones muestrales.

FIGURA 7.10
Distribución en forma de J

Las tres distribuciones poblacionales no normales parecen satisfacer el teorema de límite central (CLT); las distribuciones muestrales de medias muestrales parecen ser aproximadamente normales para las tres cuando se usaron muestras de tamaño 30. Ahora considere la figura 7.11, que muestra una población normalmente distribuida y las tres distribuciones muestrales. Con la población normal, las distribuciones muestrales de las medias muestrales para los todos los tamaños muestrales parecen ser normales. Por tanto, hemos visto un fenómeno sorprendente: no importa cuál sea la forma de una población, la distribución de medias muestrales es normal o se hace aproximadamente normal cuando n se hace grande lo suficiente.

FIGURA 7.11
Distribución normal

El lector debe observar otro punto: la media muestral se hace menos variable cuando aumenta el tamaño muestral. Nótese que cuando n aumenta de 2 a 30, todas las distribuciones se hacen más angostas y más altas.

SECCIÓN 7.3 EJERCICIOS

7.17 El siguiente ejercicio indica que se debe realizar una simulación en la que se toman muestras de tamaño 4 de una población aproximadamente normal, donde $\mu = 65.15$ y $\sigma = 2.754$.

a. Dé un clic en "1" para "# Samples." Observe los cuatro valores de datos y su media. Cambie "lento" a "lote" y tome al menos 1000 muestras usando el "500" por "# Samples."

b. ¿Cuál es la media para las 1001 medias muestrales? ¿Qué tan cercana está a la media poblacional, μ?

c. Compare la desviación estándar muestral contra la desviación estándar poblacional, σ. ¿Qué le pasa a la desviación estándar muestral? Compárela con $\sigma/\sqrt{n}$, que es $2.754/\sqrt{4}$.

d. ¿El histograma de medias muestrales tiene una forma aproximadamente normal?

e. Relate lo que encuentre en la distribución de medias muestrales (SDSM).

7.18 En el siguiente ejercicio se debe simular que se tiene una población sesgada, donde $\mu = 6.029$ y $\sigma = 10.79$.

a. Cambie la "# Observations per sample" a "4". Usando lotes y 500, tome 1000 muestras de tamaño 4.

b. Compare la media y desviación estándar para las medias muestrales con μ y σ. Compare la desviación estándar muestral con $\sigma/\sqrt{n}$, que es $10.79/\sqrt{4}$. ¿El histograma tiene una forma aproximadamente normal? Si no es así, ¿de qué forma es?

c. Usando el botón "borrar" cada vez, repita las instrucciones de las partes a y b para muestras de tamaño 25, 100 y 1000. Haga una tabla de lo que encuentre para cada tamaño muestral.

d. Relacione lo que encuentre con la SDSM y el teorema de límite central (CLT).

7.19 a. ¿Cuál es la medida total del área para cualquier distribución de probabilidad?

b. Justifique el enunciado "$\bar{x}$ se hace menos variable cuando n aumenta."

7.20 Si una población tiene una desviación estándar de 25 unidades, ¿cuál es el error estándar de la media si se seleccionan muestras de tamaño 16? ¿muestras de tamaño 36? ¿muestras de tamaño 100?

7.21 Cierta población tiene una media de 500 y una desviación estándar de 30. Muchas muestras de tamaño 36 se seleccionan en forma aleatoria y se calculan las medias.

a. ¿Qué valor esperaría hallar para la media de todas estas medias muestrales?

b. ¿Qué valor esperaría hallar para la desviación estándar de todas estas medias muestrales?

c. ¿Qué forma esperaría que tenga la distribución de todas estas medias muestrales?

7.22 Un artículo de abril de 2004, de HearTheIssues.com, dijo que los norteamericanos tienen un promedio de 2.24 televisores por familia (*fuente:* Nielsen Media Research). Si la desviación estándar para el número de televisores en una familia norteamericana es 1.2 y se selecciona una muestra aleatoria de 80 familias, la media de esta muestra pertenece a una distribución muestral.

a. ¿Cuál es la forma de esta distribución muestral?

b. ¿Cuál es la media de esta distribución muestral?

c. ¿Cuál es la desviación estándar de esta distribución muestral?

7.23 Un artículo de 2004, de HearTheIssues.com, dijo que los norteamericanos ven un promedio de 4.0 horas de televisión por persona por día (*fuente:* Nielsen Media Research). Si la desviación estándar para el número de horas de televisión que se ve por día es 2.1, y se selecciona una muestra aleatoria de 250 norteamericanos, la media de esta muestra pertenece a una distribución muestral.

a. ¿Cuál es la forma de esta distribución muestral?

b. ¿Cuál es la media de esta distribución muestral?

c. ¿Cuál es la desviación estándar de esta distribución muestral?

7.24 Según *The World Factbook,* 2004, el porcentaje de fertilidad total (número medio estimado de hijos nacidos por mujer) para Madagascar es 5.7. Suponga que la desviación estándar del porcentaje de fertilidad total es 2.6. El número medio de niños para una muestra de 200 mujeres seleccionadas al azar es un valor de muchos que forman la SDSM.

a. ¿Cuál es el valor medio para esta distribución muestral?

b. ¿Cuál es la desviación estándar para esta distribución muestral?

c. Describa la forma para esta distribución muestral.

7.25 El USDA, Economics and Statistics System de la Universidad de Cornell, publica un *Poultry Yearbook* en el que presenta una lista mensual, trimestral y trimestral de datos acerca de la industria de aves de corral. El libro del año 2004 presenta una lista del consumo anual de carne de pavo de 17.71 libras por persona. Suponga que la desviación estándar para el consumo de pavo por persona es 6.3 libras. El peso medio de pavo consumido, para una muestra de 150 personas seleccionadas al azar, es un valor de muchos que forman la distribución de medias muestrales (SDSM).

a. ¿Cuál es el valor medio para esta distribución muestral?

b. ¿Cuál es la desviación estándar para esta distribución muestral?

c. Describa la forma de esta distribución muestral.

7.26 Un investigador desea tomar una muestra aleatoria sencilla de alrededor de 5% del cuerpo estudiantil de cada una de dos escuelas. La universidad tiene aproximadamente 20 000 estudiantes y el colegio tiene unos 5000. Identifique cada uno de lo siguiente como verdadero o falso y justifique su respuesta.

a. La variabilidad muestral es la misma para ambas escuelas.

b. La variabilidad muestral para la universidad es más alta que la del colegio.

c. La variabilidad muestral para la universidad es más baja que la del colegio.

d. No es posible expresar ninguna conclusión acerca de la variabilidad muestral sin conocer los resultados del estudio.

7.27 a. Use computadora para seleccionar al azar 100 muestras de tamaño 6 de una población normal con media $\mu = 20$ y desviación estándar $\sigma = 4.5$.

b. Encuentre la media $\bar{x}$ para cada una de las 100 muestras.

c. Usando las 100 medias muestrales, construya un histograma, encuentre la media $\bar{\bar{x}}$, y encuentre la desviación estándar $s_{\bar{x}}$.

d. Compare los resultados de la parte c con los tres enunciados hechos en la SDSM.

MINITAB (Versión 14)

a. Use los comandos Normal RANDOM DATA de la página 101, sustituyendo generar con 100, guardar con C1-C6, media con 20, y desviación estándar con 4.5.

b. Use los comandos ROW STATISTICS de la página 368, sustituyendo variables de entrada con C1-C6 y guardar resultados con C7.

c. Use los comandos HISTOGRAM de la página 61 para los datos en C7. Para ajustar el histograma, seleccione Binning with midpoint y midpoint positions 12.8:27.2/1.8. Use los comandos

MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de C7.

Excel

a. Use los comandos Normal RANDOM NUMBER GENERATION de la página 110, sustituyendo número de variables con 6, número de números aleatorios con 100, media con 20, y desviación estándar con 4.5.

b. Active la celda G1.

```
Choose:     Insert function, fx > Statistical >
            AVERAGE > OK
Enter:      Number1: (A1:F1 or select cells)
Drag:       Bottom right corner of average value
            box down to give other averages
```

c. Use los comandos RANDOM NUMBER GENERATION Patterned Distribution del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 12.8, el último valor con 27.2, los intervalos con 1.8, y el rango de salida con H1. Use los comandos HISTOGRAM de la página 61 con la columna G como el rango de entrada y la columna H como el rango de depósito. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de la columna G.

TI/83-84 Plus

a. Use los comandos Integer RANDOM DATA y STO de la página 101, sustituyendo Enter con 20.4.5,100). Repita cinco veces más los comandos precedentes, guardando datos en L2, L3, L4, L5 y L6, respectivamente.

b.
```
Enter:      (L1 + L2 + L3 + L4 + L5 + L6)/6
Choose:     STOO  L7 (use ALPHA key for the 'L' or
            use 'MEAN')
```

c.
```
Choose:     2nd > STAT
            PLOT > 1:Plot1
Choose:     Window
Enter:      12.8, 27.2, 1.8,
            0, 40, 5, 1
Choose:     Trace > > >
Choose:     STAT > CALC >
            1:1-VAR STATS > 2nd > LIST
Select:     L7
```

7.28 a. Use una computadora para seleccionar al azar 200 muestras de tamaño 24 de una población normal con media $\mu = 20$ y desviación estándar $\sigma = 4.5$.

b. Encuentre la media $\bar{x}$ para cada una de las 200 muestras.

c. Usando las 200 medias muestrales, construya un histograma, encuentre la media $\bar{x}$, y encuentre la desviación estándar $s_{\bar{x}}$.

d. Compare los resultados de la parte c con los tres enunciados hechos en la SDSM y el teorema de límite central (CLT) de la página 370.

e. Compare estos resultados con los obtenidos en el ejercicio 7.27. Específicamente, ¿qué efecto tuvo el aumento en el tamaño muestral de 6 a 24? ¿Qué efecto tuvo el aumento de 100 muestras a 200 muestras?

PARA SU INFORMACIÓN Si usa computadora, vea el ejercicio 7.27.

7.4 Aplicación de la distribución de medias muestrales

Cuando la distribución de medias muestrales está normalmente distribuida, o normalmente distribuida en forma aproximada, podremos contestar preguntas de probabilidad con ayuda de la distribución normal estándar (tabla 3 del Apéndice B).

EJEMPLO 7.5 Conversión de información $\bar{x}$ en puntajes z

Considere una población normal con $\mu = 100$ y $\sigma = 20$. Si se selecciona una muestra aleatoria de 16, ¿cuál es la probabilidad de que esta muestra tenga un valor medio entre 90 y 110? Esto es, ¿cuál es $P(90 < \bar{x} < 110)$??

SOLUCIÓN Como la población está normalmente distribuida, la distribución muestral de las $\bar{x}$'s está normalmente distribuida. Para determinar probabilidades asociadas con una distribución normal, necesitaremos convertir el enunciado $P(90 < \bar{x} < 110)$ a un enunciado de probabilidad que comprenda el **puntaje *z*.** Esto nos permitirá usar la tabla 3 del Apéndice B, la tabla de distribución normal estándar. La distribución muestral se ilustra en la figura, donde el área sombreada representa $P(90 < \bar{x} < 110)$.

La fórmula para hallar el puntaje z correspondiente a un valor conocido de $\bar{x}$ es

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \tag{7.1}$$

La media y error estándar de la media son $\mu_{\bar{x}} = \mu$ y $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. Por tanto, escribiremos de nuevo la fórmula (7.1) en términos de μ, σ y n:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{7.2}$$

Regresando al ejemplo y aplicando la fórmula (7.2), encontramos:

puntaje z para $\bar{x} = 90$: $$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{90 - 100}{20/\sqrt{16}} = \frac{-10}{5} = \mathbf{-2.00}$$

puntaje z para $\bar{x} = 110$: $$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{110 - 100}{20/\sqrt{16}} = \frac{10}{5} = \mathbf{2.00}$$

Por tanto,

$$P(90 < \bar{x} < 110) = P(-2.00 < z < 2.00) = 2(0.4772) = \mathbf{0.9544}$$

Antes de ver más ejemplos, consideremos lo que se quiere dar a entender por $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. Para demostrarlo, supongamos que σ = 20 y utilicemos una distribución de muestras de tamaño 4. Ahora $\sigma_{\bar{x}}$ es $20/\sqrt{4}$, o 10, y aproximadamente 95% (0.9544) de todas las medias muestrales deben estar dentro del intervalo de 20 abajo y 20 arriba de la media poblacional (no más de 2 desviaciones estándar de la media poblacional). No obstante, si aumenta el tamaño muestral a 16, $\sigma_{\bar{x}}$ se convierte en $20/\sqrt{16} = 5$ y aproximadamente 95% de la distribución muestral debe estar dentro de 10 unidades de la media, y así sucesivamente. Cuando aumenta el tamaño muestral, el tamaño $\sigma_{\bar{x}}$ se hace más pequeño de modo que la distribución de medias muestrales se hace mucho más angosto. La figura 7.12 ilustra lo que ocurre a la distribución de $\bar{x}$'s cuando aumenta el tamaño de las muestras individuales.

FIGURA 7.12
Distribuciones de medias muestrales

Recuerde que el área (probabilidad) bajo la curva normal es siempre exactamente 1. En consecuencia, a medida que se reduce el ancho de la curva, la altura tiene que aumentar para mantener esta área.

EJEMPLO 7.6 Cálculo de probabilidades para la estatura media de niños de kindergarten

Los niños de *kindergarten* tienen estaturas que en forma aproximada están normalmente distribuidas, alrededor de una media de 39 pulgadas y una desviación estándar de 2 pulgadas. Se toma una muestra aleatoria de tamaño 25 y se calcula la media $\bar{x}$. ¿Cuál es la probabilidad de que este valor medio se encuentre entre 38.5 y 40.0 pulgadas?

SOLUCIÓN Buscamos hallar $P(38.5 < \bar{x} < 40.0)$. Los valores de $\bar{x}$, 38.5 y 40.0, deben convertirse en puntajes z (necesario para uso de la tabla 3 del apéndice B) usando $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$:

$$\bar{x} = 38.5: \qquad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{38.5 - 39.0}{2/\sqrt{25}} = \frac{-0.5}{0.4} = \mathbf{-1.25}$$

$$\bar{x} = 40.0: \qquad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{40.0 - 39.0}{2/\sqrt{25}} = \frac{1.0}{0.4} = \mathbf{2.50}$$

Por tanto,

$$P(38.5 < \bar{x} < 40.0) = P(-1.25 < z < 2.50) = 0.3944 + 0.4938 = \mathbf{0.8882}$$

EJEMPLO 7.7 Cálculo de límites de estatura media para el 90% central de niños de kindergarten

Use las estaturas de niños de *kindergarten* dadas en el ejemplo 7.6. ¿Dentro de qué límites cae el 90% central de la distribución de medias muestrales para muestras de tamaño 100?

SOLUCIÓN Las dos herramientas con las que tenemos que trabajar son la fórmula (7.2) y la tabla 3 del apéndice B. La fórmula relaciona los valores clave de la población con los valores clave de la distribución muestral, y la tabla 3 relaciona áreas con puntajes z. Primero, usando la tabla 3, encontramos que el 0.9000 central está limitado por $z = \pm 1.65$.

z	...	0.04		0.05	...
⋮					
1.6	...	0.4495	*0.4500*	0.4505	...
⋮					

PARA SU INFORMACIÓN
Recuerde: Si el valor está exactamente a la mitad, use la z más grande.

En segundo término, usamos la fórmula (7.2), $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$:

$$z = -1.65: \quad -1.65 = \frac{\bar{x} - 39.0}{2/\sqrt{100}}$$
$$\bar{x} - 39 = (-1.65)(0.2)$$
$$\bar{x} = 39 - 0.33$$
$$= \mathbf{38.67}$$

$$z = 1.65: \quad 1.65 = \frac{\bar{x} - 39.0}{2/\sqrt{100}}$$
$$\bar{x} - 39 = (1.65)(0.2)$$
$$\bar{x} = 39 + 0.33$$
$$= \mathbf{39.33}$$

Entonces,

$$\boldsymbol{P(38.67 < \bar{x} < 39.33) = 0.90}$$

Por tanto, 38.67 pulgadas y 39.33 pulgadas son los límites que captan el 90% central de las medias muestrales.

SECCIÓN 7.4 EJERCICIOS

7.29 Considere una población normal con $\mu = 43$ y $\sigma = 5.2$. Calcule el puntaje z para una $\bar{x}$ de 46.5 de una muestra de tamaño 16.

7.30 Considere una población con $\mu = 43$ y $\sigma = 5.2$.

a. Calcule el puntaje z para una $\bar{x}$ de 46.5 de una muestra de tamaño 35.

b. ¿Podría usarse este puntaje z en el calculo de probabilidades usando la tabla 3 del Apéndice B? ¿Por qué sí o por qué no?

7.31 En el ejemplo 7.5, explique cómo se obtuvo 0.4772 y qué es.

7.32 ¿Cuál es la probabilidad de que la muestra de niños de *kindergarten* del ejemplo 7.6 tenga una estatura media de 39.75 pulgadas?

7.33 Se ha de Seleccionar una muestra aleatoria de tamaño 36 de una población que tiene una media $\mu = 50$ y una desviación estándar σ de 10.

a. Esta muestra de 36 tiene un valor medio de $\bar{x}$, que pertenece a una distribución muestral. Encuentre la forma de esta distribución muestral.

b. Encuentre la media de esta distribución muestral.

c. Encuentre el error estándar de esta distribución muestral.

d. ¿Cuál es la probabilidad de que esta media muestral se encuentre entre 45 y 55?

e. ¿Cuál es la probabilidad de que la media muestral tenga un valor mayor a 48?

f. ¿Cuál es la probabilidad de que la media muestral se encuentre a no más de 3 unidades de la media?

7.34 La panadería local hornea diariamente más de mil piezas de pan de una libra, y los pesos de estas piezas varían. El peso medio es 1 libra y 1 onza, o 482 gramos. Suponga que la desviación estándar de los pesos es 18 gramos y se selecciona al azar una muestra de 40 piezas.

a. Esta muestra de 40 tiene un valor medio de $\bar{x}$, que pertenece a una distribución muestral. Encuentre la forma de esta distribución muestral.

b. Encuentre la media de esta distribución muestral.

c. Encuentre el error estándar de esta distribución muestral.

d. ¿Cuál es la probabilidad de que esta media muestral se encuentre entre 475 y 495?

e. ¿Cuál es la probabilidad de que la media muestral tenga un valor menos a 478?

f. ¿Cuál es la probabilidad de que la media muestral se encuentre dentro de 5 gramos de la media?

7.35 Considere la población aproximadamente normal de estaturas de estudiantes universitarios hombres con media $\mu = 69$ pulgadas y desviación estándar $\sigma = 4$ pulgadas. Se obtiene una muestra aleatoria de 16 estaturas.

a. Describa la distribución de x, estatura de estudiantes universitarios hombres.

b. Encuentre la proporción de estudiantes universitarios hombres cuya estatura sea mayor a 70 pulgadas.

c. Describa la distribución de $\bar{x}$, la media de muestras de tamaño 16.

d. Encuentre la media y desviación estándar de la distribución $\bar{x}$.

e. Encuentre $P(\bar{x} > 70)$. f. Encuentre $P(\bar{x} < 67)$.

7.36 La cantidad de llenado (peso de contenido), puesta en un frasco de vidrio de salsa para espagueti, está normalmente distribuida con una media de $\mu = 850$ gramos y desviación estándar $\sigma = 8$ gramos.

a. Describa la distribución de x, es decir, la cantidad de llenado por frasco.

b. Encuentre la probabilidad de que un frasco seleccionado al azar contenga entre 848 y 855 gramos.

c. Describa la distribución de $\bar{x}$, o sea el peso medio para una muestra de 24 de estos frascos de salsa.

d. Encuentre la probabilidad de que una muestra aleatoria de 24 frascos tenga un peso medio entre 848 y 855 gramos.

7.37 Las estaturas de los niños de *kindergarten* mencionados en el ejemplo 7.6 (p. 379) están normalmente distribuidos, en forma aproximada, con $\mu = 39$ y $\sigma = 2$.

a. Si al azar se selecciona un niño de *kindergarten*, ¿cuál es la probabilidad de que tenga una estatura entre 38 y 40 pulgadas?

b. Se usa un grupo de 30 de estos niños como muestra. ¿Cuál es la probabilidad de que la media del grupo, $\bar{x}$, sea entre 38 y 40 pulgadas?

c. Si al azar se selecciona un niño de *kindergarten* en particular, ¿cuál es la probabilidad de que mida más de 40 pulgadas?

d. Se usa un grupo de 30 de estos niños como muestra. ¿Cuál es la probabilidad de que la media del grupo, $\bar{x}$, sea de más de 40 pulgadas?

7.38 WageWeb (http://www.wageweb.com/health1.htm) es un servicio de HRPD1 y da información de compensación sobre más de 170 puestos de referencia en recursos humanos. El informe de octubre de 2003 indicó que los gerentes de relaciones laborales ganan un salario anual medio de $86,700. Suponga que los salarios anuales están normalmente distribuidos y tienen una desviación estándar de $8850.

a. ¿Cuál es la probabilidad de que un gerente de relaciones laborales seleccionado al azar haya ganado más de $100 000 en 2003?

b. Se toma una muestra de 20 gerentes de relaciones laborales y se informa de sus salarios anuales. ¿Cuál es la probabilidad de que el salario anual medio muestral caiga entre $80 000 y $90 000?

7.39 Con base en 53 años de datos compilados por el National Climatic Data Center (http://lwf.ncdc.noas.gov/oa/climate/online/ccd/avgwind.html), el promedio de velocidad de vientos en Honolulu, Hawai, es de 11.3 millas por hora, hasta junio de 2004. Suponga que las velocidades del viento están normalmente distribuidas, en forma aproximada, con una desviación estándar de 3.5 millas por hora.

a. Encuentre la probabilidad de que la velocidad del viento en cualquier lectura rebase las 13.5 mph.

b. Encuentre la probabilidad de que la media de una muestra aleatoria de nueve lecturas rebase 13.5 mph.

c. ¿Piensa usted que la suposición de normalidad es razonable? Explique.

d. ¿Qué efecto piensa usted que la suposición de normalidad tuvo en las respuestas a las partes a y b? Explique.

7.40 TIMSS 2003 (Trends in Internacional Mathematics and Science Study) se concentró en los éxitos, en matemáticas y ciencias, de estudiantes de octavo grado en todo el mundo. Un total de 45 países (incluyendo Estados Unidos) participaron en el estudio. La calificación media del examen de matemáticas para estu-

diantes norteamericanos fue 504, con una desviación estándar de 88.

Fuente: http://nces.ed.gv/timss/TIMSS03Tables

Suponga que las calificaciones están normalmente distribuidas y que se toma una muestra de 150 estudiantes.

a. Encuentre la probabilidad de que la calificación media TIMSS, para un grupo de estudiantes de octavo grado seleccionados al azar, sea entre 495 y 510.

b. Encuentre la probabilidad de que la calificación media TIMSS, para un grupo de estudiantes de octavo grado seleccionados al azar, sea menor de 520.

c. ¿Piensa usted que la suposición de normalidad es razonable? Explique.

7.41 Según el artículo "Only in America" (Sólo en Estados Unidos) de junio de 2004 de *Reader's Digest*, la cantidad promedio que gasta una persona de 17 años en un baile de gala de su escuela es $638. Suponga que las cantidades gastadas están normalmente distribuidas con una desviación estándar de $175.

a. Encuentre la probabilidad de que el costo medio de asistir a un baile de gala de secundaria, para 36 estudiantes de 17 años seleccionados al azar, está entre $550 y $700.

b. Encuentre la probabilidad de que el costo medio de asistir a un baile de gala de secundaria, para 36 estudiantes de 17 años seleccionados al azar, es mayor de $750.

c. ¿Piensa usted que la suposición de normalidad es razonable? Explique.

7.42 WageWeb (http://www.wageweb.com/health1.htm) da información de compensación y servicios sobre más de 160 posiciones. Hasta el 1 de octubre, 2003, el salario promedio nacional para una enfermera registrada (RN) era de $47,858. Suponga que la desviación estándar es $7750.

a. Encuentre la probabilidad de que la media de una muestra de 100 de estas enfermeras sea menos de $45 000.

b. Encuentre la probabilidad de que la media muestral de una muestra de 100 de estas enfermeras sea entre $46 000 y $48 000.

c. Encuentre la probabilidad de que la media muestral de una muestra de 100 de estas enfermeras sea mayor a $50 000.

d. Explique por qué la suposición de normalidad acerca de la distribución de salarios no estuvo comprendida en la solución a las partes a-c.

7.43 Con referencia al Ejemplo 7.6 (p. 379), ¿qué estatura limitaría al 25% inferior de todas las muestras de tamaño 25?

7.44 Se seleccionó una popular linterna eléctrica que usa dos baterías tamaño D, y varios de los mismos modelos fueron comprados para probar la "vida de uso continuo" de baterías D. Cuando se instalaron baterías nuevas, cada una de las linternas se encendió y se tomó nota del tiempo. Cuando la linterna ya no producía luz, se anotó de nuevo el tiempo. Los datos resultantes de la "vida" de baterías Rayovac tuvo una media de 21.0 horas (*fuente:* http://www.rayovac.com). Suponga que estos valores tienen una distribución normal con una desviación estándar de 1.38 horas.

a. ¿Cuál es la probabilidad de que una batería Rayovac Seleccionada al azar tenga una vida de prueba de entre 20.5 y 21.5 horas?

b. ¿Cuál es la probabilidad de que cuatro baterías Rayovac seleccionadas al azar tengan una vida de prueba de entre 20.5 y 21.5 horas?

c. ¿Cuál es la probabilidad de que 16 baterías Rayovac seleccionadas al azar tengan una vida de prueba de entre 20.5 y 21.5 horas?

d. ¿Cuál es la probabilidad de que 64 baterías Rayovac seleccionadas al azar tengan una vida de prueba de entre 20.5 y 21.5 horas?

e. Describa el efecto que tuvo el aumento el tamaño de la muestra para las partes b-d.

7.45 a. Encuentre $P(4 < \bar{x} < 6)$ para una muestra aleatoria de tamaño 4 extraída de una población normal con $\mu = 5$ y $\sigma = 2$.

b. Use computadora para general al azar 100 muestras, cada una de tamaño 4, de una distribución de probabilidad con $\mu = 5$ y $\sigma = 2$. Calcule la media, $\bar{x}$, para cada muestra.

c. ¿Cuántas de las medias muestrales de la parte b tienen valores entre 4 y 6? ¿Qué porcentaje es ese?

d. Compare las respuestas a las partes a y c, y explique cualesquiera diferencias que se presenten.

MINITAB (Versión 14)

a. Introduzca los números 4 y 6 en C1. Use los comandos CUMULATIVE NORMAL PROBABILITY DISTRIBUTION de la página 329, sustituyendo la media con 5, la desviación estándar con 1 ($2/\sqrt{4}$), la columna de entrada con C1, y la memoria opcional con C2. Encuentre CDF(6)-CDF(4)

b. Use los comandos Normal RANDOM DATA de la página 327, sustituyendo generar con 100, guardar con C3-C6, media con 5, y desviación estándar con 2. Use los comandos ROW STATISTICS de la página 368, sustituyendo variables de entrada con C3-C6 y guardar resultado en con C7.

c. Use los comandos HISTOGRAM de la página 61 para los datos en C7. Seleccione Labels, Data Labels, Label Type; use niveles de valor *y*. Para ajustar el histograma, seleccione Binning with midpoint y midpoint positions 0:10/1.

Excel

a. Introduzca los números 4 y 6 en la columna A. Active la celda B1. Use los comandos CUMULATIVE NORMAL DISTRIBUTION de la página 329, sustituyendo X con A1:A2. Encuentre CDF(6)-CDF(4).

b. Use los comandos Normal RANDOM NUMBER GENERATION de la página 328, sustituyendo el número de variables con 4, el número de números aleatorios con 100, media con 5, desviación estándar con 2, y rango de salida con C1. Active la celda G1. Use los comandos AVERAGE INSERT FUNCTION del ejercicio 7.13b de la página 368, sustituyendo Number1 con C1:F1.

c. Use los comandos RANDOM NUMBER GENERATION Patterned Distribution del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 0, el último valor con 9, los intervalos con 1, y el rango de salida con H1. Use los comandos HISTOGRAM de la página 61 con la columna G como el rango de entrada, la columna H como el rango de depósito, y la columna I como el rango de salida.

TI-83/84 Plus

a. Use los comandos CUMULATIVE NORMAL PROBABILITY de la página 330, sustituyendo Enter con 4,6,5,1). (La desviación estándar es 1; de $2/\sqrt{4}$.)

b. Use los comandos Normal RANDOM DATA y STO de la página 328, sustituyendo el Enter con 5,2,100). Repita estos comandos tres veces más, guardando datos en L2, L3 y L4, respectivamente.

```
Choose:      STAT > EDIT > 1:Edit
Highlight:   L5 (column heading)
Enter:       (L1+L2+L3+L4)/4
```

c. Use los comandos HISTOGRAM y TRACE de la página 62 para contar. Introduzca 0,9,1 0,45,1 para la Ventana.

7.46 a. Encuentre $P(46 < \bar{x} < 55)$ para un tamaño muestral aleatorio 16 sacada de una población normal con media de $\mu = 50$ y desviación estándar $\sigma = 10$.

b. Use computadora para generar en forma aleatoria 200 muestras, cada una de tamaño 16, de una distribución de probabilidad normal con media $\mu = 50$ y desviación estándar $\sigma = 10$. Calcule la media, $\bar{x}$, para cada muestra.

c. ¿Cuántas de las medias muestrales de la parte b tienen valores entre 46 y 55? ¿Qué porcentaje es ese?

d. Compare las respuestas a las partes a y c, y explique cualesquiera diferencias que se presenten.

PARA SU INFORMACIÓN Si usa computadora, vea el ejercicio 7.45.

REPASO DEL CAPÍTULO

En retrospectiva

En los capítulos 6 y 7 hemos aprendido a usar la distribución de probabilidad normal estándar. Ahora tenemos dos fórmulas para calcular un puntaje z:

$$z = \frac{x - \mu}{\sigma} \quad \text{y} \quad z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

El lector debe tener cuidado para distinguir entre estas dos fórmulas. La primera da el puntaje estándar cuando tenemos valores individuales de una distribución normal (valores x). La segunda fórmula se refiere a medias muestrales (valor $\bar{x}$). La clave para distinguir entre las fórmulas es decidir si el problema se refiera a una x individual o a una media muestral $\bar{x}$. Si se refiere a valores individuales de x, usamos la primera fórmula, como se presenta en el capítulo 6; si el problema se refiere a una media muestral, $\bar{x}$, usamos la segunda fórmula y procedemos como se ilustró en este capítulo.

El propósito fundamental para considerar lo que ocurre cuando una población se muestrea repetidamente, como se vio en este capítulo, es formar distribuciones muestrales. La distribución muestral se utiliza entonces para describir la variabilidad que ocurre de una muestra a la siguiente. Una vez que se conozca el patrón de variabilidad y se entienda para una estadística muestral específica, estamos en aptitud de hacer predicciones acerca del parámetro poblacional correspondiente con una medida de qué tan precisa es la predicción. La distribución de medias muestrales (SDSM) y el teorema de límite central ayudan a describir la distribución para medias muestrales. Empezaremos a hacer inferencias acerca de medias poblacionales en el capítulo 8.

Hay otras razones para muestreo repetido. Por lo común se usan muestras repetidas en el campo del control de producción, en el que se toman muestras para determinar si un producto es del tamaño o cantidad apropiados. Cuando la estadística muestral no satisface los estándares, es necesario un ajuste mecánico de la maquinaria. El ajuste es seguido entonces por otro muestreo para asegurarse que el proceso de producción está bajo control.

El "error estándar de la_____" es el nombre que se usa para la desviación estándar de la distribución muestral para cualquier estadística que se mencione en el espacio en blanco. En este capítulo nos hemos interesado del error estándar de la media. No obstante, también pudimos trabajar con el error estándar de la proporción, mediana o cualquier otra estadística.

El estudiante debe estar ahora familiarizado con el concepto de distribución muestral y, en particular, con la distribución de medias muestrales. En el capítulo 8 empezaremos a hacer pronósticos acerca de los valores de parámetros poblacionales.

Vocabulario y conceptos clave

teorema de límite central (p. 370)
distribución de frecuencia (p. 364)
distribución de probabilidad (p. 364)
muestra aleatoria (p. 365)
muestreo repetido (p. 366)
distribución muestral (p. 373)
distribución de medias muestrales (pp. 363, 369)
error estándar de la media (p. 370)
puntaje z (p. 377)

Objetivos de aprendizaje

✓ Entender lo que es una distribución de una estadística muestral y que la distribución se obtiene de muestras repetidas, todas del mismo tamaño.	pp. 363-364, EJ 7.1
✓ Ser capaz de formar una distribución muestral para una media, mediana, o rango con base en una población pequeña, finita.	EJ. 7.1, Ej. 7.6, 7.7
✓ Entender que una distribución muestral es una distribución de probabilidad para una estadística muestral.	EJ.7.2

✓ Entender y ser capaz de presentar y describir la distribución de medias muestrales y el teorema de límite central.	pp. 369-370, EJ. 7.4
✓ Entender y ser capaz de explicar la relación entre la distribución de medias muestrales y el teorema de límite central.	pp. 369-371, Ej. 7.17, 7.18, 7.21
✓ Determinar y ser capaz de explicar el efecto del tamaño muestral en el error estándar de la media.	pp. 373-375, Ej. 7.20, 7.26, 7.47
✓ Entender cuándo y cómo puede usarse la distribución normal para hallar probabilidades correspondientes a medias muestrales.	EJ. 7.5
✓ Calcular, describir e interpretar puntajes z correspondientes a valores conocidos de $\bar{x}$.	EJ. 7.6, EJ 7.7, Ej. 7.29, 7.30, 7.48
✓ Calcular puntajes z y probabilidades para aplicaciones de la distribución de medias muestrales	Ej. 7.33, 7.35

Ejercicios del capítulo

7.47 Si una población tiene una desviación estándar de 18.2 unidades, ¿cuál es el error estándar de la media si se seleccionan muestras de tamaño 9? ¿muestras de tamaño 25? ¿muestras de tamaño 49? ¿muestras de tamaño 100?

7.48 Considere una población normal con $\mu = 24.7$ y $\sigma = 4.5$.

a. Calcule el puntaje z para una x de 21.5.

b. Calcule el puntaje z para una $\bar{x}$ de 21.5 de una muestra de tamaño 25.

c. Explique cómo es que 21.5 puede tener puntajes z tan diferentes.

7.49 La Decana de Enfermería dice a estudiantes que se registran para el grupo de ingreso que, 1 año después de graduarse, las graduadas de universidad pueden esperar ganar un ingreso semanal medio de $675. Suponga que lo dicho por la Decana es verdadero y que los salarios semanales, 1 año después de la graduación, están normalmente distribuidos con una desviación estándar de $85. Si una graduada se selecciona al azar:

a. Describa la distribución de los salarios semanales percibidos 1 año después de la graduación.

b. ¿Cuál es la probabilidad de que la graduada seleccionada gane entre $550 y $825?

Si se selecciona una muestra aleatoria de 25 graduadas:

c. Describa la distribución de salarios medios semanales percibidos 1 año después de la graduación.

d. ¿Cuál es la probabilidad de que la media muestral sea entre $650 y $705?

e. ¿Por qué el puntaje z se usa para contestar las partes b y d?

f. ¿Por qué la fórmula para puntaje z se usa en la parte d diferente de la usada en la parte b?

7.50 Los diámetros de manzanas Rojas Deliciosas de cierta huerta están normalmente distribuidos con una media de 2.63 pulgadas y una desviación estándar de 0.25 de pulgada.

a. ¿Qué porcentaje de manzanas de esta huerta tienen diámetros menores a 2.25 pulgadas?

b. ¿Qué porcentaje de las manzanas de esta huerta es de más de 2.56 pulgadas de diámetro?

Se recolecta una muestra aleatoria de 100 manzanas, y el diámetro medio obtenido es $\bar{x} = 2.56$.

c. Si se toma otra muestra de 100, ¿cuál es la probabilidad de que la media muestral sea de más de 2.56 pulgadas?

d. ¿Por qué se usa el puntaje z para contestar las partes a-c?

e. ¿Por qué se usa la fórmula para puntaje z en la parte c diferente del que se usa en las partes a y b?

7.51 a. Encuentre un valor para e tal que 95% de las manzanas del ejercicio 7.50 se encuentren dentro de e unidades de la media, 2.63. Esto es, encuentre e tal que $P(2.63 - e < x < 2.63 + e) = 0.95$..

b. Encuentre un valor para E tal que 95% de las muestras de 100 manzanas tomadas de la huerta del ejercicio 7.50 tengan valores medios dentro de E unidades de la media, 2.63. Esto es, encuentre E tal que $P\,(2.63 - E < \bar{x} < 2.63 + E\,) = 0.95$..

7.52 Los norteamericanos gastan miles de millones de dólares en servicio de veterinarios cada año, pronosticado a llegar a $31 mil millones este año. Los servicios de salud ofrecidos a animales rivalizan con los que se dan a humanos, con una cirugía típica constante de $1700 a $3000 o más. En 2003, en promedio, los dueños de perros gastaron $196 en gastos de veterinarios en los 12 meses anteriores.

Fuente: American Pet Products Manufacturers Association

Suponga que el gasto anual de dueños de perros en servicio de veterinarios está normalmente distribuido con una media de $196 y una desviación estándar de $95.

a. ¿Cuál es la probabilidad de que un dueño de perro, seleccionado al azar de una población, gastara más de $300 en veterinario en 2003?

b. Suponga que se realiza una encuesta a 300 dueños de perros, y a cada uno se le pide informar el total de la cuenta pagada en 2003 en veterinarios. ¿Cuál es la probabilidad de que el gasto anual medio de esta muestra caiga entre $200 y $225?

c. Es probable que la suposición de una distribución normal en esta situación sea confusa. Explique por qué y qué efecto tuvo esto en las respuestas.

7.53 El gerente de una tienda Marketview, consciente de las estadísticas, registra el número de clientes que entran por la puerta todos los días. Años de registros muestran que el número medio de clientes por día es de 586, con una desviación estándar de 165. Suponga que el número de clientes está normalmente distribuido.

a. ¿Cuál es la probabilidad de que en cualquier día determinado, el número de clientes pase de 1000?

b. Si se seleccionan al azar 20 días, ¿cuál es la probabilidad de que la media de esta muestra sea menos de 550?

c. La suposición de normalidad nos permite calcular las probabilidades; no obstante, ésta puede no ser una suposición razonable. Explique por qué y cómo afecta eso a las probabilidades halladas en las partes a y b.

7.54 Un estudio de la Universidad de Michigan, como se ve en *Newsweek* (25 de marzo, 2002), expresó que los hombres promedian 16 horas de trabajo doméstico por semana (un alza de un promedio de 12 horas en 1965). Si suponemos que el número de horas en las que los hombres se ocupan del trabajo doméstico cada semana está normalmente distribuido con una desviación estándar de 5.4 horas, ¿cuál es la probabilidad de que el número medio de horas de trabajo doméstico, para una muestra de 20 hombres seleccionados al azar, sea entre 15 y 18 horas?

7.55 Un embarqu e de barras de acero será aceptado si la resistencia a la ruptura de una muestra aleatoria de 10 barras de acero es mayor a 250 libras por pulgada cuadrada. En años pasados, la resistencia a la ruptura de esas barras ha tenido una media de 235 y una varianza de 400.

a. Suponiendo que las resistencias a la ruptura están normalmente distribuidas, ¿cuál es la probabilidad de que una barra de acero seleccionada al azar tenga una resistencia a la ruptura en el rango de 245 a 255 libras por pulgada cuadrada?

b. ¿Cuál es la probabilidad de que el embarque sea aceptado?

7.56 El 15 de abril de 2002, un articulo de la revista *Time* dijo que la edad promedio para mujeres cuando se casan en Estados Unidos es ahora de 25 años. Si se supone que la desviación estándar es de 3.2 años, encuentre la probabilidad de que una muestra aleatoria de 40 mujeres norteamericanas muestren una edad media, al casarse, una edad menor o igual a 24 años.

7.57 Un fabricante de bombillas eléctricas dice que sus bombillas tienen una vida media de 700 horas, con una desviación estándar de 120 horas. Usted compra 144 de estas bombillas y decide que debe comprar más si la vida media de su muestra actual pasa de 680 horas. ¿Cuál es la probabilidad de que no compre otra vez de este fabricante?

7.58 Un fabricante de llantas dice (con base en años de experiencia con sus llantas) que la distancia media recorrida es de 35 000 millas y la desviación estándar es 5000 millas. Una agencia consumidora selecciona al azar 100 de estas llantas y encuentra una media muestral de 31 000. ¿Debe la agencia consumidora dudar del dicho del fabricante?

7.59 Para muestras grandes, la suma muestral ($\sum x$) tiene una distribución aproximadamente normal. La media de la suma muestral es $n \cdot \mu$ y la desviación estándar es $\sqrt{n} \cdot \sigma$. La distribución de ahorros por cuenta para una institución de ahorros y préstamos tiene una media igual a \$750 y una desviación estándar igual a \$25. Para una muestra de 50 de estas cuentas, encuentre la probabilidad de que la suma de las 50 cuentas pase de \$38 000.

7.60 Los pesos de equipaje para pasajeros que usan una línea aérea en particular están normalmente distribuidos con una media de 20 libras y una desviación estándar de 4 libras. Si el límite del peso total de equipaje es 2125 libras, ¿cuál es la probabilidad de que el límite sea rebasado por 100 pasajeros?

7.61 Una empresa de transportes por carretera entrega aparatos electrodomésticos para una gran operación de venta al menudeo. Los paquetes (o cajas) tienen un peso medio de 300 libras y una varianza de 2500.

a. Si un camión puede transportar 4000 libras y necesita recoger 25 aparatos, ¿cuál es la probabilidad de que los 25 aparatos tengan un peso agregado (global) mayor a la capacidad del camión? Suponga que los 25 aparatos representan una muestra aleatoria.

b. Si el camión tiene una capacidad de 8000 libras, ¿cuál es la probabilidad de que pueda transportar todo el lote de 25 aparatos?

7.62 Una empresa disquera de música pop desea la distribución de ciertas secciones de cortes de sus discos, para tener un promedio de 2 minutos y 15 segundos (135 segundos) y una desviación estándar de 10 segundos, para que los manipuladores de discos tengan tiempo suficiente para comerciales dentro de cada periodo de 5 minutos. La población de tiempos para cortes está normalmente distribuida, en forma aproximada, con sólo un sesgo insignificante a la derecha. Usted acaba de tomar el tiempo a los cortes en una nueva versión y ha encontrado que los 10 cortes promedian 140 segundos.

a. ¿Qué porcentaje del tiempo será el promedio de 140 segundos o más, si la nueva versión se selecciona al azar?

b. Si la empresa disquera desea 10 cortes para promediar 140 segundos menos que el 5% del tiempo, ¿cuál debe ser la media poblacional, dado que la desviación estándar permanece en 10 segundos?

7.63 Simulemos la distribución muestral relacionada con el interés del manipulador de discos para "duración de corte" del ejercicio 7.62.

a. Use computadora para generar al azar 50 muestras, cada una de tamaño 10, de una distribución normal con media de 135 y desviación estándar 10. Encuentre el "total muestral" y la media muestral para cada muestra.

b. Usando las 50 medias muestrales, construya un histograma y encuentre su media y desviación estándar.

c. Usando los 50 "totales" muestrales, construya un histograma y encuentre su media y desviación estándar.

d. Compare los resultados obtenidos en las partes b y c. Explique cualesquiera diferencias que se presenten.

MINITAB (Versión 14)

a. Use los comandos Normal RANDOM DATA de la página 327, sustituyendo generar con 50, guardar con C1-C10, media con 135, y desviación estándar con 10. Use los comandos ROW STATISTICS de la página 368, seleccionando Sum y sustituyendo variables de entrada con C1-C10 y guardar resultado en con C11. Use los comandos ROW STATISTICS, de nuevo seleccionando Mean y luego sustituyendo variables de entrada con C1-C10 y guardar resultado en con C12.

b. Use los comandos HISTOGRAM de la página 61 para los datos en C12. Para ajustar el histograma, seleccione Binning with midpoint. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos en C12.

c. Use los comandos HISTOGRAM de la página 61 para los datos en C11. Para ajustar el histograma, seleccione Binning with midpoints. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos en C11.

d. Use los comandos DISPLAY DESCRIPTIVE STATISTICS de la página 98 para los datos en C11 y C12.

Excel

a. Use los comandos Normal RANDOM NUMBER GENERATION de la página 328, sustituyendo número de variables con 10, número de números aleatorios con 50, media con 135, y desviación estándar con 10.

Active la celda K1.

```
Choose:  Insert function, fx > All > SUM > OK
Enter:   Number1: (A1:J1 or select cells)
Drag:    Bottom right corner of sum value box down to
         give other sums
```

Active la celda L1. Use los comandos AVERAGE INSERT FUNCTION del ejercicio 7.13b de la página 368, sustituyendo Number1 con A1:J1.

b. Use los comandos RANDOM NUMBER GENERACION Patterned Distribution del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 125.4, el último valor con 144.6, los intervalos con 3.2, y el rango de salida con M1. Use los comandos HISTOGRAM de la página 61 con la columna L como el rango de entrada y la columna M como el rango de depósito. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de la columna L.

c. Use los comandos RANDOM NUMBER GENERATION Patterned Distribution del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 1254, el último valor con 1446, los intervalos con 32, y el rango de salida con M20. Use los comandos HISTOGRAM de la página 61 con la columna L como el rango de entrada y las celdas M20-¿ como el rango de depósito. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de la columna K.

d. Use los comandos DESCRIPTIVE STATISTICS de la página 98 para los datos de las columnas K y L.

7.64 a. Encuentre la media y desviación estándar de x para una distribución de probabilidad binomial con $n = 16$ y $p = 0.5$.

b. Use computadora para construir la distribución de probabilidad e histograma para el experimento de probabilidad binomial con $n = 16$ y $p = 0.5$.

c. Use computadora para generar al azar 200 muestras de tamaño 25 de una distribución de probabilidad binomial con $n = 16$ y $p = 0.5$. Calcule la media de cada muestra.

d. Construya un histograma y encuentre la media y desviación estándar de las 200 medias muestrales.

e. Compare la distribución de probabilidad de x hallada en la parte b y la distribución de frecuencia de $\bar{x}$ en la parte d. ¿Su información apoya el teorema de límite central? Explique.

MINITAB (Versión 14)

a. Use los comandos MAKE PATTERNED DATA del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 0, el último valor con 16, y los intervalos con 1. Use los comandos BINOMIAL PROBABILITY DISTRIBUTIONS de la página 292, sustituyendo n con 16, p con 0.5, columna de entrada con C1, y memoria opcional con C2. Use los comandos Scatterplot with Connect Line de la página 155, sustituyendo Y con C2 y X con C1.

b. Use los comandos BVINOMIAL RANDOM DATA de la página 303, sustituyendo generar con 200, guardar con C3-C27, número de intentos con 16, y probabilidad con 0.5. Use los comandos ROW STATISTICS para una media de la página 368, sustituyendo variables de entrada con C3-C27 y guardar resultado en con C28. Use los comandos HISTOGRAM de la página 61 para los datos en C28. Para ajustar el histograma, selecciona Binning with midpoints. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos en C28.

Excel

a. Introduzca de 0 a 16 en la columna A. Continúe con los comandos de probabilidad binomial de la página 292, usando $n = 16$ y $p = 0.5$. Active las columnas A y B; luego continúe con:

```
Choose:  Chart Wizard > Column > 1st picture > Next >
         Series
Choose:  Series 1 > Remove
Enter:   Category (x)axis labels: (A1:A17 or select 'x
         value' cells)
Choose:  Next > Finish
```

b. Use los comandos Binomial RANDOM NUMBER GENERATION del ejercicio 5.97 de la página 304, sustituyendo número de variables con 25, número de números aleatorios con 200, valor p con 0.5, número de intentos con 16, y rango de salida con C1. Active la celda BB1. Use los comandos AVERAGE INSERT FUNCTION del ejercicio 7.13b de la página 368, sustituyendo Number1 con C1: AA1.

c. Use los comandos RANDOM NUMBER GENERACION Patterned Distribution del ejercicio 6.66 de la página 336, sustituyendo el primer valor con 6.8, el último valor con 9.2, los intervalos con 0.4, y el rango de salida con CC1. Use los comandos HISTOGRAM de la página 61 con la columna BB como el rango de entrada y la columna CC como el rango de depósito. Use los comandos MEAN y STANDARD DEVIATION de las páginas 74 y 88 para los datos de la columna BB.

7.65 a. Encuentre la media y desviación estándar de x para una distribución de probabilidad binomial con $n = 200$ y $p = 0.3$.

b. Use computadora para construir la distribución de probabilidad e histograma para la variable aleatoria x del experimento de probabilidad binomial con $n = 200$ y $p = 0.3$.

c. Use computadora para generar al azar 200 muestras de tamaño 25 de una distribución de probabilidad binomial con $n = 200$ y $p = 0.3$. Calcule la media $\bar{x}$ de cada muestra.

d. Construya un histograma y encuentre la media y desviación estándar de las 200 medias muestrales.

e. Compare la distribución de probabilidad de x hallada en la parte b y la distribución de frecuencia de $\bar{x}$ en la parte d. ¿Su información apoya al teorema de límite central? Explique.

PARA SU INFORMACIÓN Use los comandos del ejercicio 7.64, haciendo los ajustes necesarios.

7.66 Se selecciona al azar una muestra de 144 valores de una población con media , igual a 45 y desviación estándar, α, igual a 18.

a. Determine el intervalo (valor más pequeño al valor más grande) dentro del cual esperaría usted que se encuentre una media muestral.

b. ¿Cuál es la cantidad de desviación desde la media para una media muestral de 45.3?

c. ¿Cuál es la desviación máxima que ha tomado en cuenta usted en su respuesta a la parte a?

d. ¿Cómo está relacionada esta desviación máxima con el error estándar de la media?

Proyecto del capítulo

275 millones de norteamericanos

Como se dijo en "El problema del muestreo" de la Sección 7.1, "275 millones de norteamericanos" (p. 361), la meta fundamental de una encuesta es alcanzar los mismos resultados que hubiéramos obtenido si se entrevistara a todas las personas de la población. Saber que entrevistar a todas las personas de una población es casi imposible, para la mayor parte de poblaciones, promueve la importancia de una buena muestra representativa. Además, ahora tenemos la distribución de medias muestrales y el teorema de límite central para ayudarnos a hacer predicciones acerca de la población con el uso de la muestra. La siguiente sección Trabajando en el contenido del capítulo 7 nos permite reunir estos conceptos nuevos.

Trabajando en el contenido del capítulo 7

7.67 Se ha recolectado una segunda muestra de 100 edades del censo del año 200 de Estados Unidos y la presentamos aquí.

14	6	59	64	39	12	8	34	27	4
16	18	17	33	56	60	65	73	53	43
26	42	60	87	58	42	82	21	35	64
58	53	36	66	63	66	39	62	58	49
31	27	39	35	12	28	28	20	3	54
41	41	63	39	37	23	79	43	28	17
12	45	52	10	11	32	32	23	86	61
50	27	19	15	3	51	5	36	83	39
35	44	59	30	31	69	40	16	40	66
15	55	32	4	43	41	23	46	61	30

a. ¿Cómo describiría usted gráficamente los datos muestrales precedentes de "edades"? Construya la gráfica.

b. Usando la gráfica que construyó en la parte a, describa la forma de la distribución de datos muestrales.

c. ¿Qué tan bien es que la muestra describe la población de edades del censo de 2000 que se ve en la Sección 7.1? Explique usando pantallas gráficas.

d. ¿Cómo describiría usted numéricamente los precedentes datos muestrales de "edades"? Calcule las estadísticas.

e. ¿Cómo se comparan las estadísticas calculadas en la parte d con los parámetros del censo del año 2000 dados en la Sección 7.1?

f. (opcional) Si usted completó los ejercicios 7.1 y 7.2, ¿cómo se comparan su gráfica y estadísticas con las construidas y calculadas en los ejercicios 7.1 y 7.2, usando una muestra diferente de 100 edades?

g. ¿Es normal la distribución de edades para la población de norteamericanos en la Sección 7.1? ¿Es aproximadamente normal?

h. ¿Aplica la distribución de medias muestrales (SDSM) a muestras tomadas de esta población? Explique.

i. ¿Aplica el teorema de límite central (CLT) a muestras tomadas de esta población? Explique.

j. Describe la SDSM para muestras de tamaño 100. Asegúrese de incluir centro, dispersión y forma.

k. Compare sus resultados de las partes a y d con las respuestas teóricas de la parte j. Asegúrese de incluir centro, dispersión y forma.

l. Describa la SDSM para muestras de tamaño 30. Asegúrese de incluir centro, dispersión y forma.

m. Describa la SDSM de muestreo para muestras de tamaño 1000. Asegúrese de incluir centro, dispersión y forma.

n. Relacione lo que encuentre usted en las partes j, l y m con la SDSM y el CLT.

Para su estudio

7.68 El siguiente ejercicio consiste en simular que se toman muestras de tamaño 50 de la población de edades de norteamericanos del censo 2000, donde $\mu = 36.5$ y $\sigma = 22.5$ y la forma es sesgada a la derecha.

a. Dé un clic en "1" para # Simples." Observe los 50 valores de datos y su media. Cambie "lento" a "lote" y tome al menos 1000 muestras de tamaño 50.

b. ¿Cuál es la media de las medias muestrales? ¿Qué tan cercana está a la media poblacional?

c. ¿Cuál es la desviación estándar de las medias muestrales?

d. Con base en la SDSM (como se describe en la Sección 7.3), ¿qué debe esperarse para la desviación estándar de medias muestrales? ¿Qué tan cercana fue su desviación estándar respecto a la parte c?

e. ¿Qué forma tiene el histograma de las 1000 medias?

f. Relacione lo que encuentre usted con la SDSM y el CLT.

Examen de práctica del capítulo

PRIMERA PARTE: Conocimiento de las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

7.1 Una distribución muestral **es** una distribución que pone en lista todas las estadísticas muestrales que describen una muestra particular.

7.2 Los histogramas de **todas** las distribuciones muestrales son simétricas.

7.3 La media de la distribución muestral de las $\bar{x}$'s es igual a la media de la **muestra.**

7.4 El error estándar de la media es la desviación estándar de la población **de la cual se han tomado las muestras.**

7.5 El error estándar de la media **aumenta** cuando aumenta el tamaño muestral.

7.6 La forma de la distribución de medias muestrales es siempre la de una distribución **normal**.

7.7 Una distribución de **probabilidad** de una estadística muestral es una distribución de todos los valores de esa estadística que se obtuvieron de todas las muestras posibles.

7.8 La distribución de medias muestrales nos da una descripción de las tres características de una distribución muestral de **medianas** muestrales.

7.9 Una muestra de **frecuencia** se obtiene en forma tal que todas las muestras posibles de un tamaño dado tienen igual probabilidad de ser seleccionadas.

7.10 **No necesitamos** tomar muestras repetidas para usar el concepto de la distribución muestral.

SEGUNDA PARTE: Aplicación de conceptos

7.11 Se cree que las longitudes de truchas de lago en el Conesus Lake tienen una distribución normal con una media de 15.6 pulgadas y una desviación estándar de 3.8 pulgadas.

a. Kevin irá de pesca mañana al Conesus Lake. Si pesca una trucha, ¿cuál es la probabilidad de que mida menos de 15.0 pulgadas de largo?

b. Si el bote de pesca del capitán Brian llevará mañana 10 personas a pescar en el Conesus Lake, y pescan una muestra aleatoria de 16 truchas, ¿cuál es la probabilidad de que la longitud media de su pesca total mida menos de 15 pulgadas?

7.12 Se dice que los encendedores de cigarrillos fabricados por EasyVice Company tienen una vida útil media de 20 meses, con una desviación estándar de 6 meses. La garantía de devolución del dinero permite devolver el encendedor si no dura al menos 12 meses a partir de la fecha de compra.

a. Si la vida útil de estos encendedores está distribuida, ¿qué porcentaje de encendedores será devuelto a la compañía?

b. Si se prueba una muestra aleatoria de 25 encendedores, ¿cuál es la probabilidad de que la vida útil media muestral sea de más de 18 meses?

7.13 Se cree que los remaches de aluminio producidos por Rivets Forever, Inc. tienen resistencias al corte que están distribuidas alrededor de una media de 13.75, con una desviación estándar de 2.4. Si esta información es verdadera y se prueba una muestra de 64 de estos remaches en cuanto a su resistencia al corte, ¿cuál es la probabilidad de que la resistencia media sea entre 13.6 y 14.2?

TERCERA PARTE: Comprensión de conceptos

7.14 "Dos cabezas son mejores que una." Si esto es cierto, entonces ¿qué tan buenas serían varias cabezas? Para averiguarlo, una profesora de estadística trazó una línea en el pizarrón y pidió a su grupo que estimaran la longitud de esa línea a la pulgada más cercana. Ella recolectó las estimaciones, que variaron de 33 a 61 pulgadas, y calculó el valor medio. Reportó que la media fue de 42.25 pulgadas y luego midió la línea que resultó ser de 41.75 pulgadas de largo. ¿Demuestra esto que "varias cabezas son mejores que una"? ¿Qué teoría estadística apoya este suceso? Explique.

7.15 La distribución de medias muestrales es más que sólo una distribución de los valores medios que se presentan desde numerosas muestras repetidas tomadas de la misma población. Describa qué otra condición específica debe satisfacerse para tener una distribución de medias muestrales.

7.16 El Estudiante A dice "Una distribución muestral de las desviaciones estándar nos dice cómo varía la desviación estándar de una muestra a otra." El Estudiante B dice "Una distribución poblacional dice eso." ¿Quién tiene razón? Justifique su respuesta.

7.17 El Estudiante A dice que es el "tamaño de cada muestra empleada" y el Estudiante B dice que es el "número de muestras empleadas" lo que determina la dispersión de una distribución muestral empírica. ¿Quién tiene razón? Justifique su elección.

Trabajando con sus propios datos

Poniendo la probabilidad a trabajar

La distribución de medias muestrales y el teorema de límite central son muy importantes para el desarrollo del resto de este curso. La prueba, que requiere el uso de cálculo, no está incluida en este libro pero la verdad de la SDSM y el CLT se puede demostrar teóricamente y por experimentación. Las siguientes actividades pueden ayudar a verificar ambos enunciados.

A La población

Considere la población teórica que contiene los números 0, 3 y 6 en iguales proporciones.

1 a. Construya la distribución de probabilidad teórica para la toma de un solo número, con restitución, de esta población.

b. Trace un histograma de esta distribución de probabilidad.

c. Calcule la media, μ, y la desviación estándar, σ, para esta población.

B La distribución muestral, teóricamente

Estudiemos la distribución muestral teórica formada por las medias de todas las posibles muestras de tamaño 3 que puedan sacarse de una población dada.

2. Construya una lista que muestra todas las posibles muestras de tamaño 3 que puedan ser sacadas de esta población. (Hay 27 posibilidades.)
3. Encuentre la media de cada una de las 27 posibles muestras de la lista de la respuesta a la pregunta 2.
4. Construya la distribución de probabilidad (la distribución muestral teórica de medias muestrales) para estas 27 medias muestrales.
5. Construya un histograma para esta distribución muestral de medias muestrales.
6. Calcule la media $\mu\bar{x}$ y el error estándar de la media $\sigma\bar{x}$ usando la distribución de probabilidad hallada en la pregunta 4.
7. Demuestre que los resultados hallados en las preguntas 1c, 5 y 6 apoyan las tres afirmaciones hechas por la distribución muestral de medias muestrales y el teorema de límite central. Cite valores específicos para apoyar sus conclusiones.

C La distribución muestral, empíricamente

Veamos ahora si la distribución muestral de medias muestrales y el teorema de límite central se pueden verificar empíricamente, es decir, ¿se cumple cuando la distribución muestral está formada por las medias muestrales que resultan de varias muestras aleatorias?

8. Saque una muestra aleatoria de tamaño 3 de la población dada. Haga una lista de su muestra de tres números y calcule la media para esta muestra.

Puede usar computadora para generar sus muestras. Puede tomar tres "etiquetas" idénticas numeradas 0, 3 y 6, ponerlas en un "sombrero," y sacar su muestra usando restitución entre cada toma. También puede usar dados; sea un 0 representado por 1 y 2; 3 por 3 y 4; y 6, por 5 y 6. También es posible usar números aleatorios para simular la toma de sus muestras, o bien, puede sacar su muestra de la lista de muestras aleatorias que aparecen al final de esta sección. Describa el método que decida usar. (Pida ayuda a su profesor.)

9. Repita la pregunta ocho 49 veces más, de modo que tenga un total de 50 medias muestrales que han resultado de muestras de tamaño 3.

10. Construya una distribución de frecuencia de las 50 medias muestrales halladas en las preguntas 8 y 9.
11. Construya un histograma de la distribución de frecuencia de medias muestrales observadas.
12. Calcule la media $\bar{\bar{x}}$ y desviación estándar $s_{\bar{x}}$, de la distribución de frecuencia formada por las 50 medias muestrales.
13. Compare los valores observados de $\bar{\bar{x}}$ y $s_{\bar{x}}$ con los valores de $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$. ¿Están de acuerdo? ¿La distribución empírica de $\bar{x}$ se parece a la teórica?

A continuación aparecen 100 muestras aleatorias de tamaño 3 que fueron generadas por computadora:

6 3 0	0 3 0	6 6 0	3 3 6	6 6 3	6 3 3
0 0 3	3 0 6	3 3 0	3 6 6	0 3 0	6 6 3
6 6 6	0 3 0	6 3 6	0 6 3	6 0 3	6 3 3
6 0 0	3 0 6	6 3 3	3 3 0	3 3 0	3 3 3
3 3 3	3 0 0	6 6 6	3 3 6	0 0 6	0 6 3
6 6 6	0 0 6	3 3 0	0 6 6	0 0 3	6 6 3
0 0 6	0 0 6	6 6 6	6 3 6	6 6 0	3 0 0
3 6 6	6 3 0	3 6 3	3 0 0	3 3 6	0 6 0
3 0 0	0 3 6	6 3 3	6 0 6	3 3 6	6 0 3
0 3 6	3 6 3	6 6 3	6 6 0	3 3 3	3 0 0
6 3 0	6 6 0	0 3 0	6 6 0	3 6 6	0 3 6
6 3 3	0 3 0	6 6 0	6 6 3	6 6 0	3 0 3
3 6 3	3 6 0	0 0 6	0 3 3	3 6 6	0 3 6
0 6 0	6 0 0	0 6 0	0 6 6	0 3 3	0 3 6
3 3 6	3 3 3	3 3 6	6 3 6	3 3 3	3 6 6
6 3 3	3 0 0	3 0 6	6 0 3	3 6 6	6 0 3
0 3 3	6 3 0	0 3 6	0 3 6		

CAPÍTULO

8 Introducción a la inferencia estadística

8.1 ¿La gente era menos alta en otros tiempos?

8.2 La naturaleza de la estimación

8.3 Estimación de la media μ (σ conocida)

8.4 La naturaleza de la prueba de hipótesis

8.5 Prueba de hipótesis para la media μ (σ conocida): un acercamiento al valor probabilístico

8.6 Prueba de hipótesis para la media μ (σ conocida): un enfoque clásico

© Christa Renee/Getty Images

8.1 ¿La gente era menos alta en otros tiempos?

¿LA GENTE ERA MENOS ALTA EN OTROS TIEMPOS?

La estatura promedio para un inglés de principios del siglo XVII era aproximadamente 5'6"; para una inglesa del siglo XVII era de 5'½". Si bien la estatura promedio en Inglaterra permaneció prácticamente sin cambio en los siglos XVII y XVIII, los colonizadores norteamericanos eran más altos. Los promedios para norteamericanos de la época actual son un poco más de 5'9" para hombres y 5'3¾" para mujeres. Las razones principales para esta diferencia son una mejor nutrición, consumo notablemente mayor de carne y leche, y uso de antibióticos.

Fuente: http://www.plimoth.org/Library/l-short.htm

El National Center for Health Statistics (NCHS) da información estadística que guía acciones y políticas para mejorar la salud del pueblo norteamericano. Datos recientes del NCHS dan la estatura promedio de mujeres en Estados Unidos de 63.7 pulgadas, con una desviación estándar de 2.75 pulgadas.

Una muestra aleatoria de 50 mujeres profesionales de la salud dieron las siguientes estaturas.

65.0	66.0	64.0	67.0	59.0	69.0	66.0	69.0	64.0	61.5
63.0	62.0	63.0	64.0	72.0	66.0	65.0	64.0	67.0	68.0
70.0	63.0	63.0	68.0	58.0	60.0	63.5	66.0	64.0	62.0
64.5	69.0	63.5	69.0	62.0	58.0	66.0	68.0	59.0	56.0
64.0	66.0	65.0	69.0	67.0	66.5	67.5	62.0	70.0	62.0

¿Espera usted que la media de esta muestra aleatoria de 50 mujeres sea exactamente igual a la media poblacional de 63.7 pulgadas dada por el NCHS? Si la media muestral es mayor a 63.7 pulgadas, ¿significa esto que ahora son incluso más altas? Después de completar el capítulo 8 sabremos cómo contestar estas preguntas; podemos investigar más a fondo "¿La gente era menos alta en otros tiempos?" en el proyecto del capítulo de la página 469.

SECCIÓN 8.1 EJERCICIOS

8.1 a. ¿Qué población fue muestreada para obtener los datos de estatura que aparecen en la sección 8.1?

b. Describa los datos muestrales usando la media y desviación estándar, así como cualquier otra estadística que ayude a describir la muestra. Construya un histograma y comente sobre la forma de la distribución.

8.2 a. ¿Cómo está relacionada la distribución de los datos de la estatura muestral, de la sección 8.1, con la distribución de la población y la distribución de medias muestrales?

b. Usando las técnicas del capítulo 7, encuentre los límites que acotarían el 90% central de la distribución de medias muestrales para muestras de tamaño 50, seleccionadas al azar, de la población de estaturas de mujeres con una media conocida de 63.7 pulgadas y una desviación estándar de 2.75 pulgadas.

c. En el histograma trazado en el ejercicio 8.1, trace una recta vertical en la media poblacional de 63.7 y trace un segmento de recta horizontal que muestre el intervalo hallado en la parte b. La media muestral hallada en la parte b del ejercicio 8.1, ¿cae en el intervalo? Explique lo que esto significa.

d. Usando las técnicas del capítulo 7, encuentre $P(\bar{x} \geq 64.7)$ para una muestra aleatoria de 50 sacada de una población con una media conocida de 63.7 pulgadas y una desviación estándar de 2.75 pulgadas. Explique lo que significa el valor resultante.

e. La muestra de 50 valores de datos de estatura, ¿parece pertenecer a la población describa por el NCHS? Explique.

El objetivo de la estadística inferencial es usar la información contenida en los datos muestrales para aumentar nuestro conocimiento de la población muestreada. Aprenderemos a hacer dos tipos de inferencias: (1) estimar el valor de un parámetro poblacional y (2) probar una hipótesis. La distribución de medias muestrales (SDSM) es la clave para hacer estas inferencias, como se muestra en la figura 8.1.

En este capítulo nos ocupamos de preguntas acerca de la media poblacional, usando dos métodos que suponen que el valor de la desviación estándar poblacional es una cantidad conocida. Esta suposición se ve raras veces en problemas prácticos, pero hará más sencilla nuestra primera mirada a las técnicas de inferencia.

FIGURA 8.1 Lugar donde la distribución muestral cabe en el proceso estadístico

8.2 La naturaleza de la estimación

¿SABÍA USTED...?

Chocohólicos

El chocolate proviene del árbol de cacao. Cada fruto semejante a un melón contiene de 20 a 50 granos. Se necesitan unos 400 granos para hacer una libra de chocolate. En Estados Unidos se consumen 11.6 libras de chocolate por persona por año.

Una empresa fabrica remaches para usarlos en la construcción de aviones. Una característica de extrema importancia es la "resistencia al corte" de cada remache. Los ingenieros de la compañía deben vigilar la producción para asegurarse que la resistencia al corte de los remaches satisfaga las especificaciones requeridas. Para lograr esto, toman una muestra y determinan la resistencia media al corte de la muestra. Con base en esta información muestral, la compañía puede estimar la resistencia media al corte para todos los remaches que produce.

Se selecciona una muestra aleatoria de 36 remaches, y cada uno de ellos se prueba en cuanto a su resistencia al corte. La media muestral resultante es $\bar{x} = 924.23$ lb. Con base en esta muestra, decimos que "pensamos que la resistencia media al corte de todos estos remaches es de 924.23 libras".

Notas:

1. La resistencia al corte es la fuerza requerida para romper un material en una acción "cortante". Obviamente, el fabricante no va a probar todos los remaches porque la prueba destruye cada uno de los remaches que se prueben. Por tanto, se prueban muestras y la información acerca de la muestra debe usarse para hacer inferencias acerca de la población de todos estos remaches.
2. En todo el capítulo 8 trataremos la desviación estándar, σ, como una cantidad conocida, o dada, y nos concentraremos en aprender los procedimientos para hacer inferencias estadísticas acerca de la media poblacional, μ. Por tanto, para continuar con la explicación de inferencias estadísticas, supondremos que $\sigma = 18$ para los remaches específicos descritos en nuestro ejemplo.

Estimación puntual para un parámetro: es un número designado para estimar un parámetro cuantitativo de una población, por lo general el valor de la correspondiente **estadística muestral**.

Esto es, la media muestral, $\bar{x}$, es la estimación puntual (valor de un solo número) para la media, μ, de la población muestreada. Para nuestro ejemplo de remaches, 924.23 es la estimación puntual para μ, la resistencia media al corte de todos los remaches.

La calidad de esta estimación puntual debe ser cuestionada. ¿Es exacta la estimación? ¿Es probable que la estimación sea alta? ¿O baja? ¿Otra muestra daría el mismo resultado? ¿Otra muestra daría una estimación de casi el mismo valor? ¿O un valor que es muy diferente? ¿Cómo se miden "casi el mismo" o "muy diferente"? La calidad de un procedimiento (o método) de estimación aumenta grandemente si la estadística muestral es *menos variable* e *insesgada.* La variabilidad de una estadística es medida por el error estándar de su distribución muestral. La media muestral se puede hacer menos variable al reducir su error estándar, $\sigma/\sqrt{n}$, lo cual requiere el uso de una muestra más grande debido a que cuando n aumenta, el error estándar disminuye.

Estadística insesgada: es una estadística muestral cuya distribución muestral tiene un valor medio igual al valor del parámetro poblacional que se estima. Una estadística que no es insesgada es una **estadística sesgada**.

La figura 8.2 ilustra el concepto de ser insesgado y el efecto de variabilidad de la estimación puntual. El valor A es el parámetro que se estima, y los puntos representan posibles valores estadísticos de la distribución muestral de la estadística. Si A representa la media poblacional verdadera, μ, entonces los puntos representan posibles medias muestrales de la distribución muestral $\bar{x}$.

FIGURA 8.2 Efectos de variabilidad y tendencia

La figura 8.2(a), (c),(d) y (f) muestran estadísticas sesgadas; (a) y (d) presentan distribuciones de muestreo cuyos valores medios son menores que el valor del parámetro, en tanto que (c) y (f) presentan distribuciones de muestreo cuyos valores medios son mayores que el parámetro. La figura 8.2(b) y (e) presentan distribuciones muestrales que parecen tener un valor medio igual al valor del parámetro; por tanto, son insesgados. La figura 8.2(a), (b) y (c) presentan más variabilidad, en tanto que (d), (e) y (f) muestran menos variabilidad en las distribuciones muestrales. El diagrama (e) representa la mejor situación, un estimador que es insesgado (en el objetivo) y tiene baja variabilidad (todos los valores son cercanos al objetivo).

La media muestral, $\bar{x}$, es una estadística insesgada porque el valor medio de la distribución de medias muestrales, $\mu_{\bar{x}}$, es igual a la media poblacional, μ. (Recuerde que la distribución de medias muestrales tiene una media $\mu_{\bar{x}} = \mu$.) Por tanto, la estadística muestral $\bar{x} = 924.23$ es una estimación puntual insesgada para la resistencia media de todos los remaches que se fabrican en nuestro ejemplo.

Las medias muestrales varían en valor y forman una distribución muestral en la que no todas las muestras resultan en valores $\bar{x}$ iguales a la media poblacional. Por tanto, no debemos esperar que esta muestra de 36 remaches produzca una estimación puntual (media muestral), que sea exactamente igual a la media μ de la población muestreada. Debemos, no obstante, esperar que la estimación puntual sea más bien cercana en valor a la media poblacional. La distribución de medias muestrales (SDSM) y el teorema de límite central (CLT) dan la información necesaria para describir qué tan cercana se espera que sea la estimación puntual, $\bar{x}$, a la media poblacional, μ.

Recuerde que aproximadamente 95% de la distribución normal está a no más de 2 desviaciones estándar de la media y que el CLT describe la distribución de medias muestrales como que es casi normal cuando las muestras son suficientemente grandes. En general, las muestras de tamaño 36 de poblaciones de variables como la resistencia de remaches se consideran suficientemente grandes. En consecuencia, debemos anticipar que 95% de todas las muestras aleatorias seleccionadas de una

población con media desconocida μ y desviación estándar $\sigma = 18$ tendrán medias $\bar{x}$ entre

$$\mu - 2(\sigma_{\bar{x}}) \quad \text{y} \quad \mu + 2(\sigma_{\bar{x}})$$

$$\mu - 2\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{y} \quad \mu + \left(\frac{\sigma}{\sqrt{n}}\right)$$

$$\mu - 2\left(\frac{18}{\sqrt{36}}\right) \quad \text{y} \quad \mu + 2\left(\frac{18}{\sqrt{36}}\right)$$

$$\mu - 6 \quad \text{y} \quad \mu + 6$$

Esto sugiere que 95% de todas las muestras aleatorias de tamaño 36 seleccionadas de la población de remaches debe tener una media $\bar{x}$ entre $\mu - 6$ y $\mu + 6$. La figura 8.3 muestra el 95% central de la distribución, las acotaciones del intervalo que cubre el 95% y la media μ.

FIGURA 8.3
Distribución muestral de $\bar{x}$'s, desconocida μ

Ahora reunamos toda esta información en la forma de un *intervalo de confianza.*

Estimación de intervalo: intervalo limitado por dos valores y empleado para estimar el valor de un parámetro poblacional. Los valores que limitan este intervalo son estadísticas calculadas de la muestra que se usa como la base para la estimación.

Nivel de confianza 1 − α: porción de todas las estimaciones de intervalo que incluyen el parámetro en estudio.

Intervalo de confianza: estimación de intervalo con un nivel especificado de confianza.

Para construir el intervalo de confianza, usaremos la estimación puntual $\bar{x}$ como el valor central de un intervalo, en forma muy semejante a como usamos la media μ como valor central para hallar el intervalo que capta el 95% central de la distribución $\bar{x}$ de la figura 8.3.

Para nuestro ejemplo de remaches, podemos hallar los límites a un intervalo centrado en $\bar{x}$:

$$\bar{x} - 2(\sigma_{\bar{x}}) \quad \text{a} \quad \bar{x} + 2(\sigma_{\bar{x}})$$

$$924.23 - 6 \quad \text{a} \quad 924.23 + 6$$

El intervalo resultante es 918.23 a 930.23

Este nivel de confianza asignado a este intervalo es aproximadamente 95%, o 0.95. Los límites del intervalo son dos múltiplos ($z = 2.0$) del error estándar de la media

muestral y, al ver la tabla 3 del apéndice B, podemos determinar con más precisión el nivel de confianza de 0.9544. Reuniendo toda esta información, expresamos la estimación como un intervalo de confianza: **918.23 a 930.23** *es el 95.44% intervalo de confianza para la resistencia media al corte de los remaches.* O bien, en forma abreviada: **918.23 A 930.23**, *el 95.44% de intervalo de confianza para* μ.

CASO PRÁCTICO 8.1 Géiser Old Faithful del parque Yellowstone

PARA SU INFORMACIÓN
Visite el sitio Old Faithful Web Cam. ¿Cuándo se ha predicho que ocurrirá la siguiente erupción?

Cortesía National Park Service, Yellowstone National Park, http://www.nps.gov/yell.oldfaithfulcam.htm

Bienvenido al Old Faithful WebCam.

Las predicciones para la hora de la siguiente erupción del Old Faithful son hechas por los guardias que usan una fórmula que toma en cuenta la duración de la erupción anterior. La fórmula ha demostrado ser precisa, más o menos 10 minutos, 90% de las veces. A las 10:35 A.M. del 20 de agosto de 2005, la predicción de la hora fijada para la siguiente erupción fue:

Erupción pronosticada: 11:27 A.M. +/– 10 minutos

Observe la hora en la que se tomó la fotografía: 11:20:48 A.M. (*Fuente:* http://www.nps.gov/yell/oldfaithfulcam.htm)

SECCIÓN 8.2 EJERCICIOS

8.3 Explique la diferencia entre una estimación puntual y una estimación de intervalo.

8.4 Identifique cada valor numérico por "nombre" (por ejemplo, media, varianza) y por símbolo (por ejemplo, $\bar{x}$):

a. La estatura media de 24 alumnas de secundaria es 4'11".

b. La desviación estándar para calificaciones de IQ es 16.

c. La varianza entre las calificaciones de examen en el examen de la semana pasada fue 190.

d. La estatura media de todos los cadetes que han entrada a West Point es 69 pulgadas.

8.5 Se obtuvo una muestra aleatoria de la cantidad pagada (en dólares) por un viaje en taxi del centro de la ciudad al aeropuerto:

15 19 17 23 21 17 16 18 12 18 20 22 15 18 20

Use los datos para hallar una estimación puntual para cada uno de los parámetros siguientes:

a. Media b. varianza c. desviación estándar

8.6 El número de autobombas propiedad del departamento de bomberos se obtuvo de una muestra aleatoria, tomada de los perfiles de departamentos de bomberos de todo Estados Unidos (*Firehouse*/junio 2003).

29 8 7 33 21 26 6 11 4 54 7 4

Use los datos para hallar una estimación puntual para cada uno de los parámetros siguientes:

a. Media b. varianza c. desviación estándar

8.7 En cada diagrama de la parte inferior de la página, I y II representan distribuciones muestrales de dos estadísticas que podrían usarse para estimar un parámetro. En cada caso, identifique la estadística que usted piensa sería el mejor estimador y describa por qué fue esa su elección.

8.8 Suponga que hay dos estadísticas que servirán como estimador para el mismo parámetro. Una de ellas es sesgada y la otra es insesgada.

a. Si todo lo demás es igual, explique por qué generalmente preferiría usted un estimador insesgado y no uno sesgado.
b. Si una estadística es insesgada, ¿es seguro que es un buen estimador? ¿Por qué sí o por qué no? ¿Qué otras consideraciones deben tomarse en cuenta?
c. Describa una situación que podría ocurrir en la que la estadística sesgada podría ser una mejor opción como estimador que la estadística insesgada.

8.9 El uso de una muestra enormemente grande no resuelve el problema de calidad para un estimador. ¿Qué problemas anticipa usted con muestras muy grandes?

8.10 Ser insesgada y tener una variabilidad pequeña son dos características deseables de una estadística si se va a usar como estimador. Describa la forma en que la distribución de medias muestrales (SDSM) trata estas dos propiedades cuando estime la media de una población.

8.11 Explique por qué el error estándar de medias muestrales es 3 para el ejemplo de remaches de la página 399.

8.12 a. Verifique que un nivel de confianza de 95% requiera un intervalo de 1.96 de la desviación estándar.
b. Verifique que el nivel de confianza para un intervalo de 2 desviaciones estándar sea 95.44%.

8.13 Encuentre el nivel de confianza asignado a una estimación de intervalo de la media formada usando los intervalos siguientes:

a. $\bar{x} - 1.28 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 1.28 \cdot \sigma_{\bar{x}}$
b. $\bar{x} - 1.44 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 1.44 \cdot \sigma_{\bar{x}}$
c. $\bar{x} - 1.96 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 1.96 \cdot \sigma_{\bar{x}}$
d. $\bar{x} - 2.33 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 2.33 \cdot \sigma_{\bar{x}}$

8.14 Encuentre el nivel de confianza asignado a una estimación de intervalo de la media formada usando los intervalos siguientes:

a. $\bar{x} - 1.15 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 1.15 \cdot \sigma_{\bar{x}}$
b. $\bar{x} - 1.65 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 1.65 \cdot \sigma_{\bar{x}}$
c. $\bar{x} - 2.17 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 2.17 \cdot \sigma_{\bar{x}}$
d. $\bar{x} - 2.58 \cdot \sigma_{\bar{x}}$ a $\bar{x} + 2.58 \cdot \sigma_{\bar{x}}$

Figura para el ejercicio 8.7

8.15 Una muestra de 25 de 174 proyectos de inversión reveló que 19 de ellos estaban valuados en $17,320 cada uno y 6 en $20 200 cada uno. De los datos de la muestra, estime el valor total de la inversión para todos los proyectos.

8.16 Un vendedor de estampillas desea comprar una colección de estampillas que se piensa contiene aproximadamente 7000 estampillas individuales y alrededor de 4000 sobres de primer día. Diseñe un plan que pueda usarse para estimar el valor de la colección.

8.17 Usando la información de la erupción del Old Faithful del caso práctico 8.1 de la página 400:

a. ¿Qué significa "11:27 A.M. +/– 10 minutos? Explique.

b. ¿Ocurrió la erupción durante el intervalo pronosticado?

c. ¿Qué significa "90% del tiempo"? Explique.

8.18 Una buscadora de personal estima que si usted es contratado para trabajar para la compañía de ella, y se dedica toda una semana a trabajar en la posición de representante de ventas por comisiones que ella ofrece, usted ganará "$525 más o menos $250, 80% del tiempo". Ella agrega, "¡Todo depende de usted!".

a. ¿Qué significa "$525 más o menos $250"?

b. ¿Qué significa "80% del tiempo"?

c. Si usted gana $300 a los $10 más cercanos casi todas las semanas, ¿ella le habrá dicho la verdad? Explique.

8.3 Estimación de la media μ (σ conocida)

En la sección 8.2 exploramos las ideas básicas de la estimación: estimación puntual, estimación de intervalo, nivel de confianza e intervalo de confianza. Estas ideas básicas están relacionadas entre sí y se usan en toda la estadística cuando una inferencia pide una estimación. En esta sección formalizamos el proceso de estimación de intervalo, como se aplica para estimar la media poblacional μ, con base en una muestra aleatoria bajo la restricción de que la desviación estándar poblacional σ es un valor conocido.

La distribución de medias muestrales y el CLT nos da la información que necesitamos para asegurar que se satisfagan las *suposiciones* necesarias para estimar una media poblacional.

La suposición para estimar una media μ usando una σ conocida: La distribución muestral de $\bar{x}$ tiene una distribución normal.

Nota: la palabra *suposiciones* es un tanto inapropiada. No quiere decir que "suponemos" que algo es la situación y continuamos, sino que debemos estar seguros que las condiciones expresadas por las suposiciones existan antes que apliquemos un método estadístico particular.

La información necesaria para asegurar que esta suposición (o condición) se satisface está contenida en distribución de medias muestrales (SDSM) y en el teorema de límite central (CLT) (vea el capítulo 7, pp. 369-370):

La distribución de medias muestrales $\bar{x}$ está distribuida alrededor de una media igual a μ, con un error estándar igual a $\sigma/\sqrt{n}$; y (1) si la población muestreada al azar está normalmente distribuida, entonces $\bar{x}$ está normalmente distribuida para todos los tamaños muestrales, o (2) si la población muestreada al azar no está normalmente distribuida, entonces $\bar{x}$ está normalmente distribuida en forma aproximada para tamaños muestrales que sean lo suficiente grandes.

Por tanto, podemos satisfacer la suposición requerida ya sea (1) sabiendo que la población muestreada está normalmente distribuida o (2) usando una muestra aleatoria que contenga una cantidad suficientemente grande de datos. La primera posibilidad es obvia. O sabemos lo suficiente acerca de la población para saber que está normalmente distribuida, o no lo sabemos. La segunda forma de satisfacer la suposición es al aplicar el CLT. Una inspección de varias gráficas de los datos muestrales debe dar una indicación del tipo de distribución que posee la población. El CLT se puede aplicar a muestras más pequeñas (por ejemplo $n = 15$ o mayores), cuando los datos dan una fuerte indicación de una distribución unimodal que en forma aproximada es simétrica. Si hay evidencia de alguna asimetría en los datos, entonces el tamaño muestral necesita ser mucho mayor (quizá $n \geq 50$). Si los datos dan evidencia de una distribución extremadamente sesgada o en forma de J, el teorema de límite central (CLT) todavía aplicará si la muestra es suficientemente grande. En casos extremos "suficientemente grande" puede ser irreal o impracticablemente grande.

PARA SU INFORMACIÓN
Solicite ayuda de un profesional de estadística cuando trabaje con datos extremadamente sesgados.

Nota: no hay regla dura y rápida que defina "suficientemente grande"; el tamaño muestral que es "suficientemente grande" varía en gran medida según la distribución de la población.

El intervalo de confianza $1 - \alpha$ para la estimación de la media μ se encuentra usando la fórmula (8.1).

Intervalo de confianza para la media

$$\bar{x} - z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{a} \quad \bar{x} + z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) \tag{8.1}$$

FIGURA 8.4
Coeficiente de confianza $z(\alpha/2)$

A continuación veamos las partes de la fórmula del intervalo de confianza:

1. $\bar{x}$ es la estimación puntual y el punto central del intervalo de confianza.
2. $z(\alpha/2)$ es el **coeficiente de confianza.** Es el número de múltiplos del error estándar necesario para formular una estimación de intervalo del ancho correcto para tener un nivel de confianza de $1 - \alpha$. La figura 8.4 muestra la relación entre el nivel de confianza $1 - \alpha$ (la parte central de la distribución), $\alpha/2$ (el "área a la derecha" empleada con la notación de valor crítico), y el coeficiente de confianza $z(\alpha/2)$ (cuyo valor se encuentra usando la tabla 4B del apéndice B). Alfa, α, es la primera letra del alfabeto griego y representa la parte asociada con las colas de la distribución.
3. $\sigma/\sqrt{n}$ es el **error estándar de la media,** o la desviación estándar de la distribución de medias muestrales.
4. $z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right)$ es la mitad del ancho del intervalo de confianza (el producto del coeficiente de confianza y el error estándar) y se denomina **error máximo de estimación, *E*.**
5. $\bar{x} - z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right)$ recibe el nombre de **límite inferior de confianza** (LCL) y $\bar{x} + z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right)$ se llama **límite superior de confianza** (UCL) para el intervalo de confianza.

PARA SU INFORMACIÓN
Básicamente, el intervalo de confianza es "estimación puntual ± error máximo".

El procedimiento de estimación está organizado en un proceso de cinco pasos que tomará en cuenta toda la información precedente y producirá la estimación puntual y el intervalo de confianza.

EL INTERVALO DE CONFIANZA: UN PROCEDIMIENTO DE CINCO PASOS

Paso 1 El inicio:
Describir el parámetro poblacional de interés.

Paso 2 Criterios de intervalo de confianza:
a. Comprobar las suposiciones.
b. Identificar la distribución de probabilidad y la fórmula a usar.
c. Expresar el nivel de confianza, $1 - \alpha$.

Paso 3 La evidencia muestral:
Recolectar la información muestral.

Paso 4 El intervalo de confianza:
a. Determinar el coeficiente de confianza.
b. Hallar el error máximo de estimación.
c. Hallar los límites inferior y superior de confianza.

Paso 5 Los resultados:
Expresar el intervalo de confianza.

El ejemplo 8.2 ilustrará este procedimiento de cinco pasos para el intervalo de confianza.

EJEMPLO 8.2 Construir un intervalo de confianza para la distancia media de viaje suburbano en una dirección

El alumnado de los colegios de numerosas ciudades es considerado una "población de viajeros de cercanías". La oficina de actividades estudiantiles desea obtener una respuesta a la pregunta ¿Qué distancia viaja a diario (en una dirección) el estudiante promedio de una ciudad a su colegio? (En general, la "distancia de viaje de un estudiante promedio" es la "distancia media" recorrida por todos los estudiantes que viajan a cercanías.) Se identificó una muestra aleatoria de 100 estudiantes que viajan a cercanías, y se obtuvo la distancia que en una dirección viajan. La distancia media muestral resultante fue 10.22 millas.

Estime la distancia media en una dirección recorrida por todos los estudiantes de cercanías usando (a) la estimación puntual y (b) el intervalo de confianza de 95%. (Use $\sigma = 6$ millas.)

SOLUCIÓN

(a) La estimación puntual para la distancia media en una dirección es **10.22 millas** (la media muestral.)

(b) Usamos el procedimiento de cinco pasos para hallar el intervalo de confianza de 95%.

PASO 1 El inicio:
Describir el parámetro poblacional de interés.
La μ media de las distancias recorridas en una dirección por todos los estudiantes universitarios de cercanías es el parámetro de interés.

PASO 2 Criterios de intervalo de confianza:
a. Comprobar las suposiciones.
σ se conoce. Es muy probable que la variable "distancia recorrida" tenga una distribución sesgada porque la gran mayoría de estudiantes viajarán entre 0 y 25 millas, con menos de ellos que recorran más de 25 millas. Un tamaño muestral de 100 debe ser suficiente-

mente grande para que el teorema de límite central (CLT) satisfaga la suposición; la distribución muestral $\bar{x}$ es aproximadamente normal.

b. Identificar la distribución de probabilidad y la fórmula a usar.
La distribución normal estándar, z, se usará para determinar el coeficiente de confianza, y la fórmula (8.1) con $\sigma = 6$.

c. Expresar el nivel de confianza, $1 - \alpha$.
La pregunta pide 95% de confianza, o $1 - \alpha = 0.95$.

PASO 3 La evidencia muestral:
Recolectar la información muestral.
La información muestral se da en el enunciado del problema: $n = 100$, $x = 10.22$.

PASO 4 El intervalo de confianza:

a. Determinar el coeficiente de confianza.
El coeficiente de confianza se encuentra usando la tabla 4B:

Una parte de la tabla 4B

	α	...	0.05	
Nivel de confianza:	$z(\alpha/2)$	...	**1.96**	→ Coeficiente de confianza: $z(\alpha/2) = 1.96$
$1 - \alpha = 0.95$ 0	$1 - \alpha$	...	0.95	

b. Hallar el error máximo de estimación.
Use el error máximo, parte de la fórmula (8.1):

$$E = z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) = 1.96\left(\frac{6}{\sqrt{100}}\right) = (1.96)(0.6) = 1.176$$

c. Hallar los límites inferior y superior de confianza.
Usando la estimación puntual, $\bar{x}$, del paso 3 y el error máximo, E, del paso 4, encontramos los límites del intervalo de confianza:

$$\bar{x} - z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{a} \quad \bar{x} + z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$10.22 - 1.176 \quad \text{a} \quad 10.22 + 1.176$$

$$9.044 \quad \text{a} \quad 11.396$$

$$9.04 \quad \text{a} \quad 11.40$$

PASO 5 Los resultados:
Expresar el intervalo de confianza.
9.04 a 11.40, el 95% de intervalo de confianza para μ. Esto es, con 95% de confianza podemos decir, "La distancia media en una dirección es entre 9.04 y 11.40 millas".

Veamos otro ejemplo del procedimiento de estimación.

EJEMPLO 8.3 Construcción de un intervalo de confianza para el tamaño medio de una partícula

El "tamaño de partícula" es una propiedad importante en pintura de látex y se supervisa durante la producción como parte del proceso de control de calidad. Se tomaron trece mediciones de tamaño de partícula usando el disco de Dwight P. Joyce, y la media muestral fue de 3978.1 angstroms (donde 1 angstrom [1 Å] = 10^{-8} cm). El tamaño de partícula, x, está normalmente distribuido con una desviación estándar $\sigma = 200$ angstroms. Encuentre el 98% de intervalo de confianza para el tamaño medio de partícula para este lote de pintura.

SOLUCIÓN

PASO 1 El inicio:

Describir el parámetro poblacional de interés.

El tamaño medio de partícula, μ, para el lote de pintura del cual se sacó la muestra.

PASO 2 Criterios de intervalo de confianza:

a. Comprobar las suposiciones.

σ se conoce. La variable "tamaño de partícula" está normalmente distribuida; por tanto, la distribución de medias muestrales es normal para todos los tamaños muestrales.

b. Identificar la distribución de probabilidad y la fórmula a usar.

La variable normal estándar, z, y la fórmula (8.1) con $\sigma = 200$.

c. Expresar el nivel de confianza, $1 - \alpha$.

98%, o $1 - \alpha = 0.98$.

PASO 3 La evidencia muestral.

Recolectar la información muestral: $n = 13$ y $\bar{x} = 3978.1$.

PASO 4 El intervalo de confianza:

a. Determinar el coeficiente de confianza.

El coeficiente de confianza se encuentra usando la tabla 4B: $z(\alpha/2) = z(0.01) = 2.33$.

Una parte de la tabla 4B

	α	...	0.02	
	$z(\alpha/2)$	...	**2.33**	→ Coeficiente de confianza: $z(\alpha/2) = \mathbf{2.33}$
Nivel de confianza: $1 - \alpha = 0.98$ →	$1 - \alpha$	...	0.98	

b. Encontrar el error máximo de estimación.

$$E = z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) = 2.33\left(\frac{200}{\sqrt{13}}\right) = (2.33)(55.47) = 129.2$$

c. Encontrar los límites inferior y superior de confianza.

Usando la estimación puntual, $\bar{x}$, del paso 3 y el error máximo, E, del paso 4b, encontramos los límites de intervalo de confianza:

$$\bar{x} - z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{a} \quad \bar{x} + z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$3978.1 - 129.2 = 3848.9 \quad \text{a} \quad 3978.1 + 129.2 = 4107.3$$

PASO 5 Los resultados:
Expresar el intervalo de confianza.
3848.9 a 4107.3, el 98% de intervalo de confianza para μ. Con 98% de confianza podemos decir, "El tamaño medio de partícula es entre 3848.9 y 4107.3 angstroms".

Veamos de nuevo el concepto de "nivel de confianza". Se definió como la probabilidad de que la muestra a ser seleccionada producirá límites de intervalo que contienen el parámetro.

EJEMPLO 8.4 Demostración del significado de un intervalo de confianza

Los números aleatorios de un solo dígito, como los de la tabla 1 del apéndice B, tienen un valor medio $\mu = 4.5$ y una desviación estándar σ 2.87 (ver ejercicio 5.32, p. 282). Saque una muestra de 40 números de un solo dígito de la tabla 1 y construya el intervalo de confianza del 90% para la media. ¿El intervalo resultante contiene el valor esperado de μ, 4.5? Si fuéramos a seleccionar otra muestra de 40 números aleatorios de un solo dígito de la tabla 1, ¿obtendríamos el mismo resultado? ¿Qué podría ocurrir si seleccionamos un total de 15 muestras diferentes y construimos el intervalo de confianza del 90% para cada una? ¿Esperaríamos que el valor esperado para μ, es decir, 4.5, estuviera contenido en todas ellas? ¿Deberíamos esperar que los 15 intervalos de confianza contengan 4.5? Considere la definición de "nivel de confianza"; dice que, a la larga, 90% de las muestras resultarán en límites que contengan μ. En otras palabras, 10% de las muestras no contendrán μ. Veamos lo que ocurre.

Primero necesitamos tratar las suposiciones; si éstas no se satisfacen, no podemos esperar que ocurran el 90% y el 10%. Sabemos: (1) que la distribución de números aleatorios de un solo dígito es rectangular (definitivamente no normal), (2) que la distribución de números aleatorios de un solo dígito es simétrica alrededor de su media, (3) que la distribución $\bar{x}$ para muestras muy pequeñas ($n = 5$) del ejemplo 7.2 (pp. 364-365) mostró una distribución que pareció ser aproximadamente normal y (4) que no debe hacer asimetría. Por tanto, parece razonable suponer que $n = 40$ es suficientemente grande para que aplique el teorema de límite central (CLT).

La primera muestra aleatoria se sacó de la tabla 1 del apéndice B:

TABLA 8.1

Muestra aleatoria de números de un solo dígito

2	8	2	1	5	5	4	0	9	1
0	4	6	1	5	1	1	3	8	0
3	6	8	4	8	6	8	9	5	0
1	4	1	2	1	7	1	7	9	3

Las estadísticas muestrales son $n = 40$, $\sum x = 159$, y $\bar{x} = 3.98$. He aquí el intervalo de confianza de 90% resultante:

$$\bar{x} \pm z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right): \quad 3.98 \pm 1.65\left(\frac{2.87}{\sqrt{40}}\right)$$

$$3.98 \pm (1.65)(0.454)$$

$$3.98 \pm 0.75$$

$$3.98 - 0.75 = 3.23 \quad \text{a} \quad 3.98 + 0.75 = 4.73$$

3.23 a 4.73, el intervalo de confianza de 90% para μ

La figura 8.5 muestra este intervalo de confianza, en límites, y la μ media esperada.

FIGURA 8.5
Intervalo de confianza del 90%

El valor esperado para la media, 4.5, cae dentro de los límites del intervalo de confianza para esta muestra. Seleccionemos ahora 14 muestras aleatorias más de la tabla 1 del apéndice B, cada una de tamaño 40.

La tabla 8.2 es una lista de la media desde la primera muestra y las medias obtenidas desde las 14 muestras aleatorias adicionales de tamaño 40. Los intervalos de confianza del 90% para la estimación μ basadas en cada una de las 15 muestras aparecen en la tabla 8.2, y se muestran en la figura 8.6.

TABLA 8.2

Quince muestras de tamaño 40

Número de muestra	Media muestral $\bar{x}$	Estimación de intervalo de confianza de 90% para μ	Número de muestra	Media muestral $\bar{x}$	Estimación de intervalo de confianza de 90% para μ
1	3.98	3.23 a 4.73	9	4.08	3.33 a 4.83
2	4.64	3.89 a 5.39	10	5.20	4.45 a 5.95
3	4.56	3.81 a 5.31	11	4.88	4.13 a 5.63
4	3.96	3.21 a 4.71	12	5.36	4.61 a 6.11
5	5.12	4.37 a 5.87	13	4.18	3.43 a 4.93
6	4.24	3.49 a 4.99	14	4.90	4.15 a 5.65
7	3.44	2.69 a 4.19	15	4.48	3.73 a 5.23
8	4.60	3.85 a 5.35			

FIGURA 8.6
Intervalos de confianza de la tabla 8.2

Vemos que 86.7% (13 de 15) de los intervalos contienen μ y 2 de las 15 muestras (muestra 7 y muestra 12) no contienen μ. Los resultados aquí son "típicos"; una experimentación repetida podría resultar en cualquier número de intervalos que contengan 4.5. No obstante, a la larga, debemos esperar aproximadamente que $1 - \alpha = 0.90$ (o 90%) de las muestras resulten en límites que contengan 4.5 y aproximadamente 10% que no contengan 4.5.

CASO PRÁCTICO 8.5 La nieve que se derrite de las Rocallosas produce menos agua

Cuando la nieve se derrite se convierte en agua, a veces más agua que en otros tiempos. La gráfica siguiente compara el contenido de agua de nieve de dos regiones de Estados Unidos, donde por lo general cae la misma cantidad de nieve anualmente, pero el contenido de agua es muy diferente. Hay varias estimaciones puntuales para el promedio incluido en la gráfica.

LA NIEVE QUE SE DERRITE DE LAS ROCALLOSAS PRODUCE MENOS AGUA

Datos de Marcia Staimer, USA TODAY; *fuente: The USA TODAY Weather Almanac*, disponible en librerías de Vintage Books; © 1994 USA TODAY reimpreso con permiso.

INSTRUCCIONES DE TECNOLOGÍA: INTERVALO DE CONFIANZA PARA μ MEDIA CON UNA σ DETERMINADA

MINITAB (Versión 14)

Introduzca los datos en C1; luego continúe con:

```
Choose:     Stat > Basic Statistics > 1-Sample Z
Enter:      Variables: C1
            Standard deviation: σ
Select:     Options
Enter:      Con?dence Level: 1 − α (ex.: 0.95 or 95.0)
Select:     Alternative: not equal > OK > OK
```

Excel Introduzca los datos en la columna A; luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Z-Test: Mean > OK
Enter:      Input Range: (A1:A20 or select cells) > OK
            Standard Deviation (SIGMA): σ > OK
            Alpha: α (ex.: 0.05) > OK
```

TI-83/84 PLUS Introduzca los datos en L1; luego continúe con lo siguiente, introduciendo los valores apropiados y destacando Calculate:

```
Choose:     STAT > TESTS > 7:Zinterval
```

Tamaño muestral

El intervalo de confianza tiene dos características básicas que determinan su calidad: su nivel de confianza y su ancho. Se prefiere que el intervalo tenga un alto nivel de confianza y sea preciso (angosto) al mismo tiempo. Cuanto más alto sea el nivel de confianza, es más probable que el intervalo contenga el parámetro, y cuanto más angosto el intervalo, más precisa es la estimación. No obstante, estas dos propiedades parecen trabajar una contra la otra, porque parecería que un intervalo más angosto tendería a tener una menor probabilidad y un intervalo más ancho sería más preciso. La parte de error máximo de la fórmula del intervalo de confianza especifica la relación de que se trata.

Error máximo de estimación

$$E = z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right) \tag{8.2}$$

Esta fórmula tiene cuatro componentes: (1) el error máximo E, de la mitad del ancho del intervalo de confianza; (2) el coeficiente de confianza, $z(\alpha/2)$, que está determinado por el nivel de confianza; (3) el tamaño muestral, n, y (4) la desviación estándar, σ. La desviación estándar σ no es de interés en ese análisis porque es una constante (la desviación estándar de una población no cambia en valor). Esto nos deja tres factores. Una inspección a la fórmula (8.2) indica lo siguiente: aumentar el nivel de confianza hará más grande el coeficiente de confianza y, por tanto, requiere que aumente el error máximo o que el tamaño muestral aumente; disminuir el error máximo requerirá que el nivel de confianza disminuya o que el tamaño muestral aumente, y disminuir el tamaño muestral forzará el error máximo a hacerse más grande o que disminuya el nivel de confianza. Tenemos un juego de "tirar de la cuerda en tres direcciones", como se ve en la figura 8.7.

FIGURA 8.7 "Tirar de la cuerda en tres direcciones" entre $1 - \alpha$, n, y E

PARA SU INFORMACIÓN
Cuando el denominador aumenta, el valor del quebrado disminuye.

Un aumento o disminución a cualquiera de los tres factores tiene un efecto en uno o en los otros dos factores. El trabajo del experto en estadística es "equilibrar" el nivel de confianza, el tamaño muestral, y el error máximo para que resulte un intervalo aceptable.

Veamos un ejemplo de esta relación en acción.

EJEMPLO 8.6 Determinación del tamaño muestral para un intervalo de confianza

PARA SU INFORMACIÓN
Las instrucciones para usar la tabla 4B se muestran en la página 403.

Determine el tamaño muestral necesario para estimar el peso medio de todos los niños de segundo grado, si deseamos tener una variación de no más de 1 libra con 95% de confianza. Suponga una distribución normal y que la desviación estándar de los pesos de los niños es 3 libras.

SOLUCIÓN El nivel deseado de confianza determina el coeficiente de confianza: el coeficiente de confianza se encuentra usando la tabla 4B: $z(\alpha/2) = z(0.025) =$ **1.96.**

El error máximo deseado es $E = 1.0$. Ahora estamos listos para usar la fórmula del error máximo:

$$E = z(\alpha/2)\left(\frac{\sigma}{\sqrt{n}}\right): \qquad 1.0 = 1.96\left(\frac{3}{\sqrt{n}}\right)$$

$$\text{Despejando la } n: \qquad 1.0 = \frac{5.88}{\sqrt{n}}$$

$$\sqrt{n} = 5.88$$

$$n = (5.88)^2 = 34.57 = \mathbf{35}$$

Por tanto, $\boldsymbol{n} = \mathbf{35}$ es el tamaño muestral necesario y deseamos un intervalo de confianza de 95% con un error máximo no mayor a una libra.

Nota: cuando despejamos el tamaño muestral n, se acostumbra redondear al siguiente entero más grande, sin importar qué fracción (o decimal) resulte.

El uso de la fórmula (8.2) de error máximo se puede hacer un poco más fácil si reescribimos la fórmula de modo que exprese n en términos de los otros valores.

Tamaño muestral

$$n = \left(\frac{z(\alpha/2) \cdot \sigma}{E}\right)^2 \qquad (8.3)$$

Si el error máximo se expresa como múltiplo de la desviación estándar σ, entonces el valor real de σ no se necesita para calcular el tamaño muestral.

EJEMPLO 8.7

Determinación del tamaño muestral sin un valor conocido de sigma (σ)

Encuentre el tamaño muestral necesario para estimar la media poblacional a no más de $\frac{1}{5}$ de una desviación estándar con 99% de confianza.

SOLUCIÓN Determine el coeficiente de confianza (usando la tabla 4B): $1 - \alpha = 0.99$, $z(\alpha/2) = 2.58$. El error máximo deseado es $E = \frac{\sigma}{5}$. Ahora estamos listos para usar la fórmula (8.3) de tamaño muestral:

$$n = \left(\frac{z(\alpha/2) \cdot \sigma}{E}\right)^2: \quad n = \left(\frac{(2.58) \cdot \sigma}{\sigma/5}\right)^2 = \left(\frac{(2.58\sigma)(5)}{\sigma}\right)^2 = [(2.58)(5)]^2$$
$$= (12.90)^2 = 166.41 = \mathbf{167}$$

SECCIÓN 8.3 EJERCICIOS

8.19 Discuta las condiciones que deben existir antes que podamos estimar la media poblacional usando las técnicas de intervalo de la fórmula (8.1).

8.20 Determine el valor del coeficiente de confianza $z(\alpha/2)$ para cada una de las situaciones descritas:

a. $1 - \alpha = 0.90$ b. $1 - \alpha = 0.95$

8.21 Determine el valor del coeficiente de confianza $z(\alpha/2)$ para cada una de las situaciones descritas:

a. 98% de confianza b. 99% de confianza

8.22 Determine el nivel de confianza dado el coeficiente de confianza $z(\alpha/2)$ para cada situación:

a. $z(\alpha/2) = 1.645$ b. $z(\alpha/2) = 1.96$

c. $z(\alpha/2) = 2.575$ d. $z(\alpha/2) = 2.05$

8.23 Dada la información, la población muestreada está normalmente distribuida, $n = 16$, $\bar{x} = 28.7$, y $\sigma = 6$:

a. Encuentre el 0.95 de intervalo de confianza para μ.

b. ¿Se satisfacen las suposiciones? Explique.

8.24 Dada la información, la población muestreada está normalmente distribuida, $n = 55$, $\bar{x} = 78.2$, y $\sigma = 12$:

a. Encuentre 0.98 de intervalo de confianza para μ.

b. ¿Se satisfacen las suposiciones? Explique.

8.25 Dada la información, $n = 86$, $\bar{x} = 128.5$, y $\sigma = 16.4$:

a. Encuentre 0.90 de intervalo de confianza para μ.

b. ¿Se satisfacen las suposiciones? Explique.

8.26 Dada la información, $n = 22$, $\bar{x} = 72.3$, y $\sigma = 6.4$:

a. Encuentre 0.99 de intervalo de confianza para μ.

b. ¿Se satisfacen las suposiciones? Explique.

8.27 Con base en el intervalo de confianza formado en el ejercicio 8.25, dé el valor para cada uno de lo siguiente:

a. Estimación puntual

b. Coeficiente de confianza

c. Error estándar de la media

d. Error máximo de estimación, E

e. Límite inferior de confianza

f. Límite superior de confianza

8.28 Con base en el intervalo de confianza formado en el ejercicio 8.24, dé el valor para cada uno de lo siguiente:

a. Estimación puntual

b. Coeficiente de confianza

c. Error estándar de la media

d. Error máximo de estimación, E

e. Límite inferior de confianza

f. Límite superior de confianza

8.29 Con sus propias palabras, describa la relación entre lo siguiente:

a. Media muestral y estimación puntual

b. Tamaño muestral, desviación estándar muestral y error estándar

c. Error estándar y error máximo

8.30 Con sus propias palabras, describa la relación entre la estimación puntual, el nivel de confianza, el error máximo y el intervalo de confianza.

8.31 El siguiente **ejercicio** demuestra el efecto que tiene el nivel de confianza $(1 - \alpha)$ en el ancho de un intervalo de confianza. Considere muestrear de una población donde $\mu = 300$ y $\sigma = 80$.

a. Ajuste el cursor para un nivel de confianza a 68%. Dé un clic en "sample!" para construir un intervalo de confianza de 68%. Observe los límites superior e inferior de confianza y calcule el ancho del intervalo. Usando "animate!" construya muchas muestras y tome nota del porcentaje de intervalos que contenga la media verdadera de 300. Dé un clic en "stop" y "reset".

b. Ajuste el cursor para un nivel de confianza a 95%. Dé un clic en "sample!" para construir un intervalo de confianza de 95%. Observe los límites superior e inferior de confianza y calcule el ancho del intervalo. Usando "animate!" construya muchas muestras y tome nota del porcentaje de intervalos que contenga la media verdadera de 300. Dé un clic en "stop" y "reset".

c. Ajuste el cursor para un nivel de confianza a 99%. Dé un clic en "sample!" para construir un intervalo de confianza de 99%. Observe los límites superior e inferior de confianza y calcule el ancho del intervalo. Usando "animate!" construya muchas muestras y tome nota del porcentaje de intervalos que contenga la media verdadera de 300. Dé un clic en "stop".

d. Usando la información recolectada en las partes a-c, ¿qué efecto tiene el nivel de confianza en el ancho del intervalo? ¿Por qué está ocurriendo?

8.32 Discuta el efecto que cada uno de lo siguiente tiene en el intervalo de confianza:

a. Estimación puntual

b. Nivel de confianza

c. Tamaño muestral

d. Variabilidad de la característica que se mide

8.33 Una máquina produce piezas con longitudes que están normalmente distribuidas con $\sigma = 0.5$. Una muestra de 10 piezas tiene una longitud media de 75.92.

a. Encuentre la estimación puntual para μ.

b. Encuentre el error máximo de 98% de confianza de estimación para μ.

c. Encuentre el intervalo de confianza de 98% para μ.

8.34 Se obtiene una muestra de edades de 60 estudiantes de una escuela nocturna para estimar la edad media de estudiantes de escuelas nocturnas. $\bar{x} = 25.3$ años. La varianza de la población es 16.

a. Dé una estimación puntual para μ.

b. Encuentre el intervalo de confianza de 95% para μ.

c. Encuentre el intervalo de confianza de 99% para μ.

8.35 Doscientos peces capturados en el Cayuga Lake tuvieron una longitud media de 14.3 pulgadas. La desviación estándar poblacional es 2.5 pulgadas.

a. Encuentre el intervalo de confianza de 90% para la longitud media poblacional.

b. Encuentre el intervalo de confianza de 98% para la longitud media poblacional.

8.36 El Eurostar fue el primer tren internacional de Europa, diseñado para aprovechar el túnel submarino que conecta Inglaterra con la Europa continental. Transporta casi 800 pasajeros y en ocasiones alcanza una velocidad máxima de más de 190 millas por hora (http://www.o-keating.com/hsr/eurostar.htm). Suponga que la desviación estándar de la velocidad de este tren es 19 mph en el curso de todos los viajes en una y otra dirección, y que la velocidad del tren está normalmente distribuida. Suponga que las lecturas de velocidad se hacen durante los siguientes 20 viajes del Eurostar y la velocidad media de estas mediciones es 184 millas por hora.

a. ¿Cuál es la variable en estudio?

b. Encuentre la estimación del intervalo de confianza de 90% para la velocidad media.

c. Encuentre la estimación del intervalo de confianza de 95% para la velocidad media.

8.37 En 2003, Trends International Mathematics and Science Study (TIMSS) examinó la capacidad de estudiantes de octavo grado en matemáticas y ciencias. La calificación media en matemáticas para la muestra de estudiantes de octavo grado en Estados Unidos fue 504, con un error estándar de 8.4. Construya un intervalo de confianza de 95% para la calificación media de matemáticas para todos los estudiantes de octavo grado en Estados Unidos.

8.38 Alrededor del 67% de casados dice que consulta con su esposa antes de gastar $352, el promedio de la cantidad que dicen que consultan entre sí antes de hacer un gasto.

Fuente: Yankelovich Partner for Lutheran Brotherhood

a. Con base en la información precedente, ¿qué puede usted concluir acerca de la variable, cantidad en dólares, que requiera consulta con una esposa antes de hacer el gasto? ¿Qué es la cantidad de $352?

Se realizó una encuesta de 500 casados de una zona cercana y se dio una media muestral de $289.75.

b. Construya un intervalo de confianza de 0.98 para la cantidad media para todos los casados. Use $\sigma =$ $600.

c. Con base en las respuestas a las partes a y b, ¿qué puede usted concluir acerca de la cantidad media en dólares, que los casados discuten con sus esposas antes de gastar en la zona cercana, en comparación con la población general?

8.39 Cierto ajuste a una máquina cambiará la longitud de las piezas que produce, pero no afectará la desviación estándar. La longitud de las piezas está normalmente distribuida, y la desviación estándar es 0.5 mm. Después de hacer un ajuste, se toma una muestra aleatoria para determinar la longitud media de las piezas que ahora se produce. Las longitudes resultantes son como sigue:

75.3	76.0	75.0	77.0	75.4	76.3	77.0	74.9	76.5	75.8

a. ¿Cuál es el parámetro de interés?

b. Encuentre la estimación puntual para la longitud media de todas las piezas que ahora se producen.

c. Encuentre el intervalo de confianza de 0.99 para μ.

8.40 El peso atómico para una muestra de referencia de plata se midió en el National Institute of Standards and Technology (NIST), usando dos espectrómetros de masas casi idénticos. Este proyecto fue emprendido conjuntamente con la determinación de la constante de Faraday. A continuación veamos 48 observaciones:

107.8681568	107.8681465	107.8681572

Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: StatLib, http://lib.stat.cmu.edu/datasets/

Observe que los datos difieren sólo en los lugares decimales quinto, sexto y séptimo. Casi todas las computadoras redondearán los datos y sus resultados calculados; así, la variación aparentemente se pierde. Las estadísticas se pueden calcular usando sólo los últimos tres dígitos de cada uno de los valores de datos (es decir, 107.8681568 se convertirá en 568). Algebraicamente esta codificación se parece a esto:

Peso atómico codificado = (Peso atómico − 107.8681000) × 100 000 00.

a. Construya una gráfica de datos codificados. ¿Cómo aparece la codificación en la gráfica?

b. Encuentre la media y desviación estándar de los datos codificados.

c. Convierta las respuestas halladas en la parte b a unidades originales.

d. Determine si la información tiene una distribución aproximadamente normal. Presente su caso.

e. ¿Aplican la distribución de medias muestrales (SDSM) y el teorema de límite central (CLT)? Explique.

f. ¿Se conoce sigma?

g. Si la meta es hallar el intervalo de confianza de 95% para el valor medio de todas las observaciones, ¿qué haría usted?

h. Encuentre el intervalo de confianza del 95% para el valor medio de todas estas observaciones. Justifique su método.

8.41 La fuerza requerida para extraer un corcho de una botella de vino es una propiedad importante del corcho. Si la fuerza es demasiado pequeña, es probable que el corcho no sea buen protector del vino que se encuentra en el interior de la botella; si es demasiado grande, será difícil quitarlo. Ninguna es deseable. Se cree que los corchos No. 9 del caso práctico 6.15 (p. 330) tienen una fuerza de extracción que está normalmente distribuida con una desviación estándar de 36 newtons.

a. Se selecciona, para prueba, una muestra de 20 botellas escogidas al azar.

Fuerza de extracción en Newtons

296	338	341	261	250	347	336	297	279	297
259	334	281	284	279	266	300	305	310	253

Encuentre el intervalo de confianza de 98% para la fuerza media de extracción.

b. Durante una prueba diferente, al azar se seleccionó y probó una muestra de ocho botellas.

Furza de extracción en Newtons

331.9	312.0	289.4	303.6	346.9	308.1	346.9	276.0

Encuentre el intervalo de confianza de 98% para la fuerza media de extracción.

c. ¿Qué efecto tuvieron las dos medias muestrales diferentes en las respuestas a las partes a y b? Explique.

d. ¿Qué efecto tuvieron los dos tamaños muestrales diferentes en las respuestas a las partes a y b? Explique.

e. Se dijo que la fuerza media de extracción era de 310 newtons. ¿Cualquiera de ellas presentó una razón suficiente para dudar de la verdad de la afirmación? Explique.

8.42 "Los costos de colegiatura suben de nuevo" (19 de octubre, 2005), artículo que apareció en el sitio web de la CNN Money, dio las últimas cifras del Consejo de Universidades sobre el costo anual de colegiatura, cuotas y pensión completa. El promedio de la cantidad total para universidades privadas es de $27 516 y $11 354 para universidades públicas.

Fuente: http://money.cnn.com/2004/10/18/pf/college/college_costs/

En un esfuerzo por comparar los mismos costos en el estado de Nueva York, en todo el estado se seleccionó al azar una muestra de 32 estudiantes de primer año de universidades privadas, y se seleccionaron 32 estudiantes más de universidades públicas. La muestra de las privadas resultó en una media de $27 436 y la media muestral de universidades públicas fue $11 147.

a. Suponga que las colegiaturas anuales para universidades privadas tienen una distribución agrupada y la desviación estándar es $1800. Encuentre el intervalo de confianza del 95% para los costos medios de colegiatura en el estado de Nueva York.

b. Suponga que las cuotas anuales para universidades públicas tienen una distribución agrupada y la desviación estándar es $1200. Encuentre el intervalo de confianza del 95% para los costos medios de universidades en el estado de Nueva York.

c. ¿Cómo se comparan los costos universitarios del estado de Nueva York con los valores del Consejo Universitario? Explique cómo difieren.

d. Compare los intervalos de confianza hallados en las partes a y b y describa el efecto que tuvieron las dos medias muestrales diferentes en las respuestas resultantes.

e. Compare los intervalos de confianza hallados en las partes a y b y describa el efecto que tuvieron las dos desviaciones estándar diferentes en las respuestas resultantes.

8.43 "La nieve de las Rocallosas produce menos agua" (caso práctico 8.5) indica "14.36 pulgadas" y "5.07 pulgadas" como estadísticas y las usa como estimaciones puntuales. Describa por qué estos números son estadísticas y por que son también estimaciones puntuales.

8.44 Con el uso de computadora o calculadora, al azar seleccione una muestra de 40 números de un solo dígito y encuentre el intervalo de confianza del 90%

para μ. Repita varias veces, observando si 4.5 está o no está cada vez en el intervalo. Consulte el ejemplo 8.4, página 407. Describa sus resultados.

PARA SU INFORMACIÓN Use comandos para generar datos de enteros en la página 407; luego continúe con los comandos del intervalo de confianza de las páginas 409-410.

8.45 Encuentre el tamaño muestral necesario para estimar μ de una población normal con $\sigma = 3$ adentro de 1 unidad al nivel de confianza de 98%.

8.46 ¿Qué tan grande debe tomarse una muestra si la media poblacional debe estimarse con 99% de confianza a una variación de no más de $75? La población tiene una desviación estándar de $900.

8.47 Un compañía de alta tecnología desea estimar el número medio de años de educación universitaria que sus empleados han terminado. Una buena estimación de la desviación estándar para el número de años de universidad es 1.0. ¿Qué tan grande tiene que ser una muestra para estimar μ a no más de 0.5 de un año con 99% de confianza?

8.48 Al medir el tiempo que tarda un componente de un producto para pasar de una estación de trabajo a la siguiente, un ingeniero ha estimado que la desviación estándar es 5 segundos.

a. ¿Cuántas medidas deben tomarse para estar 95% seguros que el error máximo de estimación no rebasará de 1 segundo?

b. ¿Qué tamaño muestral se requiere para un error máximo de 2 segundos?

8.49 Las nuevas computadoras mini-portátiles pueden tener tanta potencia de computación como las máquinas de varias veces su tamaño, pero pesan menos de 3 libras. ¿Qué tan grande tendría que ser una muestra para estimar el peso medio poblacional, si el error máximo de estimación debe ser de 0.4 a 1 desviación estándar con 95% de confianza?

8.50 La imagen de la biblioteca pública está cambiando de manera constante, y sus servicios en línea continúan creciendo. El uso de la página inicial de la biblioteca creció 17% durante los últimos 12 meses. Se ha estimado que la duración promedio actual de una visita a la página inicial de la biblioteca es aproximadamente 20 minutos. La biblioteca desea tomar una muestra para estimar en forma estadística esta media. ¿Qué tan grande debe ser la muestra para estimar la media con una variación de no más de 0.3 de 1 desviación estándar con 0.98 de confianza?

Fuente: http://library.loganutah.org/library/annual04/annualreport2004.htm

8.4 La naturaleza de la prueba de hipótesis

Todos los días tomamos decisiones; algunas de éstas son de gran importancia, otras son insignificantes en apariencia, pero todas siguen el mismo patrón básico. Ponderamos las alternativas; entonces, con base en nuestras creencias y preferencias, así como en cualquier otra evidencia que tengamos a mano, llegamos a una decisión y tomamos la acción apropiada. La prueba de hipótesis estadística sigue un proceso muy semejante, excepto que comprende información estadística. En esta sección desarrollamos muchos de los conceptos y actitudes de la prueba de hipótesis, al tiempo que vemos diversas situaciones de toma de decisiones sin usar ninguna estadística.

Un amigo organiza una fiesta (fiesta por el supertazón, fiesta por regresar a casa desde la universidad, o cualquier otra fiesta) y usted ha sido invitado; debe tomar una decisión: ir o no ir, así de sencillo; bien, quizá vaya, excepto que irá sólo si le convencen de que la fiesta va a ser más divertida que las típicas de su amigo. Además, definitivamente no desea ir si va a haber otro desastre. Usted ha tomado

la posición de que "la fiesta será un desastre" y no va a menos que le convenzan. Su amigo le asegura, "garantizado, la fiesta será una gran diversión" ¿Vas o no vas?

El proceso de decisión empieza por identificar **algo de interés** y luego formular **dos hipótesis** acerca de éste.

Hipótesis: enunciado de que algo es verdadero.

La frase de su amigo, "La fiesta será una gran diversión", es una hipótesis. La posición de usted, "La fiesta será un desastre", también es una hipótesis.

Prueba de hipótesis estadística: proceso por el que se toma una decisión entre dos hipótesis opuestas. Las dos hipótesis opuestas se formulan de modo que cada una es la negación de la otra. (Una de ellas es siempre verdadera, y la otra es siempre falsa.) Entonces se prueba una hipótesis esperando que pueda demostrar ser un suceso improbable, lo cual implica que es probable que la otra hipótesis sea la verdad.

Las dos hipótesis comprendidas para tomar una decisión se conocen como la *hipótesis nula* y la *hipótesis alternativa.*

Hipótesis nula,* H_o: la hipótesis que probaremos. Generalmente, ésta es la exposición de que un parámetro poblacional tiene un valor específico. La hipótesis nula recibe ese nombre porque es el "punto de partida" para la investigación. (La frase "no hay diferencia" se usa con frecuencia en su interpretación.)

Hipótesis alternativa, H_a: enunciado acerca del mismo parámetro poblacional que se usa en la hipótesis nula. Generalmente, éste es un enunciado que especifica que el parámetro poblacional tiene un valor diferente, en alguna forma, del valor dado en la hipótesis nula. El rechazo de la hipótesis nula implicará la probable verdad de esta hipótesis alternativa.

Respecto a la fiesta de su amigo, las dos hipótesis o puntos de vista opuestos son "La fiesta será una gran diversión" y "La fiesta será un desastre." ¿Cuál enunciado se convierte en hipótesis nula, y cuál en hipótesis alternativa?

Determinar el enunciado de la hipótesis nula y el de la hipótesis alternativa es un paso muy importante. La *idea básica* de la prueba de hipótesis es para la evidencia de tener una probabilidad de "refutar" la hipótesis nula. La hipótesis nula es el enunciado que la evidencia podría refutar. *Su interés* (creencia o resultado deseado), como persona que hace la prueba, se expresa en la hipótesis alternativa. Como persona que toma la decisión, usted piensa que la evidencia demostrará la factibilidad de su "teoría" al demostrar la *improbabilidad* de la verdad de la hipótesis nula. La

*Utilizamos la notación H_0 para la hipótesis nula para contrastarla con H_a para la hipótesis alternativa. Otros textos pueden utilizar H_0 (cero subscrito) en lugar de H_o y H_1 en lugar de H_a.

hipótesis alternativa se conoce a veces como *hipótesis de investigación,* porque representa lo que el investigador espera que encontrar como "verdadero."

Debido a que la "evidencia" (quién va a la fiesta, qué van a servir, etcétera) puede demostrar sólo la improbabilidad de que la fiesta sea un desastre, su posición inicial, "La fiesta será un desastre", se convierte en hipótesis nula. Lo dicho por su amigo, "La fiesta será una gran diversión", se convierte entonces en hipótesis alternativa.

H_o: "La fiesta será un desastre" vs. H_a: "La fiesta será una gran diversión"

Los siguientes ejemplos ilustran la formación y la relación entre hipótesis nula y alternativa.

EJEMPLO 8.8 **Escritura de hipótesis**

Usted está probando un nuevo diseño de bolsas de aire que se usan en automóviles, y está preocupado de que pudieran no abrir adecuadamente. Exprese las hipótesis nula y alternativa.

SOLUCIÓN Las dos posibilidades opuestas son "Las bolsas abren adecuadamente" y "Las bolsas no abren adecuadamente". Unas pruebas podrían producir evidencia que desacredita la hipótesis "Las bolsas abren adecuadamente" más su preocupación es que "Las bolsas no abren adecuadamente". Por tanto, "Las bolsas no abren adecuadamente" se convertirían en la hipótesis alternativa y "Las bolsas abren adecuadamente" sería la hipótesis nula.

La hipótesis alternativa puede ser el enunciado que el experimentador desea demostrar como verdadero.

EJEMPLO 8.9 **Escritura de hipótesis**

Un ingeniero desea demostrar que la nueva fórmula que acaba de ser perfeccionada resulta en una pintura de secado más rápido. Exprese las hipótesis nula y alternativa.

SOLUCIÓN Las dos posibilidades opuestas son "seca más rápido" y "no seca más rápido". Como el ingeniero desea demostrar "seca más rápido", la hipótesis alternativa es "La pintura hecha con la nueva fórmula seca más rápido" y la hipótesis nula es "La pintura hecha con la nueva fórmula no seca más rápido".

En ocasiones, podría ser razonable esperar que la evidencia nos lleve a un rechazo de la hipótesis nula. Tal es el caso del ejemplo 8.10.

EJEMPLO 8.10 **Escritura de hipótesis**

Usted sospecha que un detergente de marca supera la marca de detergente de la tienda, y usted desea probar los dos detergentes porque preferiría comprar la marca más barata. Exprese las hipótesis nula y alternativa.

SOLUCIÓN La sospecha de usted, "El detergente de marca supera la marca de detergente de la tienda", es la razón para la prueba y por tanto se convierte en hipótesis alternativa.

H_o: "No hay diferencia en el rendimiento del detergente."

H_a: "El detergente de marca supera la marca de detergente de la tienda."

No obstante, como consumidor, usted espera no rechazar la hipótesis nula por razones presupuestales.

CASO PRÁCTICO 8.11

Evaluación de técnicas de enseñanza

RESUMEN: ESTE ESTUDIO PRUEBA EL EFECTO DE UNA RECOLECCIÓN DE TAREAS Y HACE PREGUNTAS SOBRE CALIFICACIONES DE EXAMEN

La hipótesis para este estudio es que un profesor puede mejorar el rendimiento de un estudiante (calificaciones de examen) si influye sobre la probabilidad percibida de recompensa de esfuerzo del estudiante. Un profesor logra esto al asignar tareas (técnicas de enseñanza) que son una parte de las calificaciones de un estudiante y son percibidas por estudiantes como un medio de mejorar su calificación en el grupo. El estudiante es motivado a aumentar su esfuerzo para completar las tareas que deben también mejorar la comprensión del material del curso. El resultado final esperado es de mejores calificaciones de examen. La hipótesis nula para este estudio es:

H_o: Las técnicas de enseñanza no tienen efecto significativo en las calificaciones de examen de estudiantes.

Fuente: David R. Vruwink y Janon R. Ott, *The Accounting Review*, Vol. LXII, No. 2 abril 1987. Reimpreso con permiso.

Antes de regresar a nuestro ejemplo acerca de la fiesta, necesitamos ver los cuatro posibles resultados que podrían manifestarse de que la hipótesis nula sea verdadera o falsa, y de que la decisión sea "rechazar H_o" o "no rechazar H_o". La tabla 8.3 muestra estos cuatro posibles resultados.

Se presenta una **decisión correcta tipo A** cuando la hipótesis nula es verdadera y decidimos en su favor. Ocurre una **decisión correcta tipo B** cuando la hipótesis nula es falta y la decisión es en oposición a la hipótesis nula. Se comete un **error tipo I** cuando se rechaza una hipótesis nula verdadera, es decir, cuando la hipótesis nula es verdadera pero decidimos en su contra. Se comete un **error tipo II** cuando decidimos a favor de una hipótesis nula que en realidad es falsa.

TABLA 8.3

Cuatro posibles resultados en una prueba de hipótesis

	Hipótesis nula	
Decisión	Verdadera	Falsa
No rechazar H_o	Decisión correcta tipo A	Error tipo II
Rechazar H_o	Error tipo I	Decisión correcta tipo B

EJEMPLO 8.12 Descripción de los posibles resultados y acciones resultantes (en pruebas de hipótesis)

Describa los cuatro posibles resultados y las acciones resultantes que ocurrirían para la prueba de hipótesis en el ejemplo 8.10.

SOLUCIÓN

Recuerde: H_o: "No hay diferencia en el rendimiento del detergente."
H_a: "El detergente de marca supera la marca de detergente de la tienda."

	Hipótesis nula es verdadera	Hipótesis nula es falsa
No rechazar H_0	**Decisión correcta tipo A** Verdad de situación: No hay diferencia entre los detergentes. Conclusión: Se determinó que no había diferencia. Acción: El consumidor compró el detergente más barato, ahorrando dinero y obteniendo los mismos resultados.	**Error tipo II** Verdad de situación: El detergente de marca es mejor. Conclusión: Se determinó que no había diferencia. Acción: El consumidor compró el detergente más barato, ahorrando dinero y obteniendo malos resultados.
Rechazar H_0	**Error tipo I** Verdad de situación: No hay diferencia entre los detergentes. Conclusión: Se determinó que el detergente de marca era mejor. Acción: El consumidor compró el detergente de marca, gastando dinero extra sin obtener mejores resultados.	**Decisión correcta tipo B** Verdad de situación: El detergente de marca es mejor. Conclusión: Se determinó que el detergente de marca era mejor. Acción: El consumidor compró el detergente de marca, gastando más dinero y obteniendo mejores resultados.

Notas:

1. La verdad de la situación no se conoce antes de tomar la decisión, se llega a una conclusión, y las acciones resultantes tienen lugar. La verdad de H_o puede no ser conocida nunca.
2. El error tipo II resulta con frecuencia en lo que representa una "oportunidad perdida"; perdida en esta situación es la probabilidad de usar un producto que da mejores resultados.

Cuando se toma una decisión, sería bueno siempre tomar la decisión correcta. Esto, sin embargo, no es posible en estadística porque tomamos nuestras decisiones con base en información muestral. Lo mejor que podemos esperar es controlar la probabilidad con la que ocurre un error. La probabilidad asignada al error tipo I es $\boldsymbol{\alpha}$. La probabilidad del error tipo II es $\boldsymbol{\beta}$. Vea la tabla 8.4.

Para controlar estos errores asignamos una pequeña probabilidad a cada uno de ellos. Los valores de probabilidad para α y β que se usan con más frecuencia son 0.01 y 0.05. La probabilidad asignada a cada error depende de su gravedad; cuando más grave sea el error, menos dispuestos estamos a hacer que ocurra y

TABLA 8.4

Probabilidad con la que se toman decisiones

Error en decisión	Tipo	Probabilidad	Decisión correcta	Tipo	Probabilidad
Rechazo de una verdadera H_0	I	α	Rechazo de una verdadera H_0	A	$1 - \alpha$
No rechazar una H_0 falsa	II	β	No rechazar una H_0 falsa	B	$1 - \beta$

por tanto se asignará una probabilidad más pequeña. α y β son probabilidades de errores, cada una bajo condiciones separadas, y no se pueden combinar. En consecuencia, no es posible determinar una sola probabilidad para tomar una decisión incorrecta. Del mismo modo, las dos decisiones correctas están claramente separadas y cada una tiene su propia probabilidad; $1 - \alpha$ es la probabilidad de una decisión correcta cuando la hipótesis nula es verdadera, y $1 - \beta$ es la probabilidad de una decisión correcta cuando la hipótesis nula es falsa. $1 - \beta$ recibe el nombre de *potencia de la prueba estadística*, porque es la medida de la capacidad de una prueba de hipótesis para rechazar una hipótesis nula falsa, una característica muy importante.

Nota: cualquiera que sea el resultado de la prueba de hipótesis, nunca se puede estar seguro de que se tome una decisión correcta.

Volvamos a los dos posibles errores en decisión que podrían ocurrir en el ejemplo 8.10. La mayoría de las personas se molestarían si encontraran que estuvieron gastando dinero de más por un detergente que no era mejor que una marca más barata. Del mismo modo, se molestarían si encontraran que podrían haber comprado un detergente mejor. La evaluación de la gravedad relativa de estos errores exige saber si ésta es su lavandería personal o un negocio de lavandería profesional, cuánto más dinero extra cuesta un detergente de marca, etcétera.

Hay una interrelación entre la probabilidad del error tipo I (α), la probabilidad del error tipo II (β), y el tamaño austral (n). Ésta es muy semejante a la interrelación entre nivel de confianza, error máximo y tamaño muestral que vimos en las páginas 410-411. La figura 8.8 muestra el juego de "tirar de la cuerda en tres direcciones" entre α, β y n. Si cualquiera de las tres aumenta o disminuye, tiene un efecto en una o en las otras dos. El trabajo del experto en estadística es "equilibrar" los tres valores de α, β y n para alcanzar una situación aceptable de prueba.

FIGURA 8.8 Juego de "tirar de la cuerda en tres direcciones" entre, α, β, y n.

Si α se reduce, entonces β debe aumentar o n debe aumentar; si β se reduce, entonces α aumenta o n debe aumentar; si n se reduce, entonces α aumenta o β aumenta. Las opciones para α, β y n definitivamente no son arbitrarias. En este punto de nuestro estudio, sólo se darán el tamaño muestral, n, y α, P(error tipo I) y se usarán para completar una prueba de hipótesis. β, P(error tipo II), se investiga con más detalle en los ejercicios de esta sección pero no se utilizarán en la introducción a la prueba de hipótesis.

¿SABÍA USTED...?

¿Puede la señora decir la diferencia?

A fines de la década de 1920 se hizo la pregunta siguiente, ¿puede la señora saber la diferencia entre leche que se vierta en una taza de té o de té que se vierta en una taza de leche? A la señora se le dieron al azar dos tazas, una de cada una, en pares, y ella correctamente las identificó todas. Si ella adivinaba, la probabilidad de adivinar correctamente era 0.5. Entonces, hacemos una hipótesis de que ella adivinó y vemos la evidencia muestral. Ella identificó muy bien las 10 tazas que se le ofrecieron. ¿Cuál es la probabilidad de adivinar correctamente 10 veces consecutivas? Éste es el valor *p* de Fisher. ¿Es probable que ella estaba adivinando y las identificó correctamente 10 veces consecutivas? (1 en 1024 = 0.00098.)

El tamaño muestral, n, se explica por sí mismo; por tanto, veamos el papel que desempeña α.

Nivel de significación α: probabilidad de cometer el error tipo I.

El establecimiento del nivel de significación puede ser considerado como una "decisión gerencial." Típicamente, alguien en un cargo directivo determina el nivel de probabilidad con el que está dispuesto a correr el riesgo de cometer un error tipo I.

En este punto del procedimiento de hipótesis, se recolecta y resume la evidencia y se calcula el valor de una *estadística de prueba.*

Estadística de prueba: es una variable aleatoria cuyo valor se calcula de los datos muestrales y se usa para tomar la decisión "rechazar H_o" o "no rechazar H_o".

El valor calculado de la estadística de prueba se usa en coordinación con una regla de decisión para determinar ya sea "rechazar H_o" o "no rechazar H_o." Esta **regla de decisión** debe establecerse antes de recolectar los datos; especifica la forma en que se llegará a la decisión.

De nuevo vamos a la fiesta de su amigo: usted tiene que ponderar la historia de las fiestas de su amigo, la hora y lugar, otros que vaya, etcétera, contra sus propios criterios y luego toma su decisión. Como resultado de la decisión acerca de la hipótesis nula ("La fiesta será un desastre"), usted tomará la acción apropiada; va o no va a la fiesta.

Para completar la prueba de hipótesis, necesitará escribir una conclusión que cuidadosamente describa el significado de la decisión respecto a la intención de la prueba de hipótesis.

La conclusión:

a. Si la decisión es "rechazar H_o", entonces la conclusión debe ser expresada en forma semejante a "Hay suficiente evidencia en el nivel de significación α para demostrar que…(el significado de la hipótesis alternativa)".
b. Si la decisión es "no rechazar H_o", entonces la conclusión debe ser expresada en forma semejante a "No hay suficiente evidencia en el nivel de significación α para demostrar que… (el significado de la hipótesis alternativa)".

Cuando escriba la decisión y la conclusión, recuerde que (1) la decisión es acerca de H_o y (2) la conclusión es un enunciado acerca de si se sostuvo la opinión de H_a. Esto es consistente con la "actitud" de todo el procedimiento de prueba de hipótesis. La hipótesis nula es el enunciado de que es "a prueba", y por tanto la decisión debe ser acerca de ella. La opinión de la hipótesis alternativa es la idea que ocasionó la necesidad de una decisión. Por tanto, la pregunta que llevó a la hipótesis alternativa debe ser contestada cuando se escriba la conclusión.

Debemos siempre recordar que cuando se toma la decisión, nada ha sido demostrado. Ambas decisiones pueden llevar a errores: "no rechazar H_o" podría ser un error tipo II (la falta de evidencia suficiente ha llevado, más de una vez, a no asistir a grandes fiestas), y "rechazar H_o" podría ser un error tipo I (más de una persona ha decidido ir a una fiesta que fue un desastre).

SECCIÓN 8.4 EJERCICIOS

8.51 Usted está probando un nuevo sistema detonador para explosivos y le preocupa que el sistema no sea confiable. Exprese las hipótesis nula y alternativa.

8.52 Con referencia al caso práctico 8.11, exprese la hipótesis del profesor, la hipótesis alternativa.

8.53 Exprese las hipótesis nula y alternativa para cada uno de lo siguiente:

a. Usted está investigando una queja de que "el correo de entrega inmediata tarda demasiado" para su entrega.

b. Usted desea demostrar que las personas encuentran más cómodo el nuevo diseño de un sillón reclinable que el anterior diseño.

c. Usted está tratando de demostrar que el humo de cigarrillos afecta la calidad de vida de una persona.

d. Usted está probando una nueva fórmula para un acondicionador de cabello y espera demostrar que es eficaz en casos de "puntas partidas".

8.54 Exprese las hipótesis nula y alternativa para cada uno de lo siguiente:

a. Usted desea demostrar un aumento en compra y venta de casas unifamiliares este año, en comparación con el porcentaje del año pasado.

b. Usted está probando una nueva receta para un pastel de queso "bajo en grasas" y espera encontrar que su sabor no es tan bueno como el del queso tradicional.

c. Usted está tratando de demostrar que las lecciones de música tienen un efecto positivo en la autoestima de un niño.

d. Usted está investigando la relación entre el género de una persona y el automóvil que conduce; específicamente desea demostrar que los hombres tienden a manejar vehículos del tipo de camiones más que las mujeres.

8.55 Usando el ejemplo de la fiesta de su amigo (pp. 417 y 422) con H_o: "La fiesta será un desastre" contra H_a: "La fiesta será una gran diversión," describa las cuatro posibles decisiones y las acciones resultantes como se describe en el ejemplo 8.12.

8.56 Cuando se inspecciona un paracaídas, el inspector está buscando cualquier cosa que pudiera indicar que el paracaídas no se abra.

a. Exprese las hipótesis nula y alternativa.

b. Describa los cuatro posibles resultados que puedan aparecer, dependiendo de la verdad de la hipótesis nula y la decisión a la que haya llegado.

c. Describa la gravedad de los dos posibles errores.

8.57 Cuando una doctora en el lugar de un accidente grave inspecciona a cada víctima, administra la asistencia médica apropiada a todas las víctimas, a menos que esté segura que la víctima está muerta.

a. Exprese las hipótesis nula y alternativa.

b. Describa los cuatro posibles resultados que puedan aparecer, dependiendo de la verdad de la hipótesis nula y la decisión a la que haya llegado.

c. Describa la gravedad de los dos posibles errores.

8.58 Un proveedor de materiales de construcción para carreteras dice que puede suministrar una mezcla de asfalto que hará menos resbalosos los caminos húmedos que se pavimenten con sus materiales. Un contratista general que construye caminos desea probar lo dicho por el proveedor. La hipótesis nula es "Los caminos pavimentados con esta mezcla de asfalto son menos resbalosos que los pavimentados con otro asfalto". La hipótesis alternativa es "Los caminos pavimentados con esta mezcla de asfalto son menos resbalosos que los pavimentados con otro asfalto".

a. Describa el significado de los dos posibles tipos de errores que puedan ocurrir en la decisión cuando se complete la prueba de esta hipótesis.

b. Describa la forma en que la hipótesis nula, como se indicó previamente, es un "punto de partida" para la decisión a tomar acerca del asfalto.

8.59 Usando la información del ejercicio 8.55, describa cómo es que el error tipo II del ejemplo de la fiesta representa una "oportunidad perdida".

8.60 Describa las acciones que resultarían en un error tipo I y un error tipo II si se probaran cada una de las siguientes hipótesis nulas. (Recuerde, la hipótesis alternativa es la negación de la hipótesis nula.)

a. H_o: La mayoría de norteamericanos está a favor de leyes contra armas de asalto.

b. H_o: Las opciones en el menú de comida rápida no son bajas en sal.

c. H_o: Este edificio no debe ser demolido.

d. H_o: No hay despilfarro en el gasto del gobierno.

8.61 Describa la acción que resultaría en una decisión correcta tipo A y una decisión correcta tipo B, si se probaran cada una de las hipótesis nulas del ejercicio 8.60.

8.62 Describa la acción que resultaría en una decisión correcta tipo A y una decisión correcta tipo B, si se probaran las hipótesis para el nuevo sistema detonador de explosivos del ejercicio 8.51.

8.63 Considere la hipótesis nula del caso práctico 8.11, "H_o: Las técnicas de enseñanza no tienen efecto significativo en las calificaciones de exámenes de estudiantes". Describa las acciones que resultarían en un error tipo I y un error tipo II si H_o se probara.

8.64 Considere la hipótesis nula del ejercicio Práctico 8.11, "H_o: Las técnicas de enseñanza no tienen efecto significativo en las calificaciones de exámenes de estudiantes". Describa las acciones que resultarían en una decisión correcta tipo A y una decisión correcta tipo B si H_o se probara.

8.65 a. Si la hipótesis nula es verdadera, ¿qué error de decisión podría cometerse?

b. Si la hipótesis nula es falsa, ¿qué error de decisión podría cometerse?

c. Si se toma la decisión "rechazar H_o", ¿qué error de decisión podría haberse cometido?

d. Si se toma la decisión "no rechazar H_o", ¿qué error de decisión podría haberse cometido?

8.66 La directora de una agencia de publicidad está preocupada por la efectividad de un comercial de televisión.

a. ¿Qué hipótesis nula está probando si comete un error tipo I cuando erróneamente dice que el comercial es eficaz?

b. ¿Qué hipótesis nula está probando si comete un error tipo II cuando erróneamente dice que el comercial es eficaz?

8.67 La directora de una agencia de publicidad está preocupada por la efectividad de un comercial de televisión.

a. ¿Qué hipótesis nula está probando si toma una decisión correcta tipo A cuando correctamente dice que el comercial no es eficaz?

b. ¿Qué hipótesis nula está probando si toma una decisión correcta tipo B cuando correctamente dice que el comercial no es eficaz?

8.68 Un político está preocupado por ganar una elección próxima.

a. ¿Qué hipótesis nula está probando si comete un error tipo I cuando erróneamente dice que ganará la elección?

b. ¿Qué hipótesis nula está probando si comete un error tipo II cuando erróneamente dice que ganará la elección?

8.69 a. Si a α se le asigna un valor de 0.001, ¿qué decimos acerca del error tipo I?

b. Si a α se le asigna un valor de 0.05, ¿qué decimos acerca del error tipo I?

c. Si a α se le asigna un valor de 0.10, ¿qué decimos acerca del error tipo I?

8.70 a. Si a β se le asigna un valor de 0.001, ¿qué decimos acerca del error tipo II?

b. Si a β se le asigna un valor de 0.05, ¿qué decimos acerca del error tipo II?

c. Si a β se le asigna un valor de 0.10, ¿qué decimos acerca del error tipo II?

8.71 a. Si la hipótesis nula es verdadera, ¿la probabilidad de un error de decisión se identifica con qué nombre?

b. Si la hipótesis nula es falsa, ¿la probabilidad de un error de decisión se identifica con qué nombre?

8.72 Suponga que se va a llevar a cabo una prueba de hipótesis usando $\alpha = 0.05$. ¿Cuál es la probabilidad de cometer un error tipo I?

8.73 Explique por qué α no siempre es la probabilidad de rechazar la hipótesis nula.

8.74 Explique cómo es que asignando una pequeña probabilidad a un error controla la probabilidad de que ocurra.

8.75 La conclusión es parte de la prueba de hipótesis que comunica los hallazgos de la prueba al lector. Como tal, necesita especial atención de modo que el lector reciba una imagen precisa de los hallazgos.

a. Con todo cuidado describa la "actitud" del experto en estadística y el comunicado de la conclusión cuando la decisión sea "rechazar H_o".

b. Con todo cuidado describa la "actitud" y el comunicado de la conclusión cuando la decisión sea "no rechazar H_o".

8.76 Encuentre la potencia de una prueba cuando la probabilidad del error tipo II es:

a. 0.01 b. 0.05 c. 0.10

8.77 Se sabe que una población normalmente distribuida tiene una desviación estándar de 5, pero su media está en duda. Se ha dicho que es $\mu = 80$ o $\mu = 90$, y la siguiente prueba de hipótesis se ha diseñado para solucionar el debate. La hipótesis nula, H_o: $\mu = 80$, se probará usando un valor de datos seleccionado al azar y comparándolo con el valor crítico de 86. Si el valor de datos es mayor o igual a 86, la hipótesis nula será rechazada.

a. Encuentre α, la probabilidad del error tipo I.

b. Encuentre β, la probabilidad del error tipo II.

8.78 Suponga que el debate del ejercicio 8.77 ha de solucionarse usando una muestra de tamaño 4; encuentre α y β.

8.79 Usted es un inspector de control de calidad y está en aptitud de tomar la decisión de si un gran embarque de tapones de corcho para botellas de vino no espumoso pasa una inspección. Una vez que inspeccione el número obligatorio en la forma aprobada, tomará una decisión de aceptar o rechazar el lote.

La parte 1 de la inspección requiere que usted seleccione al azar 32 corchos y mida tres dimensiones físicas del tapón cilíndrico según procedimientos definidos.

Límites de especificación

Diámetro	24 mm ± 0.5 mm
Ovalización	≤0.7 mm
Longitud	45 mm ± 0.7 mm

Nivel de calidad aceptable (AQL)

El lote se acepta si no más de dos corchos presentan un resultado inferior o superior a los límites de especificación.

El lote puede ser rechazado si tres corchos o más presentan un resultado inferior o superior a los límites de especificación.

Fuente: http://www.codiliege.org

A continuación veamos los resultados de inspeccionar la muestra obligatoria. (Todas las mediciones son en milímetros.)

Corcho	1	2	3	4	5	6	7	8
Diámetro	24.51	24.13	24.28	24.27	23.79	24.11	24.08	23.66
Ovalización	0.20	0.88	0.38	0.20	0.29	0.14	0.20	0.32
Longitud	44.89	44.69	45.36	44.94	44.65	45.50	44.86	44.67
Corcho	9	10	11	12	13	14	15	16
Diámetro	24.41	24.08	24.02	23.94	23.71	24.18	24.13	24.30
Ovalización	0.03	0.43	0.50	0.43	0.51	0.46	0.53	0.14
Longitud	45.13	44.92	44.88	45.14	44.87	44.67	45.01	44.86
Corcho	17	18	19	20	21	22	23	24
Diámetro	23.78	24.01	24.03	24.10	23.77	24.28	23.85	24.39
Ovalización	0.07	0.32	0.34	0.23	0.76	0.39	0.47	0.43
Longitud	45.12	45.21	45.70	44.95	44.27	45.23	45.29	44.98
Corcho	25	26	27	28	29	30	31	32
Diámetro	24.27	23.92	24.23	24.17	23.77	24.40	24.31	23.85
Ovalización	0.20	0.47	0.23	0.23	0.28	0.34	0.56	0.05
Longitud	44.80	45.06	45.38	45.11	44.75	45.42	45.04	44.53

a. Determine el número de corchos que pasan la parte 1 de la inspección.

b. Exprese la decisión y explique cómo llegó a ella.

c. Elabore un breve informe por escrito que resuma los requisitos y sus hallazgos y decisión.

8.80 Como inspector de control de calidad del ejercicio 8.79, usted está listo para la segunda fase de la inspección.

La parte 2 requiere que el porcentaje de humedad de 20 tapones de corcho se determine mientras se siga el procedimiento prescrito.

Límites de especificación

Valor nominal: 6%

Límites de especificación: ±2% (es decir, de 4 a 8%)

Nivel de calidad aceptable (AQL)

El lote se acepta si no más de dos corchos presentan un resultado inferior o superior a los límites de especificación.

El lote puede ser rechazado si tres o más corchos presentan un resultado inferior o superior a los límites de especificación.

Fuente: http://www.codiliege.org.

A continuación se citan tres muestras diferentes, cada una tomada de lotes diferentes. Revise los resultados muestrales y conteste por separado estas preguntas para cada muestra.

Muestra 1:	5	5	6	3	7	6	6	7	8	6
	6	7	5	7	6	6	7	6	4	5
Muestra 2:	1	6	6	8	6	5	7	6	10	6
	7	5	7	6	5	6	6	8	5	9
Muestra 3:	5	7	3	5	5	5	6	5	9	3
	5	7	7	9	7	8	5	10	8	9

a. Construya una gráfica de puntos de los datos.

b. Aplique la totalidad de leyendas a la gráfica de puntos y circule los puntos que representen porcentajes inferiores o superiores a los límites de especificación.

c. Exprese la decisión y explique cómo llegó a ella.

d. Elabore un breve informe por escrito que resuma los requisitos y sus hallazgos, así como la decisión para cada muestra.

8.5 Prueba de hipótesis para la media μ (σ conocida): un acercamiento al valor probabilístico

En la sección 8.4 estudiamos los conceptos y buena parte del razonamiento que hay detrás de una prueba de hipótesis cuando se observan ejemplos no estadísticos. En esta sección vamos a formalizar el procedimiento de prueba de hipótesis como se aplica a enunciados respecto a la media μ de una población bajo la restricción de que σ, la desviación estándar poblacional, es un valor conocido.

Suposición para pruebas de hipótesis alrededor de la media μ usando una σ conocida: La distribución muestral de $\bar{x}$ tiene una distribución normal.

La información que necesitamos para asegurar que esta suposición se satisfaga está contenida en la distribución muestral de medias muestrales y en el teorema de límite central (CLT) (vea el capítulo 7, pp. 369-370):

La distribución de medias muestrales $\bar{x}$ está distribuida alrededor de una media igual μ, con un error estándar igual $\sigma/\sqrt{n}$; y (1) si la población muestreada al azar está normalmente distribuida, entonces $\bar{x}$ está normalmente distribuida para todos los tamaños muestrales, o (2) si la población muestreada al azar no está normalmente distribuida, $\bar{x}$ entonces está normalmente distribuida en forma aproximada para tamaños muestrales suficientemente grandes.

La prueba de hipótesis es un procedimiento paso a paso, bien organizado, que se emplea para tomar una decisión. Por lo general se emplean dos formatos diferentes para pruebas de hipótesis. El *acercamiento al valor probabilística*, o simplemente *acercamiento al valor p,* es el proceso de prueba de hipótesis que ha ganado popularidad en años recientes, en gran medida como resultado de la comodidad y la capacidad de "triturar números" de la computadora. Este acercamiento está organizado como procedimiento de cinco pasos.

LA PRUEBA DE HIPÓTESIS DE VALOR PROBABILÍSTICO: UN PROCEDIMIENTO DE CINCO PASOS

Paso 1 El inicio:

a. Describir el parámetro poblacional de interés.
b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

Paso 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones.
b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
c. Determinar el nivel de significación, α.

Paso 3 La evidencia muestral:

a. Recolectar la información muestral.
b. Calcular el valor de la estadística de prueba.

Paso 4 La distribución de probabilidad:

a. Calcular el valor p para la estadística de prueba.
b. Determinar si el valor p es o no es menor que α.

Paso 5 Los resultados:

a. Expresar la decisión acerca de H_o.
b. Expresar la conclusión acerca de H_a.

Un fabricante de aviones comerciales compra remaches para usarles en el ensamble de aviones. Todo proveedor de remaches que desee vender remaches al fabricante de aviones debe demostrar que sus remaches satisfacen las especificaciones requeridas. Una de éstas es, "La resistencia media al corte de todos los remaches, μ, es al menos de 925 libras". Cada vez que el fabricante de aviones compra remaches, está preocupado porque la resistencia media pueda ser menor que la especificación de 925 libras.

PARA SU INFORMACIÓN
Considere las consecuencias de usar remaches débiles.

Nota: cada remache individual tiene una resistencia al corte, que se determina midiendo la fuerza necesaria para cortar ("romper") el remache. Es evidente que no todos los remaches se pueden probar. Por tanto, se probará una muestra de remaches y una decisión acerca de la resistencia media de todos los remaches no probados estará basada en la media de los muestreados y probados.

Paso 1 El inicio:

a. **Describir el parámetro poblacional de interés.**
El parámetro poblacional de interés es la media, μ, la resistencia media al corte (o fuerza media requerida para cortar) de los remaches considerados para compra.

PARA SU INFORMACIÓN

En las páginas 417 y 418 se dan más instrucciones específicas.

b. Exprese la hipótesis nula (H_o) y la hipótesis alternativa (H_a). La hipótesis nula y la hipótesis alternativa se formulan al inspeccionar el problema o enunciado a investigar y formular primero dos enunciados opuestos acerca de la μ media. Por ejemplo, estos dos enunciados opuestos son (A) "La resistencia media al corte es menor a 925" ($\mu < 925$, el interés del fabricante de aviones), y (B) "La resistencia media al corte es al menos 925" ($\mu = 925$, el dicho del proveedor de remaches y especificación del fabricante de aviones).

Nota: la ley de la tricotomía de álgebra expresa que dos valores numéricos deben estar relacionados en exactamente una de las tres posibles relaciones: $<$, $=$, o $>$. Estas tres posibilidades deben considerarse en las dos hipótesis opuestas para que las dos hipótesis sean negaciones una de la otra. Las tres posibles combinaciones de signos e hipótesis se muestran en la tabla 8.5. Recuerde que la hipótesis nula asigna un valor específico al parámetro en cuestión, y por tanto "es igual a" siempre será parte de la hipótesis nula.

TABLA 8.5

Los tres posibles enunciados de hipótesis nula y alternativa

Hipótesis nula	Hipótesis alternativa
1. Mayor o igual a ($\geq$)	Menor que ($<$)
2. Menor o igual a ($\leq$)	Mayor que ($>$)
3. Igual a ($=$)	No igual de ($\neq$)

El parámetro de interés, la media poblacional μ, está relacionada con el valor 925. El enunciado (A) se convierte en la hipótesis alternativa:

$$H_a\text{: } \mu < 925 \text{ (la media es menor que 925)}$$

Este enunciado representa el interés del fabricante de aviones y dice, "Los remaches no satisfacen las especificaciones requeridas". El enunciado (B) se convierte en la hipótesis nula:

$$H_o\text{: } \mu = 925\ (\geq) \text{ (la media es al menos 925)}$$

Esta hipótesis representa la negación del interés del fabricante de aviones y dice, "Los remaches satisfacen las especificaciones requeridas".

Nota: escribiremos la hipótesis nula con sólo el signo igual, con lo que se expresa el valor exacto asignado. Cuando "igual" se parea con "menor que" o con "mayor que", el símbolo combinado se escribe además de la hipótesis nula como recordatorio de que los tres signos han sido considerados en estos dos enunciados opuestos.

Antes de continuar con nuestro ejemplo, veamos tres ejemplos que demuestran la formulación de las hipótesis nula y alternativa abarcando la media poblacional μ. Los ejemplos 8.13 y 8.14 demuestran cada uno una hipótesis alternativa de "una cola".

EJEMPLO 8.13

Escritura de hipótesis nula y alternativa (situación de una cola)

Suponga que la Environmental Protection Agency (EPA) estaba demandando a la ciudad de Rochester por no cumplir ésta con los estándares de monóxido de carbono. Específicamente, la EPA desea demostrar que el nivel medio de monóxido de carbono en el aire del centro de la ciudad de Rochester es peligrosamente alto, más de 4.9 partes por millón. Exprese las hipótesis nula y alternativa.

SOLUCIÓN Para expresar las dos hipótesis, primero necesitamos identificar el parámetro poblacional en cuestión: "el nivel medio de monóxido de carbono en Rochester". El parámetro μ está siendo comparado con el valor 4.9 partes por millón, el valor específico de interés. La EPA está cuestionando el valor de μ y desea demostrar que es mayor a 4.9 (es decir, $\mu > 4.9$). Las tres posibles relaciones —(1) $\mu < 4.9$, (2) $\mu = 4.9$, y (3) $\mu > 4.9$— deben arreglarse para formar dos enunciados opuestos: uno expresa la posición de la EPA, "El nivel medio es mayor a 4.9 ($\mu > 4.9$)", y el otro expresa la negación, "El nivel medio no es mayor a 4.9 ($\mu \leq 4.9$)". Uno de los dos enunciados se convertirá en la hipótesis nula, H_o, y el otro se convertirá en la hipótesis alternativa, H_a.

Recuerde que hay dos reglas para formar las hipótesis: (1) la hipótesis nula dice que el parámetro en cuestión tiene un valor especificado ("H_o debe contener el signo igual"), y (2) la opinión de la EPA se convierte en la hipótesis alternativa ("mayor a"). Ambas reglas indican que:

$$H_o\text{: } \mu = 4.9\ (\leq) \qquad \text{y} \qquad H_a\text{: } \mu > 4.9$$

EJEMPLO 8.14

Escritura de hipótesis nula y alternativa (situación de una cola)

Un ingeniero desea demostrar que las aplicaciones de pintura hechas con una nueva fórmula secan y están listas para la siguiente capa en un tiempo medio menor a 30 minutos. Exprese las hipótesis nula y alternativa para esta situación de prueba.

SOLUCIÓN El parámetro de interés es el tiempo medio de secada por aplicación, y 30 minutos es el valor especificado. $\mu < 30$ corresponde a "El tiempo medio es menor a 30", mientras que $\mu \geq 30$ corresponde a la negación, "El tiempo medio no es menor a 30". Por tanto, las hipótesis son

$$H_o\text{: } \mu = 30\ (\geq) \qquad \text{y} \qquad H_a\text{: } \mu < 30$$

El ejemplo 8.15 demuestra una hipótesis alternativa de "dos colas".

EJEMPLO 8.15

Escritura de hipótesis nula y alternativa (situación de dos colas)

La satisfacción en el trabajo es muy importante para la productividad de trabajadores. Un cuestionario estándar sobre satisfacción en el trabajo fue aplicado por oficiales sindicales, a una muestra de trabajadores de una línea de montaje de una gran fábrica, con la esperanza de demostrar que la calificación media de los trabajadores

de una línea de montaje en este cuestionario sería diferente de la media establecida de 68. Exprese las hipótesis nula y alternativa.

SOLUCIÓN O bien, la calificación media de satisfacción en el trabajo es diferente de 68 ($\mu \neq 68$) o es igual a 68 ($\mu = 68$). Por tanto,

$$H_o: \mu = 68 \qquad \text{y} \qquad H_a: \mu \neq 68$$

Notas:

1. La hipótesis alternativa se conoce como de "dos colas" cuando H_a es "diferente a".
2. Cuando "menor a" se combina con "mayor a," se convierten en "diferentes de".

El punto de vista del experimentador afecta en gran medida la manera en que se forman las hipótesis. Por lo general, el experimentador está tratando de demostrar que el valor del parámetro es diferente del valor especificado. Así, con frecuencia el experimentador espera ser capaz de rechazar la hipótesis nula para que su teoría se justifique. Los ejemplos 8.13, 8.14 y 8.15 también representan los tres posibles arreglos para las relaciones $<$, $=$ y $>$ entre el parámetro μ y un valor especificado.

La tabla 8.6 es una lista de frases comunes adicionales que se emplean en frases e indica sus negaciones y la hipótesis en la que se usará cada frase. De nuevo, observe que "igual a" está siempre en la hipótesis nula. También observe que la negación de "menor a" es "mayor o igual a". Considere la negación como "todos los otros" del conjunto de tres signos.

TABLA 8.6

Frases comunes y sus negaciones

H_0: ($\geq$) vs.	H_a: ($<$)	H_0: ($\leq$) vs.	H_a: ($>$)	H_0: ($=$) vs.	H_a: ($\neq$)
Al menos	Menor a	A lo sumo	Más que	Es	No es
No menor a	Menor a	No más de	Más que	No diferente de	Diferente de
No menor a	Menor a	No mayor que	Mayor que	Igual que	No igual a

Una vez establecidas las hipótesis, nula y alternativa, trabajaremos bajo la suposición de que la hipótesis nula es un enunciado verdadero hasta que haya evidencia suficiente para rechazarlo. Esta situación podría compararse con un juicio en la sala de un tribunal, donde se supone que el acusado es inocente (H_o: El acusado es inocente vs. H_a: El acusado no es inocente) hasta que se haya presentado evidencia suficiente para demostrar que la inocencia es totalmente increíble ("fuera de toda duda razonable"). En la conclusión de la prueba de hipótesis, tomaremos una de dos posibles decisiones. Decidiremos en oposición a la hipótesis nula y diremos que "rechazamos H_o" (esto corresponde a la "sentencia" del acusado en un juicio), o decidiremos de acuerdo con la hipótesis nula y diremos que "no rechazamos H_o" (esto corresponde a "no sentenciar" o una "absolución" del acusado en un juicio).

Regresemos al ejemplo de los remaches que interrumpimos en la página 428 y continuemos con el paso 2. Recuerde que

$$H_o: \mu = 925\ (\geq)\ \text{(al menos 925)} \qquad H_a: \mu < 925\ \text{(menor a 925)}$$

PASO 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones:

Suponga que la desviación estándar de la resistencia al corte de remaches se sabe por experiencias pasadas que es $\sigma = 18$. Las variables como la resistencia al corte por lo general tienen una distribución agrupada; por tanto, una muestra de tamaño 50 debe ser suficientemente grande para que el teorema de límite central (CLT) aplique, y asegurar para que la distribución de medias muestrales (SDSM) esté normalmente distribuida.

b. Identifique la distribución de probabilidad y la estadística de prueba a usar.

La distribución de probabilidad normal estándar se usa porque se espera que $\bar{x}$ tenga una distribución normal.

Para una prueba de hipótesis de μ, deseamos comparar el valor de la media muestral con el valor de la media poblacional como se indicó en la hipótesis nula. Esta comparación se logra usando la estadística de prueba en la fórmula (8.4):

Estadística de prueba para media

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{8.4}$$

El valor calculado resultante se identifica como $z\bigstar$ ("z estrella"), porque se espera que tenga una distribución normal estándar cuando la hipótesis nula sea verdadera y las suposiciones se hayan satisfecho. La $\bigstar$ ("estrella") es para recordarnos que éste es un valor calculado de la estadística de prueba.

La estadística de prueba a usar es $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ con $\sigma =$

c. Determine el nivel de significación, α.

Establecer α se describió como una decisión gerencial en la sección 8.4. Para ver lo que está involucrado para determinar α, la probabilidad del error tipo I, para nuestro ejemplo de remaches, empezamos por identificar los cuatro posibles resultados, sus significados, y la acción relacionada con cada uno de ellos.

PARA SU INFORMACIÓN
Hay más de esto, pero esperamos que el lector capte la idea.

PARA SU INFORMACIÓN
α se asignará en el enunciado de ejercicios.

El error tipo I sucede cuando se rechaza una hipótesis nula verdadera. Esto ocurriría cuando el fabricante probó los remaches que no llenaron las especificaciones y los rechazó. Sin duda, esto llevaría a que los remaches no sean comprados aun cuando satisfagan las especificaciones. Para que el gerente establezca un nivel de significación es necesaria una información relacionada, ¿qué tan pronto se necesita una nueva provisión de remaches? por ejemplo. Si se necesitan mañana y éste es el único vendedor con existencia disponible, esperar una semana para hallar remaches aceptables podría ser muy costoso; por tanto, rechazar remaches buenos podría ser considerado como un error grave. Por otra parte, si los remaches no se necesitan sino hasta el mes próximo, entonces este error podría no ser muy grave. Sólo el gerente conocerá todas las ramificaciones y, por tanto, la entrada del gerente es muy importante aquí.

Después de mucha consideración, el gerente asigna el nivel de significación: $\alpha = 0.05$.

PASO 3 La evidencia muestral:

a. Recolectar la información muestral.

La muestra debe ser aleatoria sacada de la población cuya media μ esté siendo cuestionada. Se selecciona una muestra aleatoria de 50 remaches, se prueba cada remache y se calcula la resistencia muestral media al corte: $\bar{x} = 921.18$ y $n = 50$.

b. Calcular el valor de la estadística de prueba.
La evidencia muestral ($\bar{x}$ y n hallada en el paso 3a) se convierte en seguida en el **valor calculado de la estadística de prueba, $z\bigstar$,** usando la fórmula (8.4). (μ es 925 de H_o, y $\sigma = 18$ es una cantidad conocida.) Tenemos

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}: \quad z\bigstar = \frac{921.18 - 925.0}{18/\sqrt{50}} = \frac{-3.82}{2.5456} = -1.50$$

PASO 4 La distribución de probabilidad:

a. Calcular el valor *p* para la estadística de prueba

Valor de probabilidad, o valor *p*: es la probabilidad de que la estadística de prueba pueda tener el valor que tiene, o un valor más extremo (en la dirección de la hipótesis alternativa), cuando l hipótesis nula es verdadera. (*Nota:* El símbolo **P** se usará para representar el valor *p*, es especial en situaciones de álgebra.)

Trace un bosquejo de la distribución normal estándar y localice $z\bigstar$ (hallada en el paso 3b) en ella. Para identificar el área que representa el valor p, vea el signo de la hipótesis alternativa. Para esta prueba, la hipótesis alternativa indica que estamos interesados en la parte de la distribución muestral que sea *"menor que"* $z\bigstar$. Por tanto, el valor p es el área que se encuentra a la *izquierda* de $z\bigstar$. Aplique sombreado a esta área.

Para hallar el valor p, es posible usar cualquiera de los tres métodos señalados aquí. El método que se use no es lo importante, porque cada uno de ellos es sólo la herramienta de opción para ayudar a hallar el valor p.

PARA SU INFORMACIÓN
En las páginas 317-320 se dan instrucciones completas para usar la tabla 3.

Método 1: use la tabla 3 del apéndice B para determinar el área en tabla relacionado con $z = 1.50$; luego calcule el valor p al restar de 0.5000:

$$\text{valor } p = P(z < z\bigstar) = P(z < -1.50) = P(z > 1.50) = 0.5000 - 0.4332 = \mathbf{0.0668}$$

PARA SU INFORMACIÓN
El estudiante usará sólo uno de estos tres métodos equivalentes.

Método 2: use la tabla 5 del apéndice B y la propiedad de simetría: la tabla 5 se forma para permitir leer el valor p directamente desde la tabla. Como $P(z < -1.50) = P(z > 1.50)$, simplemente localice $z\bigstar = 1.50$ en la tabla 5 y lea el valor p:

$$P(z < -1.50) = \mathbf{0.0668}$$

PARA SU INFORMACIÓN
Las instrucciones para utilizar este comando de computadora se dan en las páginas 329-330. ¡Haga la prueba! A ver si obtiene el mismo resultado.

Método 3: use la función de probabilidad acumulativa en una computadora o calculadora para hallar el valor p:

$$P(z < -1.50) = \mathbf{0.0668}$$

b. Determine si el valor p es o no es menor a α.
El valor p (0.0668) no es menor a α (0.05).

PASO 5 Los resultados:

a. Exprese la decisión acerca de H_o.
El valor p, ¿es suficientemente pequeño para indicar que la evidencia muestral es muy poco probable en caso que sea verdadera la hipótesis nula? Para tomar la decisión, necesitamos saber la *regla de decisión*.

Regla de decisión:

a. Si el valor p es *menor o igual al* nivel de significación α, entonces la decisión debe ser **rechazar H_o.**

b. Si el valor p es *mayor al* nivel de significación α, entonces la decisión debe ser **no rechazar H_o.**

PARA SU INFORMACIÓN

En la página 422 se da información específica acerca de cómo escribir la conclusión.

Decisión acerca de H_o: no rechazar H_o.

b. Expresar la conclusión acerca de H_a.

No hay suficiente evidencia en el nivel de significación 0.05 para demostrar que la resistencia media al corte de los remaches es menor a 925. "No sentenciamos" es la hipótesis nula. En otras palabras, es probable que ocurra una media muestral de sólo 921.18 (definida por α) cuando el verdadero valor medio poblacional sea 925.0 y $\bar{x}$ sea normalmente distribuida. La acción resultante del gerente sería comprar los remaches.

Nota: cuando la decisión alcanzada sea "no rechazar H_o". simplemente significa "por falta de mejor información, actuar como si la hipótesis nula fuera verdadera". ("Aceptar H_o" es un nombre inapropiado.)

Antes de ver otro ejemplo, veamos los procedimientos para hallar el valor p. El valor p está representado por el área bajo la curva de la distribución de probabilidad para la estadística de prueba que sea más extrema que el valor calculado de la estadística de prueba. Hay tres casos separados, y la dirección (o signo) de la hipótesis alternativa es la clave. La tabla 8.7 señala el procedimiento para los tres casos.

TABLA 8.7

Búsqueda de valores p

Caso 1 H_a contiene "$>$" "Cola derecha"	El valor p es el *área a la derecha de $z\bigstar$* valor $p = P(z > z\bigstar)$	Valor p en cola derecha
Caso 2 H_a contiene "$<$" "Cola izquierda"	El valor p es el *área a la izquierda de $z\bigstar$* el área de la cola izquierda es el nombre del área de la cola derecha limitada por la $z\bigstar$ positiva; por lo tanto, valor $p = P(z < z\bigstar) = P(z > \lvert z\bigstar\rvert)$	Valor p en cola izquierda
Caso 3 H_a contiene "$\neq$" "Dos colas"	El valor p es el *área total de ambas colas* valor $p = P(z < -\lvert z\bigstar\rvert) + P(z > \lvert z\bigstar\rvert)$ $z\bigstar$ puede estar en cualquier cola, y como ambas áreas son iguales, encuentre la probabilidad de un cola y multiplíquela por dos. Así, valor $p = 2 \times P(z > \lvert z\bigstar\rvert)$	Valor p en dos colas

Veamos un ejemplo que comprende el procedimiento de dos colas.

EJEMPLO 8.16 Prueba de hipótesis de dos colas

¿SABÍA USTED...?

El conflicto Fisher y Neyman

Sus diferencias se centraron en sus métodos para probar hipótesis. Ambos métodos se inician con una hipótesis nula y usan la misma estadística de prueba; no obstante, Neyman y Pearson usan un conjunto de riesgo de error que Fisher no utiliza. El procedimiento Neyman/Pearson sigue un método lógico deductivo básico de suponer que la hipótesis es verdadera, y a continuación busca una evidencia que contradiga las suposiciones. El procedimiento Fiser determina la probabilidad del suceso para los datos que resultan y emplea el valor de esa probabilidad para evaluar los datos. Una pequeña probabilidad muestra "ausencia de probabilidad" para que hayan ocurrido los datos bajo una hipótesis nula verdadera. La probabilidad de que la hipótesis nula sea correcta es otra historia, que exige un procedimiento de Bayes y lleva a otra disputa académica más.

Durante años, numerosas empresas grandes en cierta ciudad han utilizado la Agencia de Empleos Kelley para hacer pruebas a empleados en perspectiva. La prueba de selección de empleo que se usa ha resultado históricamente en calificaciones normalmente distribuidas alrededor de una media de 82 y una desviación estándar de 8. La Agencia Brown ha creado una nueva prueba que es más rápida y más fácil de aplicar y, por tanto, es menos costosa. Brown dice que los resultados de su prueba son iguales a los obtenidos en la prueba de Kelley. Muchas de las compañías están considerando un cambio de la Agencia Kelley a la Agencia Brown para reducir costos, pero no están dispuestos a cambiar si los resultados de la prueba de Brown tienen un valor medio diferente. Una empresa independiente de pruebas examinó a 36 empleados prospectos con la prueba de Brown y resultó una media de 79. Determine el valor p asociado con esta prueba de hipótesis. (Suponga que $\sigma = 8$.)

SOLUCIÓN

PASO 1 El inicio:

a. Describir el parámetro poblacional de interés.
La media poblacional μ, la media de todas las calificaciones de prueba usando la prueba de la Agencia Brown.

b. Exprese la hipótesis nula (H_o) y la hipótesis alternativa (H_a).
Los resultados de la prueba de la Agencia Brown "serán diferentes" (el interés) si la calificación media de prueba no es igual a 82. Serán "iguales" si la media es igual a 82. Por tanto,

$$H_o: \mu = 82 \text{ (los resultados de prueba tienen la misma media)}$$

$$H_a: \mu \neq 82 \text{ (los resultados de prueba tienen una media diferente)}$$

PASO 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones.
σ se conoce. Si las calificaciones de la prueba de Brown están distribuidas igual que las calificaciones de la prueba de Kelley, estarán normalmente distribuidas y la distribución muestral será normal para todos los tamaños muestrales.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
La distribución de probabilidad normal estándar y la estadística de prueba $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ se usarán con $\sigma = 8$.

c. Determinar el nivel de significación, α.
El nivel de significación se omite porque la pregunta pide el valor p y no una decisión.

PASO 3 La evidencia muestral:

a. Recolectar la información muestral: $n = 36$, $\bar{x} = 79$.

b. Calcular el valor de la estadística de prueba.

μ es 82 de H_o; $\sigma = 8$ es una cantidad conocida. Tenemos

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}: \quad z\bigstar = \frac{79 - 82}{8/\sqrt{36}} = \frac{-3}{1.3333} = -2.25$$

PASO 4 Distribución de probabilidad:

a. Calcular el valor *p* para la estadística de prueba.

Como la hipótesis alternativa indica una prueba de dos colas, debemos hallar la probabilidad asociada con ambas colas. El valor p se encuentra al duplicar el área de una cola (vea la tabla 8.7, p. 433). Como $z\bigstar = -2.25$, el valor $|z\bigstar| = 2.25$.

El valor $p = 2 \times P(z > |z\bigstar|) = 2 \times P(z > 2.25)$.

De la tabla 3: valor $p = 2 \times P(z > 2.25) = 2 \times (0.5000 - 0.4878) = 2(0.0122) = 0.0244$.

o bien

De la tabla 5: valor $p = 2 \times P(z > 2.25) = 2(0.0122) = 0.0244$.

o bien

Use la función de probabilidad acumulativa en una computadora o calculadora: valor $p = 2 \times P(z < -2.25) = 0.0244$.

½ valor *p* | tabla de valor

b. Determinar si el valor *p* es o no es menor a *α*.

No es posible una comparación; no se dio un valor α en el enunciado de la pregunta.

Para su información

Vea las instrucciones en la página 432.

PASO 5 Los resultados:

El valor p para esta prueba de hipótesis es 0.0244. Cada compañía individual decidirá ahora si continúa usando los servicios de la Agencia Kelley o cambia a la Agencia Brown. Cada una necesitará establecer el nivel de significación que mejor se ajuste a su propia situación y luego tomar una decisión usando la regla de decisión descrita previamente.

La *idea fundamental del valor p* es expresar el grado de credibilidad en la hipótesis nula:

- Cuando el valor p es minúsculo (como de 0.0003) la hipótesis nula sería rechazada por todos, porque los resultados muestrales son muy poco probables para una verdadera H_o.
- Cuando el valor p sea bastante pequeño (como de 0.012), la evidencia contra H_o es muy fuerte y H_o será rechazada por muchos.
- Cuando el valor p empieza a aumentar (de 0.02 a 0.08, por ejemplo), hay demasiada probabilidad de que datos como la muestra involucrada puedan haber ocurrido incluso si H_o fuera verdadera, y el rechazo de H_o no es una decisión fácil.
- Cuando el valor p es grande (0.15 o más), los datos para nada son improbables si la H_o es verdadera y nadie rechazará la H_o.

Las *ventajas del acercamiento al valor p* son como sigue: (1) Los resultados del procedimiento de prueba se expresan en términos de una escala continua de probabilidad de 0.0 a 1.0, más que sólo un "rechazo" o "no rechazo". (2) Un valor p puede presentarse y el usuario de la información puede decidir acerca de la fuerza de la evidencia en cuanto a su propia situación. (3) Las computadoras

pueden hacer todos los cálculos y presentar el valor p, eliminando así la necesidad de tablas.

La *desventaja del acercamiento al valor p* es la tendencia de las personas a diferir la determinación del nivel de significación. No debe permitirse que esto ocurra, porque es entonces posible que alguien establezca el nivel de significación después del hecho, dejando abierta la posibilidad de que resulte la decisión "preferida". Es probable que esto sea importante sólo cuando el valor p reportado caiga en el rango de "opción dura" (0.02 a 0.08, por ejemplo), como se describió líneas antes.

PARA SU INFORMACIÓN
¿Antes que usted apostara, sus oponentes le mostraron sus cartas?

EJEMPLO 8.17 Prueba de hipótesis de dos colas con datos muestrales

Según los resultados del ejercicio 5.32 (p. 282), la media de números aleatorios de un solo dígito es 4.5 y la desviación estándar es $\sigma = 2.87$. Saque una muestra aleatoria de 40 números de un solo dígito de la tabla 1 del apéndice B y pruebe la hipótesis, "La media de números de un solo dígito de la tabla 1 es 4.5". Use $\alpha = 0.10$.

SOLUCIÓN

PASO 1 El inicio:

a. Describir el parámetro poblacional de interés.
El parámetro poblacional de interés es la media μ de la población de números de un solo dígito de la tabla 1 del apéndice B.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

$$H_o: \mu = 4.5 \text{ (la media es 4.5)}$$

$$H_a: \mu \neq 4.5 \text{ (la media no es 4.5)}$$

PASO 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones.
σ se conoce. Las muestras de tamaño 40 deben ser suficientemente grandes para satisfacer el teorema de límite central (CLT); vea la discusión de este problema en la página 426.

b. Identificar la distribución de probabilidad y estadística de prueba a usar.
Usamos la distribución de probabilidad normal estándar, y la estadística de prueba es $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$; $\sigma = 2.87$.

c. Determinar el nivel de significación, α.
$\alpha = 0.10$ (dada en el enunciado del problema).

PASO 3 La evidencia muestral:

a. Recolectar la información muestral.
Esta muestra aleatoria se tomó de la tabla 1 del apéndice B.

2	8	2	1	5	5	4	0	9	1	0	4	6	1
5	1	1	3	8	0	3	6	8	4	8	6	8	
9	5	0	1	4	1	2	1	7	1	7	9	3	

De la muestra: $\bar{x} = 3.975$ y $n = 40$.

b. Calcular el valor de la estadística de prueba.
Usamos la fórmula (8.4), y μ es 4.5 de H_o, y $\sigma = 2.87$:

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}: \quad z\bigstar = \frac{3.975 - 4.50}{2.87/\sqrt{40}} = \frac{-0.525}{0.454} = -1.156 = -1.16$$

PASO 4 Distribución de probabilidad:

a. Calcular el valor *p* para la estadística de prueba.
Como la hipótesis alternativa indica una prueba de dos colas, debemos hallar la probabilidad asociada con ambas colas. El valor p se encuentra duplicando el área de una cola. Como $z\bigstar = -1.16$, el valor de $|z\bigstar| = 1.16$.
El valor $p = 2 \times P(z > |z\bigstar|)$:

$$\mathbf{P} = 2 \times P(z > 1.16) = 2 \times (0.5000 - 0.3770)$$
$$= 2(0.1230) = 0.2460$$

b. Determinar si el valor *p* es o no es menor a α.
El valor p (0.2460) es mayor a α (0.10).

PASO 5 Los resultados:

a. Expresar la desición acerca de H_o: no rechazar H_o.

b. Expresar la conclusión acerca de H_a.
La media muestral observada no es significativamente diferente de 4.5 en el nivel 0.10 de significación.

Suponga que fuéramos a tomar otra muestra de tamaño 40 de la tabla 1. ¿Obtendríamos los mismos resultados? Suponga que tomamos una tercera muestra y una cuarta. ¿Qué resultados podríamos esperar? ¿Qué es lo que mide el valor p del ejemplo 8.17? La tabla 8.8 es una lista (1) de las medias obtenidas de 50 muestras aleatorias diferentes de tamaño 40 que se tomaron de la tabla 1 del apéndice B, (2)

TABLA 8.8

a. Las medias de 50 muestras aleatorias tomadas de la tabla 1 del apéndice B

3.850	5.075	4.375	4.675	5.200	4.250	3.775	4.075	5.800	4.975
4.225	4.125	4.350	4.925	5.100	4.175	4.300	4.400	4.775	4.525
4.225	5.075	4.325	5.025	4.725	4.600	4.525	4.800	4.550	3.875
4.750	4.675	4.700	4.400	5.150	4.725	4.350	3.950	4.300	4.725
4.975	4.325	4.700	4.325	4.175	3.800	3.775	4.525	5.375	4.225

b. Valores $z\bigstar$ correspondientes a las 50 medias

−1.432	1.267	−0.275	0.386	1.543	−0.551	−1.598	−0.937	2.865	1.047
−0.606	−0.826	−0.331	0.937	1.322	−0.716	−0.441	−0.220	0.606	0.055
−0.606	1.267	−0.386	1.157	0.496	0.220	0.055	0.661	0.110	−1.377
0.551	0.386	0.441	−0.220	1.432	0.496	−0.331	−1.212	−0.441	0.496
1.047	−0.386	0.441	−0.386	−0.716	−1.543	−1.598	0.055	1.928	−0.606

c. Valores *p* correspondientes a las 50 medias

0.152	0.205	0.783	0.700	0.123	0.582	0.110	0.349	0.004	0.295
0.545	0.409	0.741	0.349	0.186	0.474	0.659	0.826	0.545	0.956
0.545	0.205	0.700	0.247	0.620	0.826	0.956	0.509	0.912	0.168
0.582	0.700	0.659	0.826	0.152	0.620	0.741	0.226	0.659	0.620
0.295	0.700	0.659	0.700	0.474	0.123	0.110	0.956	0.054	0.545

los 50 valores de $z\bigstar$ correspondientes a las 50 $\bar{x}$, y (3) sus 50 valores p correspondientes. La figura 8.9 muestra un histograma de los 50 valores $z\bigstar$.

FIGURA 8.9 Los 50 valores de $z\bigstar$ de la tabla 8.8

El histograma muestra que seis valores de $z\bigstar$ fueron menores a −1.16 y siete valores fueron mayores a 1.16. Esto significa que 13 de las 50 muestras, o sea 26%, tienen valores medios más extremos que la media ($\bar{x} = 3.975$) del ejemplo 8.17. Esta frecuencia relativa observada de 0.26 representa una vista empírica del valor p. Nótese que el valor empírico para el valor p (0.26) es muy semejante al valor p calculado de 0.2460. Comprueba la lista de valores p; ¿encuentra que 13 de los 50 valores p son menores a 0.2460? ¿Cuáles muestras resultaron en $|z\bigstar| > 1.16$? ¿Qué muestras resultaron en un valor p mayor a 0.2460? ¿Cómo se comparan?

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA UNA MEDIA μ CON UNA σ DETERMINADA

MINITABA (Versión 14) Introduzca los datos en C1; luego continúe con:

```
Choose:     Stat > Basic Statistics > 1-Sample Z
Enter:      Samples in columns: C1
            Standard deviation: σ
            Test mean: μ
Select:     Options
Select:     Alternative: less than or not equal to or greater than
            > OK > OK
```

Excel Introduzca los datos en la columna A; luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Z-Test: Mean > OK
Enter:      Input Range: (A1:A20 or select cells)
            Hypothesized Mean: μ
            Standard Deviation (SIGMA): σ > OK
Gives p-values for both one-tailed and two-tailed tests.
```

TI-83/84 Plus Introduzca los datos en L1; luego continúe con lo siguiente, introduciendo los valores apropiados y destacando Calculate:

Choose: **STAT > TESTS > 1:Z-Test**

PARA SU INFORMACIÓN
El acercamiento al valor *p* fue "hecho" para la computadora.

La solución MINITAB al ejemplo de los remaches, usado en esta sección (pp. 427-428, 430-433), se muestra aquí:

```
One-sample Z: C1
Test of mu = 925.00 vs < 925.00
The assumed standard deviation = 18.0
N        Mean        StDev       SE Mean        Z          P
50       921.18      17.58       2.546        −1.50      0.0668
```

Cuando se usa computadora, todo lo restante se deja al estudiante para que tome la decisión y escriba la conclusión.

SECCIÓN 8.5 EJERCICIOS

8.81 En el ejemplo inicial de la página 427, el fabricante de aviones que compra los remaches está preocupado por que éstos no satisfagan la especificación de resistencia media. Exprese las hipótesis nula y alternativa del fabricante de aviones.

8.82 El profesor Hart no cree una frase que escuchó: "El peso medio de muchachas universitarias es 54.4 kg". Exprese las hipótesis nula y alternativa que usaría para desafiar esta frase.

8.83 Exprese las hipótesis nula y alternativa empleada para probar cada una de las frases siguientes:

a. El tiempo medio de reacción es mayor a 1.25 segundos.
b. La calificación media en ese examen es menor a 335.
c. El precio medio de venta de casas en la región no es $230 000.
d. El peso medio de jugadores de fútbol colegial no es más de 210 libras.
e. El salario medio por hora para una niñera es $9 a lo sumo.

8.84 Exprese la hipótesis nula H_o y la hipótesis alternativa H_a que se usarían para una prueba de hipótesis relacionada a cada una de las frases siguientes:

a. La edad media de estudiantes inscritos en clases vespertinas en cierta universidad es mayor a 26 años.
b. El peso medio de paquetes enviados en Air Express durante el mes pasado fue menor a 36.7 libras.
c. La vida útil media de lámparas fluorescentes es al menos 1600 horas.
d. La resistencia media de soldaduras hechas con un nuevo proceso es diferente de 570 libras por unidad de área, que es la resistencia media de soldaduras hechas con el proceso anterior.

8.85 Identifique los cuatro posibles resultados y describa la situación de que se trata en cada resultado, respecto a la prueba y compra de remaches del fabricante de aviones. ¿Cuál es el error más grave: el error tipo I o el tipo II? Explique.

8.86 Un fabricante desea probar la hipótesis de que "cambiando la fórmula de su pasta dental, dará a sus usuarios una mejor protección". La hipótesis nula representa la idea de que "el cambio no va a mejorar la protección", y la hipótesis alternativa es "el cambio va a mejorar la protección". Describa el significado de los dos posibles tipos de errores que pueden ocurrir en la decisión cuando se realice la prueba de la hipótesis.

8.87 Suponga que deseamos probar la hipótesis de que el cargo medio por hora, para reparaciones de automóviles, es al menos $60 por hora en los talleres de reparación de una ciudad vecina. Explique las condiciones que existirían si cometemos un error de decisión al cometer un error tipo I. ¿Qué hay de un error tipo II?

8.88 Describa la forma en que la hipótesis nula, como se indica en el ejemplo 8.14 (p. 429), es un "punto de partida" para la decisión a tomarse acerca del tiempo de secado para pintura hecha con la nueva fórmula.

8.89 Suponga que z es la estadística de prueba y calcule el valor de z★ para cada uno de lo siguiente:

a. H_o: $\mu = 10$, $\sigma = 3$, $n = 40$, $\bar{x} = 10.6$

b. H_o: $\mu = 120$, $\sigma = 23$, $n = 25$, $\bar{x} = 126.2$

c. H_o: $\mu = 18.2$, $\sigma = 3.7$, $n = 140$, $\bar{x} = 18.93$

d. H_o: $\mu = 81$, $\sigma = 13.3$, $n = 50$, $\bar{x} = 79.6$

8.90 Suponga que z es la estadística de prueba y calcule el valor de z★ para cada uno de lo siguiente:

a. H_o: $\mu = 51$, $\sigma = 4.5$, $n = 40$, $\bar{x} = 49.6$

b. H_o: $\mu = 20$, $\sigma = 4.3$, $n = 75$, $\bar{x} = 21.2$

c. H_o: $\mu = 138.5$, $\sigma = 3.7$, $n = 14$, $\bar{x} = 142.93$

d. H_o: $\mu = 815$, $\sigma = 43.3$, $n = 60$, $\bar{x} = 799.6$

8.91 Hay sólo dos posibles decisiones que pueden resultar de una prueba de hipótesis.

a. Expresar las dos posibles decisiones.

b. Describir las condiciones que llevarán a cada una de las dos decisiones identificadas en la parte a.

8.92 a. ¿A qué decisión se llega cuando el valor p es mayor que α?

b. ¿A qué decisión se llega cuando α es mayor que el valor p?

8.93 Para cada uno de los siguientes pares de valores, exprese la decisión que ocurrirá y por qué.

a. valor $p = 0.014$, $\alpha = 0.02$

b. valor $p = 0.118$, $\alpha = 0.05$

c. valor $p = 0.048$, $\alpha = 0.05$

d. valor $p = 0.064$, $\alpha = 0.10$

8.94 Para cada uno de los siguientes pares de valores, exprese la decisión que ocurrirá y por qué.

a. valor $p = 0.018$, $\alpha = 0.01$

b. valor $p = 0.033$, $\alpha = 0.05$

c. valor $p = 0.078$, $\alpha = 0.05$

d. valor $p = 0.235$, $\alpha = 0.10$

8.95 El valor p calculado para una prueba de hipótesis es 0.084. ¿Qué decisión ocurriría acerca de la hipótesis nula en lo siguiente?

a. La prueba de hipótesis se completa en el nivel de significación de 0.05.

b. La prueba de hipótesis se completa en el nivel de significación de 0.10.

8.96 a. Se ha de completar una prueba de hipótesis de una cola en el nivel de significación de 0.05. ¿Qué valores calculados de p causarán un rechazo de H_o?

b. Se ha de completar una prueba de hipótesis de dos colas en el nivel de significación de 0.02. ¿Qué valores calculados de p causarán una decisión "no rechazar H_o"?

8.97 **El siguiente ejercicio** es una pantalla de un simulador que estima el valor p para una prueba de hipótesis de una cola al simular que toma muchas muestras. La prueba de hipótesis es para una H_o: $\mu = 1500$ contra H_a:

$\mu < 1500$. Se ha tomado una muestra de 24 y la media muestral es 1451.

a. Dé un clic en "10" para "# of samples". Observe las medias muestrales y la probabilidad de ser menor a 1451 si la media verdadera es realmente 1500.

b. Cambie a "Batch" y simule 1000 muestras más. ¿Cuál es la probabilidad de ser menor a 1451? Éste es su valor p estimado.

c. ¿Cómo se muestra su valor estimado p en el histograma formado por la toma de muchas muestras? Explique lo que significa este valor p respecto a la prueba.

d. Si el nivel de significación fuera 0.01, ¿cuál sería su decisión?

8.98 El siguiente ejercicio se realiza en un simulador que estima el valor p para una prueba de hipótesis de dos colas al simular que toma muchas muestras. La prueba de hipótesis dada es para H_o: $\mu = 4$ contra H_a: $\mu \neq 4$. Se ha tomado una muestra de 100 y la media muestral es 3.6.

a. Dé un clic en "10" para "# of samples". Observe las medias muestrales y que la probabilidad sea menor a 3.6 o mayor a 4.4. ¿Por qué estamos incluyendo el mayor a 4.4?

b. Cambie a "Batch" y simule 1000 muestras más. ¿Cuál es la probabilidad de que sea menor a 3.6 o mayor a 4.4? Éste es su valor p estimado.

c. ¿Cómo se muestra su valor estimado p en el histograma formado por la toma de muchas muestras? Explique lo que significa este valor p respecto a la prueba.

d. Si el nivel de significación fuera 0.05, ¿cuál sería su decisión?

8.99 Describa, en sus propias palabras, qué mide el valor p.

8.100 a. Calcule el valor p, dada H_a: $\mu = 45$ y $z\bigstar = -2.3$

b. Calcule el valor p, dada H_a: $\mu = 58$ y $z\bigstar = 1.8$.

8.101 Calcule el valor p, dada H_a: $\mu \neq 245$ y $z\bigstar = 1.1$.

8.102 Encuentre la estadística $z\bigstar$ y el valor p para cada una de las siguientes situaciones.

a. H_o: $\mu = 22.5$, H_a: $\mu > 22.5$; $\bar{x} = 24.5$, $\sigma = 6$, $n = 36$

b. H_o: $\mu = 200$, H_a: $\mu < 200$; $\bar{x} = 192.5$, $\sigma = 40$, $n = 50$

c. H_o: $\mu = 12.4$, H_a: $\mu \neq 2.4$; $\bar{x} = 11.52$, $\sigma = 2.2$, $n = 16$

8.103 Calcule el valor p para cada uno de lo siguiente:

a. H_o: $\mu = 10$, H_a: $\mu > 10$, $z\bigstar = 1.48$

b. H_o: $\mu = 105$, H_a: $\mu < 105$, $z\bigstar = -0.85$

c. H_o: $\mu = 13.4$, H_a: $\mu \neq 13.4$, $z\bigstar = 1.17$

d. H_o: $\mu = 8.56$, H_a: $\mu < 8.56$, $z\bigstar = -2.11$

e. H_o: $\mu = 110$, H_a: $\mu \neq 110$, $z\bigstar = -0.93$

8.104 Calcule el valor p para cada uno de lo siguiente:

a. H_o: $\mu = 20$, H_a: $\mu < 20$; $\bar{x} = 17.8$, $\sigma = 9$, $n = 36$

b. H_o: $\mu = 78.5$, H_a: $\mu > 78.5$; $\bar{x} = 79.8$, $\sigma = 15$, $n = 100$

c. H_o: $\mu = 1.587$, H_a: $\mu \neq 1.587$; $\bar{x} = 1.602$, $\sigma = 0.15$, $n = 50$

8.105 Encuentre el valor de $z\bigstar$ para cada uno de lo siguiente:

a. H_o: $\mu = 35$ contra H_a: $\mu > 35$ cuando el valor $p = 0.0582$

b. H_o: $\mu = 35$ contra H_a: $\mu < 35$ cuando el valor $p = 0.0166$

c. H_o: $\mu = 35$ contra H_a: $\mu \neq 35$ cuando el valor $p = 0.0042$

8.106 La hipótesis nula, H_o: $\mu = 48$, fue probada contra la hipótesis alternativa, H_a: $\mu > 48$. Resultó una muestra de 75 en un valor p calculado de 0.102. Si $\sigma = 3.5$, encuentre el valor de la media muestral, $\bar{x}$.

8.107 La hipótesis nula, H_o: $\mu = 16$, fue probada contra la hipótesis alternativa, H_a: $\mu < 16$. Resultó una muestra de 50 en un valor p calculado de 0.017. Si $\bar{x} = 14$, encuentre el valor de la desviación poblacional estándar.

8.108 Usando la solución MINITAB para el ejemplo de los remaches, como se ilustra en la página 439, describa la forma en que MINITAB halló cada uno de los seis valores numéricos que reportó como resultados.

8.109 La siguiente salida de computadora se utilizó para completar una prueba de hipótesis.

```
TEST OF MU = 525.00 VS MU < 525.00
THE ASSUMED SIGMA = 60.0
N     MEAN      STDEV    SE MEAN    Z        P VALUE
38    512.14    64.78    9.733      −1.32    0.093
```

a. Exprese las hipótesis nula y alternativa.
b. Si la prueba se completa usando $\alpha = 0.05$, ¿a qué decisión y conclusión se llega?
c. Verifique el valor del error estándar de la media.

8.110 Usando la salida de computadora e información del ejercicio 8.109, determine el valor de lo siguiente:
a. Valor hipotético de media poblacional
b. Media muestral
c. Desviación estándar poblacional
d. Estadística de prueba

8.111 La siguiente salida de computadora se utilizó para completar una prueba estadística

```
TEST OF MU = 6.250 VS MU not = 6.250
THE ASSUMED SIGMA = 1.40
N     MEAN      STDEV    SE MEAN    Z        P VALUE
78    6.596     1.273    0.1585     2.18     0.029
```

a. Exprese las hipótesis nula y alternativa.
b. Si la prueba se completó usando $\alpha = 0.05$, ¿a qué decisión y conclusión se llegó?
c. Verifique el valor del error estándar de la media
d. Encuentre los valores para Σx y Σx^2.

8.112 Usando la salida de computadora e información del ejercicio 8.111, determine el valor de lo siguiente:
a. Valor hipotético de media poblacional
b. Media muestral
c. Desviación estándar poblacional
d. Estadística de prueba

8.113 Según el artículo de *USA Today* "Las computadoras portátiles se acercan poco a poco a la popularidad de las PC" (6 de febrero, 2005), la PC de escritorio está terminando y se ha iniciado la era de las portátiles. Un estudio del investigador Current Analysis encontró que los precios de las portátiles bajaron a un promedio de \$1211 durante los días feriados de diciembre de 2004. El precio medio de compra se situó en \$1240. Lo que encontró el investigador está basado en el precio de compra de una muestra de 35 computadoras portátiles adquiridas en fechas recientes. Suponiendo que $\sigma = \$66.75$, ¿el investigador tiene suficiente evidencia para publicar lo que encontró, o podrían sus resultados muestrales haber ocurrido realmente por casualidad?
a. Describa el parámetro de interés.
b. Exprese las hipótesis nula y alternativa.
c. Calcule el valor de $z\bigstar$ y encuentre el valor p.
d. Exprese su decisión y conclusión usando $\alpha = 0.001$.

8.114 Uno de los mejores indicadores de la salud de un bebé es el peso que tenga al nacer. En Estados Unidos, las madres que viven en la pobreza por lo general tienen bebés con menor peso al nacer que las que no viven en la pobreza. Aun cuando el peso promedio al nacer para bebés nacidos en Estados Unidos es aproximadamente 3300 gramos, el peso al nacer para bebés de mujeres que viven en la pobreza es de 2800 gramos con una desviación estándar de 500 gramos. Recientemente, un hospital local introdujo un innovador y nuevo programa de cuidado prenatal para reducir el número de bebés con bajo peso al nacer, y que nazcan en el hospital. Al final del primer año, se recolectaron los pesos al nacer de 25 bebés seleccionados al azar; todos los bebés nacieron de mujeres que vivían en la pobreza y participaron en el programa. Su peso medio al nacer era de 3075 gramos. La pregunta formulada para usted, investigador, es, "¿Ha habido una mejoría significativa en los pesos al nacer de bebés nacidos de mujeres pobres?" Use $\alpha = 0.02$.

Fuente: http://www.ccnmtl.columbia.edu/projects/qmss/t_one.html

a. Defina el parámetro.
b. Exprese las hipótesis nula y alternativa.
c. Especifique los criterios de prueba de hipótesis.
d. Presente la evidencia muestral.
e. Encuentre la información de distribución de probabilidad.
f. Determine los resultados.

8.115 La propietaria de una cadena local de tiendas de abarrotes está siempre tratando de reducir al mínimo el tiempo que tardan sus clientes en salir. En el pasado, ella ha dirigido numerosos estudios de los tiempos de salida, y éstos han mostrado una distribución normal con un tiempo medio de 12 minutos y una desviación estándar de 2.3 minutos. Ella ha puesto en práctica un nuevo horario para cajeras, con la esperanza de reducir el tiempo medio de salida. Una muestra aleatoria de 28 clientes que visitan su tienda esta semana resultó en una media de 10.9 minutos. ¿Tiene ella suficiente evidencia para decir que el tiempo medio de salida esta semana fue menor a 12 minutos? Use $\alpha = 0.02$.

8.116 El tamaño promedio de una casa en 2003 era de 2320 pies cuadrados según la National Association of Home Builders y publicado en *USA Today* Snapshots (27 de noviembre, 2004). Los constructores de casas de una ciudad del noreste piensan que el tamaño promedio de casas continúa en aumento cada año. Para probar su dicho, se seleccionó una muestra aleatoria de 45 casas nuevas que revelaron un promedio de 2490 pies cuadrados. Suponiendo que la desviación estándar poblacional es aproximadamente de 450 pies cuadrados, ¿hay evidencia de que el tamaño promedio ha aumentado desde la cifra de 2003? Use un nivel de significación de 0.05.

8.117 De dulces a joyería a flores, se esperaba que el consumidor promedio gastara \$104.63 para el día de la madre en 2005, según el artículo "Mamá tiene más este año" de *Democrat & Chronicle* (7 de mayo, 2005). Los comerciantes locales pensaron que este promedio era demasiado alto para su zona y contrataron una agencia para que realizara un estudio. Se tomó una muestra aleatoria de 60 consumidores en una zona comercial bajo techo de la localidad, el sábado antes del día de la madre, y se obtuvo una cantidad media muestral de \$94.27. Si σ = \$29.50, ¿la muestra proporciona suficiente evidencia para apoyar lo dicho por los comerciantes al nivel de significación de 0.05?

8.118 Imagine que usted es un cliente que vive en la zona comercial descrita en el ejercicio 8.117 y que necesita comprar un regalo para el día de la madre. Identifique los cuatro posibles resultados y describa la situación de que se trata con cada resultado, respecto a la cantidad promedio gastada en un regalo del día de la madre. ¿Cuál es el error más grave: el error tipo I o el tipo II? Explique.

8.119 ¿Quién dice que cuanto más se gasta en un reloj de pulsera, más precisa será la hora que marque el reloj? Alguien dice que se puede comprar un reloj de cuarzo por menos de \$25 y que mantiene un tiempo tan preciso como los relojes que cuestan cuatro veces más. Supongamos que la precisión promedio para todos los relojes que se venden hoy en día, cualquiera que sea su precio, no varía más de 19.8 segundos por mes con una desviación estándar de 9.1 segundos. Se toma una muestra aleatoria de 36 relojes de cuarzo de un precio menor a \$25, y una comprobación de su precisión deja ver un error medio austral de 22.7 segundos por mes. Con base en esta evidencia, complete la prueba de hipótesis de H_o: $\mu = 20$ vs. H_a: $\mu > 20$ al nivel de significación de 0.05 usando el método de valor de probabilidad.

a. Defina el parámetro.
b. Exprese las hipótesis nula y alternativa.
c. Especifique los criterios de prueba de hipótesis.
d. Presente la evidencia muestral.
e. Encuentre la información de distribución de probabilidad.
f. Determine los resultados.

8.120 El National Health and Nutrition Examination Survey (NHANES) indica que más adultos norteamericanos tienen exceso de peso o son obesos, lo que se define como tener un índice de masa corporal (BMI) de 25 o más. Los datos de los Centro de Control y Prevención de Enfermedades (CDC) indican que, para mujeres de 35 a 55 años, el BMI es 25.12 con una desviación estándar de 5.3. En un estudio similar que examinó a especialistas cardiovasculares mujeres, registradas en Estados Unidos y que estaban dentro del mismo rango de edades, resultaron las siguientes puntuaciones de BMI:

22 28 26 19

••• Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: "An Assessment of Cardiovascular Risk Behaviors of Registered Cardiovascular Technologists", Disertación por la doctora. Susan Wambold, Universidad de Toledo, 2002

Pruebe la afirmación de que las especialistas cardiovasculares tienen un BMI promedio más bajo que la población en general. Use $\alpha = 0.05$.

a. Describa el parámetro de interés.
b. Exprese las hipótesis nula y alternativa.
c. Calcule el valor para z★ y encuentre el valor p.
d. Exprese su decisión y conclusión usando a = 0.05.

8.121 Use una computadora o calculadora para seleccionar 40 números aleatorios de un solo dígito. Encuentre la media muestral, $z\bigstar$, y el valor p para probar H_o: $\mu = 4.5$ contra una alternativa de dos colas. Repita varias veces como en la tabla 8.8. Describa lo que encuentre.

PARA SU INFORMACIÓN Use comandos para generar datos de enteros en la página 101; luego continúe con los comandos de prueba de hipótesis de las páginas 438-439.

8.122 Use computadora o calculadora para seleccionar 36 números aleatorios de una distribución normal con media de 100 y desviación estándar de 15. Encuentre la media muestral, $z\bigstar$, y valor p para probar una prueba de hipótesis de dos colas de $\mu = 100$. Repita varias veces como en la tabla 8.8. Describa lo que encuentre.

PARA SU INFORMACIÓN Use comandos para generar datos de enteros en las páginas 327-328, luego continúe con los comandos de prueba de hipótesis de las páginas 438-439.

8.6 Prueba de hipótesis para la media μ (σ conocida): un enfoque clásico

En la sección 8.4 examinamos los conceptos y buena parte del razonamiento que hay tras una prueba de hipótesis y vimos ejemplos no estadísticos. En esta sección vamos a formalizar el procedimiento de prueba de hipótesis como se aplica a enunciados respecto a la media μ de una población bajo la restricción de que σ, la desviación estándar de población, es un valor conocido.

Suposición para pruebas de hipótesis acerca de la media μ usando una σ conocida: la distribución muestral de $\bar{x}$ tiene una distribución normal.

La información que necesitamos para asegurar que esta suposición se satisface está contenida en la distribución muestral de medias muestrales, así como en el teorema de límite central (vea capítulo 7, pp. 369-370).

La distribución de medias muestrales $\bar{x}$ está distribuida alrededor de una media igual a μ, con un error estándar igual a $\sigma/\sqrt{n}$; y (1) si la población muestreada al azar está normalmente distribuida, entonces $\bar{x}$ está normalmente distribuida para todos los tamaños muestrales, o (2) si la población muestreada al azar no está normalmente distribuida, entonces $\bar{x}$ está normalmente distribuida en forma aproximada para tamaños muestrales que sean suficientemente grandes.

La prueba de hipótesis es un bien organizado procedimiento de paso a paso que se emplea para tomar una decisión. Por lo general se emplean dos formados diferentes para pruebas de hipótesis. El *método clásico* es el proceso de prueba de hipótesis que ha disfrutado de popularidad durante muchos años. Este método, o acercamiento, está organizado en un procedimiento de cinco pasos.

LA PRUEBA CLÁSICA DE HIPÓTESIS: UN PROCEDIMIENTO DE CINCO PASOS

Paso 1 El inicio:

a. Describir el parámetro poblacional de interés.
b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

Paso 2 Criterios de la prueba de hipótesis:

a. Comprobar las suposiciones.
b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
c. Determinar el nivel de significación, α.

Paso 3 La evidencia muestral:

a. Recolectar la información muestral.
b. Calcular el valor de la estadística de prueba.

Paso 4 Distribución de probabilidad:

a. Determinar la región crítica y valor(es) crítico(s).
b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

Paso 5 Los resultados:

a. Expresar la decisión acerca de H_o.
b. Expresar la conclusión acerca de H_a.

Un fabricante de aviones comerciales compra remaches para usarlos en el ensamble de aviones. Todo proveedor de remaches que desee vender remaches al fabricante de aviones debe demostrar que sus remaches satisfacen las especificaciones requeridas. Una de éstas es, "La resistencia media al corte de todos los remaches, μ, es al menos de 925 libras". Cada vez que el fabricante de aviones compra remaches, está preocupado por que la resistencia media pueda ser menor que la especificación de 925 libras.

Nota: cada remache individual tiene una resistencia al corte, que se determina midiendo la fuerza necesaria para cortar ("romper") el remache. Es evidente que no todos los remaches se pueden probar. Por tanto, se probará una muestra de remaches y una decisión acerca de la resistencia media de todos los remaches no probados estará basada en la media de los muestreados y probados.

Paso 1 El inicio:

a. **Describir el parámetro poblacional de interés.**
El parámetro poblacional de interés es la media μ, la resistencia media al corte (o fuerza media necesaria para cortar) de los remaches considerados para compra.

b. **Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a)**
La hipótesis nula y la hipótesis alternativa se formulan por inspección del problema o enunciado a ser investigado, y formular primero dos enunciados opuestos acerca de la media μ. Para nuestro ejemplo, estos dos enunciados opuestos son: (A) "La resistencia media al corte es menor a 925" ($\mu < 925$, la preocupación del fabricante de aviones), y (B) "La resistencia media al corte es al menos 925" ($\mu = 925$, el dicho del proveedor de remaches y especificación del fabricante de aviones).

Nota: la ley de la tricotomía de álgebra expresa que dos valores numéricos deben estar relacionados en exactamente una de las tres posibles relaciones: $<$, $=$, o $>$.

PARA SU INFORMACIÓN
En las páginas 417-418 se da información más específica.

Estas tres posibilidades deben considerarse en las dos hipótesis opuestas para que las dos hipótesis sean negaciones una de la otra. Las tres posibles combinaciones de signos e hipótesis se muestran en la tabla 8.9. Recuerde que la hipótesis nula asigna un valor específico al parámetro en cuestión, y por tanto "es igual a" siempre será parte de la hipótesis nula.

TABLA 8.9

Los tres posibles enunciados de hipótesis nula y alternativa

Hipótesis nula	Hipótesis alternativa
1. Mayor o igual a ($\geq$)	Menor que ($<$)
2. Menor o igual a ($\leq$)	Mayor que ($>$)
3. Igual a ($=$)	No igual a ($\neq$)

El parámetro de interés, la media poblacional μ, está relacionada con el valor 925. El enunciado (A) se convierte en hipótesis alternativa:

$$H_a\text{: } \mu < 925 \text{ (la media es menor que 925)}$$

Este enunciado representa el interés del fabricante de aviones y dice, "Los remaches no satisfacen las especificaciones requeridas". El enunciado (B) se convierte en la hipótesis nula:

$$H_o\text{: } \mu = 925\ (\geq) \text{ (la media es al menos 925)}$$

Esta hipótesis representa la negación del interés del fabricante de aviones y dice, "Los remaches satisfacen las especificaciones requeridas".

Nota: escribiremos la hipótesis nula con sólo el signo igual, con lo que se expresa el valor exacto asignado. Cuando "igual" se parea con "menor que" o con "mayor que", el símbolo combinado se escribe además de la hipótesis nula como recordatorio de que los tres signos han sido considerados en estos dos enunciados opuestos.

Antes de continuar con nuestro ejemplo, veamos tres ejemplos que demuestran la formulación de las hipótesis nula y alternativa abarcando la media poblacional μ. Los ejemplos 8.18 y 8.19 demuestran cada uno una hipótesis alternativa de "una cola".

EJEMPLO 8.18 Escritura de hipótesis nula y alternativa (situación de una cola)

A un grupo defensor del consumidor le gustaría rechazar el dicho de un fabricante de automóviles de que un modelo específico va a promediar 24 millas por galón de gasolina. Específicamente, al grupo le gustaría demostrar que el recorrido medio en millas por galón es considerablemente menor a 24. Exprese las hipótesis nula y alternativa.

SOLUCIÓN Para expresar dos hipótesis, primero necesitamos identificar el parámetro poblacional en cuestión: "el recorrido medio en millas alcanzado por este modelo de auto". El parámetro μ está siendo comparado con el valor de 24 millas por galón, que es el valor específico de interés. Los defensores están cuestionando el valor de μ y desean demostrar que es menor a 24 (es decir, $\mu = 24$). Hay tres posibles relaciones: (1) $\mu < 24$, (2) $\mu = 24$, y (3) $\mu > 24$. Estos tres casos deben arreglarse para formar dos enunciados opuestos: uno expresa que los defensores están tratando de demostrar que "El nivel medio es menor a 24 ($\mu < 24$)", mientras que la "negación" es "El nivel medio no es menor a 24 ($\mu \geq 24$)". Uno de estos dos enunciados se convertirá en la hipótesis nula H_o, y el otro se convertirá en la hipótesis alternativa H_a.

Nota: recuerde que hay dos reglas para formar las hipótesis: (1) la hipótesis nula expresa que el parámetro en cuestión tiene un valor especificado ("H_o debe contener el signo igual"), y (2) la opinión del grupo de defensa del consumidor se convierte en la hipótesis alternativa "menor que". Ambas reglas indican:

$$H_o: \mu = 24\ (\geq) \quad \text{y} \quad H_a: \mu < 24$$

EJEMPLO 8.19

Escritura de hipótesis nula y alternativa (situación de una cola)

Suponga que la EPA está demandando a una gran empresa manufacturera por no satisfacer los lineamientos federales de emisiones contaminantes. De manera específica, la EPA está diciendo que la cantidad media de dióxido de azufre en el aire es peligrosamente alta, más de 0.09 partes por millón. Exprese las hipótesis nula y alternativa para esta situación de prueba.

SOLUCIÓN El parámetro de interés es la cantidad media de dióxido de azufre en el aire, y 0.09 partes por millón es el valor especificado. $\mu > 0.09$ corresponde a "La cantidad media es mayor a 0.09", mientras que $\mu \leq 0.09$ corresponde a la negación, "La cantidad media no es mayor a 0.09". Por lo tanto, las hipótesis son

$$H_o: \mu = 0.09\ (\leq) \quad \text{y} \quad H_a: \mu > 0.09$$

El ejemplo 8.20 demuestra una hipótesis alternativa de "dos colas".

EJEMPLO 8.20

Escritura de hipótesis nula y alternativa (situación de dos colas)

La satisfacción en el trabajo es muy importante para la productividad de trabajadores. Un cuestionario estándar sobre satisfacción en el trabajo fue aplicado por oficiales sindicales, a una muestra de trabajadores de una línea de montaje de una gran fábrica, con la esperanza de demostrar que la calificación media de los trabajadores de una línea de montaje en este cuestionario sería diferente de la media establecida de 68. Exprese las hipótesis nula y alternativa.

SOLUCIÓN O bien, la calificación media de satisfacción en el trabajo es diferente de 68 ($\mu \neq 68$) o es igual a 68 ($\mu = 68$). Por tanto,

$$H_o: \mu = 68 \quad \text{y} \quad H_a: \mu \neq 68$$

Notas:

1. La hipótesis alternativa se denomina de "dos colas" cuando H_a es "no igual".
2. Cuando "menor a" se combina con "mayor que," se convierten en "no igual a".

El punto de vista del experimentador afecta en gran medida la manera en que se forman las hipótesis. Por lo general, el experimentador está tratando de demostrar que el valor del parámetro es diferente del valor especificado. Así, con frecuencia el experimentador espera ser capaz de rechazar la hipótesis nula para que su teoría se justifique. Los ejemplos 8.18, 8.19 y 8.20 también representan los tres posibles arreglos para las relaciones <, = y > entre el parámetro μ y un valor especificado.

La tabla 8.6 es una lista de frases comunes adicionales que se emplean en frases e indica sus negaciones y la hipótesis en la que se usará cada frase. De nuevo, observe que "igual a" está siempre en la hipótesis nula. También observe que la negación de "menor a" es "mayor o igual a". Considere la negación como "todos los otros" del conjunto de tres signos.

TABLA 8.10

Frases comunes y sus negaciones

H_0: ($\geq$) vs.	H_a: ($<$)	H_0: ($\leq$) vs.	H_a: ($>$)	H_0: ($=$) vs.	H_a: ($\neq$)
Al menos	Menor que	A lo sumo	Más que	Es	No es
No menos de	Menor que	No más de	Más que	No diferente de	Diferente de
No menor a	Menor que	No mayor que	Mayor que	Igual que	No igual a

Una vez establecidas las hipótesis, nula y alternativa, trabajaremos bajo la suposición de que la hipótesis nula es un enunciado verdadero hasta que haya evidencia suficiente para rechazarlo. Esta situación podría compararse con un juicio en la sala de un tribunal, donde se supone que el acusado es inocente (H_o: El acusado es inocente vs. H_a: El acusado no es inocente) hasta que se haya presentado evidencia suficiente para demostrar que la inocencia es totalmente increíble ("fuera de toda duda razonable"). En la conclusión de la prueba de hipótesis, tomaremos una de dos posibles decisiones. Decidiremos en oposición a la hipótesis nula y diremos que "rechazamos H_o" (esto corresponde a la "sentencia" del acusado en un juicio), o decidiremos de acuerdo con la hipótesis nula y diremos que "no rechazamos H_o" (esto corresponde a "no sentenciar" o una "absolución" del acusado en un juicio).

Regresemos al ejemplo de los remaches que interrumpimos en la página 446 y continuemos con el paso 2. Recuerde que

$$H_o: \mu = 925\ (\geq)\ \text{(al menos 925)} \quad H_a: \mu < 925\ \text{(menor a 925)}$$

PASO 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones:

Suponga que la desviación estándar de la resistencia al corte de remaches se sabe por experiencias pasadas que es $\sigma = 18$. Las variables como la resistencia al corte por lo general tienen una distribución agrupada; por tanto, una muestra de tamaño 50 debe ser suficiente-

mente grande para que el teorema de límite central (CLT) aplique, y asegurar para que la distribución de medias muestrales (SDSM) esté normalmente distribuida.

b. Identifique la distribución de probabilidad y la estadística de prueba a usar.

La distribución de probabilidad normal estándar se usa porque se espera que $\bar{x}$ tenga una distribución normal.

Para una prueba de hipótesis de μ, deseamos comparar el valor de la media muestral con el valor de la media poblacional como se indicó en la hipótesis nula. Esta comparación se logra usando la estadística de prueba en la fórmula (8.4):

Estadística de prueba para media

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{8.4}$$

El valor calculado resultante se identifica como $z\bigstar$ ("z estrella"), porque se espera que tenga una distribución normal estándar cuando la hipótesis nula sea verdadera y las suposiciones se hayan satisfecho. La $\bigstar$ ("estrella") es para recordarnos que éste es un valor calculado de la estadística de prueba.

La estadística de prueba a usar es $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

c. Determine el nivel de significación, α.

Establecer α se describió como una decisión gerencial en la sección 8.4. Para ver lo que está involucrado para determinar α, la probabilidad del error tipo I, para nuestro ejemplo de remaches, empezamos por identificar los cuatro posibles resultados, sus significados, y la acción relacionada con cada uno de ellos.

El error tipo I sucede cuando se rechaza una hipótesis nula verdadera. Esto ocurriría cuando el fabricante probó los remaches que no llenaron las especificaciones y los rechazó. Sin duda, esto llevaría a que los remaches no sean comprados aun cuando satisfagan las especificaciones. Para que el gerente establezca un nivel de significación es necesaria una información relacionada, ¿qué tan pronto se necesita una nueva provisión de remaches? por ejemplo. Si se necesitan mañana y éste es el único vendedor con existencia disponible, esperar una semana para hallar remaches aceptables podría ser muy costoso; por tanto, rechazar remaches buenos podría ser considerado como un error grave. Por otra parte, si los remaches no se necesitan sino hasta el mes próximo, entonces este error podría no ser muy grave. Sólo el gerente conocerá todas las ramificaciones y, por tanto, la entrada del gerente es muy importante aquí.

PARA SU INFORMACIÓN
Hay más de esto, pero esperamos que el lector capte la idea.

PARA SU INFORMACIÓN
α se asignará en el enunciado de ejercicios.

Después de mucha consideración, el gerente asigna el nivel de significación: $\alpha = 0.05$.

PASO 3 La evidencia muestral:

a. Recolectar la información muestral.

La muestra debe ser aleatoria sacada de la población cuya media μ esté siendo cuestionada. Se selecciona una muestra aleatoria de 50 remaches, se prueba cada remache y se calcula la resistencia muestral media al corte: $\bar{x} = 921.18$ y $n = 50$.

b. Calcular el valor de la estadística de prueba.
La evidencia muestral ($\bar{x}$ y n hallada en el paso 3a) se convierte enseguida en el valor calculado de la estadística de prueba, $z\bigstar$, usando la fórmula (8.4). (μ es 925 de H_o, y $\sigma = 18$ es una cantidad conocida.) Tenemos

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}: \quad z\bigstar = \frac{921.18 - 925.0}{18/\sqrt{50}} = \frac{-3.82}{2.5456} = -1.50$$

PASO 4 La distribución de probabilidad:

a. Determinar la región crítica y valor(es) crítico(s)

La variable z normal estándar es nuestra estadística de prueba para esta prueba de hipótesis.

Región crítica: es el conjunto de valores para la estadística de prueba que nos llevará a rechazar la hipótesis nula. El conjunto de valores que no están en la región crítica se denomina **región no crítica** (a veces llamada *región de aceptación*.)

Recuerde que estamos trabajando bajo la suposición de que la hipótesis nula es verdadera. Así, estamos suponiendo que la resistencia media al corte de todos los remaches de la población muestreada es 925. Si éste es el caso, entonces cuando seleccionamos una muestra aleatoria de 50 remaches, podemos esperar que esta media muestral, $\bar{x}$, sea parte de una distribución normal que se centra en 925 y tener un error estándar de $\sigma/\sqrt{n} = 18/\sqrt{50}$, o sea aproximadamente 2.55. Alrededor de 95% de los valores de media muestral serán mayores a 920.8 (un valor de 1.65 errores estándar debajo de la media: 925 − (1.65)(2.55) − 920.8). Así, si H_o es verdadera y $\mu = 925$, entonces esperamos que $\bar{x}$ sea mayor a 920.8 aproximadamente 95% del tiempo y menos de 920.8 sólo 5% del tiempo.

Si, no obstante, el valor de $\bar{x}$ que obtenemos de nuestra muestra es menor a 920.9, por ejemplo 919.5, tendremos que escoger. Podría ser que: (A)ese valor de $\bar{x}$ (919.5) sea miembro de la distribución con media de 925 aun cuando tiene muy baja probabilidad de que se presente (menor a 0.05), o bien, (B) $\bar{x} = 919.5$ es miembro de una distribución muestral cuya media es menor a 925, que la haría un valor que es más probable que ocurra.

¿SABÍA USTED...?

Disputas en método

Las estadísticas no son sólo matemáticas. Hay diferentes modos de abordar inferencias estadísticas y diferentes modos de interpretar lo que nos dicen los datos. Cuanto más significativas sean las diferencias, es más probable que haya acalorados desacuerdos entre quienes tengan puntos de vista opuestos. Una de estas disputas surgió en 1935 en una discusión en la Royal Statistical Society cuando R. A. Fisher desafió a Jerzy Neyman para que se pusiera al corriente en el tema que estaban tratando. La disputa se centró en el uso de intervalos de confianza y el método para probar hipótesis de Pearson y Neyman, contra los intervalos y concepto de valores *p* de Fischer en pruebas de significación. La enemistad duró hasta la muerte de Fisher en 1962.

En estadística, "apostamos" a que "ocurra lo más probable" y consideramos que la segunda opción (B) es la correcta. Así, la cola del lado izquierdo de la distribución z se convierte en la región crítica, y el nivel de significación α se convierte en la medida de su área.

PARA SU INFORMACIÓN
En las páginas 338-340 se da información acerca de la notación de valor crítico, $z(\alpha)$.

PARA SU INFORMACIÓN
Se usará sombreado para identificar la región crítica.

Valor(es) crítico(s): valor(es) "primero" o "frontera" de la(s) región(es) crítica(s).

El valor crítico para nuestro ejemplo es $-z(0.05)$ y tiene el valor de -1.65, como se ve en la tabla 4a del apéndice B.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.
Gráficamente esta determinación se muestra al localizar el valor de $z\star$ en el dibujo del paso 4a.
El valor calculado de z, $z\star = -1.50$, **no es está en la región crítica** (está en la parte no sombreada de la figura).

Paso 5 El resultado:

a. Expresar la decisión acerca de H_o
Para tomar la decisión, necesitamos conocer la *regla de decisión.*

Regla de decisión

a. Si la estadística de prueba cae *dentro de la región crítica*, entonces la decisión debe ser **rechazar H_o**. (El valor crítico es parte de la región crítica.)

b. Si la estadística de prueba *no está en la región crítica*, entonces la decisión debe ser **no rechazar H_o**.

La decisión es: no rechazar H_o.

PARA SU INFORMACIÓN
En la página 422 se da información específica acerca de escribir la conclusión.

a. Expresar la conclusión acerca de H_a.
No hay suficiente evidencia al nivel de significación 0.05 que demuestre que los remaches tienen una resistencia media al corte menor a 925. "No lo sentenciamos" es la hipótesis nula. En otras palabras, una media muestral de sólo 921.18 no es nada probable que ocurra (como lo define α) cuando el valor verdadero de la media poblacional es 925.0. Por tanto, la acción resultante sería comprar los remaches.

Antes de ver otro ejemplo, resumamos brevemente algunos de los detalles que hemos visto hasta aquí:

1. La hipótesis nula especifica un valor particular de un parámetro poblacional.
2. La hipótesis alternativa puede tomar tres formas. Cada una de las formas dicta una situación específica de la(s) región(es) crítica(s), como se muestra en la tabla siguiente.
3. Para numerosas pruebas de hipótesis, el signo de la hipótesis alternativa "apunta" en la dirección en la que está situada la región crítica. (Considere que el signo no igual [$\neq$] es al tiempo menor que [<] y mayor que [>], con lo que apunta en ambas direcciones.)

	Signo de la hipótesis alternativa		
	$<$	$\neq$	$>$
Región crítica	Una región Lado izquierdo **Prueba de una cola**	Dos regiones La mitad a cada lado **Prueba de dos colas**	Una región Lado derecho **Prueba de una cola**

El valor asignado a α se denomina *nivel de significación* de la prueba de hipótesis. Alfa no puede ser interpretada como que no sea el riesgo (o probabilidad) de rechazar la hipótesis nula cuando en realidad es verdadera. Raras veces podremos determinar si la hipótesis nula es verdadera o falsa; decidiremos sólo "rechazar H_o" o "no rechazar H_o". La frecuencia relativa con la que rechazamos una hipótesis verdadera es α, pero nunca conoceremos la frecuencia relativa con la que cometemos un error de decisión. Las dos ideas son muy diferentes, es decir, un error tipo I y un error en decisión son dos cosas completamente diferentes. Recuerde que hay dos tipos de errores: tipo I y tipo II.

Veamos otra prueba de hipótesis, una de las cuales comprende un procedimiento de dos colas.

EJEMPLO 8.21 Prueba de hipótesis de dos colas

Se ha dicho que el peso medio de muchachas estudiantes en una universidad es 54.4 kg. El profesor Hart no cree esto y se propone demostrar que el peso medio no es 54.4 kg. Para probar la frase, recolecta una muestra aleatoria de 100 pesos de entre las estudiantes y resulta una media muestral de 53.75 kg. ¿Es esto suficiente evidencia para que el profesor Hart rechace el enunciado? Use $\alpha = 0.05$ y $\sigma = 5.4$ kg.

SOLUCIÓN

PASO 1 El inicio:

a. Describir el parámetro poblacional de interés.

El parámetro poblacional de interés es la media μ, el peso medio de todas las estudiantes de la universidad.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

El peso medio es igual a 54.4 kg, o el peso medio no es igual a 54.4 kg.

H_o: $\mu = 54.4$ kg (peso medio es 54.4)

H_a: $\mu \neq 54.4$ (peso medio no es 54.4)

(Recuerde: $\neq$ es < y > juntos.)

Paso 2 Criterios de la prueba de hipótesis:

a. Comprobar las suposiciones.

σ se conoce. Los pesos de un grupo de mujeres adultas está por lo general distribuido normalmente en forma aproximada; por tanto, una muestra de $n = 100$ es grande lo suficiente para permitir se aplique el teorema de límite central (CLT).

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.

La distribución de probabilidad normal estándar y la estadística de prueba $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ se usarán; $\sigma = 5.4$.

c. Determinar el nivel de significación, α.

$\alpha = 0.05$ (dado en el enunciado del problema).

PASO 3 Evidencia muestral:

a. Recolectar la información muestral: $\bar{x} = 53.75$ y $n = 100$.

b. Calcular el valor de la estadística de prueba.

Use la fórmula (8.4), información de H_o: $\mu = 54.4$, y $\sigma = 5.4$ (conocida):

$$z\bigstar = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}: \quad z\bigstar = \frac{53.75 - 54.4}{5.4/\sqrt{100}} = \frac{-0.65}{0.54} = -1.204 = -1.20$$

PASO 4 Distribución de probabilidad:

a. Determinar la región crítica y valor(es) crítico(s).

La región crítica es la cola izquierda y la cola derecha porque los valores mayor y mayor de la media muestral sugieren que la hipótesis nula es errónea. El nivel de significación se dividirá en dos, con 0.025 siendo la medida de cada cola. Los valores críticos se encuentran en la tabla 4B del apéndice B: $\pm z(0.025) = \pm 1.96$. (Las instrucciones de la tabla 4B están en la página 403.)

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

El valor calculado de z, $z\bigstar = -1.20$, no está en la región crítica (se muestra en la región sombreada en la figura siguiente).

PASO 5 Los resultados:

a. Expresar la decisión acerca de H_o: no rechazar H_o.

b. Expresar la conclusión acerca de H_a:
No hay suficiente evidencia al nivel de significación de 0.05 para demostrar que las estudiantes mujeres tienen un peso medio diferente de 54.4 kg dicho. En otras palabras, no hay evidencia estadística para apoyar la opinión del profesor Hart.

EJEMPLO 8.22 Prueba de hipótesis de dos colas con datos muestrales

Según los resultados del ejercicio 5.32 (p. 282), la media de números aleatorios de un solo dígito es 4.5 y la desviación estándar es $\sigma = 2.87$. De la tabla 1 del apéndice B saque una muestra aleatoria de 40 números de un solo dígito y pruebe la hipótesis, "La media de los números de un solo dígito de la tabla 1 es 4.5". Use $\alpha = 0.10$.

SOLUCIÓN

PASO 1 El inicio:

a. Describir el parámetro poblacional de interés.
El parámetro de interés es la media μ de la población de números de un solo dígito de la tabla 1 del apéndice B.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

H_o: $\mu = 4.5$ (la media es 4.5)

H_a: $\mu \neq 4.5$ (la media no es 4.5)

PASO 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones.
σ se conoce. Las muestras de tamaño 40 deben ser grandes lo suficiente para satisfacer el CLT; vea la discusión de este problema en la página 426.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
Usamos la distribución de probabilidad normal estándar y la estadística de prueba $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$; $\sigma = 2.87$.

c. Determinar el nivel de significación, α.
$\alpha = 0.10$ (dado en el enunciado del problema).

PASO 3 Evidencia muestral:

a. Recolectar la información muestral.
Esta muestra aleatoria fue sacada de la tabla 1 del apéndice B.

TABLA 8.11

Muestra aleatoria de números de un solo dígito

2	8	2	1	5	5	4	0	9	1
0	4	6	1	5	1	1	3	8	0
3	6	8	4	8	6	8	9	5	0
1	4	1	2	1	7	1	7	9	3

Las estadísticas muestrales son $\bar{x} = 3.975$ y $n = 40$.

b. Calcular el valor de la estadística de prueba.
Use la fórmula (8.4), información de H_o: $\mu = 4.5$, y $\sigma = 2.87$:

$$z\star = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}: \quad z\star = \frac{3.975 - 4.50}{2.87/\sqrt{40}} = \frac{-0.525}{0.454} = -1.156 = -1.16$$

PASO 4 Distribución de probabilidad:

a. Determinar la región crítica y valor(es) crítico(s).
Se usará una región crítica de dos colas, y 0.05 será el área en cada cola. Los valores críticos son $\pm z(0.05) = \pm 1.65$.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.
El valor calculado de z, $z\star = -1.16$, no está en la región crítica (área sombreada).

PASO 5 El resultado:

a. Expresar la decisión acerca de H_o: no rechazar H_o.

b. Expresar la conclusión acerca de H_a.
La media muestral observada no es significativamente diferente de 4.5 al nivel de significación 0.10.

Suponga que fuéramos a tomar otra muestra de tamaño 40 de la tabla 1. ¿Obtendríamos los mismos resultados? Suponga que tomamos una tercera muestra y una cuarta. ¿Qué resultados podríamos esperar? ¿Cuál es el nivel de significación? Si, su valor es 0.10, pero, ¿qué mide? La tabla 8.12 es una lista de medias obtenidas de 20 muestras aleatorias diferentes de tamaño 40 que se tomaron de la tabla 1 del apéndice B. También se citan en lista el valor calculado de $z\star$ que corresponde a cada $\bar{x}$ y la decisión que cada uno dictaría. Los 20 puntajes z calculados se presentan en la figura 8.10. Nótese que 3 de las 20 muestras (15%) nos llevaron a rechazar la hipótesis nula, aun cuando sabemos que la hipótesis nula es verdadera para esta situación. ¿Puede el lector explicar esto?

TABLA 8.12

Veinte muestras aleatorias de tamaño 40 tomadas de la tabla 1 del apéndice B

Número de muestra	Media muestral $\bar{x}$	z calculada $z\star$	Decisión alcanzada	Número de muestra	Media muestral $\bar{x}$	z calculada $z\star$	Decisión alcanzada
1	4.62	+0.26	No rechazar H_o	11	4.70	+0.44	No rechazar H_o
2	4.55	+0.11	No rechazar H_o	12	4.88	+0.83	No rechazar H_o
3	4.08	−0.93	No rechazar H_o	13	4.45	−0.11	No rechazar H_o
4	5.00	+1.10	No rechazar H_o	14	3.93	−1.27	No rechazar H_o
5	4.30	−0.44	No rechazar H_o	15	5.28	+1.71	Rechazar H_o
6	3.65	−1.87	Rechazar H_o	16	4.20	−0.66	No rechazar H_o
7	4.60	+0.22	No rechazar H_o	17	3.48	−2.26	Rechazar H_o
8	4.15	−0.77	No rechazar H_o	18	4.78	+0.61	No rechazar H_o
9	5.05	+1.21	No rechazar H_o	19	4.28	−0.50	No rechazar H_o
10	4.80	+0.66	No rechazar H_o	20	4.23	−0.61	No rechazar H_o

FIGURA 8.10
Puntajes *z* de la tabla 8.12

Nota: recuerde que α es la probabilidad de que "rechazamos H_o" cuando en realidad es un enunciado verdadero. Por tanto, podemos anticipar que un error tipo I ocurrirá a α en el momento en que se pruebe una hipótesis nula verdadera. En la situación empírica precedente, observamos un porcentaje de rechazo de 15%. Si fuéramos a repetir este experimento muchas veces, la proporción de muestras que llevaría a un rechazo variaría, pero la frecuencia de rechazo relativa observada debería ser aproximadamente α o 10 por ciento.

SECCIÓN 8.6 EJERCICIOS

8.123 En el ejemplo de la página 445, el fabricante de aviones que está comprando remaches está preocupado de que éstos no puedan satisfacer la especificación de resistencia media. Exprese las hipótesis nula y alternativa del fabricante de aviones.

8.124 El profesor Hart no cree en el enunciado "la distancia media recorrida diariamente por los estudiantes no residentes en nuestra universidad es más de 9 millas". Exprese las hipótesis nula y alternativa que usaría para desafiar este enunciado.

8.125 Exprese las hipótesis nula y alternativa empleada para probar cada una de las frases siguientes:

a. El tiempo medio de reacción es menor a 1.25 segundos.

b. La puntuación media en ese examen de calificación es diferente de 335.

c. El precio medio de venta de casas en la zona no es más de $230 000.

8.126 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usarían para una prueba de hipótesis de los siguientes enunciados:

a. La edad media de jóvenes frecuentan la zona comercial es menor a 16 años.

b. La estatura media de jugadores profesionales de baloncesto es mayor a 6'6".

c. La caída de elevación media para pistas de esquí en centros de esquiar del este es al menos 285 pies.

d. El diámetro medio de remaches es de no más de 0.375 pulgadas.

e. El nivel medio de colesterol de estudiantes universitarios hombres es diferente de 200 mg/dL.

8.127 Suponga que usted desea probar la hipótesis de que "el contenido medio de sal de comidas congeladas `lite´ (de dieta) es más de 350 mg por porción". Un promedio de 350 mg es una cantidad aceptable de sal por porción; por tanto, usted lo usa como estándar. La hipótesis nula es "El contenido promedio es de no más de 350 mg" ($\mu = 350$). La hipótesis alternativa es "El contenido promedio es más de 350 mg" ($\mu > 350$).

a. Describa las condiciones que existirían si su decisión resulta en un error tipo I.

b. Describa las condiciones que existirían si su decisión resulta en un error tipo II.

8.128 Identifique los cuatro posibles resultados y describa la situación involucrada en cada resultado respecto a la prueba y compra de remaches del fabricante de aviones. ¿Cuál es el error más grave: el tipo I o el tipo II? Explique.

8.129 Suponga que usted desea probar la hipótesis de que el cargo medio mínimo en su zona, por una llamada de servicio a domicilio de un plomero, es a lo sumo \$85. Explique las condiciones que existirían si comete un error de decisión al cometer un error tipo I. ¿Qué hay si es un error tipo II?

8.130 Describa la forma en que la hipótesis nula del ejemplo 8.21 es un "punto de partida" para la decisión a tomarse acerca del peso medio de todas las estudiantes de la universidad.

8.131 a. ¿Cuál es la región crítica?

b. ¿Cuál es el valor crítico?

8.132 a. ¿A qué decisión se llega cuando la estadística de prueba cae en la región crítica?

b. ¿A qué decisión se llega cuando la estadística de prueba cae en la región no crítica?

8.133 Debido a que el tamaño del error tipo I siempre se puede hacer más pequeño al reducir el tamaño de la región crítica, ¿por qué no siempre escogemos regiones críticas que hacen una α extremadamente pequeña?

8.134 Calcule la estadística de prueba $z\bigstar$, dada H_0: $\mu = 356$, $\sigma = 17$, $\bar{x} = 354.3$, y $n = 120$.

8.135 Encuentre la región crítica y valor(es) para H_a: $\mu < 19$ y $\alpha = 0.01$.

8.136 Encuentre la región crítica y valor(es) para H_a: $\mu > 34$ y $\alpha = 0.02$.

8.137 Determine la región crítica y valores críticos para z que se usarían para probar la hipótesis nula al nivel de significación determinado, como se describe en cada uno de lo siguiente:

a. H_o: $\mu = 20$, H_a: $\mu \neq 20$, $\alpha = 0.10$

b. H_o: $\mu = 24$ ($\leq$), H_a: $\mu > 24$, $\alpha = 0.01$

c. H_o: $\mu = 10.5$ ($\geq$), H_a: $\mu < 10.5$, $\alpha = 0.05$

d. H_o: $\mu = 35$, H_a: $\mu \neq 35$, $\alpha = 0.01$

8.138 Determine la región crítica y valores críticos empleados para probar las siguientes hipótesis nulas:

a. H_o: $\mu = 55$ ($\geq$), H_a: $\mu < 55$, $\alpha = 0.02$

b. H_o: $\mu = -86$ ($\geq$), H_a: $\mu < -86$, $\alpha = 0.01$

c. H_o: $\mu = 107$, H_a: $\mu \neq 107$, $\alpha = 0.05$

d. H_o: $\mu = 17.4$ ($\leq$), H_a: $\mu > 17.4$, $\alpha = 0.10$

8.139 La hipótesis nula, H_o: $\mu = 250$, fue probada contra la hipótesis alternativa, H_a: $\mu < 250$. Resultó una muestra de $n = 85$ en una estadística de prueba calculada de $z\bigstar = -1.18$. Si $\sigma = 22.6$, encuentre el valor de la media muestral, $\bar{x}$. Encuentre la suma de los datos muestrales, Σx.

8.140 Encuentre el valor de $\bar{x}$ para cada uno de lo siguiente:

a. H_o: $\mu = 580$, $z\bigstar = 2.10$, $\sigma = 26$, $n = 55$

b. H_o: $\mu = 75$, $z\bigstar = -0.87$, $\sigma = 9.2$, $n = 35$

8.141 El valor calculado de la estadística de prueba es en realidad el número de errores estándar que la media muestral difiere del valor hipotético de μ en la hipótesis nula. Suponga que la hipótesis nula es H_o: $\mu = 4.5$, σ se sabe que es 1.0, y la muestra de tamaño 100 resulta en $\bar{x} = 4.8$.

a. ¿A cuántos errores estándar está $\bar{x}$ sobre 4.5?

b. Si la hipótesis alternativa es H_a: $\mu > 4.5$ y $\alpha = 0.01$, ¿rechazaría usted H_o?

8.142 Considere la prueba de hipótesis donde las hipótesis son H_o: $\mu = 26.4$ y H_a: $\mu < 26.4$. Una muestra de tamaño 64 se selecciona al azar y da una media muestral de 23.6.

a. Si se sabe que $\sigma = 12$, ¿cuántos errores estándar debajo de $\mu = 26.4$ está la media muestral, $\bar{x} = 23.6$?

b. Si $\alpha = 0.05$, ¿rechazaría usted H_o? Explique.

8.143 Hay sólo dos posibles decisiones como resultado de una prueba de hipótesis.

a. Exprese las dos posibles decisiones.

b. Describa las condiciones que llevarán a cada una de las dos decisiones identificadas en la parte a.

8.144 a. ¿Qué proporción de la distribución de probabilidad está en la región crítica, siempre que sea correcta la hipótesis nula?

b. ¿Qué error podría cometerse si la estadística de prueba cae en la región crítica?

c. ¿Qué proporción de la distribución de probabilidad está en la región no crítica, siempre que no sea correcta la hipótesis nula?

d. ¿Qué error podría cometerse si la estadística de prueba cae en la región no crítica?

8.145 La siguiente salida de computadora se utilizó para completar una prueba de hipótesis.

```
TEST OF MU = 15.0000 VS MU not = 15.0000
THE ASSUMED SIGMA = 0.50
N       MEAN        STDEV       SE MEAN       Z
30      15.6333     0.4270      0.0913        6.94
```

a. Exprese las hipótesis nula y alternativa.

b. Si la prueba se completa usando $\alpha = 0.01$, ¿a qué decisión y conclusión se llega?

c. Verifique el valor del error estándar de la media.

8.146 Usando la salida de computadora e información del ejercicio 8.145, determine el valor de lo siguiente:

a. Valor hipotético de media poblacional

b. Media muestral

c. Desviación estándar poblacional

d. Estadística de prueba

8.147 La siguiente salida de computadora se utilizó para completar una prueba de hipótesis.

```
TEST OF MU = 72.00 VS MU > 72.00
THE ASSUMED SIGMA = 12.0
N       MEAN        STDEV       SE MEAN       Z
36      75.2        11.87       2.00          1.60
```

a. Exprese las hipótesis nula y alternativa.

b. Si la prueba se completa usando $\alpha = 0.05$, ¿a qué decisión y conclusión se llega?

c. Verifique el valor del error estándar de la media.

8.148 Usando la salida de computadora e información del ejercicio 8.147, determine el valor de lo siguiente:

a. Valor hipotético de media poblacional

b. Media muestral

c. Desviación estándar poblacional

d. Estadística de prueba

8.149 El departamento de salud de Texas publicó los resultados estatales para el Emergency Medical Services Certification Examination. Los datos para quienes toman el examen de paramédico por primera vez dieron una calificación promedio de 79.68, (de 100 posibles), con una desviación estándar de 9.06. Suponga que una muestra aleatoria de 50 personas que toman el examen dan una calificación media de 81.05. ¿Hay evidencia suficiente para concluir que "la población de la cual se tomó esta muestra aleatoria, en promedio, obtuvo una calificación más alta que la indicada"? Use $\alpha = 0.05$.

8.150 Según el artículo "Millones [de personas] perdieron en cuentas de salud" del *Democrat & Chronicle* (26 de diciembre, 2004), se estima que cada año se pierden $210 millones de dólares porque empleados no usan todo el dinero con el que contribuyeron para sus cuentas de gastos médicos flexibles. La cantidad promedio puesta en estas cuentas anualmente es de alrededor de $1000.

Suponga que en 2005 se tomó una muestra aleatoria de 150 empleados de un negocio local, y que se depositó una cantidad promedio de $925 en sus cuentas de gastos flexibles. Pruebe la hipótesis de que no

hubo cambio significativo en las cantidades promedio de gastos flexibles a pesar de que se perdieron millones de dólares. Suponga que σ = \$307 por año. Use α = 0.05.

a. Defina el parámetro.
b. Exprese las hipótesis nula y alternativa.
c. Especifique los criterios de prueba de hipótesis.
d. Presente la evidencia muestral.
e. Encuentre la información de distribución de probabilidad.
f. Determine los resultados.

8.151 Las cifras de la Nacional Thoroughbred Racing Association muestran que la edad promedio de sus 3 millones de aficionados comprometidos, que asisten a varias carreras cada año, es ahora de 55 años de edad. A pesar de una campaña promocional de \$30 millones de dólares al año de la asociación y sus miembros, destinada a fanáticos más jóvenes y mujeres, su cálculo es que en un plazo de 10 años la edad promedio de aficionados será cercana a 65.

Fuente: http://money.cnn.com/2002/05/03/commentary/column_sportsbiz/horse_racing/

Suponga que se toma una muestra aleatoria de 35 patrocinadores de la pista de carreras de los Finger Lakes, y que la media muestral es 52.7 años. Si σ = 8 años, ¿la muestra da suficiente evidencia para apoyar la cifra actual de edad promedio de la asociación de carreras? Use un nivel de significación de 0.10.

8.152 Una compañía de seguros contra incendios pensó que la distancia media de una casa al departamento de bomberos más cercano, en un suburbio de Chicago, era de al menos 4.7 millas. De conformidad con eso ha fijado sus tarifas de seguro contra incendio. Los miembros de la comunidad salieron a demostrar que la distancia media era menos de 4.7 millas. Esto, pensaron, convencería a la compañía de seguros a bajar sus tarifas. Al azar identificaron 64 casas y midieron la distancia al departamento de bomberos más cercano a cada una. La media muestral resultante fue 4.4. Si σ = 2.4 millas, ¿la muestra da evidencia suficiente para apoyar el dicho de la comunidad al nivel de significación de α = 0.05?

8.153 La duración de juegos de béisbol de las ligas mayores está normalmente distribuida en forma aproximada y promedia 2 horas 50.1 minutos, con una desviación estándar de 21.0 minutos. Se ha dicho que los juegos de béisbol de los Yanquis de Nueva York duran, en promedio, más que los juegos de los otros equipos de ligas mayores. Para probar la verdad de esta afirmación, se identificó al azar una muestra de ocho juegos de los Yanquis y por cada uno de ellos obtuvo el "tiempo de juego" (en minutos):

199	196	202	213	187	169	169	188

Fuente: MLB.com

Al nivel de significación de 0.05, ¿estos datos muestran suficiente evidencia para concluir que el tiempo medio de los juegos de béisbol de los Yanquis es mayor que el de otros equipos de béisbol de ligas mayores?

8.154 El gerente de Air Express piensa que los pesos de paquetes enviados recientemente son menores que en otros tiempos. Los registros muestran que en el pasado, los paquetes habían tenido un peso medio de 36.5 libras y una desviación estándar de 14.2 libras. Una muestra aleatoria de los registros de embarque del mes pasado dio los siguientes 64 valores de datos:

32.1	41.5	16.1	8.9	36.2	12.3	28.4	40.4
45.5	15.2	26.5	13.3	23.5	33.7	18.3	16.3
15.4	39.7	50.3	14.8	44.4	47.7	45.8	52.3
48.4	10.4	59.9	5.5	6.7	17.1	20.0	28.1
48.1	29.5	22.9	47.8	24.8	20.1	40.1	12.6
24.3	43.3	32.4	57.7	42.9	36.7	15.5	46.4
51.3	38.6	39.4	27.1	55.7	37.7	39.4	55.5
26.9	15.7	32.3	47.8	33.2	29.1	31.1	34.5

¿Es esto suficiente evidencia para rechazar la hipótesis nula a favor de lo dicho por el gerente? Use α = 0.01.

8.155 ¿Bebe usted la cantidad recomendada de agua al día? La mayoría de norteamericanos no la toma ya que, en promedio, toman de 4.6 porciones de 8 onzas (1 oz = 0.029 de litro) de agua al día.

Fuente: http://www.bottledwater.org/public/summary.htm

Una muestra de 42 profesionales de la educación se seleccionó al azar y se observó su consumo de agua durante un periodo de 24 horas; la cantidad media consumida fue 39.3 onzas. Suponiendo que la cantidad de agua consumida al día por adultos está normalmente distribuida y la desviación estándar es 11.2 onzas, ¿hay suficiente evidencia para demostrar que los profesionales de la educación consumen, en promedio, más agua al día que el promedio nacional? Use α = 0.05.

8.156 La cantidad recomendada de agua que una persona debe tomar es ocho porciones de 8 oz por día.

a. ¿La muestra de profesionales de la educación del ejercicio 8.155 muestra suficiente evidencia de que consumen, en promedio, significativamente menos agua al día que la cantidad recomendada? Use $\alpha = 0.05$.

b. El valor del puntaje z calculado en la parte a es poco común. ¿En qué forma es poco común, y qué significa?

8.157 Use computadora o calculadora para seleccionar 40 números aleatorios de un dígito. Encuentre la media muestral y z★. Usando $\alpha = 0.05$, exprese la decisión para probar H_o: $\mu = 4.5$ contra una alternativa de dos colas. Repita varias veces como en la tabla 8.12. Describa lo que encuentre después de varios intentos.

PARA SU INFORMACIÓN

Use comandos para generar datos enteros de la página 101; luego continúe con los comandos de prueba de hipótesis de las páginas 438-439.

8.158 Use computadora o calculadora para seleccionar 36 números aleatorios de una distribución normal con media de 100 y desviación estándar 15. Encuentre la media muestral y z★ para probar una prueba de hipótesis de dos colas de $\mu = 100$. Usando $\alpha = 0.05$, exprese la decisión. Repita varias veces como en la tabla 8.12. Describa lo que encuentre.

PARA SU INFORMACIÓN

Use comandos para generar datos de las páginas 327-328; luego continúe con los comandos de prueba de hipótesis de las páginas 438-439.

REPASO DEL CAPÍTULO

En retrospectiva

En este capítulo se presentaron dos formas de inferencia: estimación y prueba de hipótesis. Se pueden usar por separado y, con frecuencia, así es. Parece natural, sin embargo, que el rechazo de una hipótesis nula sea seguido por un intervalo de confianza. (Si el valor que se afirma es erróneo, a veces buscamos una estimación para el valor verdadero.)

Estas dos formas de inferencia son muy diferentes, pero están relacionadas entre sí; hay cierta cantidad de cruce entre el uso de las dos inferencias. Por ejemplo, supongamos que usted había muestreado y calculado un intervalo de confianza de 90% para la media de una población. El intervalo fue 10.5 a 15.6. Entonces alguien dice que la verdadera media es 15.2. El intervalo de confianza de usted se puede comparar con esta última afirmación. Si el valor dicho cae dentro de su estimación de intervalo, usted no rechazaría la hipótesis nula de que $\mu = 15.2$ a un nivel de significación de 10% en una prueba de dos colas. Si el valor afirmado (16.0, por ejemplo) cae fuera del intervalo, usted rechazaría la hipótesis nula de que $\mu = 16.0$ en $\alpha = 0.10$ en una prueba de dos colas. Si se requiere una prueba de una cola, o si usted prefiere un valor de α, debe usarse una prueba de hipótesis separada.

Numerosos usuarios de estadística (en especial quienes venden un producto) dirán que sus resultados estadísticos demuestran que su producto es superior. Pero, recuerde, la prueba de hipótesis no *demuestra* ni *desaprueba* nada. La decisión a la que se llegue en una prueba de hipótesis tiene probabilidades asociadas con las cuatro diferentes situaciones. Si "no rechazar H_o" es la decisión, es posible que haya ocurrido un error. Además, si "rechazar H_o" es la decisión a la que se llegue, es posible que éste sea un error. Ambos errores tienen probabilidades mayores a cero.

En este capítulo hemos restringido nuestro análisis de inferencias a la media de una población para la cual se conoce la desviación estándar. En los capítulos 9 y 10 estudiaremos inferencias acerca de la media poblacional, y eliminaremos la restricción acerca del valor conocido para desviación estándar. También veremos inferencias acerca de la proporción de parámetros, varianza y desviación estándar.

Vocabulario y conceptos clave

alfa (α) (pp. 403, 420)
beta (β) (p. 420)
coeficiente de confianza (p. 403)
conclusión (pp. 422, 433, 452)
confianza, (p. 404)
criterios de prueba (pp. 431, 448)
decisión correcta tipo A (p. 419)
decisión correcta tipo B (p. 419)
error estándar de media (p. 403)
error máximo de estimación (pp. 403, 410)
error tipo I (p. 419)
error tipo II (p. 419)
estadística de prueba (pp. 422, 431, 449)
estadística insesgada (p. 397)
estadística muestral (p. 397)
estadísticas sesgadas (p. 397)
estimación (p. 397)
estimación de intervalo (p. 399)
estimación puntual (p. 397)
valor p (p. 432)
hipótesis (p. 417)
hipótesis alternativa (pp. 417, 428, 445)
hipótesis nula (pp. 417, 428, 445)
intervalo de confianza (p. 399)
límite inferior de confianza (p. 403)
límite superior de confianza (p. 403)
nivel de confianza (p. 399)
nivel de significación (pp. 422, 431, 449)
parámetro (p. 397)
procedimiento clásico (p. 444)
procedimiento de valor p (p. 427)
prueba de hipótesis (p. 417)
región crítica (p. 450)
región no crítica, (p. 450)
regla de decisión (pp. 422,
suposiciones (pp. 402, 426, 444)
tamaño muestral (p. 410)
valor calculado ($z\bigstar$) (pp. 432, 450)
valor crítico (p. 451)
$z(\alpha)$ (pp. 402, 451)

Objetivos de aprendizaje

✓ Entender la diferencia entre estadística descriptiva y estadísticas inferenciales.	p. 4, Ejer. 1.6, p. 396
✓ Entender que una estadística insesgada tiene una distribución muestral con una media que es igual al parámetro poblacional que se estime.	pp. 397-398
Respecto a intervalos de confianza:	
✓ Entender que un intervalo de confianza es un estimado de intervalo de un parámetro poblacional, con un grado de certidumbre, empleado cuando se desconoce el parámetro poblacional.	p. 399
✓ Entender que una estimación puntual para un parámetro poblacional es el valor de la estadística muestral correspondiente.	p. 397, Ejer. 8.5
✓ Entender que el nivel de confianza es la proporción de los intervalos a largo plazo, que contendrá los parámetros poblacionales verdaderos, con base en muestreo repetido.	EJ. 8.4
✓ Entender y ser capaz de describir los componentes clave para un intervalo de confianza: estimación puntual, nivel de confianza, coeficiente de confianza, máximo error de estimación, límite inferior de confianza, y límite superior de confianza.	p. 403, Ejer. 8.25, 8.27, 8.163
✓ Entender que la suposición para un intervalo de confianza para μ usando una σ conocida es que la distribución muestral de $\bar{x}$ tiene una distribución normal. Con base en esta suposición, se utilizará la distribución z normal estándar.	pp. 402-403

✓ Calcular, describir e interpretar un intervalo de confianza para la media poblacional, μ.	EJ. 8.2, Ejer. 8.33
✓ Calcular tamaños muestrales pedidos para construir intervalos de confianza con niveles de confianza variables y errores aceptables.	pp. 410-412, Ejer. 8.45, 8.49
Respecto a pruebas de hipótesis:	
✓ Entender que una prueba de hipótesis se emplea para tomar una decisión acerca del valor de un parámetro poblacional.	p. 417
✓ Entender y ser capaz de definir hipótesis nula y alternativa.	p. 417
✓ Entender y ser capaz de describir los dos tipos de error en una prueba de hipótesis, tipo I y tipo II. Entender que la probabilidad de estos errores son α y β, respectivamente.	pp. 419-421, Ejer. 8.57
✓ Entender y ser capaz de describir los dos tipos de decisiones correctas en una prueba de hipótesis, tipo A y tipo B.	pp. 419-421, Ejer. 8.57
✓ Comprender y ser capaz de describir la relación que existe entre las cuatro posibles respuestas de una prueba de hipótesis, las dos clases de errores y los dos tipos de decisiones correctas.	pp. 419-421, Ejer. 8.57
✓ Demostrar y entender las tres posibles combinaciones para las hipótesis nula y alternativa.	pp. 428-430, Ejer. 8.83, pp. 446-448, Ejer. 8.125
✓ Entender que la suposición para una prueba de hipótesis para μ usando una σ conocida es que la distribución muestral de $\bar{x}$ tiene una distribución normal. Con base en esta suposición, se utilizará la distribución z normal estándar.	pp. 426, 444
✓ Calcular y entender el valor de la estadística de prueba. Calcular el valor p para la estadística de prueba y/o determinar la región crítica y valor(es) crítico(s).	pp. 432-433, Ejer. 8.102, pp. 450-452, Ejer. 8.134, 8.137
✓ Entender y ser capaz de describir que un valor p y/o una región crítica es respecto a una prueba de hipótesis.	pp. 432-433, 435, 450-452
✓ Determinar y conocer el formato apropiado para expresar una decisión en una prueba de hipótesis.	pp. 422, 433, 451
✓ Entender y ser capaz de expresar la conclusión para una prueba de hipótesis.	pp. 422, 433, 452, Ejer. 8.179, 8.185, 8.186

Ejercicios del capítulo

8.159 Se toma una muestra de 64 mediciones de una población continua, y se encuentra que la media muestral es 32.0. Se sabe que la desviación estándar de la población es 2.4. Una estimación de intervalo ha de hacerse de la media con un nivel de confianza de 90%. Exprese o calcule los siguientes elementos.

a. $\bar{x}$ b. σ c. n d. $1 - \alpha$ e. $z_{(\alpha/2)}$ f. $\sigma_{\bar{x}}$ g. E(máximo error de estimación)

h. Límite superior de confianza

i. Límite inferior de confianza

8.160 Suponga que a un intervalo de confianza se le asigna un nivel de confianza de $1 - \alpha = 95\%$. ¿Cómo se usa el 95% para construir el intervalo de confianza? Si $1 - \alpha$ se cambia a 90%, ¿qué efecto tendría esto en el intervalo de confianza?

8.161 El elemento promedio de voluntario de ambulancia tiene 45 años de edad y tiene 8 años de servicio, de acuerdo con el artículo "Trabajadores de ambulancia no pagados podrían obtener 'pensión'" del *Democrat & Chronicle* (23 de enero, 2005). Las estadísticas citadas se basaron en el escuadrón voluntario de ambulancias de Penfield, de 80 miembros. Si este escuadrón es considerado representativo de todos los escuadrones voluntarios de ambulancias del interior del estado de Nueva York, determine un intervalo de confianza de 95% para la edad media de todos los miembros voluntarios de ambulancia del interior del estado. Suponga que la desviación estándar poblacional es 7.8 años.

8.162 La desviación estándar de una población normalmente distribuida es igual a 10. Se selecciona el tamaño muestral de 25 y se encuentra que su media es 95.

a. Encuentre un intervalo de confianza de 80% para μ.

b. ¿Cuál sería el intervalo de confianza de 80% para una muestra de tamaño 100?

c. ¿Cuál sería el intervalo de confianza de 80% para una muestra de tamaño 25 con una desviación estándar de 5 (en lugar de 10)?

8.163 Los pesos de cajas llenas de cierta clase de cereal están normalmente distribuidos con una desviación estándar de 0.27 onzas. Una muestra de 18 cajas seleccionada al azar produjo un peso medio de 9.87 onzas.

a. Encuentre el intervalo de confianza de 95% para el verdadero peso medio de una caja de este cereal.

b. Encuentre el intervalo de confianza de 99% para el verdadero peso medio de una caja de este cereal.

c. ¿Qué efecto tuvo el aumento del nivel de confianza en el ancho del intervalo de confianza?

8.164 Se piensa que los tiempos de espera (en horas) en un popular restaurante están normalmente distribuidos en forma aproximada, con una varianza de 2.25 durante periodos de gran actividad.

a. Una muestra de 20 clientes reveló un tiempo medio de espera de 1.52 horas. Construya el intervalo de confianza de 95% para la media poblacional.

b. Suponga que la media de 1.52 horas había resultado de una muestra de 32 clientes. Encuentre el intervalo de confianza de 95%.

c. ¿Qué efecto tiene un tamaño muestral más grande en el intervalo de confianza?

8.165 Una muestra aleatoria de las calificaciones de 100 solicitantes para puestos de mecanógrafa en una gran empresa de seguros mostró una calificación media de 72.6. El preparador de la prueba aseguró que los solicitantes calificados deberían promediar 75.0.

a. Determine el intervalo de confianza de 99% para la calificación media de todas las solicitantes en la compañía de seguros. Suponga que la desviación estándar de calificaciones de prueba es 10.5.

b. ¿Puede la compañía de seguros concluir que está obteniendo personal calificado (medido por esta prueba)?

8.166 ¿Está usted preocupado por identificar un robo? Si usted hace negocios casi siempre en línea, puede estar en una posición más segura, como lo publicó el 30 de enero de 2005 el *USA Today* en su artículo "Pocas posibilidades para robo en línea". En una encuesta telefónica hecha por Better Business Bureau y Javelin Strategy & Research, el promedio de pérdidas por robos de identidad en línea fue \$551 contra \$4543 por robo de identidad en papel. Si el promedio en línea estuvo basado en una encuesta de 60 personas seleccionadas al azar de quienes habían experimentado un robo de identidad en línea, encuentre el intervalo de confianza de 95% para la cantidad media de robo de identidad en línea. Suponga que la desviación estándar es \$180.

8.167 El tiempo que tarda un juego de béisbol de ligas mayores es de interés para muchos aficionados. Para estimar el "tiempo medio de juego", se identificó una muestra aleatoria de 48 juegos de la Liga Nacional y se obtuvo el "tiempo de juego" (en minutos) por cada juego. La media muestral resultante fue de 2 horas y 49.1 minutos, y la historia del béisbol indica que la variable del tiempo de juego tiene una desviación es-

tándar de 21 minutos. Construya el intervalo de confianza de 98% para el tiempo medio para todos los juegos de la Liga Nacional.

8.168 Un gran pedido de los corchos No. 9 descritos en el caso práctico 6.15 (p. 330) está a punto de ser embarcado. La inspección final de control de calidad incluye una estimación de la ovalidad media (ovalización; fuera de redondez) de los corchos. El diámetro de cada corcho se mide en varios lugares, y la diferencia entre los diámetros máximo y mínimo es la medida de ovalidad para cada corcho. Después de años de medir corchos, el fabricante está seguro que la ovalidad tiene una distribución agrupada con una desviación estándar de 0.10 mm. Se toma una muestra aleatoria de 36 corchos del lote y se determina la ovalidad para cada uno.

0.32	0.27	0.24	0.31	0.20	0.38	0.32	0.11	0.25
0.22	0.35	0.20	0.28	0.17	0.36	0.28	0.38	0.17
0.34	0.06	0.43	0.13	0.39	0.15	0.18	0.13	0.25
0.20	0.16	0.26	0.47	0.21	0.19	0.34	0.24	0.20

a. La especificación de fuera de redondez es "menor a 1.0 mm". ¿Le parece que este pedido satisface la especificación si se miden corchos individuales? Explique.

b. La hoja de certificación que acompaña al embarque incluye un intervalo de 95% de confianza para la ovalidad media. Construya el intervalo de confianza.

c. Explique lo que el intervalo de confianza hallado en la parte b nos dice acerca de este embarque de corchos.

8.169 Esta salida de computadora presenta una muestra simulada de tamaño 25 generada al azar de una población normal con $\mu = 130$ y $\sigma = 10$. Se utiliza entonces un comando de intervalo de confianza para establecer un intervalo de confianza de 95% para μ.

```
116.187 119.832 121.782 122.320 141.436 129.197 119.172
120.713 135.765 131.153 122.307 126.155 137.545 141.154
123.405 143.331 121.767 109.742 140.524 150.600 121.655
127.992 136.434 139.768 125.594
```

N	MEAN	STDEV	SE MEAN	95.0 PERCENT C.I.
25	129.02	10.18	2.00	(125.10, 132.95)

a. Exprese el intervalo de confianza que resultó.

b. Verifique los valores reportados para el error estándar de la media y los límites del intervalo.

8.170 Use computadora y genere 50 muestras aleatorias, cada una de tamaño $n = 25$, de una distribución de probabilidad normal con $\mu = 130$ y $\sigma = 10$.

a. Calcule el intervalo de confianza de 95% con base en cada media muestral.

b. ¿Qué proporción de estos intervalos de confianza contiene $\mu = 130$?

c. Explique lo que representa la proporción hallada en la parte b.

8.171 Una compañía farmacéutica desea estimar el tiempo medio de respuesta para que un suplemento reduzca la presión sanguínea. ¿Qué tan grande debe ser la muestra que tomen para estimar el tiempo medio de respuesta en no más de 1 semana al 99% de confianza? Suponga $\sigma = 3.7$ semanas.

8.172 Un fabricante de automóviles desea estimar el recorrido medio en millas por galón de gasolina en su nuevo modelo compacto. ¿Cuántas carreras muestrales deben realizarse para asegurar que la estimación es precisa con una variación de no más de 0.3 millas por galón (mpg) al 95% de confianza? (Suponga $\sigma = 1.5$.)

8.173 La gerente de un criadero de peces desea estimar la longitud media de su trucha de 3 años producida en el criadero. Ella desea hacer un intervalo de confianza preciso a una variación de no más de $\frac{1}{3}\mu$ de la variación estándar. ¿Qué tan grande debe ser la muestra que ella necesita tomar?

8.174 Estamos interesados en estimar la vida media de un nuevo producto. ¿Qué tan grande debe ser la muestra que necesitamos tomar para estimar la media a no más de $\frac{1}{10}$ de la desviación estándar con 90% de confianza?

8.175 Suponga que se lleva a cabo una prueba de hipótesis usando el método de valor p y se asigna un nivel de significación de $\alpha = 0.01$.

a. ¿Cómo se usa el 0.01 para completar la prueba de hipótesis?

b. Si α se cambia a 0.05, ¿qué efecto tendría esto en el procedimiento de prueba?

8.176 Suponga que se lleva a cabo una prueba de hipótesis usando el método clásico y se asigna un nivel de significación de $\alpha = 0.01$.

a. ¿Cómo se usa el 0.01 para completar la prueba de hipótesis?

b. Si α se cambia a 0.05, ¿qué efecto tendría esto en el procedimiento de prueba?

8.177 La media esperada de una población continua es 100, y su desviación estándar es 12. Una muestra de 50 mediciones da una media muestral de 96. Usando un nivel de significación de 0.01, ha de hacerse una prueba para decidir entre "la media poblacional es 100" o "la media poblacional es diferente de 100". Exprese o encuentre cada uno de lo siguiente:

a. H_o b. H_a c. α d. μ (con base en H_o)

e. $\bar{x}$ f. σ g. $\sigma_{\bar{x}}$ *h.* z★, puntaje z para $\bar{x}$

i. Valor p j. Decisión

k. Trace la curva normal estándar y localice z★ y el valor p.

8.178 La media esperada de una población continua es 200, y su desviación estándar es 15. Una muestra de 80 mediciones da una media muestral de 205. Usando un nivel de significación de 0.01, ha de hacerse una prueba para decidir entre "la media poblacional es 200" o "la media poblacional es diferente de 200". Exprese y encuentre cada uno de lo siguiente:

a. H_o b. H_a c. α d. $z(\alpha/2)$

e. $\bar{x}$ (con base en H_o) f. $\bar{x}$ g. σ *h.* $\sigma_{\bar{x}}$

i. z★, puntaje z para $\bar{x}$ j. decisión

k. Trace la curva normal estándar y localice $\alpha/2$, $z(\alpha/2)$, la región crítica, y z★.

8.179 Un sistema de aspersión de prados y jardines está diseñado para tener un arranque retardado, es decir, hay un retardo desde el momento en que inicia hasta que empieza a salir el agua. Los tiempos de retardo forman una distribución normal con media de 45 segundos y desviación estándar de 8 segundos. Varios clientes se han quejado que el tiempo de retardo es considerablemente más largo de lo dicho por el fabricante. El ingeniero del sistema ha seleccionado una muestra aleatoria de 15 sistemas instalados y ha obtenido un tiempo de retardo de cada sistema. La media muestral es 50.1 segundos. Usando $\alpha = 0.02$, ¿hay evidencia significativa para demostrar que los clientes podrían tener razón de que el tiempo medio de retardo es más de 45 segundos?

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

8.180 La librería de la universidad informa a futuros estudiantes que el costo promedio de sus libros de texto es \$90 por libro, con una desviación estándar de \$15. Los estudiantes de ciencias de ingeniería piensan que el costo promedio de sus libros es más alto que el promedio para todos los estudiantes. Para probar lo dicho por la librería contra la alternativa de los estudiantes de ingeniería, éstos recolectan una muestra aleatoria de tamaño 45.

a. Si usan $\alpha = 0.05$, ¿cuál es el valor crítico de la estadística de prueba?

b. Los datos muestrales de los estudiantes de ingeniería están resumidos por $n = 45$ y $\Sigma x = 4380.30$. ¿Es esto suficiente evidencia para apoyar su opinión?

8.181 Un proceso de manufactura produce cojinetes de bolas con diámetros que tienen una distribución normal y una desviación estándar de $\sigma = 0.04$ cm. Los cojinetes de bolas que tienen diámetros que son demasiado pequeños o demasiado grandes son indeseables. Para probar la hipótesis nula de que $\mu = 0.50$ cm, se selecciona al azar una muestra de 25 y se encuentra que la media muestral es 0.51.

a. Diseñe hipótesis nula y alternativa tales que el rechazo de la hipótesis nula implicará que los cojinetes de bolas sean indeseables.

b. Usando la regla de decisión establecida en la parte a, ¿cuál es el valor p para los resultados muestrales?

c. Si la regla de decisión de la parte a se usa con $\alpha = 0.02$, ¿cuál es el valor crítico para la estadística de prueba?

8.182 Después de efectuar un gran número de pruebas durante un largo tiempo, un fabricante de cuerdas ha encontrado que sus cuerdas tienen una resistencia media a la ruptura de 300 libras, y una desviación estándar de 24 libras. Suponga que estos valores son μ y σ. Se piensa que mediante el uso de un proceso de

alta velocidad inventado recientemente, la resistencia media a la ruptura se ha reducido.

a. Diseñe hipótesis nula y alternativa tales que el rechazo de la hipótesis nula implicará que la resistencia media a la ruptura se reduzca.

b. Usando la regla de decisión establecida en la parte a, ¿cuál es el valor p asociado con rechazar la hipótesis nula cuando 45 pruebas resultan en una media muestral de 295?

c. Si la regla de decisión de la parte a se usa con $\alpha = 0.01$, ¿cuál es el valor crítico para la estadística de prueba y qué valor de $\bar{x}$ corresponde a aquél si se usa una muestra de tamaño 45?

8.183 Una abeja obrera sale de la colmena regularmente hacia unas flores y otras fuentes de polen y néctar antes de regresar a la colmena para entregar su carga. El proceso se repite varias veces al día para alimentar abejas jóvenes y apoyar la producción de miel y cera de la colmena. La obrera puede llevar un promedio de 0.0113 gramos de polen y néctar por viaje, con una desviación estándar de 0.0063 gramos. Fuzzy Drone está entrando al negocio de la miel y cera con una nueva variedad de abejas italianas que se dice son capaces de llevar cargas más grandes de polen y néctar que la abeja común. Después de instalar tres colmenas, Fuzzy aisló 200 abejas antes y después de su viaje de regreso y cuidadosamente pesó sus cargas. El peso medio muestral del polen y néctar fue de 0.0124 gramos. ¿Pueden las abejas de Fuzzy llevar una mayor carga de polen y néctar que el resto de la población de abejas? Complete la prueba de hipótesis apropiada al nivel de significación de 0.01.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

8.184 El peso promedio para una niña de 10 años de edad en 2002 fue de 88.0 libras, según el Centro Nacional de Estadística de la Salud. Este promedio rebasó en más de 10 libras la cantidad de 1966. Suponga que se toma una muestra aleatoria de 35 niñas de 10 años y que la media muestral es 90.5 libras. Suponiendo que la desviación estándar poblacional sea de 10.7 libras, ¿hay evidencia desde esta muestra de que el peso promedio para niñas de 10 años ha aumentado desde 2002? Complete la prueba de hipótesis apropiada al nivel de significación de 0.05.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

8.185 En un gran supermercado, el tiempo de espera de clientes para salir está normalmente distribuido en forma aproximada con una desviación estándar de 2.5 minutos. Una muestra de tiempos de espera de 24 clientes produjo una media de 10.6 minutos. ¿Es esto evidencia suficiente para rechazar el dicho del supermercado de que el tiempo de salida de sus clientes promedia no más de 9 minutos? Complete esta prueba de hipótesis usando el nivel de significación de 0.02.

a. Resuelva usando el método de valor p.

b. Resuelva usando el método clásico.

8.186 En una empresa muy grande, las mecanógrafas fueron muestreadas para ver si sus salarios diferían entre departamentos respecto a trabajadoras de categoría similares. En una muestra de 50 de empleados de contabilidad de la compañía, el salario anual promedio fue \$16 010. La oficina de personal de la compañía insiste en que el salario promedio pagado a todas sus mecanógrafas es \$15 650 y que la desviación estándar es \$1800. Al nivel de significación de 0.05, ¿podemos concluir que los empleados de contabilidad reciben, en promedio, un salario diferente del de las mecanógrafas?

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

8.187 Jack Williams es vicepresidente de marketing para una de las compañías más grandes de gas natural de la nación. Durante los últimos cuatro años, él ha observado dos factores principales que erosionan las utilidades y ventas de la compañía. Primero, el precio promedio de petróleo crudo ha estado prácticamente sin cambio, y muchos de sus clientes industriales están quemando aceite pesado en lugar de gas natural para operar sus hornos, son considerar las mayores emisiones de contaminantes a la atmósfera. En segundo término, clientes residenciales y comerciales están todavía buscando técnicas de conservación de energía (por ejemplo instalar aislamiento extra, instalar termostatos accionados por reloj, y sellar grietas alrededor de puertas y ventanas para eliminar infiltración de aire frío). En otros años, los clientes residenciales compraban un promedio de 129.2 millones de pies

cúbicos (mpc) de gas natural de la compañía de Jack (σ = 18 mpc), con base en registros internos de facturación de la compañía, pero los ambientalistas dijeron que la conservación está reduciendo el consumo de combustible hasta en un 3% por año. Jack lo ha comisionado a usted para dirigir un control por sondeo para ver si algún cambio en el uso anual ha ocurrido, antes de su siguiente reunión con los oficiales de la corporación. Una muestra de 300 clientes seleccionados al azar de los registros de facturación revela un promedio de 127.1 mpc durante los últimos 12 meses. ¿Hay un descenso significativo en el consumo?

a. Complete la prueba de hipótesis apropiada al nivel de significación de 0.01, usando el método del valor *p* para que pueda asesorar debidamente a Jack antes de su reunión.

b. Como usted es asistente de Jack, ¿por qué es mejor que usted use el método del valor *p*?

8.188 Con un tiempo promedio de viajar en auto de alrededor de 24.3 minutos, a nivel nacional, los norteamericanos pasan ahora más de 100 horas al año desde o hacia su trabajo, según un estudio de la comunidad norteamericana de la oficina del censo de Estados Unidos. Sí, esto es más del promedio de 2 semanas de vacaciones (80 horas) que toman numerosos trabajadores durante un año.

Fuente: http://usgovinfo.about.com/od/censusandstatistics/a/commutetimes.htm

Se interrogó a una muestra aleatoria de 150 trabajadores de una gran industria cercana acerca de su tiempo de viaje al trabajo. Si se sabe que la desviación estándar es 10.7 minutos, ¿la media muestral resultante de 21.7 minutos es significativamente menor que el promedio nacional? Use $\alpha = 0.01$.

a. Resuelva usando el método del valor *p*.

b. Resuelva usando el método clásico.

8.189 Esta salida de computadora presenta una muestra simulada de tamaño 28, generada al azar de una población normal con $\mu = 18$ y $\sigma = 4$. Luego se usaron comandos de computadora para completar una prueba de hipótesis para $\mu = 18$ contra una alternativa de dos colas.

a. Exprese la hipótesis alternativa, la decisión, y la conclusión que resultó.

b. Verifique los valores informados del error estándar de media, $z\bigstar$, y el valor *p*.

```
18.7734 21.4352 15.5438 20.2764 23.2434 15.7222 13.9368
14.4112 15.7403 19.0970 19.0032 20.0688 12.2466 10.4158
 8.9755 18.0094 20.0112 23.2721 16.6458 24.6146 17.8078
16.5922 16.1385 12.3115 12.5674 18.9141 22.9315 13.3658
```

```
TEST OF MU = 18.000 VS MU not = 18.000
THE ASSUMED STANDARD DEVIATION = 4.00
N      MEAN     STDEV    SE MEAN     Z        P VALUE
28    17.217    4.053     0.756     -1.04      0.30
```

8.190. Un fabricante de neumáticos para automóvil piensa que ha perfeccionado un nuevo compuesto de caucho que tiene cualidades superiores que resisten el desgaste. Produjo una corrida de prueba de llantas hechas con este nuevo compuesto e hizo que se probaran en camino. Los valores de datos registrados fueron la cantidad de desgaste de la superficie de rodadura por 10 000 millas. En el pasado, la cantidad media de desgaste de la superficie de rodadura por 10 000 millas, para llantas de esta calidad, había sido de 0.0625 pulgadas.

La hipótesis nula a ser probada aquí es "La cantidad media de desgaste en las llantas hecha con el nuevo compuesto es la misma cantidad media de desgaste con el compuesto anterior, 0.0625 pulgadas por 10 000 millas", H_o: $\mu = 0.0625$. Podrían usarse tres posibles hipótesis alternativas: (1) H_a: $\mu < 0.0625$, (2) H_a: $\mu \neq 0.0625$, (3) H_a: $\mu > 0.0625$.

a. Explique el significado de cada una de estas tres alternativas.

b. ¿Cuál de las posibles hipótesis alternativas debe usar el fabricante, si espera concluir que "el uso del nuevo compuesto produce menos desgaste"?

8.191. De una población de media desconocida μ y una desviación estándar $\sigma = 5.0$, se seleccionó una muestra de $n = 100$ y se encontró la media muestral de 40.6. Compare los conceptos de estimación y prueba de hipótesis al completar lo siguiente:

a. Determine el intervalo de confianza de 95% para μ.

b. Complete la prueba de hipótesis que involucre H_a: $\mu \neq 40$ usando el método del valor *p* y $\alpha = 0.05$.

c. Complete la prueba de hipótesis que involucre H_a: $\mu \neq 40$ usando el método clásico y $\alpha = 0.05$.

d. En el dibujo de la curva normal estándar, localice el intervalo que represente el intervalo de confianza de la parte a; la $z\bigstar$, el valor *p*, y α de la parte b; y la $z\bigstar$ y regiones críticas de la parte c. Describa la relación entre estos tres procedimientos separados.

8.192. De una población de media desconocida μ y una desviación estándar $\sigma = 5.0$, se selecciona una muestra de $n = 100$ y se encuentra la media muestral de 41.5. Compare los conceptos de estimación y prueba de hipótesis al completar lo siguiente:

a. Determine el intervalo de confianza de 95% para μ.

b. Complete la prueba de hipótesis que involucre H_a: $\mu \neq 40$ usando el método del valor p y $\alpha = 0.05$.

c. Complete la prueba de hipótesis que involucre H_a: $\mu \neq 40$ usando el método clásico y $\alpha = 0.05$.

d. En el dibujo de la curva normal estándar, localice el intervalo que represente el intervalo de confianza de la parte a; la $z\bigstar$, el valor p, y α de la parte b; y la $z\bigstar$ y regiones críticas de la parte c. Describa la relación entre estos tres procedimientos separados.

8.193. De una población de media desconocida μ y una desviación estándar $\sigma = 5.0$, se selecciona una muestra de $n = 100$ y se encuentra la media muestral de 40.9. Compare los conceptos de estimación y prueba de hipótesis al completar lo siguiente:

a. Determine el intervalo de confianza de 95% para μ.

b. Complete la prueba de hipótesis que involucre H_a: $\mu > 40$ usando el método del valor p y $\alpha = 0.05$.

c. Complete la prueba de hipótesis que involucre H_a: $\mu > 40$ usando el método clásico y $\alpha = 0.05$.

d. En el dibujo de la curva normal estándar, localice el intervalo que represente el intervalo de confianza de la parte a; la $z\bigstar$, el valor p, y α de la parte b; y la $z\bigstar$ y regiones críticas de la parte c. Describa la relación entre estos tres procedimientos separados.

8.194. Un fabricante de mostaza, molida en piedra y estilo delicioso, utiliza una máquina de alta velocidad para llenar frascos. La cantidad de mostaza dosificada en los frascos forma una distribución normal con una media de 290 gramos y una desviación estándar de 4 gramos. Cada hora, se toma una muestra aleatoria de 12 frascos de la producción de esa hora. Si la media muestral está entre 287.74 y 292.26, se acepta la producción de esa hora; de no ser así, se rechaza y la máquina se vuelve a calibrar antes de continuar.

a. ¿Cuál es la probabilidad del error tipo I al rechazar la producción de la hora previa, cuando el peso medio de un frasco es de 290 gramos?

b. ¿Cuál es la probabilidad del error tipo II al rechazar la producción de la hora previa, cuando el peso medio de un frasco es de 288 gramos?

8.195. Todos los medicamentos deben ser aprobados por la Food and Drug Administration (FDA) antes que puedan ser vendidos por una compañía farmacéutica. La FDA debe ponderar el error de vender un medicamento que no sea eficaz, con los riesgos usuales de efectos colaterales, contra las consecuencias de no permitir la venta de un medicamento efectivo. Suponga, usando tratamiento médico estándar, que se conoce que la tasa de mortalidad (r) de cierta enfermedad es A. Un fabricante envía para aprobación un medicamento que se supone es tratamiento para esta enfermedad. La FDA establece la hipótesis para probar la tasa de mortalidad para el medicamento como (1) H_o: $r = A$, H_a: $r < A$, $\alpha = 0.005$ o (2) H_o: $r = A$, H_a: $r > A$, $\alpha = 0.005$.

a. Si $A = 0.95$, ¿cuál prueba piensa usted que debe usar la FDA? Explique.

b. Si $A = 0.05$, ¿cuál prueba piensa usted que debe usar la FDA? Explique.

8.196. El fabricante de medicamentos del ejercicio 8.195 tiene un punto de vista diferente sobre esto. Desea vender el nuevo medicamento empezando tan pronto como sea posible, de modo que pueda vencer a sus competidores del mercado y obtener buenas utilidades. Su posición es, "Vender el medicamento a menos que sea totalmente ineficaz".

a. ¿Cómo es que la compañía farmacéutica establecería la hipótesis alternativa si estuviera haciendo la prueba? H_a: $r < A$, H_a: $r \neq A$, o H_a: $r > A$. Explique.

b. ¿La tasa de mortalidad ($A = 0.95$ o $A = 0.05$) del tratamiento existente, afecta la alternativa? Explique.

8.197. Use computadora para generar 50 muestras aleatorias, cada una de tamaño $n = 28$, de una distribución de probabilidad normal con $\mu = 18$ y $\sigma = 4$.

a. Calcule la $z\bigstar$ correspondiente a cada media muestral.

b. Respecto al método del valor p, encuentre la proporción de 50 valores de $z\bigstar$ que sean "más extremos" que la $z = -1.04$ que ocurrieron en el ejercicio 8.189 (H_a: $\mu \neq 18$). Explique lo que representa esta proporción.

c. Respecto al método clásico, encuentre los valor críticos para una prueba de dos colas usando $\alpha = 0.01$; encuentre la proporción de 50 valores $z\bigstar$ que caigan en la región crítica. Explique lo que representa esta proporción.

8.198 Use computadora para generar 50 muestras aleatorias, cada una de tamaño $n = 28$, de una distribución de probabilidad normal con $\mu = 19$ y $\sigma = 4$.

a. Calcule la $z\bigstar$ correspondiente para cada media muestral que resultaría al probar la hipótesis nula $\mu = 18$.

b. Respecto al método del valor p, encuentre la proporción de 50 valores de $z\bigstar$ que sean "más extremos" que la $z = -1.04$ que ocurrieron en el ejercicio 8.189 (H_a: $\mu \neq 18$). Explique lo que representa esta proporción.

c. Respecto al método clásico, encuentre los valores críticos para una prueba de dos colas usando $\alpha = 0.01$; encuentre la proporción de 50 valores $z\bigstar$ que caigan en la región crítica. Explique lo que representa esta proporción.

Proyecto del capítulo

¿La gente era menos alta en otros tiempos?

Los datos del Centro Nacional para Estadísticas de la Salud indican que la estatura promedio de una mujer en Estados Unidos es 63.7 pulgadas, con una desviación estándar de 2.75 pulgadas. Use los datos de estaturas de mujeres de la profesión de la salud de la sección 8.1, "la gente era menos alta en otros tiempos?" (p. 395), para contestar las siguientes preguntas.

65.0	66.0	64.0	67.0	59.0	69.0	66.0	69.0	64.0	61.5
63.0	62.0	63.0	64.0	72.0	66.0	65.0	64.0	67.0	68.0
70.0	63.0	63.0	68.0	58.0	60.0	63.5	66.0	64.0	62.0
64.5	69.0	63.5	69.0	62.0	58.0	66.0	68.0	59.0	56.0
64.0	66.0	65.0	69.0	67.0	66.5	67.5	62.0	70.0	62.0

Trabajando en el contenido del capítulo 8

8.199 a. ¿Se satisfacen las suposiciones del intervalo de confianza y métodos de prueba de hipótesis de este capítulo? Explique.

b. Usando los datos muestrales y un nivel de confianza de 95%, estime la estatura media de mujeres de la profesión de la salud. Use la desviación estándar poblacional dada de 2.75 pulgadas.

c. Pruebe lo dicho de que la estatura media de mujeres de la profesión de la salud es diferente de 63.7 pulgadas, que es la estatura media para todas las mujeres de Estados Unidos. Use un nivel de significación de 0.05.

d. En el mismo histograma empleado en la parte b del ejercicio 8.1 de la página 396:

(i) Trace una recta vertical en el valor medio poblacional hipotético, 63.7.

(ii) Trace un segmento de recta horizontal que muestre el intervalo de confianza de 95% de la parte b.

e. ¿La media $\mu = 63.7$ cae en el intervalo? Explique lo que esto significa.

f. Describa la relación entre las dos rectas trazadas en su gráfica para la parte c, del ejercicio 8.2 de la página 396, y las dos rectas trazadas para la parte d de este ejercicio.

g. Con base en los resultados obtenidos antes, ¿parece que las mujeres de este estudio, en promedio, tienen la misma estatura que todas las mujeres de Estados Unidos como lo reporta el NCHS? Explique.

Para su estudio

8.200 Diseñe su propio estudio de estaturas de mujeres. Defina una población específica que usted muestreará, describa su plan de muestreo, recolecte sus datos y conteste la parte b del ejercicio 8.1 (p. 396) y las partes a, b, c y g del ejercicio 8.199, sustituyendo la profesión de saludo con su población particular. Discuta las diferencias y similitudes entre su muestra y la población y entre su muestra y la muestra de 50 mujeres profesionales de la salud.

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

8.1 **Beta** es la probabilidad de un error tipo I.

8.2 $\mathbf{1 - \alpha}$ se conoce como nivel de significación de una prueba de hipótesis.

8.3 El error estándar de la media es la desviación estándar de la **muestra seleccionada.**

8.4 El error máximo de estimación está controlado por tres factores: **nivel de confianza, tamaño muestral** y **desviación estándar.**

8.5 Alfa es la medida del área bajo la curva del puntaje estándar que se encuentra en la **región de rechazo** para H_o.

8.6 El riesgo de cometer un **error tipo I** está directamente controlado en una prueba de hipótesis al establecer un nivel para α.

8.7 No rechazar la hipótesis nula cuando es falsa es una **decisión correcta.**

8.8 Si la región no crítica en una prueba de hipótesis se hace más ancha (suponiendo que σ y n permanecen fijas), α se hace más grande.

8.9 Rechazar una hipótesis nula que es falsa es un **error tipo II.**

8.10 Para concluir que la media es mayor (o menor) que un valor afirmado, el valor de la estadística de prueba debe caer en la **región de aceptación.**

SEGUNDA PARTE: Aplicación de conceptos

Conteste todas las preguntas, mostrando todas las fórmulas, sustituciones y trabajo.

8.11 Un desafortunado cliente de una oficina de correos está frustrado con el tiempo de espera para comprar estampillas. Al registrar su queja, le dijeron, "El promedio de tiempo de espera en el pasado ha sido de alrededor de 4 minutos, con una desviación estándar de 2 minutos". El cliente recolectó una muestra de $n = 45$ clientes y encontró que la espera media era de 5.3 minutos. Encuentre el intervalo de confianza de 95% para el tiempo medio de espera.

8.12 Exprese las hipótesis nula (H_o) y alternativa (H_a) que se usarían para probar cada una de estas frases:

a. El peso medio de jugadores profesionales de futbol es más de 245 libras.

b. La cantidad media mensual de lluvia en el condado de Monroe es menor a 4.5 pulgadas.

c. El peso medio de los bates de béisbol que usan jugadores de ligas mayores no es igual a 35 onzas.

8.13 Determine el nivel de significación, estadística de prueba, región crítica y valor(es) crítico(s) que se usarían para completar cada prueba de hipótesis usando α = 0.05:

a. H_o: $\mu =$. H_o: $\mu =$ . . H_o: $\mu =$
$H : \mu <$ $H : \mu >$. $H : \mu \neq$
a $\sigma =$ a $\sigma =$. a $\sigma =$)

8.14 Encuentre cada valor:

a. $z(0.05)$ b. $z(0.01)$ c. $z(0.12)$

8.15 En el pasado, las toronjas producidas en una huerta particular habían tenido un diámetro medio de 5.50 pulgadas, y una desviación estándar de 0.6 pulgadas. El propietario piensa que la cosecha de este año es mayor que las del pasado. Recolectó una muestra aleatoria de 100 toronjas y encontró un diámetro medio muestral de 5.65 pulgadas.

a. Encuentre el valor de la estadística de prueba, z★, que corresponde a $\bar{x} = 5.65$.

b. Calcule el valor p para la hipótesis del propietario.

8.16 Un fabricante dice que sus bombillas eléctricas tienen una vida útil media de 1520 horas con una desviación estándar de 85 horas. Se selecciona una muestra aleatoria de 40 de estas bombillas

para probarlas. Si la muestra produce un valor medio de 1498.3 horas, ¿hay suficiente evidencia para decir que la vida útil media es menor que lo dicho por el fabricante? Use $\alpha = 0.01$.

TERCERA PARTE: Comprensión de los conceptos

8.17 Las tiendas de conveniencia de Sugar Creek han pedido a una firma de estadísticas hacer una encuesta a sus clientes para estimar la cantidad media gastada por cliente. De registros anteriores, se piensa que la desviación estándar es $\sigma =$ \$5. En su propuesta a Sugar Creek, la firma de expertos en estadística dice que planea basar la estimación, para la cantidad media gastada, en una muestra de tamaño 100 y usar el nivel de confianza de 95%. El director de Sugar Creek ha sugerido que el tamaño muestral se aumente a 400. Si nada de lo demás cambia, ¿qué efecto tendrá este aumento en tamaño muestral sobre lo siguiente?

a. La estimación puntual para la media

b. El error máximo de estimación

c. El intervalo de confianza

El director general desea que el nivel de confianza se aumente a 99%. Si nada de lo demás cambia, ¿qué efecto tendrá este aumento en nivel de confianza sobre lo siguiente?

d. La estimación puntual para la media

e. El error máximo de estimación

f. El intervalo de confianza

8.18 El nivel de ruido en un hospital puede ser un factor crítico que influye en la rapidez de recuperación de un paciente. Suponga que, por someterlo a discusión, una comisión de investigación ha recomendado un nivel medio máximo de ruido de 30 decibels (dB), con una desviación estándar de 10 dB. El personal de un hospital trata de muestrear una de sus salas para determinar si el nivel de ruido es significativamente más alto que el recomendado. La siguiente prueba de hipótesis se completará:

$$H_o: \mu = 30\ (\leq) \quad \text{vs.} \quad H_a: \mu > 30, \quad \alpha = 0.05$$

a. Identifique la interpretación correcta para cada hipótesis respecto a la recomendación y justifique su selección.

H_o: (1) El nivel de ruido no es significativamente más alto que el nivel recomendado, o (2) el nivel de ruido es significativamente más alto que el nivel recomendado.

H_a: (1) El nivel de ruido no es significativamente más alto que el nivel recomendado, o (2) el nivel de ruido es significativamente más alto que el nivel recomendado.

b. ¿Cuál enunciado describe mejor el error tipo I?

(i) La decisión a la que se llegó fue que el nivel de ruido está dentro del nivel recomendado, cuando en realidad estaba dentro.

(ii) La decisión a la que se llegó fue que el nivel de ruido está dentro del nivel recomendado, cuando en realidad lo rebasaba.

(iii) La decisión a la que se llegó fue que el nivel de ruido rebasa el nivel recomendado, cuando en realidad estaba dentro.

(iv) La decisión a la que se llegó fue que el nivel de ruido está dentro del nivel recomendado, cuando en realidad lo rebasaba.

c. ¿Cuál enunciado de la parte b describe mejor el error tipo II?

d. Si α se cambiara de 0.05 a 0.01, identifique y justifique el efecto (aumenta, disminuye, o sigue igual) sobre P(error tipo I) y sobre P(error tipo II).

8.19 La hipótesis alternativa recibe a veces el nombre de *hipótesis de investigación.* La conclusión es un enunciado escrito acerca de la hipótesis alternativa. Explique por qué estos dos enunciados son compatibles.

CAPÍTULO

9 Inferencias que involucran a una población

9.1 ¿Se ejercita lo suficiente todos los días?

9.2 Inferencias relacionadas con la media μ (σ desconocida)

9.3 Inferencias acerca de los éxitos de la distribución binomial

9.4 Inferencias relacionadas con la varianza y la desviación estándar

© Rachel Epstein/The Image Works

9.1 ¿Se ejercita lo suficiente todos los días?

BUENAS NOTICIAS PARA MUJERES

El National Women's Health Information Center informa, "Nunca es demasiado tarde para empezar un estilo de vida activo". No importa la edad que tenga usted, qué incapaz se sienta ni cuánto tiempo ha estado inactivo, las investigaciones demuestran que empezar un estilo de vida más activo ahora por medio de una actividad regular y de intensidad moderada puede hacer que se sienta más sano y mejore su calidad de vida. Vea a continuación lo que tiene que hacer:

Si:	Entonces:
No practica hoy día una actividad física regular,	debe empezar por incorporar unos minutos diarios de actividad física, aumentando gradualmente a 30 minutos o más de actividades de intensidad moderada.
Está ahora activo, pero a menos de los niveles recomendados,	debe esforzarse por adoptar actividades consistentes: actividad física de intensidad moderada durante 30 minutos o más por cinco días o más de la semana o bien actividad física de intensidad vigorosa durante 20 minutos o más en tres días o más de la semana.
Practica actualmente actividades de intensidad moderada al menos 30 minutos por cinco días o más a la semana,	usted puede alcanzar beneficios aún mayores en su salud si en esas actividades aumenta el tiempo o la intensidad
Practica actualmente y con regularidad actividades de intensidad vigorosa, 20 minutos o más, por tres días o más a la semana,	usted debe continuar haciendo esas actividades.

La evidencia científica hasta hoy apoya lo indicado en esta tabla. *Fuente*: Centers for Disease Control and Prevention, http://www.4woman.gov/pub/steps/Physical%20Activity.htm

El artículo recomienda diferentes cantidades y diferentes niveles de actividad, dependiendo del nivel de actividad actual de una mujer. Algunas de las recomendaciones requieren sólo 60 minutos de ejercicio a la semana. Los valores de los datos que siguen son de un estudio que examina a técnicas cardiovasculares (personas que realizan diversos procedimientos de diagnóstico cardiovascular) en cuanto a su propio ejercicio físico por semana, medido en minutos.

60	40	50	30	60	50	90	30	60	60
60	80	90	90	60	30	20	120	60	50
20	60	30	120	50	30	90	20	30	40
50	40	30	40	20	30	60	50	60	80

¿Le parece que los técnicos, en promedio, se ejercitan al menos 60 minutos por semana?

Después de completar el capítulo 9, investigue con más detalle los valores precedentes de datos y "¿Se ejercita lo suficiente todos los días?" en el Proyecto del capítulo, en la página 540.

SECCIÓN 9.1 EJERCICIOS

9.1 Considere los datos muestrales precedentes.

a. ¿Cuál es el parámetro poblacional de interés?

b. Construya un histograma de los datos.

c. ¿Qué nombre le daría usted a la forma del histograma de la parte b?

d. ¿Diría usted que el histograma de la parte b sugiere que la variable, cantidad de tiempo, no tiene una distribución normal?

9.2 Con base en los datos presentados en la sección 9.1 para las técnicas cardiovasculares:

a. Encuentre la media y desviación estándar para la cantidad de tiempo que las técnicas cardiovasculares se ejercitaron por semana.

b. ¿Cómo estimaría usted la cantidad media de tiempo que se ejercitaron por semana todas las técnicas cardiovasculares?

c. Le parece a usted que las técnicas cardiovasculares se ejercitan ¿"al menos 60 minutos por semana"? Justifique su respuesta.

9.2 Inferencias relacionadas con la media μ (σ desconocida)

Las inferencias acerca de la media poblacional μ están basadas en la media muestral $\bar{x}$ e información obtenida de la distribución muestral de medias muestrales. Recuerde que la distribución muestral de medias muestrales tiene una media μ y un **error estándar** de $\sigma/\sqrt{n}$ para todas las muestras de tamaño n, y está normalmente distribuida cuando la población muestreada tiene una distribución normal, o está normalmente distribuida en forma aproximada cuando el **tamaño muestral** es lo suficiente grande. Esto significa que la estadística de prueba $z\bigstar = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ tiene una distribución normal estándar. No obstante, cuando σ es **desconocida,** el error estándar $\sigma/\sqrt{n}$ también es desconocido. Por tanto, la desviación muestral estándar s

se usará como estimación puntual para σ. En consecuencia, un error estándar estimado de la media, $s/\sqrt{n}$, se usará y nuestra estadística de prueba se convertirá en $\dfrac{\bar{x} - \mu}{s/\sqrt{n}}$.

Cuando se usa una **σ conocida** para hacer una inferencia acerca de la media μ, una muestra da un valor para usar en las fórmulas; ese valor es $\bar{x}$. Cuando también se usa la desviación estándar muestral, la muestra da dos valores: la media muestral $\bar{x}$ y el error estándar estimado $s/\sqrt{n}$. Por tanto, la estadística z será sustituida con una estadística que toma en cuenta el uso de un error estándar estimado. Esta nueva estadística se conoce como **estadística *t* de Student.**

En 1908, W. S. Gosset, empleado de una cervecería en Irlanda, publicó un artículo acerca de esta distribución t bajo el seudónimo de "Student". Al derivar la distribución t, Gosset supuso que las muestras fueron tomadas de poblaciones normales. Aun cuando esto podría parecer restrictivo, se obtienen resultados satisfactorios cuando se seleccionan muestras grandes de muchas poblaciones no normales.

FIGURA 9.1 Uso de la estadística *z* o la estadística *t*

1. ¿Es n grande? Muestras tan pequeñas como $n = 15$ o 20 pueden ser consideradas suficientemente grandes para que se cumpla el teorema de límite central si los datos muestrales son unimodales, casi simétricos, de cola corta y sin resultados aislados. Las muestras que no son simétricas requieren tamaños muestrales más grandes, 50 siendo suficiente excepto para muestras extremadamente sesgadas. Vea el análisis en la página 403.
2. Requiere el uso de una técnica no paramétrica; vea el capítulo 14.

La figura 9.1 presenta una organización diagramático para las inferencias alrededor de la media poblacional, como se expuso en el capítulo 8 y en esta segunda sección del capítulo 9. Existen dos situaciones: σ es conocida, o σ es desconocida.

¿SABÍA USTED...?

William Gosset ("Student")

William Gosset estudió matemáticas y química en la Universidad de Oxford y al graduarse aceptó un puesto en la Cervecería Guinness de Dublín, donde encontró gran cantidad de datos recolectados respecto al proceso de elaboración de cerveza. En 1905, se encontró con Karl Pearson para discutir sus problemas estadísticos y, un año después, con la aprobación de la compañía Guinness, fue a trabajar al laboratorio biométrico de Pearson. Cuando regresó a Guinness, pasó a hacerse cargo de la Elaboración Experimental. Durante estos años escribió varios artículos científicos, que Guinness acordó permitirle publicar, siempre que usara un seudónimo y no incluyera datos de la compañía; usó el seudónimo de "Un estudiante".

Como ya dijimos antes, σ es una cantidad que casi nunca se conoce en problemas reales; por lo tanto, el error estándar casi siempre se estima por $s/\sqrt{n}$. El uso de un error estándar estimado de la media requiere el uso de la distribución t. Casi todas las inferencias prácticas acerca de la media poblacional se harán con la estadística t de Student.

La distribución t tiene las siguientes propiedades (ver también la figura 9.2):

Propiedades de la distribución t (df $>$ 2)*

1. t está distribuida con una media de cero.
2. t está distribuida simétricamente alrededor de su media.
3. t está distribuida para formar una familia de distribuciones, una distribución separada para cada número diferente de grados de libertad (df $\geq$ 1).
4. La distribución t se aproxima a la **distribución normal estándar** cuando aumenta el número de grados de libertad.
5. t está distribuida con una varianza mayor a 1, pero cuando aumentan los grados de libertad, la varianza se aproxima a 1.
6. t está distribuida para ser menos puntiaguda en la media y más gruesa en las colas que su distribución normal.

Grados de libertad, df:
Un **parámetro** que identifica cada distribución diferente de la distribución t de Student. Para los métodos presentados en este capítulo, el valor de df será el tamaño muestral menos 1: df $= n - 1$.

FIGURA 9.2
Distribuciones t de Student

El número de grados de libertad asociados con s^2 es el divisor $(n - 1)$ empleado para calcular la varianza muestral s^2 [fórmula (2.6), p. 94]; esto es, df $= n - 1$. La varianza muestral es la media de las desviaciones cuadradas. El número de grados de libertad es el "número de desviaciones no relacionadas" disponible para uso para estimar σ^2. Recuerde que la suma de las desviaciones, $\sum(x - \bar{x})$, debe ser cero. De una muestra de tamaño n, sólo la primera $n - 1$ de estas desviaciones tiene libertad de valor. Esto es, el último, o n-ésimo, valor de $(x - \bar{x})$ debe hacer exactamente cero la suma del total de n desviaciones. Como resultado de esto, se dice que la varianza promedia $n - 1$ valores de desviación cuadrada no relacionados, y este número, $n - 1$, se denominó "grados de libertad".

Aun cuando hay una distribución t separada por cada uno de los grados de libertad, df = 1, df = 2, . . . , df = 20, . . . , df = 40, etcétera, sólo ciertos **valores críticos**

*No todas las propiedades se cumplen para df = 1 y df = 2. Como no encontraremos situaciones donde df =1 o 2, estos casos especiales no se estudiarán más.

de *t* clave serán necesarios para nuestro trabajo. En consecuencia, la tabla de distribución t de Student (tabla 6 del apéndice B) es una tabla de valores críticos más que una tabla completa, como la tabla 3 es para la distribución normal estándar para z. Cuando vea la tabla 6, observará que el lado izquierdo de la tabla está identificado por "df", grados de libertad. Esta columna de la izquierda se inicia en 3 en la parte superior y es una lista consecutiva de valores df hasta 30, luego salta a 35, . . . , a "df $>$ 100" en la parte inferior. Como ya dijimos, cuando aumentan los grados de liberta, la distribución t se aproxima a las características de la distribución z normal estándar. Una vez que df sea "mayor a 100", los valores críticos de la distribución t son iguales que los valores críticos correspondientes de la distribución normal estándar como se da en la tabla 4A del apéndice B.

Uso de la tabla de distribución *t* (tabla 6, apéndice B)

Los valores críticos de la distribución t de Student, que han de usarse para construir un intervalo de confianza y para probar hipótesis, se obtendrán de la tabla 6 del apéndice B. Para hallar el valor de t, será necesario conocer dos valores de identificación: (1) df, el número de grados de libertad (que identifica la distribución de interés) y (2) α, el área bajo la curva a la derecha del valor crítico del lado derecho. Se empleará una notación muy semejante a la usada con z para identificar un valor crítico. $t(\text{df}, \alpha)$, léase "t de df, α",) es el símbolo para el valor de t con df grados de libertad y un área de α en la cola del lado derecho, como se ve en la figura 9.3.

FIGURA 9.3
Distribución *t*
Que muestra $t(\text{df}, \alpha)$

EJEMPLO 9.1 ***t* en el lado derecho de la media**

Encuentre el valor de $t(10, 0.05)$ (vea el diagrama)

SOLUCIÓN Hay 10 grados de libertad, y 0.05 es el área a la derecha del valor crítico. En la tabla 6 del apéndice B, buscamos la fila df = 10 y la columna marcada "Cantidad de α en una cola", $\alpha = 0.05$. En su intersección, vemos que $t(10, 0.05) =$ **1.81.**

Porción de tabla 6

	Cantidad de α en una cola		
df	. . .	0.05	. . .
:			
10		1.81	

$\longrightarrow$ $t(10, 0.05) = 1.81$

Para los valores de t del lado izquierdo de la media, podemos usar una de dos notaciones. El valor t que se ilustra en la figura 9.4 podría denominarse $t(\text{df}, 0.95)$, porque el área a la derecha de él es 0.95, o podría identificarse como $-t(\text{df}, 0.05)$, porque la distribución t es simétrica alrededor de su media, cero.

FIGURA 9.4
Valor t en el lado izquierdo

EJEMPLO 9.2 *t* en el lado izquierdo de la media

Encuentre el valor de $t(15, 0.95)$.

SOLUCIÓN Hay 15 grados de libertad. En la tabla 6 buscamos la columna marcada $\alpha = 0.05$ (una cola) y su intersección con la fila df = 15. La tabla nos da $t(15, 0.05) = 1.75$; por lo tanto, $t(15, 0.95) = -t(15, 0.95) = -1.75$. El valor es negativo porque está a la izquierda de la media; vea la figura.

Veamos otro ejemplo que enlaza la distribución t a percentiles.

EJEMPLO 9.3 Valores *t* que limitan un porcentaje central

Encuentre los valores de la distribución t que limitan el 0.90 central del área bajo la curva de la distribución con df = 17.

SOLUCIÓN El 0.90 central deja 0.05 para el área de cada cola. El valor de t que limita la cola derecha es $t(17, 0.05) = \mathbf{1.74}$, como se ve en la tabla 6. El valor que limita la cola del lado izquierdo es **−1.74** porque la distribución t es simétrica alrededor de su media, cero.

Si el df necesario no aparece en la lista de la columna izquierda de la tabla 6, entonces usamos el siguiente valor más pequeño de df que aparezca en la lista. Por ejemplo, $t(72, 0.05)$ se estima usando $t(70, 0.05) = 1.67$.

Casi todos los paquetes de software o calculadoras de estadística calcularán el área relacionada con el valor t específico. La siguiente figura muestra la relación entre la probabilidad acumulativa y un valor t específico para una distribución t con df grados de libertad.

INSTRUCCIONES DE TECNOLOGÍA: PROBABILIDAD ASOCIADA CON UN VALOR DE *t* ESPECIFICADO

MINITAB (Versión 14)

Probabilidad acumulativa para un valor de *t* especificado:

```
Choose:     Calc > Probability Distribution > t
Select:     Cumulative Probability
            Noncentrality parameter: 0.0
Enter:      Degrees of freedom: df
Select:     Input constant*
Enter:      t-value (ex. 1.74) > OK
```

*Seleccione la columna Input si en C1 se guardan valores *t*. Use C2 para guardado opcional. Si se necesita el área de la cola derecha, de 1 reste la probabilidad calculada.

Excel

Probabilidad en una o dos colas para un valor *t* dado:

Si han de usarse varios valores *t* (no negativos), introduzca los valores en la columna A y active B1; luego continúe con:

```
Choose:     Insert function fx > Statistical > TDIST > OK
Enter:      X: individual t-value or (A1:A5 or select 't-value' cells)*
            Deg_freedom: df
            Tails: 1 or 2 (one or two-tailed distributions) > OK
Drag*:      Bottom right corner of the B1 cell down to give other proba-
            bilities
```

Para hallar la probabilidad dentro de las dos colas o la probabilidad acumulativa para una cola, de 1 reste la probabilidad calculada.

TI-83/84 Plus

Probabilidad acumulativa para un valor de *t* especificado:

```
Choose:     2nd > DISTR > 5:tcdf(†
Enter:      −1EE99, t-value, df)
```

†Para hallar la probabilidad entre dos valores *t*, introduzca los dos valores en lugar de −1EE99 y el valor *t*.

Si se necesita el área de la cola derecha, de 1 reste la probabilidad calculada.

Procedimiento de intervalo de confianza

Ahora estamos listos para hacer inferencias acerca de la media poblacional μ usando la desviación muestral estándar. Como ya dijimos antes, el uso de la distribución *t* tiene una condición.

Suposición de inferencias acerca de la media μ cuando σ es desconocida: la población muestreada está normalmente distribuida.

El procedimiento para hacer intervalos de confianza usando la desviación muestral estándar es muy semejante al usado cuando σ es desconocida (vea pp. 402-407). La diferencia es el uso de la t de Student en lugar de la z normal estándar y el uso de s, la desviación muestral estándar, como estimación de σ. El teorema de límite central (CLT) implica que esta técnica también se puede aplicar a poblaciones no normales cuando el tamaño muestral es suficientemente grande.

Intervalo de confianza para media

$$\bar{x} - t(\text{df}, \alpha/2)\left(\frac{s}{\sqrt{n}}\right) \quad \text{a} \quad \bar{x} + t(\text{df}, \alpha/2)\left(\frac{s}{\sqrt{n}}\right), \quad \text{con df} = n - 1 \tag{9.1}$$

El ejemplo 9.4 ilustrará la formación de un intervalo de confianza que utiliza la distribución t.

EJEMPLO 9.4 Intervalo de confianza para μ con σ desconocida

Se toma una muestra aleatoria de 20 pesos de bebés nacidos en el Northside Hospital. Para la muestra, se encontraron una media de 6.87 y una desviación estándar de 1.76 libras. Estime, con 95% de confianza, el peso medio de todos los bebés nacidos en este hospital. Con base en información pasada, se supone que los pesos de recién nacidos están normalmente distribuidos.

SOLUCIÓN

PARA SU INFORMACIÓN
El procedimiento de cinco pasos del intervalo de confianza se da en la página 404

Paso 1 **El inicio:**

Describir el parámetro poblacional de interés.

μ, el peso medio de recién nacidos en el Northside Hospital.

Paso 2 **Criterios del intervalo de confianza:**

a. Comprobar las suposiciones.

La información pasada indica que la población muestreada es normal.

b. Identificar la distribución de probabilidad y la fórmula a usar. El valor de la desviación poblacional estándar, σ, es desconocida.

Se usará la distribución t de Student con la fórmula (9.1).

PARA SU INFORMACIÓN
Recuerde que los intervalos de confianza son situaciones de dos colas.

c. Exprese el nivel de confianza: $1 - \alpha = 0.95$.

Paso 3 **Evidencia muestral:**

Recolecte la información muestral: $n = 20$ $\bar{x} = 6{,}87$ y $s = 1.76$.

Paso 4 **Intervalo de confianza:**

a. Determinar los coeficientes de confianza.

PARA SU INFORMACIÓN
df se utiliza para calcular el coeficiente de confianza de la tabla 6, n se usa en la fórmula.

Como $1 - \alpha = 0.95$, $\alpha = 0.05$, y por tanto $\alpha/2 = 0.025$. También, como $n = 20$, df = 19. En la int ersección de la fila df = 19 y la columna de una cola $\alpha = 0.025$ de la tabla 6, encontramos $t(\text{df}, \alpha/2) = t(19, 0.025) = 2.09$ Vea la figura.

La información sobre el coeficiente de confianza y uso de la tabla 6 está en las páginas 477-479.

b. Encontrar el error máximo de estimación.

$$E = t(\text{df}, \alpha/2)\left(\frac{s}{\sqrt{n}}\right): \quad E = t(19, 0.025)\left(\frac{s}{\sqrt{n}}\right)$$

$$= 2.09\left(\frac{1.76}{\sqrt{20}}\right) = (2.09)(0.394) = 0.82$$

c. Hallar los límites de confianza inferior y superior.

$$\bar{x} - E \quad \text{a} \quad \bar{x} + E$$
$$6.87 - 0.82 \quad \text{a} \quad 6.87 + 0.82$$
$$6.05 \quad \text{a} \quad 7.69$$

Paso 5 **Resultados:**

Expresar el intervalo de confianza.

6.05 a 7.69, el intervalo de confianza de 95% para μ. Esto es, con 95% de confianza estimamos que el peso medio de bebés nacidos en el Northside Hospital está entre 6.05 y 7.69 libras.

INSTRUCCIONES DE TECNOLOGÍA: INTERVALO DE CONFIANZA DE $1 - \alpha$ PARA LA MEDIA μ CON σ DESCONOCIDA

MINITAB (Versión 14)

Introduzca los datos en C1; luego continúe con:

```
Choose:     Stat > Basic Statistics > 1-Sample t
Enter:      Samples in columns: C1
Select:     Options
Enter:      Confidence level: 1 − α (ex. 95.0)
Select:     Alternative: not equal > OK > OK
```

Excel

Introduzca los datos en la columna A; luego continúe con:

```
Choose:     Tools > Data Analysis Plus > t-Estimate: Mean > OK
Enter:      Input Range: (A1:A20 or select cells)
Enter:      Alpha: α (ex. 0.05) > OK
```

TI-83/84 Plus

Introduzca los datos en L1; luego continúe con lo siguiente, introduciendo los valores apropiados y destacando Calculate:

```
Choose:     STAT > TESTS > 8:Tinterval
```

La solución del MINITAB al ejemplo 9.4 se ve como lo siguiente:

```
One-Sample T: C1
Variable     N    Mean     StDev    SE Mean    95% CI
C1           20   6.870    1.760    0.394      (6.047, 7.693)
```

Procedimiento de prueba de hipótesis

La estadística t se usa para completar una prueba de hipótesis acerca de la media poblacional μ en forma muy semejante a como se usó z en el capítulo 8. En situaciones de prueba de hipótesis, usamos la fórmula (9.2) para calcular el valor de la **estadística de prueba $t\star$:**

Estadística de prueba para media

$$t\star = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{con } df = n - 1 \tag{9.2}$$

La ***t* calculada** es el número de errores estándar estimados a que $\bar{x}$ está de la media hipotética μ. Al igual que con los intervalos de confianza, el teorema de límite central (CLT) indica que la distribución t también se puede aplicar a poblaciones no normales cuando el **tamaño muestral** es suficientemente grande.

EJEMPLO 9.5 Prueba de hipótesis de una cola para μ con σ desconocida

Regresemos a la hipótesis del ejemplo 8.13 (p. 429) donde la Environmental Protection Agency (EPA) deseaba demostrar que el nivel medio de monóxido de carbono era más alto que 4.9 partes por millón. ¿Una muestra aleatoria de 22 lecturas (resultados muestrales: $\bar{x} = 5.1$ y $s = 1.17$) presenta evidencia suficiente para apoyar lo dicho por la EPA? Use $\alpha = 0.05$. Estudios previos han indicado que estas lecturas tienen una distribución aproximadamente normal.

SOLUCIÓN

PARA SU INFORMACIÓN
El procedimiento de cinco pasos para prueba de hipótesis de valor *p* se da en la página 404.

PARA SU INFORMACIÓN
Los procedimientos para escribir H_0 y H_a se analizan en las páginas 428-430.

Paso 1 El inicio:

a. Describir el parámetro poblacional de interés.
μ, el nivel medio de monóxido de carbono de aire en el centro de Rochester.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).
H_o: $\mu = 4.9$ ($\leq$) (no más alto)
H_a: $\mu > 4.9$ (más alto)

Paso 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones.
Las suposiciones se satisfacen porque la población muestreada es aproximadamente normal y el tamaño muestral es grande lo suficiente para que aplique el CLT (vea página 475).

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
σ, es desconocida; por tanto, se usará la distribución t con df $= n - 1 = 21$, y la estadística de prueba es $t\bigstar$, fórmula (9.2).

c. Determinar el nivel de significación: $\alpha = 0.05$.

Paso 3 Evidencia muestral:

a. Recolectar la información muestral: $n = 22$, $\bar{x} = 5.1$ y $s = 1.17$.

b. Calcular el valor de la estadística de prueba.
Use fórmula (9.2):

$$t\bigstar = \frac{\bar{x} - \mu}{s/\sqrt{n}} : \quad t\bigstar = \frac{5.1 - 4.9}{1.17/\sqrt{22}} = \frac{0.20}{0.2494} = 0.8018 = 0.80$$

Paso 4 Distribución de probabilidad:

Usando el procedimiento del valor *p*:

a. Calcular el valor *p* para la estadística de prueba.
Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor que".

O

Usando el procedimiento clásico:

a. Determine la región crítica y valor(es) crítico(s).
La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "mayor que".

P = $P(t\bigstar > 0.80$, con df = 21) como se muestra en la figura.

Para hallar el valor *p*, use uno de tres métodos:

1. Use la tabla 6 del apéndice B para poner límites al valor *p*: $0.10 < \mathbf{P} < 0.25$.
2. Usar la tabla 7 del apéndice B para leer directamente el valor: **P** = 0.216.
3. Use computadora o calculadora para calcular el valor *p*: **P** = 0.2163.

Detalles específicos siguen a este ejemplo.

b. Determinar si el valor *p* es o no es menor a α.

El valor *p* no es menor que α, el nivel de significación.

El valor crítico se encuentra en la intersección de la fila df = 21 y la columna de una cola 0.05 de la tabla 6: $t(21, 0.05) = 1.72$

Las instrucciones específicas se dan en las páginas 477-479.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

$t\bigstar$ no está en la región crítica, como se ve en **rojo** en la figura de líneas antes.

Paso 5 **Resultados:**

a. Expresar la decisión acerca de H_o: No rechazar H_o.

b. Expresar la conclusión acerca de H_a.

En el nivel de significación de 0.05, la EPA no tiene suficiente evidencia para demostrar que el nivel medio de monóxido de carbono es mayor a 4.9.

Cálculo del valor *p* cuando se usa la distribución *t*

Método 1: usar la tabla 6 del apéndice B para poner límites al valor p. Al inspeccionar la fila df = 21 de la tabla 6, puede determinar un intervalo dentro del cual se encuentra el valor *p*. Localice $t\bigstar$ por toda la fila marcada df = 21. Si $t\bigstar$ no aparece en la lista, localice los dos valores de tabla entre los que caiga, y lea los límites al valor *p* desde la parte superior de la tabla. En este caso, $t\bigstar = 0.80$ está entre 0.686 y 1.32; por tanto, **P** está entre 0.10 y 0.25. Use el encabezado de una cola, puesto que H_a es de una cola en esta ilustración. (Use el encabezado de dos colas cuando H_a sea de dos colas.)

Localización de P = $P(t\bigstar > 0.80$, con df = 21)

Parte de la tabla 6

	Cantidad de α en una cola			
df	0.25	P		0.10
⋮	⋮	↑		⋮
21	0.686	0.80		1.32

$0.10 < \mathbf{P} < 0.25$

La entrada de 0.686 en la tabla nos dice que $P(t > 0.686) = 0.25$, como se ve en púrpura en la figura. La entrada de 1.32 en la tabla nos dice que $P(t > 1.32) = 0.10$, como se ve en verde. Se puede ver que el valor *p* **P** (que se ilustra en azul) está entre 0.10 y 0.25. Por tanto, $0.10 < \mathbf{P} < 0.25$, y decimos que 0.10 y 0.25 son los "límites" para el valor *p*.

Método 2: usar la tabla 7 del apéndice B para leer el valor p o "poner límites" al valor p. La tabla 7 está diseñada para generar valores p dados los valores $t\star$ y df o producir límites a **P** que son más angostos de lo que produce la tabla 6.

En el ejemplo precedente, $t\star = 0.80$ y df = 21. Resulta que éstos son encabezados de fila y columna, de modo que el valor p se puede leer directamente de la tabla. Localice el valor p en la intersección de la fila $t\star = 0.80$ y la columna df = 21. El valor p para $t\star = 0.80$ con df = 21 es **0.216.**

Parte de la tabla 7

$t\star$	df	. . .	21
⋮			
0.80			**0.216**

$\longrightarrow$ $\mathbf{P} = P(t\star > 0.80, \text{ con df } = 21) = \mathbf{0.216}$

Para ilustrar cómo poner límites al valor p cuando $t\star$ y df no son los valores de encabezado, consideremos la situación donde $t\star = 2.43$ con df = 16. El $t\star = 2.43$ está entre las filas $t = 2.4$ y $t = 2.5$, en tanto que df = 16 está entre las columnas df = 15 y df = 18. Estas dos filas y dos columnas se intersecan un total de cuatro veces, es decir, en 0.015 y 0.014 en la fila $t\star = 2.4$ y en 0.012 y 0.011 en la fila $t\star = 2.5$. El valor p que buscamos está limitado por los más pequeños y más grandes de estos cuatro valores, es decir, 0.011 (inferior derecha) y 0.015 (superior izquierda). Por tanto, los límites al valor p son $0.011 < \mathbf{P} < 0.015$.

Parte de la tabla 7

$t\star$	df	. . .	15	*16*	18
⋮					
2.4			**0.015**		0.014
2.43				**P**	
2.5			**0.012**		0.011

$\longrightarrow$ $\mathbf{P} = P(t\star > 2.43, \text{ con df } = 16)$

$\mathbf{0.011 < P < 0.015}$

Método 3: si está usted haciendo la prueba de hipótesis con ayuda de una computadora o calculadora es muy probable que calculen el valor p, o puede usar los comandos de distribución de probabilidad acumulativa descritos en la página 479.

Veamos una situación de probar hipótesis de dos colas.

EJEMPLO 9.6 Prueba de hipótesis de dos colas para μ con σ desconocida

En una popular prueba de imagen propia que resulta en calificaciones distribuidas normalmente, se espera que el puntaje medio para quienes reciben asistencia pública. Se aplica la prueba a una muestra aleatoria de 28 personas que reciben asistencia pública en el condado de Emerson. Alcanzan una calificación de 62.1, y sus calificaciones tienen una desviación estándar de 5.83. Quienes reciben asistencia pública en el condado de Emerson, ¿obtienen calificación muy diferente, en promedio, de lo que se esperaba al nivel de significación de 0.02?

SOLUCIÓN

PASO 1 El inicio:

a. Describir el parámetro poblacional de interés.

μ, calificación media de prueba de imagen propia para todos los que reciben asistencia pública en el condado de Emerson.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

H_o: $\mu = 65$ (la media es 65)

H_a: $\mu \neq 65$ (la media es diferente de 65)

Paso 2 **Criterios de prueba de hipótesis:**

a. Comprobar las suposiciones.

Se espera que la prueba produzca calificaciones normalmente distribuidas; por tanto, la suposición ha quedado satisfecha; σ es desconocida.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.

La distribución t con df $= n - 1 = 27$, y la estadística de prueba es $t\bigstar$, fórmula (9.2).

c. Determinar el nivel de significación: $\alpha = 0.02$ (dada en el enunciado del problema).

Paso 3 **Evidencia muestral:**

a. Recolectar la información muestral: $n = 28$, $\bar{x} = 62.1$, y $s = 5.83$.

b. Calcular el valor de la estadística de prueba.

Use la fórmula (9.2):

$$t\bigstar = \frac{\bar{x} - \mu}{s/\sqrt{n}} : \quad t\bigstar = \frac{62.1 - 65.0}{5.83/\sqrt{28}} = \frac{-2.9}{1.1018} = -2.632 = -2.63$$

Paso 4 **Distribución de probabilidad:**

Usando el procedimiento del valor *p*:

a. Calcular el valor *p* para la estadística de prueba.

Usar ambas colas porque H_a expresa interés por valor relacionados con "diferente de".

$\mathbf{P} = P(t < -2.63) + P(t > 2.63) = 2 \cdot P(t > 2.63)$, con df = 27 como se muestra en la figura.

Para hallar el valor p, use uno de tres métodos:

1. Use la tabla 6 del apéndice B para poner límites al valor p: $0.01 < \mathbf{P} < 0.02$.
2. Use la tabla 7 del apéndice B para poner límites al valor p: $0.012 < \mathbf{P} < 0.016$.
3. Use computadora o calculadora para calcular el valor p: $\mathbf{P} = 0.0140$.

Detalles específicos siguen a este ejemplo:

b. Determinar si el valor *p* es o no es menor que α.

El valor p es menor que el nivel de significación, α.

O

Usando el procedimiento clásico:

a. Determinar la región crítica y valor(es) crítico(s):

La región crítica es ambas colas porque H_a expresa interés por valores relacionados con "diferente de". El valor crítico se encuentra en la intersección de la fila df = 27 y la columna de una cola 0.01 de la tabla 6: $t(27, 0.01) = 2.47$

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

$t\bigstar$ está en la región crítica, como se ve en **sombreado** en la figura precedente.

Paso 5 **Resultados:**

a. Expresar la decisión acerca de H_o: Rechazar H_o.

b. Expresar la conclusión acerca de H_a:

En el nivel de significación de 0.02, tenemos suficiente evidencia para concluir que quienes reciben asistencia del condado de Emerson obtienen calificación significativamente diferente, en promedio, al esperado 65.

Cálculo del valor *p* cuando se usa la distribución *t*

Método 1: usando la tabla 6, encuentre 2.63 entre dos entradas en la fila df = 27 y lea los límites a **P** del encabezado de dos cola en la parte superior de la tabla:

$$\mathbf{0.01 < P < 0.02.}$$

Método 2: generalmente, los límites hallados con el uso de la tabla 7 serán más angostos que los límites hallados usando la tabla 6. La tabla siguiente muestra cómo leer los límites de la tabla 7; encuentre $t\bigstar = 2.63$ entre dos filas y df = 27 entre dos columnas, y localice las cuatro intersecciones de estas columnas y filas. El valor de $\frac{1}{2}$ **P** está limitado por las entradas superior izquierda e inferior derecha de la tabla.

Parte de la tabla 7

	Grados de libertad			
$t\bigstar$	25	*27*	29	**P** = $2P(t\bigstar > 2.63$, con df = 27)
⋮		↓	⋮	
2.6	**0.008**		0.007	
2.63		½ **P**		**0.006 < ½ P < 0.008**
2.7	0.006		**0.006**	**0.012 < P < 0.016**

Método 3: si está haciendo usted la prueba de hipótesis con ayuda de una computadora o calculadora, es probable que calcule el valor *p* (no lo duplique). O bien, puede usar los comandos de distribución de probabilidad acumulativa descritos en la página 479.

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA μ MEDIA CUANDO σ ES DESCONOCIDA

MINITAB (Versión 14)

Introduzca los datos en C1; luego continúe con:

```
Choose:     Stat > Basic Statistics > 1-Sample t
Enter:      Samples in columns: C1
            Test mean: μ (value in H_o)
Select:     Options
Select:     Alternative: less than or not equal or greater than > OK >
            OK
```

Excel

Introduzca los datos en la columna A; luego continúe con:

```
Choose:     Tools > Data Analysis Plus > t-Test: Mean > OK
Enter:      Input Range: (A1:A20 or select cells) > OK
            Hypothesized Mean: μ
            Alpha: α (ex. 0.05) > OK
            Gives p-values and critical values for both one-tailed and
            two-tailed tests.
```

TI-83/84 Plus

Introduzca los datos en L1; luego continúe con lo siguiente, introduciendo los valores apropiados y destacando Calculate:

```
Choose:     STAT > TESTS > 2:T-Test
```

PARA SU INFORMACIÓN

Compare los resultados del MINITAB con la solución hallada en el ejemplo 9.6.

A continuación aparece la solución MINITAB al ejemplo 9.6:

```
One-Sample T: C1
Test of mu 5 65 vs not 5 65
Variable      N       Mean      StDev      SE Mean      T          P
C1            28      62.1      5.83       1.102        22.63      0.0140
```

CASO PRÁCTICO 9.7 Uso de pronombres personales, de parte de mamá, cuando habla a niños pequeños

El valor t calculado y el valor de probabilidad para cinco pruebas de hipótesis diferentes se dan en el artículo siguiente. La expresión $t(44) = 1.92$ significa $t\bigstar = 1.92$ con df = 44 y es significativa con valor $p < 0.05$. ¿Puede usted verificar los valores p? Explique.

RESUMEN

La interacción verbal de niños de 2 años ($N = 46$; 16 niñas, 30 niños) y sus madres fue grabada, trascrita y analizada en busca de pronombres personales, el número total de expresiones, la duración media de pronunciación del niño y la respuesta de la madre a las expresiones de su hijo. El uso del pronombre personal "nosotros" por parte de la madre estuvo significativamente relacionado con el comportamiento de los niños en la prueba Stanford-Binet para niños de 5 años de edad, y la escala de inteligencia Wechsler para niños de 8 años de edad. El uso de "nosotros" por parte de la madre en intercambio social-vocal, que indica un sistema de establecer una relación compartida con el niño, estuvo estrechamente conectada con su sensibilidad verbal a sus hijos. La cantidad total de voces maternales, el número de pronombres personales usados por las madres, y su sensibilidad verbal a sus hijos no estuvieron relacionados con la clase social de las madres ni los años de educación.

Las madres tendieron a usar más pronombres singulares en primera persona (yo y mi), $t(44) = 1.81$, $p < .10$ y usaron significativamente más pronombres plurales en primera persona (nosotros), $t(44) = 1.92$, $p < .05$, con niñas que con niños. Las madres también respondieron más verbalmente a sus hijas, $t(44) = 2.0$, $p < .06$.

En general, las madres hablaron más a sus hijos primogénitos, $t(44) = 3.41$, $p < .001$, y respondían más a éstos, $t(44) = 3.71$, $p < .001$. No obstante,

la proporción de pronombres personales usados cuando hablaban a sus primogénitos no fue diferente a la que usaban cuando hablaban a sus siguientes hijos.

Fuente: Dan R. Laks, Leila Beckwith y Sarale E. Cohen, *The Journal of Genetic Psychology,* 151(1), 25-32, 1900. Reimpreso con permiso de la Helen Dwight Reid Educational Foundation. Publicado por Heldref Publications, 1319 Eighteenth St. N.W., Washington, D.C., 20036-1802. Copyright © 1990.

SECCIÓN 9.2 EJERCICIOS

9.3 Haga una lista de cuatro números que totalicen "cero". ¿Cuántos números pudo seleccionar sin restricción? Explique cómo es que esto demuestra los grados de libertad.

9.4 Explique la relación entre los valores críticos hallados en la fila inferior de la tabla 6 y los valores críticos de z dados en la tabla 4A.

9.5 Encuentre:

a. $t(12, 0.01)$ b. $t(22, 0.025)$

c. $t(50, 0.10)$ d. $t(8, 0.005)$

9.6 Encuentre los valores críticos usando la tabla 6 del apéndice B:

a. $t(25, 0.05)$ b. $t(10, 0.10)$

c. $t(15, 0.01)$ d. $t(21, 0.025)$

9.7 Encuentre:

a. $t(18, 0.90)$ b. $t(9, 0.99)$

c. $t(35, 0.975)$ d. $t(14, 0.98)$

9.8 Encuentre los valores críticos usando la tabla 6 del apéndice B.

a. $t(21, 0.95)$ b. $t(26, 0.975)$

c. $t(27, 0.99)$ d. $t(60, 0.025)$

9.9 Usando la notación del ejercicio 9.8, mencione y encuentre los siguientes valores críticos de t:

9.10 Usando la notación del ejercicio 9.8, mencione y encuentre los siguientes valores críticos de t:

9.11 Encuentre los valores de t que limitan el 0.95 central de la distribución para df = 12.

9.12 Encuentre los valores de t que limitan el 0.80 central de la distribución para df = 26.

9.13 a. Encuentre el primer percentil de la distribución t de Student con 24 grados de libertad.

b. Encuentre el 95avo percentil de la distribución t de Student con 24 grados de libertad.

c. Encuentre el primer percentil de la distribución t de Student con 24 grados de libertad.

9.14 Encuentre el porcentaje de la distribución t de Student que se encuentre entre los siguientes valores:

a. df = 12 y rangos de t de −1.36 a 2.68

b. df = 15 y rangos de t de −1.75 a 2.95

9.15 Noventa por ciento de la distribución t de Student está entre $t = -1.89$ y $t = 1.89$ ¿para cuántos grados de libertad?

9.16 Noventa por ciento de la distribución t de Student está a la derecha de $t = -1.37$ ¿para cuántos grados de libertad?

9.17 Use computadora o calculadora para hallar el área situada a la izquierda de $t = -2.12$ con df = 18. Trace un dibujo que muestre la pregunta con la respuesta.

9.18 Use computadora o calculadora para hallar el área situada a la izquierda de $t = 1.12$ con df = 15. Trace un dibujo que muestre la pregunta con la respuesta.

9.19 a. Exprese dos formas en las que la distribución normal estándar y la distribución t de Student son semejantes.

b. Exprese dos formas en las que son diferentes.

9.20 La varianza para cada una de las distribuciones t de Student es igual a df/(df − 2). Encuentre la desviación estándar para una distribución t de Student con cada uno de los siguientes grados de libertad:

a. 10 b. 20 c. 30

En resumen:

d. Explique la forma en que esto verifica la propiedad 5 de las distribuciones t de la página 4.76.

9.21 Construya una estimación de intervalo de confianza para la media μ usando la información muestral $n = 24$, $\bar{x} = 16.7$, y $s = 2.6$.

9.22 En un estudio de 25 criminales condenados por delitos de antimonopolios, la edad promedio era de 54 años, con una desviación estándar de 7.5 años. Construya un intervalo de confianza de 90% sobre la verdadera edad media.

9.23 Un estudio de 3000 ciudadanos de Minnesota seleccionados al azar, de 65 años de edad o mayores, reveló que en promedio gastan $85 por mes en recetas médicas, con una desviación estándar de $50.35 por mes. Construya un intervalo de confianza de 99% para la verdadera cantidad media gastada por mes.

9.24 El tornillo Robertson para clavar fue inventado en 1908, pero ha ganado popularidad entre carpinteros norteamericanos y ebanistas caseros sólo hasta los últimos diez años. Las ventajas de estos tornillos sobre los convencionales es sorprendente, con una mayor resistencia, más fuerza de sujeción y menos resistencia para introducirlos y sacarlos. Los resultados en pruebas de resistencia publicados en el catálogo McFeely de 2005 revelaron que los tornillos Robertson No. 8 de acero y cabeza plana fallan sólo después de un promedio de aplicarles una fuerza de torsión de 46 pulgadas-libras, que es una fuerza casi 50% mayor que la de los tornillos de cabeza con ranuras en cruz y que son más comunes.

Fuente: McFeely Square Drive Screws, 2005.

Suponga que un laboratorio de pruebas independiente selecciona al azar 22 tornillos para clavar, de acero y cabeza plana, de una caja de 1000 tornillos y obtiene una torsión media de falla de 45.2 pulgadas-libras y una desviación estándar de 5.1 pulgadas-libras. Estime, con 95% de confianza, la torsión media de falla de tornillos No. 8 para madera con base en el estudio por el laboratorio independiente. Especifique el parámetro poblacional de interés, los criterios, la evidencia muestral y los límites de intervalo.

9.25 Cuando escribía un artículo sobre el alto costo de la educación universitaria, una periodista tomó una muestra aleatoria del costo de los libros de texto nuevos para un semestre. La variable aleatoria x es el costo de un libro. Sus datos muestrales pueden resumirse con $n = 41$, $\Sigma x = 3582.17$, y $\Sigma(x - \bar{x}) = 9960.336$.

a. Encuentre la media muestral, $\bar{x}$.

b. Encuentre la desviación muestral estándar, s.

c. Encuentre el intervalo de confianza de 90% para estimar el verdadero costo medio de libros de texto para el semestre, con base en esta muestra.

9.26 Los tiempos para tomar un bocadillo son considerados demasiado cortos, y es frecuente que los empleados se formen el hábito de "alargarlos". El gerente de Giant Mart identificó al azar 22 empleados y observó el tiempo de sus tiempos para tomar bocadillos (en minutos) para un día seleccionado en forma aleatoria durante la semana:

30	24	38	35	27	35	23	28	28	22	26
34	29	25	28	34	24	26	28	32	29	40

a. Muestre evidencia de que se satisfacen las suposiciones de normalidad.

b. Encuentre el intervalo de confianza de 95% para "duración media de tiempo para tomar bocadillos" en Giant Mart.

9.27 Nuestra carne o pollo para azar en estos tiempos pueden pesar 5.5 libras a los 49 días, tiempo común para vender el pollo a un procesador de carnes. Antes que el granjero productor venda sus pollos, desea estimar el peso medio de los mismos; selecciona una muestra aleatoria de 15 de los pollos actuales de 7 semanas y obtiene sus pesos (en onzas):

74.9	74.2	73.9	72.6	70.4	66.9	76.6	73.3
75.4	78.9	76.6	75.2	73.6	78.6	70.4	

a. ¿Cuántas onzas hay en 5.5 libras? ¿Cuántas libras hay en 74.9 onzas?

b. Muestre evidencia de que las suposiciones de normalidad se satisfacen.

c. Calcule las medias estadísticas muestrales y desviación estándar.

d. Estime el peso medio por pollo para asar usando un intervalo de confianza de 98%.

e. Si el granjero vende 1000 pollos para asar, construya un intervalo de confianza de 98% para el peso total en libras.

9.28 La incorporación de un nuevo acelerador se dice que reduce el tiempo de secado de pintura de látex en más de 4%. Se efectuaron varias muestras de prueba con la siguiente reducción de porcentaje en tiempo de secado.

5.2	6.4	3.8	6.3	4.1	2.8	3.2	4.7

Suponga que la reducción de porcentaje en tiempo de secado está normalmente distribuida.

a. Encuentre el intervalo de confianza de 95% para la verdadera reducción media en el tiempo de secado con base en esta muestra. (La media muestral y desviación estándar se encontraron al contestar el ejercicio 2.181, p. 129.)

b. La estimación de intervalo alcanzada en la parte a, ¿resulta en la misma conclusión que esperaba al contestar la parte c del ejercicio 2.181 para estos mismos datos?

9.29 Los ritmos cardiacos para 13 mujeres adultas fueron como sigue:

83	58	70	56	76	64	80	76	70	97	68	78	108

Verifique los resultados mostrados en la última línea de la salida del MINITAB:

```
MTB > TINTERVAL 90 PERCENT CONFIDENCE INTERVAL FOR DATA
IN C1
      N    MEAN    STDEV    SE MEAN    90% CI
C1    13   75.69   14.54    4.03       (68.50, 82.88)
```

9.30 Usando la salida de computadora del ejercicio 9.29, determine el valor de cada uno de lo siguiente:

a. Estimación puntual b. coeficiente de confianza

c. Error estándar de la media

d. Error máximo de estimación, E

e. Límite inferior de confianza

f. Límite superior de confianza

9.31 Use computadora o calculadora para construir un intervalo de confianza de 0.98 usando los datos muestrales:

6	7	12	9	10	8	5	9	7	9	6	5

9.32 James Short (1708-1768), óptico escocés, construyó los reflectores de más alta calidad de su tiempo. Fue con estos reflectores que Short obtuvo las siguientes mediciones del paralaje del Sol (en segundos de grado), con base en el tránsito de 1761 de Venus. El paralaje del Sol es el ángulo α subtendido por la Tierra, como se ve desde la superficie del Sol. (Vea el diagrama siguiente.)

8.50	8.50	7.33	8.64	9.27	9.06	9.25	9.09	8.50
8.06	8.43	8.44	8.14	7.68	10.34	8.07	8.36	9.71
8.65	8.35	8.71	8.31	8.36	8.58	7.80	7.71	8.30
9.71	8.50	8.28	9.87	8.86	5.76	8.44	8.23	8.50
8.80	8.40	8.82	9.02	10.57	9.11	8.66	8.34	8.60
7.99	8.58	8.34	9.64	8.34	8.55	9.54	9.07	

Fuente: Los datos e información descriptiva están basados en material de Stephen M. Stigler. (1977). ¿Funcionan los estimadores robustos con datos reales? *Annals of Statistics, 5,* 1055-1098.

a. Determine si es razonable una suposición de normalidad. Explique.

b. Construya un intervalo de confianza para la estimación del paralaje medio del Sol.

c. Si el verdadero valor es 8.798 segundos de grado, ¿qué sugiere el intervalo de confianza acerca de las mediciones de Short?

9.33 La información sobre economía de combustible en una calcomanía de ventanilla de una camioneta de uso múltiple (SUV), de modelo reciente, indica que su nuevo propietario puede esperar 16 mpg (millas por galón) en tránsito urbano y 20 mpg en carretera, con 18 mpg en promedio total. Se conservaron registros precisos del consumo de gasolina para uno de estos vehículos, y se recolectó una muestra aleatoria del rendimiento por tanque de gasolina:

17.6	17.7	18.1	22.0	17.0	19.4	18.9	17.4	21.0	19.2
18.3	19.1	20.7	16.7	19.4	18.2	18.4	17.1	17.4	15.8
17.9	18.0	16.3	17.5	17.3	20.4	19.1	21.0	18.1	19.0
19.6	18.9	16.8	18.2	17.6	19.1	18.0	16.8	20.9	17.9
17.7	20.3	18.6	19.0	16.5	19.4	18.6	18.6	17.3	18.7

a. Determine si es razonable una suposición de normalidad. Explique.

b. Construya un intervalo de confianza de 95% para la estimación de la distancia media en millas por galón.

c. ¿Qué sugiere el intervalo de confianza acerca de las expectativas de economía de combustible de las camionetas SUV, como se expresa en la calcomanía de ventanilla?

9.34 Los estudiantes universitarios tiran un promedio de 640 libras de basura al año, 30% de esa cantidad en el mes antes de graduarse, según el artículo de junio de 2004 "Desechos en universidades" de *Readers Digest.* El año pasado, 20 universidades recolectaron más de $100 000 para donarlos a caridad local vendiendo la "chatarra". Para estimar la cantidad de basura desechada por los estudiantes de la Universidad del Estado, 18 estudiantes fueron seleccionados al azar y cuidadosamente observados durante 1 año. Las cantidades de basura tirada (en libras) fueron como sigue:

692	563	482	413	437	456	509	347	676
944	678	392	399	742	584	492	514	758

a. Describa los datos usando un histograma y su media y desviación estándar.

b. Determine si es razonable una suposición de normalidad. Explique.

c. Construya un intervalo de confianza de 95% para la cantidad media de basura.

d. ¿Qué sugiere el intervalo de confianza acerca de la cantidad media de basura tirada por los estudiantes de la Universidad del Estado, comparada con la tirada por todos los estudiantes universitarios?

9.35 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar cada una de las siguientes frases:

a. El peso medio de abejas es al menos 11 gramos.

b. La edad media de pacientes en el Memorial Hospital no es mayor a 54 años.

c. La cantidad media de sal en barras de bocadillos de granola es diferente de 75 gramos.

9.36 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar cada una de las siguientes frases:

a. Un granjero productor de pollos en Best Broilers dice que sus pollos tienen un peso medio de 56 onzas.

b. La edad media de aviones jet comerciales de Estados Unidos es menor a 18 años.

c. El saldo medio no pagado mensualmente de cuentas de tarjeta de crédito es más de $400.

9.37 Calcule el valor de t★ para la prueba de hipótesis: H_o: $\mu = 32$, H_a: $\mu > 32$, $n = 16$, $\bar{x} = 32.93$, $s = 3.1$.

9.38 Calcule el valor de t★ para la prueba de hipótesis: H_o: $\mu = 73$, H_a: $\mu \neq 73$, $n = 12$, $\bar{x} = 71.46$, $s = 4.1$.

9.39 Determine el valor p para las siguientes pruebas de hipótesis que comprenden la distribución t con 10 grados de libertad.

a. H_o: $\mu = 15.5$, H_a: $\mu < 15.5$, $t\bigstar = -2.01$

b. H_o: $\mu = 15.5$, H_a: $\mu > 15.5$, $t\bigstar = 2.01$

c. H_o: $\mu = 15.5$, H_a: $\mu \neq 15.5$, $t\bigstar = 2.01$

d. H_o: $\mu = 15.5$, H_a: $\mu \neq 15.5$, $t\bigstar = -2.01$

9.40 Determine la región crítica y valor(es) crítico(s) que se usarían en el método clásico para probar las siguientes hipótesis nulas:

a. H_o: $\mu = 10$, H_a: $\mu \neq 10$ ($\alpha = 0.05$, $n = 15$)

b. H_o: $\mu = 37.2$, H_a: $\mu > 37.2$ ($\alpha = 0.01$, $n = 25$)

c. H_o: $\mu = -20.5$, H_a: $\mu < -20.5$ ($\alpha = 0.05$, $n = 18$)

d. H_o: $\mu = 32.0$, H_a: $\mu > 32.0$ ($\alpha = 0.01$, $n = 42$)

9.41 a. Encuentre el valor de **P** y exprese la decisión para la prueba de hipótesis del ejercicio 9.37 usando $\alpha = 0.05$.

b. Encuentre la región crítica y valor crítico, y exprese la decisión para la prueba de hipótesis del ejercicio 9.37 usando $\alpha = 0.05$.

9.42 a. Use la tabla 6 o la tabla 7 del apéndice B para hallar el valor de **P** para la prueba de hipótesis del ejercicio 9.38; exprese la decisión usando $\alpha = 0.05$.

b. Encuentre la región crítica y valores críticos para la prueba de hipótesis del ejercicio 9.38; exprese la decisión usando $\alpha = 0.05$.

9.43 Use computadora o calculadora para hallar el valor p para la prueba de hipótesis: H_o: $\mu = 32$, H_a: $\mu > 32$, $n = 16$, $\bar{x} = 32.93$, $s = 3.1$.

9.44 Use computadora o calculadora para hallar el valor p para la siguiente prueba de hipótesis: H_o: $\mu = 73$, H_a: $\mu \neq 73$, $n = 12$, $\bar{x} = 71.46$, $s = 4.1$.

9.45 Use el valor p y métodos clásicos para prueba de hipótesis y llegar a una decisión para cada una de las siguientes situaciones. Use $\mu = 0.05$.

a. H_o: $\mu = 128$, H_a: $\mu \neq 128$, $n = 15$, $t\bigstar = 1.60$

b. H_o: $\mu = 18$, H_a: $\mu > 18$, $n = 25$, $t\bigstar = 2.16$

c. H_o: $\mu = 38$, H_a: $\mu < 38$, $n = 45$, $t\bigstar = -1.73$

d. Compare los resultados de las dos técnicas para cada caso.

9.46 Con referencia al caso práctico 9.7 (p. 487):

a. Verifique que $t(44) = 1.92$ es significativa al nivel 0.05.

b. Verifique que $t(44) = 3.41$ es significativa al nivel 0.01.

c. Explique por qué $t(44) = 1.81$, $p < .10$ tiene sentido sólo si la prueba de hipótesis es de dos colas.

d. Si la prueba es de una cola, ¿qué nivel se informaría?

9.47 Un grupo de estudiantes sostiene que todos los días, el estudiante promedio debe viajar al menos 25 minutos en una dirección para llegar a la universidad. La oficina de inscripciones de la universidad obtuvo una muestra aleatoria de 31 tiempos de viaje en una dirección de estudiantes. La muestra tuvo una media de 19.4 minutos y una desviación estándar de 9.6 minutos. ¿Tiene la oficina de inscripciones suficiente evidencia para rechazar lo dicho por los estudiantes? Use $\alpha = 0.01$.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.48 Las casas de una población cercana a la universidad tienen un valor medio de \$88 950. Se supone que las casas ubicadas cerca de la universidad tienen un valor medio más alto. Para probar esta teoría, se selecciona una muestra aleatoria de 12 casas de la zona de la universidad. Su valor medio es \$92 460, y la desviación estándar es \$5200. Complete una prueba de hipótesis usando $\alpha = 0.05$. Suponga que los precios están normalmente distribuidos.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.49 Los consumidores disfrutan la meticulosa selección de mercancía hecha posible por tiendas de especialidades que sacrifican variedad por obtener mejores artículos. Considere las tiendas que venden sólo pantalones Levi Strauss. La compañía informa que una tienda de Levi completamente surtida lleva en existencia 130 pares de pantalones listos para usarse, de cualquier medida de cintura y costura interior dadas, y

está introduciendo dos líneas más de pantalones (Par Personal y Paseo Original) que dice cuadruplicará finalmente.

Fuente: Fortune, "The Customized, Digitized, Have-It-Your-Way Economy"

Suponga que una muestra aleatoria de 24 tiendas Levi se muestrea 2 meses después que se ha lanzado el proceso de introducción, y se toman inventarios en cada una de las tiendas de la muestra en cuanto a todas las tallas de pantalones vaqueros. El número medio muestral de opciones para cualquier talla determinada es 141.3, y la desviación estándar es 36.2. ¿Esta muestra de tiendas lleva en existencia una mayor selección de pantalones vaqueros, en promedio, de lo que se espera al nivel de significación de 0.01?

a. Resuelva usando el método del valor *p*.

b. Resuelva usando el método clásico.

9.50 ¿Despierto toda la noche? El deseo intenso por tomar cafeína puede causar problemas de salud a largo plazo. Tareas, trabajos y estudios pueden ser causas de que adolescentes consuman con frecuencia demasiado café. Los expertos en salud advierten que un elevado consumo de cafeína no es bueno para nadie y que sigue siendo cada vez más popular. No es el café lo que preocupa sino la cantidad de cafeína; una cantidad moderada no es para preocupar, dicen los expertos. No hay riesgos de salud si se toman tres tazas de 8 onzas de café regular, o sea unos 250 mg de cafeína al día, según el Sistema de Salud Henry Ford.

Fuente: New Expressions, http://www.newexpression.org/main/cover

Una muestra aleatoria a nivel nacional de estudiantes universitarios reveló que 24 estudiantes consumían un total de 5428 mg de cafeína al día, con una desviación estándar de 48 mg. Suponiendo que la cantidad de cafeína consumida por persona al día está normalmente distribuida, ¿hay suficiente evidencia para concluir que la cantidad media de cafeína consumida al día por estudiantes universitarios es menos de 250 mg, usando $\alpha = 0.05$?

a. Complete la prueba usando el método del valor *p*. Incluya $t\bigstar$, el valor *p*, y su conclusión.

b. Complete la prueba usando el método clásico. Incluya $t\bigstar$, los valores críticos *p*, y su conclusión.

9.51 Para probar la hipótesis nula "el peso medio de hombres adultos es de 160 libras", contra la alternativa de "el peso medio de hombres adultos rebasa las 160 libras," se obtuvieron los pesos de 16 hombres:

173	178	145	146	157	175	173	137
152	171	163	170	135	159	199	131

Suponga normalidad y verifique los resultados mostrados en el siguiente análisis del MINITAB, calculando por sí mismo los valores.

```
TEST OF MU = 160.00 VS MU > 160.00
        N      MEAN      STDEV     SE MEAN    T       P
C1      16     160.25    18.49     4.62       0.05    0.48
```

9.52 Usando la salida de computadora del ejercicio 9.51, determine los valores de los términos siguientes:

a. Valor hipotético de media poblacional

b. Media muestral

c. Desviación estándar poblacional

d. Desviación estándar muestral

e. Estadística de prueba

9.53 Use computadora o calculadora para completar la prueba de hipótesis: H_o: $\mu = 52$, H_a: $\mu < 52$, $\alpha = 0.01$ usando los datos:

45	47	46	58	59	49	46	54	53	52	47	41

9.54 El número recomendado de horas de sueño por noche es 8 horas, pero todo el mundo "sabe" que el promedio de estudiante universitario duerme menos de 7 horas. El número de horas que durmieron la noche anterior 10 estudiantes universitarios seleccionados al azar es el siguiente:

5.2	6.8	6.2	5.5	7.8	5.8	7.1	8.1	6.9	5.6

Use computadora o calculadora para completar la prueba de hipótesis: H_o: $\mu = 7$, H_a: $\mu < 7$, $\alpha = 0.05$.

9.55 Se dice que los estudiantes de cierta universidad obtendrán una calificación promedio de 35 en un examen determinado. ¿Es razonable esta afirmación si una muestra aleatoria de calificación de examen de esta universidad dio 33, 42, 38, 37, 30, 42? Complete una prueba de hip ótesis usando $\alpha = 0.05$. Suponga que los resultados del examen están normalmente distribuidos.

a. Resuelva usando el médico del valor *p*.

b. Resuelva usando el método clásico.

9.56 Se supone que la gasolina bombeada desde el oleoducto de un proveedor tiene un índice de 87.5 octanos. En 13 días consecutivos, se tomó y analizó una muestra con los resultados siguientes.

88.6	86.4	87.2	88.4	87.2	87.6	86.8	86.1	87.4	87.3	86.4	86.6	87.1

a. Si los índices de octano tienen una distribución normal, ¿hay suficiente evidencia para demostrar que estas lecturas de octanos se tomaron de gasolina con un octanaje medio significativamente menor a 87.5 al nivel de 0.05? (La media muestral y desviación estándar se encontraron al contestar el ejercicio 2.182, p. 129.)

b. La decisión estadística a la que se llegó en la parte a, ¿resulta en la misma conclusión esperada al contestar la parte c del ejercicio 2.182 para estos mismos datos?

9.57 Se sabe que la densidad de la Tierra respecto a la densidad del agua es 5.517 g/cm^3. Henry Cavendish, químico y físico inglés (1731-1810), fue el primero en medir con precisión la densidad de la Tierra. A continuación aparecen 29 mediciones tomadas por Cavendish en 1798 usando una balanza de torsión.

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65	5.57
5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39	5.42	5.47
5.63	5.34	5.46	5.30	5.75	5.68	5.85				

Fuente: Los datos e información descriptiva están basados en material de "Do robust estimators work with real data?" de Stephen M. Stigler, *Annals of Statistics, 5* (1977), 1055-1098.

a. ¿Qué evidencia tiene usted de que la suposición de normalidad es razonable? Explique.

b. La media de los datos de Cavendish, ¿es significativamente menor que el estándar reconocido hoy día? Use un nivel de significación de 0.05.

9.58 Use computadora o calculadora para completar los cálculos y la prueba de hipótesis para este ejercicio. Delco Products, división de General Motors, produce conmutadores diseñados para medir 18.810 mm en longitud total. (Un conmutador es un aparato empleado en el sistema eléctrico de un automóvil.) Los datos siguientes son las longitudes de una muestra de 35 conmutadores tomados al observar el proceso de manufactura:

18.802	18.810	18.780	18.757	18.824	18.827	18.825
18.809	18.794	18.787	18.844	18.824	18.829	18.817
18.785	18.747	18.802	18.826	18.810	18.802	18.780
18.830	18.874	18.836	18.758	18.813	18.844	18.861
18.824	18.835	18.794	18.853	18.823	18.863	18.808

Fuente: Con permiso de Delco Products Division, GMC

¿Hay suficiente evidencia para rechazar el dicho de que estas piezas satisfacen los requisitos de diseño "la longitud media es 18.810" al nivel de significación de $\alpha = 0.01$?

9.59 El acetaminofeno es un ingrediente activo que se halla en más de 600 medicamentos que se venden con receta y en mostrador, por ejemplo supresores de dolor o tos y medicinas para el resfrío. Es seguro y eficaz cuando se usa correctamente, pero tomar demasiado puede llegar a dañar el hígado.

Fuente: http://www.keetpkidshealthy.com/medicine_cabinet/acetaminophen.html

Un investigador piensa que la cantidad de acetaminofeno, en una marca particular de pastillas para el resfrío, contiene una cantidad media de acetaminofeno por pastilla que es diferente de los 600 mg que afirma el fabricante. Una muestra aleatoria de 30 pastillas tuvo un contenido medio de acetaminofeno de 596.3 mg con una desviación estándar de 4.7 mg.

a. ¿Es razonable la suposición de normalidad? Explique.

b. Construya un intervalo de confianza de 99% para la estimación del contenido medio de acetaminofeno.

c. ¿Qué sugiere el intervalo de confianza hallado en la parte b acerca del contenido medio de acetaminofeno de una pastilla? ¿Piensa usted que hay 600 mg por pastilla? Explique.

9.60 Un productor de vinos ha colocado un gran pedido de los corchos No. 9 descritos en el caso práctico 6.15 (p. 330) y está preocupado por el número de corchos que puedan tener diámetros menores. Durante el proceso de colocar el tapón, los corchos son comprimidos hasta llegar a 16 o 17 mm de diámetro para insertarlos en botellas con abertura de 18 mm. El corcho se dilata entonces para hacer el sello. El productor de

vinos desea que los corchos sean tan apretados como sea posible y, por tanto, está preocupado por cualquiera que pueda tener una medida más pequeña. El diámetro de cada corcho se mide en varios lugares, y se informa de un diámetro promedio por cada corcho. El productor de corchos ha asegurado al productor de vinos que cada corcho tiene un diámetro promedio que está dentro de las especificaciones, y que todos los diámetros promedio tienen una distribución normal con una media de 24.0 mm.

a. ¿Por qué tiene sentido que al diámetro del corcho se le asigne el promedio de varias mediciones de diámetro diferentes?

Se toma una muestra aleatoria de 18 corchos del lote a embarcar y se obtienen los diámetros (en milímetros):

23.93	23.91	23.82	24.02	23.93	24.17	23.93	23.84	24.13
24.01	23.83	23.74	23.73	24.10	23.86	23.90	24.32	23.83

b. La especificación del diámetro promedio es "24 mm + 0.6 mm/−0.4 mm". ¿Parece que este pedido satisface las especificaciones con base en un corcho individual? Explique.

c. La muestra de la parte a, ¿muestra razón suficiente para dudar de la verdad de la afirmación, de que el promedio del diámetro medio es 24.0 mm, a un nivel de significación de 0.02?

Se selecciona al azar una muestra diferente de 18 corchos y se obtienen los diámetros (en milímetros):

23.90	23.98	24.28	24.22	24.07	23.87	24.05	24.06	23.82
24.03	23.87	24.08	23.98	24.21	24.08	24.06	23.87	23.95

d. La muestra precedente, ¿muestra razón suficiente para dudar de la verdad de la afirmación, de que el promedio del diámetro medio es 24.0 mm, al nivel de significación de 0.02?

e. ¿Qué efecto tuvieron las dos medias muestrales diferentes sobre la estadística de prueba calculada en las partes c y d? Explique.

f. ¿Qué efecto tuvieron las dos desviaciones estándar muestrales diferentes sobre la estadística de prueba calculada en las partes c y d? Explique.

9.61 La longitud no es muy importante al evaluar la calidad de corchos porque tiene poco qué hacer con la efectividad de un corcho para preservar el vino. Los productores de vinos tienen varias longitudes a escoger y ordenan la longitud de corcho que prefieren (los corchos largos tienden a hacer un taponazo más fuerte cuando se descorcha la botella). No obstante, la longitud se vigila cuidadosamente porque es una cualidad especificada del corcho. Las longitudes de corchos naturales No. 9 (24 mm de diámetro por 45 mm de largo) tienen una distribución normal. Se midieron 12 corchos seleccionados al azar, al centésimo de milímetro más cercano.

44.95	44.95	44.80	44.93	45.22	44.82
45.12	44.62	45.17	44.60	44.60	44.75

a. La muestra precedente, ¿da suficiente razón para demostrar que la longitud media es diferente de 45.0 mm, al nivel de significación de 0.02?

Se toma una muestra aleatoria diferente de 18 corchos del mismo lote.

45.17	45.02	45.30	45.14	45.35	45.50	45.26	44.88	44.71
44.07	45.10	45.01	44.83	45.13	44.69	44.89	45.15	45.13

b. La muestra precedente, ¿da suficiente razón para demostrar que la longitud media es diferente de 45.0 mm, al nivel de significación de 0.02?

c. ¿Qué efecto tuvieron las dos medias muestrales diferentes sobre la estadística de prueba calculada en las partes a y b? Explique.

d. ¿Qué efecto tuvieron los tamaños muestrales diferentes sobre la estadística de prueba calculada en las partes a y b? Explique.

e. ¿Qué efecto tuvieron las dos desviaciones estándar muestrales diferentes sobre la estadística de prueba calculada en las partes a y b? Explique.

9.62 ¿Qué tan importante es la suposición, "la población muestreada está normalmente distribuida", para el uso de la distribución t de Student? Usando computadora, simule sacar 100 muestras de tamaño 10 de cada uno de tres tipos diferentes de distribuciones poblacionales, es decir, uno normal, uno uniforme y uno exponencial. Primero genere 1000 valores de datos de la población y construya un histograma para ver qué aspecto tiene la población. Luego genere 100 muestras de tamaño 10 de la misma población; cada fila representa una muestra. Calcule la media y desviación estándar para cada una de las 100 muestras. Calcule $t\star$ para cada una de las 100 muestras. Construya histogramas de las 100 medias muestrales y los 100 valores. $t\star$

Para las muestras de la población normal:

a. ¿La distribución $\bar{x}$ parece ser normal? Encuentre porcentajes para intervalos y compare con la distribución normal.

b. ¿La distribución de $t\bigstar$ parece tener una distribución t con df = 9? Encuentre porcentajes para intervalos y compárelos con la distribución t.

Para las muestras de la población rectangular o uniforme:

c. ¿La distribución $\bar{x}$ parece ser normal? Encuentre porcentajes para intervalos y compare con la distribución normal.

d. ¿La distribución de $t\bigstar$ parece tener una distribución t con df = 9? Encuentre porcentajes para intervalos y compárelos con la distribución t.

Para las muestras de la población sesgada (exponencial):

e. ¿La distribución $\bar{x}$ parece ser normal? Encuentre porcentajes para intervalos y compare con la distribución normal.

f. ¿La distribución de $t\bigstar$ parece tener una distribución t con df = 9? Encuentre porcentajes para intervalos y compárelos con la distribución t.

En resumen:

g. En cada una de las tres situaciones precedentes, la distribución muestral para $\bar{x}$ parece ser ligeramente diferente a la distribución de $t\bigstar$. Explique por qué.

h. ¿La condición de normalidad parece ser necesaria para que la estadística de prueba calculada $t\bigstar$ tenga una distribución t de Student? Explique.

9.3 Inferencias acerca de los éxitos de la distribución binomial

Quizá la inferencia más común comprende el **parámetro binomial *p*,** la "probabilidad de éxito". Sí, todos usamos esta inferencia, aun cuando sea sólo por casualidad. En miles de situaciones estamos interesados en algo que "pasa" o "no pasa". Hay sólo dos posibles resultados de interés, y ésa es la propiedad fundamental de un **experimento binomial.** El otro ingrediente necesario es el de intentos independientes múltiples. Preguntar a cinco personas si están "a favor" o "en contra" de algún problema que puede crear cinco intentos independientes; si a 200 personas se les hace la misma pregunta, pueden aparecer 200 intentos independientes; si se inspeccionan 30 artículos para ver si cada uno de ellos "exhibe una propiedad particular" o "no", habrá 30 intentos repetidos; éstas son las cualidades de una inferencia binomial.

El parámetro binomial p se define como la probabilidad de éxito en un solo intento en un experimento binomial.

Probabilidad binomial muestral

$$p' = \frac{x}{n} \tag{9.3}$$

donde la **variable aleatoria *x*** representa el número de éxitos que ocurre en una muestra formada por n intentos.

PARA SU INFORMACIÓN

En las páginas 287-290 se pueden hallar detalles completos acerca de experimentación binomial.

Recuerde que la media y desviación estándar de la variable aleatoria binomial x se encuentran con el uso de la fórmula (5.7), $\mu = np$, y la fórmula (5.8), $\sigma = \sqrt{npq}$, donde $q = 1 - p$. Se considera que la distribución de x es aproximadamente nor-

mal si n es mayor a 20 y si np y nq son ambos mayores a 5. Esta comúnmente aceptada *regla práctica* nos permite usar la **distribución normal estándar** para estimar probabilidades para la variable aleatoria binomial x, el número de éxitos en n intentos, y para hacer inferencias respecto al parámetro binomial p, la probabilidad de éxito en un intento individual.

En general, es más fácil y más significativo trabajar con la distribución de p' (la probabilidad observada de ocurrir) que con x (el número de sucesos). En consecuencia, convertiremos las fórmulas (5.7) y (5.8) de unidades de x (enteros) a unidades de proporciones (porcentajes expresados como decimales) al dividir cada fórmula entre n, como se ve en la tabla 9.1.

TABLA 9.1

Fórmulas (9.4) y (9.5)

Variable	Media	Desviación estándar
x	$\mu_x = np$ **(5.7)**	$\sigma_x = \sqrt{npq}$ **(5.8)**
para cambiar x a p', divida entre n		
$\frac{x}{n}$	$\frac{np}{n}$	$\frac{\sqrt{npq}}{n}$
p'	$\mu_{p'} = p$ **(9.4)**	$\sigma_{p'} = \sqrt{\frac{pq}{n}}$ **(9.5)**

Recuerde que $\mu_{p'} = p$ y que la *estadística muestral* p' es un **estimador insesgado para *p*.** Por tanto, la información acerca de la distribución muestral de p se resume como sigue:

Si se selecciona una muestra aleatoria de tamaño n de una población grande con $p = P$(éxito), entonces la distribución muestral de p' tiene:

1. Una media $\mu_{p'}$ igual a p
2. Un error estándar $\sigma_{p'}$ igual a $\sqrt{\frac{pq}{n}}$
3. Una distribución aproximadamente normal si n es lo suficiente grande

En la práctica, el uso de estos lineamientos asegura normalidad:

1. El tamaño muestral es mayor a 20.
2. Los productos np y nq son ambos mayores a 5.
3. La muestra está formada por menos de 10% de la población.

Ahora estamos listos para hacer inferencias acerca del parámetro poblacional p. El uso de la distribución z comprende una suposición.

Las suposiciones para inferencias acerca del parámetro binomial *p*: las n observaciones aleatorias que forman la muestra se seleccionan independientemente de una población que no está cambiando durante el muestreo.

Procedimiento de intervalo de confianza

PARA SU INFORMACIÓN
La desviación estándar de una distribución muestral se denomina "error estándar".

Las inferencias respecto al parámetro binomial poblacional p, P(éxito), se hacen usando procedimientos que cercanamente se comparan con los procedimientos de inferencia empleados para la media poblacional. Cuando estimamos la **proporción poblacional *p*,** basaremos nuestras estimaciones sobre el **estimador insesgado *p'***. La estimación puntual, la estadística austral p', se convierte en el centro del intervalo de confianza, y el error máximo de estimación es un múltiplo del **error estándar.** El **nivel de confianza** determina el coeficiente de confianza, el número de múltiplos del error estándar.

Intervalo de confianza para una proporción

$$p' - z(\alpha/2)\left(\sqrt{\frac{p'q'}{n}}\right) \quad \text{a} \quad p' + z(\alpha/2)\left(\sqrt{\frac{p'q'}{n}}\right) \tag{9.6}$$

donde $p' = \dfrac{x}{n}$ y $q' = 1 - p'$

Nótese que el error estándar $\sqrt{\dfrac{pq}{n}}$ ha sido sustituido por $\sqrt{\dfrac{p'q'}{n}}$. Como estamos estimando p, no conocemos su valor y por tanto debemos usar el mejor sustituto disponible. Ese sustituto es p', el valor observado o la estimación puntual para p. Este sustituto producirá un pequeño cambio en el error estándar o en el ancho de nuestro intervalo de confianza, siempre que n sea suficientemente grande.

El ejemplo 9.8 ilustra la formación de un intervalo de confianza para el parámetro binomial, p.

EJEMPLO 9.8

Intervalo de confianza para *p*

En una discusión acerca de los autos que manejan compañeros estudiantes, se hicieron varias aseveraciones acerca de tipos, edades, marcas, colores, etcétera. Dana decidió que él quería estimar la proporción de autos convertibles de estudiantes, de modo que al azar identificó 200 autos en el estacionamiento para estudiantes y encontró que 17 eran convertibles. Encuentre el intervalo de confianza de 90% para la proporción de autos convertibles que manejan estudiantes.

SOLUCIÓN

PARA SU INFORMACIÓN
El procedimiento de intervalo de confianza de cinco pasos se da en la página 404.

Paso 1 **El inicio:**
Describir el parámetro poblacional de interés.
p, la proporción (porcentaje) de autos convertibles de estudiantes.

Paso 2 **Criterios de intervalo de confianza:**

a. **Comprobar las suposiciones.**
La muestra se seleccionó al azar, y la respuesta de cada estudiante es independiente de las de los otros encuestados.

b. **Identificar la distribución de probabilidad y la fórmula a usar.**
La distribución normal estándar se usará con la fórmula (9.6) como la estadística de prueba. Se espera que p' sea aproximadamente normal porque:
(1) $n = 200$ es mayor a 20, y

(2) np [aproximadamente por $np' = 200(17/200) = 17$] y nq [aproximadamente por $nq' = 200(183/200) = 183$] son mayores a 5.

c. Expresar el nivel de confianza: $1 - \alpha = 0.90$.

Paso 3 **Evidencia muestral:**

Recolectar la información muestral.

$n = 200$ autos se identificaron, y $x = 17$ fueron convertibles:

$$p' = \frac{x}{n} = \frac{17}{200} = 0.085$$

PARA SU INFORMACIÓN

Las instrucciones específicas se indican en las páginas 403-405.

Paso 4 **Intervalo de confianza:**

a. Determinar el coeficiente de confianza.

Éste es el puntaje z [$z(\alpha/2)$, "z de la mitad de alfa"] que identifica el número de errores estándar necesarios para alcanzar el nivel de confianza y se encuentra usando la tabla 4 del apéndice B; $z(\alpha/2) = z(0.05) = 1.65$ (vea el diagrama).

b. Encontrar el error máximo de estimación.

Use la parte de error máximo de la fórmula (9.6):

$$E = z(\alpha/2)\left(\sqrt{\frac{p'q'}{n}}\right) = 1.65\left(\sqrt{\frac{(0.085)(0.915)}{200}}\right)$$
$$= (1.65)\sqrt{0.000389} = (1.65)(0.020) = \mathbf{0.033}$$

c. Encontrar los límites inferior y superior de confianza.

$$p' - E \quad \text{a} \quad p' + E$$
$$0.085 - 0.033 \quad \text{a} \quad 0.085 + 0.033$$
$$0.052 \quad \text{a} \quad 0.118$$

Paso 5 **Resultados:**

Expresar el intervalo de confianza.

0.052 a 0.118 es el intervalo de confianza de 90% para $p = P$(maneja convertible).

Esto es, la verdadera proporción de estudiantes que manejan convertibles es entre 0.052 y 0.118, con una confianza de 90%.

INSTRUCCIONES DE TECNOLOGÍA: INTERVALO DE CONFIANZA DE $1 - \alpha$ PARA UNA PROPORCIÓN p

MINITAB (Versión 14)

```
Choose:     Stat > Basic Statistics > 1 Proportion
Select:     Summarized Data
Enter:      Number of trials: n
            Number of events: x
Select:     Options
Enter:      Conidence level: 1 − α (ex. 95.0)
Select:     Alternative: not equal
            Use test and interval based on normal distribution. > OK > OK
```

Excel

Introduzca los datos en la columna A usando números 0 por fracasos (o no) y números 1 por éxitos (o sí); luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Z-Estimate: Proportion
Enter:      Input Range: (A2:A20 or select cells) > OK
            Code for success: 1
            Alpha: α (ex. 0.05) > OK
```

TI-83/84 Plus

```
Choose:     STAT > TESTS > A:1-PropZint
            Enter the appropriate values
            and highlight Calculate.
```

CASO PRÁCTICO 9.9 Mito y realidad al informar error de muestreo

En casi toda ocasión cuando se realiza una nueva encuesta, alguien de los medios de comunicación preguntará "¿Cuál es el margen de error para esta encuesta? Cuando los medios imprimen frases como "el margen de error es más o menos tres puntos porcentuales", dan una fuerte sugerencia de que los resultados son precisos y quedan dentro del porcentaje indicado; pretenden advertir al público sobre el error de muestreo, pero estarían mejor suponiendo que todas las encuestas y todos los sondeos de opiniones son estimaciones, que pueden estar equivocados.

En la práctica, "error aleatorio de muestreo", o probabilidad de que una muestra pura de probabilidad produzca respuestas dentro de cierta banda de porcentajes sólo debido al tamaño muestral, es uno de los más pequeños problemas de medición.

Por esta razón, incluimos (Harris) una fuerte advertencia en todas las encuestas que publicamos. Por lo general, es como sigue: en teoría, con una muestra de este tamaño, se puede decir con 95% de certidumbre que los resultados tienen una precisión estadística de más o menos _ puntos porcentuales de lo que sería si toda la población de adultos se hubiera encuestado con precisión completa. Desafortunadamente, hay otras posibles fuentes de error en todas las encuestas o sondeos que es probable sean más graves que los cálculos teóricos de error de muestreo. Incluyen rechazo a ser entrevistado (no hay respuesta), redacción de la pregunta y orden de preguntas, sesgo de quien hace la entrevista, dar valor por datos demográficos de control e investigación. Es difícil o imposible cuantificar los errores que puedan resultar de estos factores.

Si los periodistas tienen un mínimo interés en todo esto podrían preguntar, "Si hay tantas fuentes de error en encuestas, ¿por qué debemos molestarnos en leer o informar de los resultados de cualquier encuesta?". A lo cual normalmente doy dos respuestas:

1. Las encuestas bien diseñadas y bien dirigidas funcionan. Su registro general es muy bueno. La mayoría de investigadores sociales y de marketing estarían felices con el promedio de errores de pronóstico de las encuestas. No obstante, hay suficientes desastres en la historia de predicciones de elecciones para que los lectores tengan cuidado al interpretar los resultados.
2. (Y esto es más eficaz.) Expreso con otras palabras las famosas observaciones de Winston Churchill acerca de la democracia y digo, "Las encuestas son la peor forma de medir la opinión pública y el comportamiento del público, o de pronosticar elecciones —excepto todas las otras formas".

Fuente: The Polling Report, 4 de mayo, 1998, por Humprey Taylor, Presidente, Louis Harris & Assoc., Inc., http://www.pollinreport.com/sampling.htm

Determinación del tamaño muestral

Con el uso de la parte de máximo error de la fórmula del intervalo de confianza, es posible determinar el **tamaño de la muestra** que debe tomarse para estimar p con la precisión deseada. A continuación está la fórmula para el **máximo error de estimación para una proporción:**

$$E = z(\alpha/2)\left(\sqrt{\frac{pq}{n}}\right) \tag{9.7}$$

Para determinar el tamaño muestral con esta fórmula, debemos decidir sobre la calidad que deseamos para nuestro intervalo final de confianza. Esta calidad se mide en dos formas: el nivel de confianza y la precisión (angostura) del intervalo. El nivel de confianza que establecemos determina a su vez el coeficiente de confianza, $z(\alpha/2)$. La precisión deseada determinará el máximo error de estimación, E. (Recuerde que estamos estimando p, la probabilidad binomial; por tanto, E por lo general se expresa en centésimos.)

Para facilidad de uso, de la fórmula (9.7) podemos despejar n como sigue:

Tamaño muestral para un intervalo de confianza 1 $-\alpha$ de p

$$n = \frac{[z(\alpha/2)]^2 \cdot p^\star \cdot q^\star}{E^2} \tag{9.8}$$

donde $p^\star$ y $q^\star$ son valores provisionales de p y q empleados para planificación

PARA SU INFORMACIÓN
Recuerde que $q = 1 - p$.

Al inspeccionar la fórmula (9.8), podemos observar que tres componentes determinan el tamaño muestral:

1. El nivel de confianza [$1 - \alpha$, que determina el coeficiente de confianza, $z(\alpha/2)$]
2. El valor provisional de p (p^* determina el valor de $q^\star$)
3. El error máximo, E

Un aumento o disminución de uno de estos tres componentes afecta el tamaño muestral. Si el nivel de confianza se aumenta o disminuye (mientras los otros componentes se conservan constantes), entonces el tamaño muestral aumentará o disminuirá, respectivamente. Si el producto de p^* y q^* se aumenta o disminuye (con otros componentes conservados constantes), entonces el tamaño muestral aumentará o disminuirá, respectivamente. (El producto $p^* \cdot q^*$ es máximo cuando $p^* = 05$ y disminuye a medida que el valor de p^* se aleja de 0.5.) Un aumento o disminución en el error máximo deseado tendrá el efecto opuesto en el tamaño muestral, puesto que E aparece en el denominador de la fórmula. Si no hay valores provisionales para p y q, entonces use $p^* = 0.5$ y $q^* = 0.5$. Usando $p^* = 0.5$ es seguro porque da el máximo tamaño muestral de cualquier valor posible de p. Usando $p^* = 0.5$ funciona razonablemente bien cuando el verdadero valor es "cercano a 0.5" (por ejemplo, entre 0.3 y 0.7); no obstante, cuando p se acerca ya sea a 0 o a 1, ocurrirá una estimación excesiva en tamaño muestral.

EJEMPLO 9.10 Tamaño muestral para estimar *p* (sin información previa)

Determine el tamaño muestral que se requiere para estimar la verdadera proporción de estudiantes universitarios de la comunidad, que tengan ojos azules, si desea hacer una estimación con límite no mayor a 0.02 con 90% de confianza.

SOLUCIÓN

PASO 1 El nivel de confianza es $1 - \alpha = 0.90$; por tanto, el coeficiente de confianza es $z(\alpha/2) = z(0.05) = 1.65$ de la tabla 4 del apéndice B; vea el diagrama.

PASO 2 El máximo error deseado es $E = 0.02$.

PARA SU INFORMACIÓN
Cuando encuentre el tamaño muestral *n*, siempre haga redondeo al siguiente entero más grande, sin importar lo pequeño del decimal.

PASO 3 No se dio estimación para p, de modo que use $p^* = 0.5$ y $q^* = 1 - p^* = 0.5$.

PASO 4 Use la fórmula (9.8) para hallar n:

$$n = \frac{[z(\alpha/2)]^2 \cdot p^* \cdot q^*}{E^2}: \quad n = \frac{(1.65)^2 \cdot 0.5 \cdot 0.5}{(0.02)^2} = \frac{0.680625}{0.0004} = 1701.56 = \mathbf{1702}$$

EJEMPLO 9.11 Tamaño muestral para estimar *p* (con información previa)

Un fabricante de automóviles compra tornillos de un proveedor que dice que sus tornillos son aproximadamente 5% defectuosos. Determine el tamaño muestral que será necesario para estimar la verdadera proporción de tornillos defectuosos, si deseamos que nuestra estimación se encuentre dentro de $\pm$ 0.02 con 90% de confianza.

SOLUCIÓN

PASO 1 El nivel de confianza es $1 - \alpha = 0.90$; el coeficiente de confianza es $z(\alpha/2) = z(0.05) = 1.65$.

PASO 2 El máximo error deseado es $E = 0.02$.

PASO 3 Hay una estimación para p(el proveedor dice "5% defectuosos"), de modo que use $p^* = 0.05$ y $q^* = 1 - p^* = 0.95$.

PARA SU INFORMACIÓN
Sí, los cálculos del tamaño muestral se redondean siempre al siguiente entero más grande.

PASO 4 Use la fórmula (9.8) para hallar n:

$$n = \frac{[z(\alpha/2)]^2 \cdot p^* \cdot q^*}{E^2}: \quad n = \frac{(1.65)^2 \cdot 0.05 \cdot 0.95}{(0.02)^2} = \frac{0.12931875}{0.0004} = 323.3 = \mathbf{324}$$

Observe la diferencia en tamaños muestrales requerida en los ejemplos 9.10 y 9.11. La única diferencia matemática entre los problemas es el valor empleado para p^*. En el ejemplo 9.10 usamos $p^* = 0.5$, y en el ejemplo 9.11 usamos $p^* = 0.05$. Recuerde que el uso del valor provisional $p^* = 0.5$ da el máximo tamaño muestral. Como se puede ver, será una ventaja tener alguna indicación del valor esperado de p, en especial cuando p se aleja cada vez más de 0.5.

Procedimiento de prueba de hipótesis

Cuando se ha de probar el parámetro binomial p usando un procedimiento de prueba de hipótesis, usaremos una estadística de prueba que represente la diferencia entre la proporción observada y la proporción hipotética, dividida entre el error

estándar. Esta estadística de prueba se supone normalmente distribuida cuando la hipótesis nula es verdadera, cuando las suposiciones para la prueba hayan sido satisfechas, y cuando n sea suficientemente grande ($n > 20$, $np > 5$, y $nq > 5$).

PARA SU INFORMACIÓN
p' es de la muestra, p es de H_0, y $q = 1 - p$.

Estadística de prueba para una proporción

$$z\bigstar = \frac{p' - p}{\sqrt{\dfrac{pq}{n}}} \quad \text{con } p' = \frac{x}{n} \tag{9.9}$$

EJEMPLO 9.12 Prueba de hipótesis de una cola para la proporción *p*

Numerosas personas duermen hasta tarde en fines de semana para recuperarse de las "noches cortas" durante los días hábiles de la semana. El consejo de Mejor Sueño informa que 61% de nosotros duerme más de 7 horas por noche en el fin de semana. Una muestra aleatoria de 350 adultos encontró que 235 habían dormido más de 7 horas cada noche el último fin de semana. Al nivel de significación de 0.05, ¿esta evidencia demuestra que más de 61% duermen 7 horas o más por noche en un fin de semana?

SOLUCIÓN

Paso 1 El inicio:

a. Describir el parámetro poblacional de interés.
p, la proporción de adultos que duermen más de 7 horas por noche en fines de semana.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).
H_o: $p = P$(más de 7 horas de sueño) $= 0.61$ ($\leq$)(no más de 61%)
H_a: $p > 0.61$ (más de 61%)

Paso 2 Criterios de prueba de hipótesis:

a. Comprobar las suposiciones.
La muestra aleatoria de 350 adultos se encuestó independientemente.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
El tamaño normal z se usará con la fórmula (9.9). Como $n = 350$ es mayor a 20 y como $np = (350)(0.61) = 213.5$ y $nq = (350)(0.39) = 136.5$ son mayores a 5, se espera que p' se encuentre distribuida normalmente en forma aproximada.

c. Determinar el nivel de significación: $\alpha = 0.05$.

Paso 3 Evidencia muestral:

a. Recolectar la información muestral: $n = 350$ y $x = 235$:

$$p' = \frac{x}{n} = \frac{235}{350} = 0.671$$

b. Calcular el valor de la estadística de prueba.
Use la fórmula (9.9):

$$z\bigstar = \frac{p' - p}{\sqrt{\dfrac{pq}{n}}}: \quad z\bigstar = \frac{0.671 - 0.61}{\sqrt{\dfrac{(0.61)(0.39)}{350}}} = \frac{0.061}{\sqrt{0.0006797}} = \frac{0.061}{0.0261} = 2.34$$

Paso 4 Distribución de probabilidad:

Usando el procedimiento del valor *p*:

a. Calcular el valor *p* para la estadística de prueba.

Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor que".

P = valor $p = P(z > 2.34)$ como se ve en la figura.

Para hallar el valor *p*, use uno de tres métodos:

1. Use la tabla 3 del apéndice B para calcular el valor *p*: **P** = 0.5000 − 0.4904 = **0.0096.**
2. Use la tabla 5 del apéndice B para poner límites al valor *p*: 0.0094 < **P** < 0.0107.
3. Use computadora o calculadora para calcular el valor *p*: **P** = 0.0096.

Para instrucciones específicas, vea el método 3 que sigue.

b. Determinar si el valor *p* es o no es menor que *α*.

El valor *p* es menor que *α*.

O

Usando el procedimiento clásico:

a. Determinar la región crítica y valor(es) crítico(s).

La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "más de". El valor crítico se obtiene de la tabla 4A: $z(0.05)$ = **1.65.**

Las instrucciones específicas para hallar valores críticos se dan en las páginas 450-451.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

z★ está en la región crítica, como se muestra en **área sombreada** en la figura de líneas antes.

Paso 5 Resultados:

a. Expresar la decisión acerca de H_o: Rechazar H_o.

b. Expresar la conclusión acerca de H_a.

Hay suficiente razón para concluir que la proporción de adultos en la población muestreada, que duermen más de 7 horas todas las noches o los fines de semana, es significativamente mayor al 61% al nivel de significación de 0.05.

Método 3: si está usted haciendo la prueba de hipótesis con ayuda de una computadora o calculadora es muy probable que calculen el valor *p*, o puede usar los comandos de distribución de probabilidad acumulativa descritos en las páginas 329-330.

EJEMPLO 9.13 Prueba de hipótesis de dos colas para proporción *p*

Al hablar de autos que manejan compañeros estudiantes (vea ejemplo 9.8, p. 498), Tom dijo que 15% de los estudiantes manejan convertibles. Jody encuentra que esto es difícil de creer y desea comprobar la validez de lo dicho por Tom usando la muestra aleatoria de Dana. A un nivel de significación de 0.10, ¿hay suficiente evidencia para rechazar lo dicho por Tom si había 17 convertibles en su muestra de 200 autos?

SOLUCIÓN

Paso 1 El inicio:

a. Describir el parámetro poblacional de interés.

$p = P$(estudiante maneja convertible).

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).
H_o: $p = 0.15$(15% manejan convertibles)
H_a: $p \neq 0.15$ (el porcentaje es diferente a 15%)

Paso 2 **Criterios de prueba de hipótesis:**

a. Comprobar las suposiciones.
La muestra fue seleccionada en forma aleatoria, y la respuesta de cada persona es independiente de otras respuestas.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
Se usarán la z normal estándar y la fórmula (9.9). Como $n = 200$ es mayor a 20 y tanto np como nq son mayores a 5, se espera que p' esté normalmente distribuida en forma aproximada.

c. Determinar el nivel de significación: $\alpha = 0.10$.

Paso 3 **Evidencia muestral:**

a. Recolectar la información muestral: $n = 200$ y $x = 17$:

$$p' = \frac{x}{n} = \frac{17}{200} = 0.085$$

b. Calcular el valor de la estadística de prueba.
Use la fórmula (9.9):

$$z\bigstar = \frac{p' - p}{\sqrt{\frac{pq}{n}}}: \quad z\bigstar = \frac{0.085 - 0.150}{\sqrt{\frac{(0.15)(0.85)}{200}}}$$

$$= \frac{-0.065}{\sqrt{0.00064}} = \frac{-0.065}{0.022525} = -2.57$$

Paso 4 **Distribución de probabilidad:**

Usando el procedimiento de valor *p*:

a. Calcular el valor *p* para la estadística de prueba.

Use ambas colas porque H_a expresa interés en valores relacionados con "diferente de".

$\mathbf{P}$ = valor $p = P(z < -2.57) + P(z > 2.57)$
$= 2 \times P(|z| > 2.57)$ como se muestra en la figura.

Para hallar el valor p, use uno de tres métodos:

1. Use la tabla 3 del apéndice B para calcular el valor p: $\mathbf{P} = 2 \times (0.5000 - 0.4949) = 0.0102$.
2. Use la tabla 5 del apéndice B para poner límites al valor p: $0.0094 < \mathbf{P} < 0.0108$.

O

Usando el procedimiento clásico:

a. Determine la región crítica y valor(es) crítico(s).

La región crítica es de dos colas porque H_a expresa interés en valores relacionados con "diferente de". El valor crítico se obtiene de la tabla 4B: $z(0.05) = 1.65$.

Para instrucciones específicas, vea las páginas 453-454.

3. Use computadora o calculadora para calcular el valor p: **P** = 0.0102.

Para instrucciones específicas, vea la página 432.

b. Determinar si el valor p es o no es menor que α.

El valor p es menor que α.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica

z★ está en la región crítica, como se muestra en **rojo** en la figura de líneas antes.

Paso 5 **Resultados:**

a. Expresar la decisión acerca de H_o: Rechazar H_o.

b. Expresar la conclusión acerca de H_a.

Hay suficiente evidencia para rechazar lo dicho por Tom y concluir que el porcentaje de estudiantes que manejan convertibles es diferente a 15%, al nivel de significación de 0.10.

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA UNA PROPORCIÓN *p*

MINITAB (Versión 14)

```
Choose:     Stat > Basic Statistics > 1 Proportion
Select:     Summarized Data
Enter:      Number of trials: n
            Number of events: x
Select:     Options
Enter:      Test proportion: p
Select:     Alternative: less than or not equal or greater than
            Use test and interval based on normal distribution. > OK > OK
```

Excel

Introduzca los datos en la columna A usando ceros (0) por fracasos (o no) y unos (1) para éxitos (o sí); luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Z-Test: Proportion
Enter:      Input Range: (A2:A20 or select cells) > OK
            Code for success: 1
            Hypothesized Proportion: p
            Alpha: α (ex. 0.05)
Choose:     Alternative: less than or not equal or greater than > OK
            Gives p-values and critical values for both one-tailed and
            two-tailed tests.
```

Ti-83/84 Plus

```
Choose:     STAT > TESTS > 5:1-PropZTest
            Enter the appropriate values and
            highlight Calculate.
```

Relación entre intervalos de confianza y pruebas de hipótesis

Hay una relación entre intervalos de confianza y pruebas de hipótesis de dos colas, cuando el nivel de confianza y el nivel de significación suman 1. Los coeficientes de confianza y los valores críticos son lo mismo, lo cual significa que el ancho del

intervalo de confianza y el ancho de la región no crítica son lo mismo. La estimación puntual es el centro del intervalo de confianza, y la media hipotética es el centro de la región no crítica. Por tanto, si el valor hipotético de p está contenido en el intervalo de confianza, entonces la estadística de prueba estará en la región no crítica (vea la figura 9.5).

FIGURA 9.5
El intervalo de confianza contiene p

Además, si la probabilidad hipotética p no cae dentro del intervalo de confianza, entonces la estadística de prueba estará en la región crítica (vea la figura 9.6).

FIGURA 9.6
El intervalo de confianza no contiene p

Esta comparación debe usarse sólo cuando la prueba de hipótesis sea de dos colas y cuando el mismo valor de α se use en ambos procedimientos.

CASO PRÁCTICO 9.14 ¿Cara o cruz?

CARA, BÉLGICA GANA —Y GANA

Memorándum a todos los equipos que juegan en Bélgica en la Copa Mundial este año: "No los dejen jugar sus propias monedas para el tiro".

Los matemáticos dicen que es más probable que las monedas acuñadas en el corazón administrativo de la Unión Europea caigan con la cara hacia arriba que hacia abajo. Mientras los billetes que empezaron a circular en los 12 miembros de la Unión Europea el 1 de enero son todos iguales, las monedas muestran símbolos nacionales en un lado y un mapa de Europa en el otro. El rey Albert, que vemos en monedas de Bélgica, parece ser de peso ligero, según los matemáticos polacos Tomasz Gliszczynski y Waclaw Zawadowski. Los dos profesores y sus estudiantes de la Academia Podlaska en Siedlce lanzaron al aire 250 veces una moneda de un euro, y encontraron que quedó cara hacia arriba en 140 ocasiones.

"El euro está acuñado de manera asimétrica" dijo el profesor Gliszczynski, que da clases de estadística, al periódico alemán *Die Welt*. La cara de la moneda fueron "pura suerte" dijo ayer el matemático polaco. "Cuando se hicieron las monedas, se acuñaron exactamente en la misma forma en todos los lados y el metal estaba distribuido de manera homogénea", dijo Romain Coenen. "No he oído de problema alguno con las monedas." Pero una variación del experimento en la oficina del Guardián sugirió que los matemáticos polacos pueden tener razón. Cuando se lanzó al aire 250

veces, la moneda de un euro quedó cara arriba 139 veces y 111 de cruz. "Me parece muy sospechoso", dijo Barry Blight, conferencista de estadísticas en la Escuela de Economía de Londres. "Si la moneda no estuviera 'cargada', la probabilidad de obtener un resultado tan extremo sería menos del 7%."

[Vea el Ejer. 9.111]

Fuente: Charlotte Denny y Sarah Dennis, *The Guardian*, viernes 4 de enero, 2002, http://www.guardian.co.uk/euro/story/0,11306,627496,00.html

SECCIÓN 9.3 EJERCICIOS

9.63 De los 150 elementos de una muestra aleatoria, 45 se clasificaron como "éxito".

a. Explique por qué a x y a n se les asignaron los valores 45 y 150, respectivamente.

b. Determine el valor de p'. Explique la forma en que se encuentra p' y el significado de p'.

Para cada una de las siguientes situaciones, encuentre p'.

c. $x = 24$ y $n = 250$

d. $x = 640$ y $n = 2050$

e. 892 de 1280 respondieron "sí"

9.64 a. ¿Cuál es la relación entre $p = P$(éxito) y $q = P$(fracaso)? Explique.

b. Explique por qué la relación entre p y q puede ser expresada por la fórmula $q = 1 - p$.

c. Si $p = 0.6$, ¿cuál es el valor de q?

d. Si el valor de $q' = 0.273$, ¿cuál es el valor de p'?

9.65 a. ¿Parece razonable que la media de la distribución muestral de valores observados de p' deba ser p, la verdadera proporción? Explique.

b. Explique por qué p es un estimador insesgado para la población p.

9.66 Demuestre que $\frac{\sqrt{npq}}{n}$ se simplifica a $\sqrt{\frac{pq}{n}}$.

9.67 Encuentre α, el área de una cola, y los coeficientes de confianza de z que se usan con cada uno de los siguientes niveles de confianza.

a. $1 - \alpha = 0.90$ b. $1 - \alpha = 0.95$ c. $1 - \alpha = 0.99$

9.68 Encuentre α, el área de una cola, y los coeficientes de confianza de z que se usan con cada uno de los siguientes niveles de confianza.

a. $1 - \alpha = 0.80$ b. $1 - \alpha = 0.98$ c. $1 - \alpha = 0.75$

9.69 Consulte de nuevo el ejemplo 9.8. Se toma otra muestra para estimar la proporción de estudiantes que manejan convertibles. Los resultados son $n = 400$ y $x = 92$. Encuentre:

a. La estimación para el error estándar

b. El intervalo de confianza de 95%

9.70 "Usted dice tomate rojo; a quienes les gustan las hamburguesas dicen ketchup." Según un sondeo aleatorio reciente hecho por los restaurantes T.G.I. Friday's a 1027 norteamericanos, alrededor de la mitad (47%) dicen que la ketchup es su condimento preferido para hamburguesas. El sondeo mencionó un margen de error de más o menos 3.1%.

Fuente: Harris Interactive/Yankelovich Partners para restaurants T.G.I. Friday's, http://www.knoxville3.com/fridays/News/burger.htm

a. Describa cómo es que este sondeo de 1027 norteamericanos se ajusta a las propiedades de un experimento binomial. Específicamente identifique n, un intento, éxito, p, y x.

b. ¿Cuál es la estimación puntual para la proporción de todos los norteamericanos que prefieren ketchup en su hamburguesa? ¿Es un parámetro o una estadística?

c. Calcule el máximo error de confianza de 95% de estimación para un experimento binomial de 1027 intentos que resultan en una proporción observada de 0.47.

d. ¿Cómo está relacionado el error máximo, hallado en la parte c, con el margen de error de 3.1% citado en el informe del sondeo?

e. Encuentre el intervalo de confianza de 95% para la verdadera proporción p basada en un experimento binomial de 1027 intentos que resultan en una proporción observada de 0.47.

9.71 Aun cuan la mayoría de las personas están conscientes de síntomas menores de deshidratación como piel reseca y dolor de cabeza, muchos están menos informados de las causas de la deshidratación. Según una encuesta hecha por el Nutrition Information Center, los resultados de una muestra aleatoria de 3003 adultos norteamericanos mostró que 20% no sabían que la cafeína deshidrata. El sondeo cita un margen de error de más o menos 1.8%.

Fuente: Yankelovich Partners para el Nutrition Information Center del centro médico Cornell-New York Hospital y la asociación internacional de aguas embotelladas

a. Describa cómo es que este sondeo de 3003 norteamericanos se ajusta a las propiedades de un experimento binomial. Específicamente identifique n, un intento, éxito, p, y x.

b. ¿Cuál es la estimación puntual para la proporción de todos los norteamericanos que la cafeína deshidrata? ¿Es un parámetro o una estadística?

c. Calcule el máximo error de confianza de 95% de estimación para un experimento binomial de 3003 intentos que resultan en una proporción observada de 0.20.

d. ¿Cómo está relacionado el error máximo, hallado en la parte c, con el margen de error de 3.1% citado en el informe del sondeo?

e. Encuentre el intervalo de confianza de 95% para la verdadera proporción p basada en un experimento binomial de 3003 intentos que resultan en una proporción observada de 0.20.

9.72 Un banco seleccionó al azar 250 clientes de cuentas de cheques y encontró que 110 de ellos también tenían cuentas de ahorras en este mismo banco. Construya un intervalo de confianza de 95% para la verdadera proporción de clientes de cuentas de cheques que también tienen cuentas de ahorros.

9.73 En una muestra de 60 estudiantes seleccionados al azar, sólo 22 estuvieron a favor de la cantidad que se presupuestó para deportes intramuros y entre escuelas del año próximo. Construya el intervalo de confianza de 99% para la proporción de todos los estudiantes que apoyan la cantidad propuesta de presupuesto.

9.74 El 21 de diciembre de 2004, el artículo del *USA Today* "Hacia un juego problema: encuesta nacional muestra que casinos, máquinas tragamonedas y loterías atraen a jóvenes hacia hábitos potencialmente adictivos" informó de una encuesta hecha a 200 adolescentes hombres y 199 mujeres, de entre 14 y 17 años de edad. Los resultados publicados fueron que 66% de los hombres "casi nunca habían jugado" y que 37% de las mujeres "casi nunca habían jugado". La encuesta fue financiada por el 2003 Annenberg National Risk Survey of Youth.

a. Encuentre la estimación puntual, el máximo error de estimación, y el intervalo de confianza de 95% que resulta de un experimento binomial con $n = 200$ y proporción observada de éxito de 0.66.

b. Encuentre la estimación puntual, el máximo error de estimación y el intervalo de confianza de 95% que resulta de un experimento binomial con $n = 199$ y proporción observada de éxito de 0.37.

c. Compare los intervalos de confianza resultante y haga un enunciado con base en los dos intervalos que comparen hábitos de juego de adolescentes hombres y mujeres.

9.75 De los 1742 gerentes y profesionales encuestados por Management Recruiters International, en mayo de 2002, 27.8% trabajan hasta tarde 5 días a la semana en promedio. Usando un intervalo de confianza de 99% para la verdadera proporción binomial, con base en una muestra aleatoria de 1742 intentos binomiales y una proporción observada de 0.278, estime la proporción de gerentes y profesionales que trabajan hasta tarde 5 días a la semana.

9.76 Una investigación reciente sugiere que es frecuente que los niños aprendan copiando habilidades de amigos imaginarios. En un estudio de 100 niños hecho en la Universidad de Oregon y publicados en *Developmental Psychology* (*USA Today,* 20 de diciembre

de 2004, "Amigos imaginarios, beneficios reales"), 33% de los niños todavía tenían un amigo imaginario a la edad de 7. Encuentre el intervalo de confianza de 90% para la verdadera proporción de "éxito" para un experimento binomial con $n = 100$ y una proporción observada de 0.33.

9.77 Un sondeo del índice Cambridge de crédito del consumidor, realizado por teléfono a 1000 personas en todo el país, encontró que la mayoría de norteamericanos no se dejan influir fácilmente por el atractivo de puntos de recompensa o rebajas, cuando deciden usar una tarjeta de crédito o pagar al contado o con cheque. El sondeo encontró que 2 de cada 3 consumidores ni siquiera tienen tarjetas de crédito que ofrezcan puntos de recompensa o rebajas. Explique por qué estaría usted reacio a usar esta información para construir un intervalo de confianza para estimar la verdadera proporción de consumidores que no tienen tarjetas de crédito que ofrezcan puntos de recompensa o rebajas.

9.78 Construya intervalos de confianza de 90% para el parámetro binomial p para cada uno de los siguientes pares de valores. Escriba sus respuestas en la tabla.

	Proporción observada $p' = x/n$	Tamaño muestral	Límite inferior	Límite superior
a.	$p' = 0.3$	$n = 30$		
b.	$p' = 0.7$	$n = 30$		
c.	$p' = 0.5$	$n = 10$		
d.	$p' = 0.5$	$n = 100$		
e.	$p' = 0.5$	$n = 1000$		

f. Explique la relación entre las respuestas a las partes a y b.

g. Explique la relación entre las respuestas a las partes c-e.

9.79 A continuación aparecen descripciones de tres encuestas a nivel nacional.

Encuesta CNN/*USA Today*/Gallup, 21-23 de junio, 2002; $N = 1020$ adultos en todo el país; MoE ± 3. (MoE es margen de error.)

"¿Piensa usted que a los distritos escolares se les debe o no se les debe permitir revisar estudiantes de escuelas públicas, para ver si llevan drogas, antes que esos estudiantes puedan participar en actividades no atléticas?" Se les debe permitir: 70%, No se les debe permitir: 29%, Sin opinión: 1%

Encuesta CNN/*USA Today*/Gallup, 28-30 de junio, 2002; $N = 1019$ adultos en todo el país; MoE ± 3.

"Se ha propuesto que permitiría a personas poner una parte de sus impuestos de nómina al Seguro Social en cuentas personales de retiro que se invertirían en acciones y bonos. ¿Está usted a favor o en contra de esta propuesta?"

A favor: 57%, En contra: 39%, No: 4%

International Communications Research (ICR), diciembre de 2004; $N = 1021$ adultos; MoE ± 3.1.

El sondeo de la Asociación de Jabones y Detergentes (SDA) informó que 15% dijo que el factor más importante para seleccionar un producto de limpieza era la marca del producto.

Cada una de las encuestas está basada en aproximadamente 1020 adultos seleccionados al azar.

a. Calcule el máximo error de estimación de confianza de 95% para la verdadera proporción binomial basada en experimentos binomiales, con el mismo tamaño muestral y proporción observada como se indica primero en cada artículo.

b. Explique qué hizo variar los valores de los errores máximos.

c. El margen de error que se informa es por lo general el valor del máximo error redondeado al siguiente porcentaje de enteros más alto. ¿Los resultados de usted de la parte a verifican esto?

d. Explique ¿porqué la práctica de redondeo se considera "conservadora"?

e. ¿Qué valor de p debe usarse para calcular el error estándar si se desea el margen de error más conservador?

9.80 a. Si x éxitos resultan de un experimento binomial con $n = 1000$ y $p = P$(éxito), y se determina el intervalo de confianza de 95% para la verdadera probabilidad de éxito, ¿cuál es el máximo valor posible para el "máximo error de estimación"?

b. Compare el valor numérico del "máximo error de estimación" hallado en la parte a con el "margen de error" discutido en el caso práctico 9.9.

c. ¿Bajo qué condiciones son iguales? ¿No iguales?

d. Explique la forma en que los resultados de encuestas nacionales, como las de Harris y Gallup, están relacionados (similitudes y diferencias) con la técnica del intervalo de confianza estudiada en esta sección.

e. El error teórico de muestreo con un nivel de confianza se puede calcular, pero las encuestas por lo general informan sólo un "margen de error" sin probabilidad (nivel de confianza). ¿Por qué es esto?

9.81 Las reacciones adversas de drogas a medicinas legalmente prescritas están entre las principales causas de muertes relacionadas con las drogas en Estados Unidos. Suponga que usted investiga muertes relacionadas con las drogas en su ciudad, y encuentra que 223 de 250 incidencias fueron causadas por medicinas legalmente prescritas y el resto fueron el resultado el uso de drogas ilícitas. Entonces se usó el MINITAB para formar el intervalo de confianza de 98% para la proporción de muertes relacionadas con drogas que son causadas por medicinas legalmente prescritas. Verifique los resultados del MINITAB que siguen.

```
CI for One Proportion
Sample    X      N      Sample p            98% CI
1        223    250     0.892000     (0.846333, 0.937667)
```

9.82 Usando la salida del MINITAB e información del ejercicio 9.81, determine los valores de los términos siguientes:

a. Estimación puntual

b. Coeficiente de confianza

c. Error estándar de la media

d. Máximo error de estimación, E

e. Límite inferior de confianza

f. Límite superior de confianza

9.83 Karl Pearson una vez lanzó al aire una moneda 24 000 veces y registró 12 012 caras.

a. Calcule la estimación puntual para $p = P(\text{cara})$ con base en los resultados de Pearson.

b. Determine el error estándar de proporción.

c. Determine la estimación de intervalo de confianza de 95% para $p = P(\text{cara})$.

d. Debe haberle tomado muchas horas a Mr. Pearson lanzar al aire una moneda 25 000 veces. Usted puede simular 24 000 tiros de moneda usando los comandos de computadora y calculadora que siguen. (*Nota:* Un experimento de Bernoulli es como un experimento binomial de "un tiro". Esto es, un tiro de una moneda es un experimento de Bernoulli con $p = 0.5$; y 24 000 tiros de una moneda es un experimento binomial con $n = 24\,000$ o es 24 000 experimentos de Bernoulli. Código: 0 = cruz, 1 = cara. La suma de los números 1 será el número de caras en los 24 000 tiros.)

MINITAB

```
Choose Calc > Random Data > Bernoulli, entering 24000
for generate, C1 for Store in column(s) and 0.5 for
Probability of success. Sum the data and divide by
24,000.
```

Excel

```
Choose Tools > Data Analysis > Random Number Generation
> Bernoulli, entering 1 for Number of Variables, 24000
for Number of Random Numbers and 0.5 for p Value. Sum
the data and divide by 24,000.
```

TI-83/84 PLUS

```
Choose MATH > PRB > 5:randInt, then enter 0, 1, number
of trials. The maximum number of elements (trials) in a
list is 999. (slow process for large n's) Sum the data
and divide by n.
```

e. ¿Cómo se comparan los resultados simulados de usted con los de Pearson?

f. Use los comandos (parte d) y genere otro conjunto de 24 000 tiros de moneda. Compare estos resultados con los obtenidos por Pearson. También, compare las dos muestras simuladas una con otra. Explique lo que pueda concluir a partir de estos resultados.

9.84 Cuando se tira un solo dado, la probabilidad de un "uno" es 1/6, o 0.167. Simulemos 3000 tiros de un dado. (*Nota:* Un experimento de Bernoulli es como un experimento binomial de "un tiro". Esto es, un tiro de un dado es un experimento de Bernoulli con $p = 1/6$; y 3000 tiros de un dado es un experimento binomial con $n = 3000$ o es 3000 experimentos de Bernoulli. Código: 0 = 2, 3, 4, 5, o 6, y 1= 1. La suma de los números 1 será el número de unos (1) en los 3000 tiros.)

a. Use los comandos dados en el ejercicio 9.83 y una calculadora o computadora para simular el tiro de un solo dado 3000 veces.

Usando los resultados de la simulación:

b. Sume los datos y divida entre 3000. Explique lo que representa este valor.

c. Determine el error estándar de proporción.

d. Determine el intervalo de confianza de 95% para $p = P(\text{uno})$.

e. ¿Cómo se comparan los resultados de la simulación con sus expectativas? Explique.

9.85 La "regla práctica," que aparece en la página 497, indicaba que esperaríamos que la distribución muestral de p' sería aproximadamente normal cuando "$n > 20$ y tanto np como nq fueran mayores a 5". ¿Qué ocurre cuando estos lineamientos no se siguen?

a. Use el siguiente conjunto de comandos de computadora o calculadora para demostrar lo que pasa. Intente $n = 15$ y $p = 0.1$ (K1 = n y K2 = p). ¿Se ven normales las distribuciones? Explique qué ocasiona las "brechas." ¿Por qué se ven parecidos los histogramas? Intente algunas combinaciones diferentes de n (K1) y p (K2):

MINITAB

Choose Calc > Random Data > Binomial to simulate a 1000 trials for an n of 15 and a p of 0.5. Divide each generated value by n, forming a column of sample p's. Calculate a z value for each sample p by using $\mathbf{z = (p' - p)/\sqrt{p(1-p)/n}}$. Construct a histogram for the sample p's and another histogram for the z's.

Excel

Choose Tools > Data Analysis > Random Number Generation > Binomial to simulate a 1000 trials for an n of 15 and a p of 0.5. Divide each generated value by n, forming a column of sample p's. Calculate a z value for each sample p by using $\mathbf{z = (p' - p)/\sqrt{p(1-p)/n}}$. Construct a histogram for the sample p's and another histogram for the z's.

TI-83/84 Plus

Choose MATH > PRB > 7:randBin, then enter n, p, number of trials. The maximum number of elements (trials) in a list is 999. (slow process for large n's) Divide each generated value by n, forming a list of sample p's. Calculate a z value for each sample p by using $\mathbf{z = (p' - p)/\sqrt{p(1-p)/n}}$. Construct a histogram for the sample p's and another histogram for the z's.

b. Intente $n = 15$ y $p = 0.01$.

c. Intente $n = 50$ y $p = 0.03$.

d. Intente $n = 20$ y $p = 0.2$.

e. Intente $n = 20$ y $p = 0.8$.

f. ¿Qué pasa cuando la regla práctica no se sigue?

9.86 ¿Ha fallado la ley que exige el uso de casco para ciclistas? Yankelovich Partners llevó a cabo una encuesta de ciclistas en Estados Unidos. Sólo 60% de la muestra representativa nacionalmente de 1020 ciclistas informó que tenía un casco para andar en bicicleta.

Fuente: http://www.cpsc.gov/library/helmet.html

a. Encuentre un intervalo de confianza de 95% para la verdadera proporción p para un experimento binomial de 1020 intentos que resultaron en una proporción observada de 0.60. Use esto para estimar el porcentaje de ciclistas que informaron que tenían un casco.

b. Con base en los resultados de la encuesta, ¿diría usted que se cumple con la ley que requiere el uso de casco? Explique.

Suponga que usted desea llevar a cabo una encuesta en su ciudad y determinar qué porcentaje de ciclistas poseen cascos. Use la cifra nacional de 60% para su estimación inicial de p.

c. Encuentre el tamaño muestral si usted desea que su estimación esté dentro de 0.02 con 95% de confianza.

d. Encuentre el tamaño muestral si desea que su estimación esté dentro de 0.04 con confianza de 95%.

e. Encuentre el tamaño muestral si desea que su estimación esté dentro de 0.02 con confianza de 90%.

f. ¿Qué efecto tiene el cambio del máximo error en el tamaño muestral? Explique.

g. ¿Qué efecto tiene el cambio de nivel de confianza en el tamaño muestral? Explique.

9.87 Encuentre el tamaño muestral n necesario para una estimación de intervalo de 95% en el ejemplo 9.10.

9.88 Encuentre n para un intervalo de confianza de 90% para p con $E = 0.02$ usando una estimación de $p = 0.25$.

9.89 Según un artículo de USA Snapshot del 14 de mayo de 2002, 81% de automovilistas usan sus cin-

turones de seguridad. Usted desea llevar a cabo una encuesta en su ciudad para determinar qué porcentaje de automovilistas usan cinturones de seguridad. Use la cifra nacional de 81% para su estimación inicial de p.

a. Encuentre el tamaño muestral si desea que su estimación esté dentro de 0.02 con 90% de confianza.

b. Encuentre el tamaño muestral si desea que su estimación esté dentro de 0.04 con 90% de confianza.

c. Encuentre el tamaño muestral si desea que su estimación esté dentro de 0.02 con 98% de confianza.

d. ¿Qué efecto tiene el cambio del máximo error en el tamaño muestral? Explique.

e. ¿Qué efecto tiene el cambio de nivel de confianza en el tamaño muestral? Explique.

9.90 El cáncer pulmonar es la principal causa de muertes por cáncer en hombres y mujeres en Estados Unidos, Canadá y China. En otros países, el cáncer pulmonar es la causa número uno de muertes por cáncer en hombres y segunda y tercera causa entre mujeres. Sólo alrededor del 14% de las personas en las que se desarrolla cáncer sobreviven 5 años.

Fuente: eMedicine Consumer Journal, 20 de junio, 2002, vol. 3, número 6.

Suponga que usted desea ver si este porcentaje de supervivencia todavía fuera verdadero. ¿Qué tan grande debe ser la muestra que necesitaría tomar para estimar la verdadera proporción que sobrevive durante 5 años después del diagnóstico, para estar dentro de 1% con 95% de confianza? (Use el 14% como valor de p.)

9.91 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas frases:

a. Más de 60% de todos los estudiantes en nuestra universidad trabajan a tiempo parcial durante el año académico.

b. No más de un tercio de fumadores están interesados en dejar de fumar.

c. Una mayoría de electores votará para el presupuesto escolar este año.

d. Al menos tres cuartos de los árboles en nuestro condado fueron seriamente dañados por la tormenta.

e. Los resultados muestran que la moneda no se tiró imparcialmente.

9.92 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas frases:

a. La probabilidad de que nuestro equipo gane esta noche es menor de 0.50.

b. Al menos 50% de todos los padres piensan castigar a sus hijos cuando se lo merecen.

c. A lo sumo, 80% de los invitados asistirá a la boda.

d. Los números de un dígito generados por computadora no parecen ser igualmente probables con respecto a ser pares o nones.

e. A menos de la mitad de los clientes les gusta la nueva pizza.

9.93 Calcule la estadística de prueba $z\bigstar$ usada para probar lo siguiente:

a. H_o: $p = 0.70$ vs. H_a: $p > 0.70$, con la muestra $n = 300$ y $x = 224$

b. H_o: $p = 0.50$ vs. H_a: $p < 0.50$, con la muestra $n = 450$ y $x = 207$

c. H_o: $p = 0.35$ vs. H_a: $p \neq 0.35$, con la muestra $n = 280$ y $x = 94$

d. H_o: $p = 0.90$ vs. H_a: $p > 0.90$, con la muestra $n = 550$ y $x = 508$

9.94 Encuentre el valor de **P** para cada una de las pruebas de hipótesis del ejercicio 9.93; exprese la decisión usando $\alpha = 0.05$.

9.95 Determine el valor p para cada una de las siguientes situaciones de prueba de hipótesis.

a. H_o: $p = 0.5$, H_a: $p \neq 0.5$, $z\bigstar = 1.48$

b. H_o: $p = 0.7$, H_a: $p \neq 0.7$, $z\bigstar = -2.26$

c. H_o: $p = 0.4$, H_a: $p > 0.4$, $z\bigstar = 0.98$

d. H_o: $p = 0.2$, H_a: $p < 0.2$, $z\bigstar = -1.59$

9.96 Encuentre la región crítica, valores críticos, para cada una de las pruebas de hipótesis del ejercicio 9.93; exprese la decisión usando $\alpha = 0.05$.

9.97 Determine los criterios de prueba que se usarían para probar las siguientes hipótesis cuando z se usa como estadística de prueba y se emplea el método clásico.

a. H_o: $p = 0.5$ y H_a: $p > 0.5$, con $\alpha = 0.05$

b. H_o: $p = 0.5$ y H_a: $p \neq 0.5$, con $\alpha = 0.05$

c. H_o: $p = 0.4$ y H_a: $p < 0.4$, con $\alpha = 0.10$

d. H_o: $p = 0.7$ y H_a: $p > 0.7$, con $\alpha = 0.01$

9.98 La variable aleatoria binomial, x, puede usarse como estadística de prueba cuando se prueben hipótesis acerca del parámetro binomial, p, cuando n es pequeña (15, por ejemplo, o menos). Use la tabla 2 del apéndice B y determine el valor p para cada una de las siguientes situaciones.

a. H_o: $p = 0.5$, H_a: $p \neq 0.5$, donde $n = 15$ y $x = 12$

b. H_o: $p = 0.8$, H_a: $p \neq 0.8$, donde $n = 12$ y $x = 4$

c. H_o: $p = 0.3$, H_a: $p > 0.3$, donde $n = 14$ y $x = 7$

d. H_o: $p = 0.9$, H_a: $p < 0.9$, donde $n = 13$ y $x = 9$

9.99 La variable aleatoria binomial, x, puede usarse como estadística de prueba cuando se prueben hipótesis acerca del parámetro binomial, p. Cuando n es pequeña (15, por ejemplo, o menos), la tabla 2 del apéndice B da las probabilidades para cada valor de x separadamente, haciendo así innecesario estimar probabilidades de la variable aleatoria discreta binomial con la variable normal estándar continua z. Use la tabla 2 para determinar el valor de α para cada uno de lo siguiente:

a. H_o: $p = 0.5$ y H_a: $p > 0.5$, donde $n = 15$ y la región crítica es $x = 12, 13, 14, 15$

b. H_o: $p = 0.3$ y H_a: $p < 0.3$, donde $n = 12$ y la región crítica es $x = 0, 1$

c. H_o: $p = 0.6$ y H_a: $p \neq 0.6$, donde $n = 10$ y la región crítica es $x = 0, 1, 2, 3, 9, 10$

d. H_o: $p = 0.05$ y H_a: $p > 0.05$, donde $n = 14$ y la región crítica es $x = 4, 5, 6, 7, \ldots, 14$

9.100 Use la tabla 2 del apéndice B y determine la región crítica empleada para probar cada una de las siguientes hipótesis. (*Nota:* Como x es discreta, seleccione regiones críticas que no excedan el valor de α dado.)

a. H_o: $p = 0.5$ y H_a: $p > 0.5$, donde $n = 15$ y $\alpha = 0.05$

b. H_o: $p = 0.5$ y H_a: $p \neq 0.5$, donde $n = 14$ y $\alpha = 0.05$

c. H_o: $p = 0.4$ y H_a: $p < 0.4$, donde $n = 10$ y $\alpha = 0.10$

d. H_o: $p = 0.7$ y H_a: $p > 0.7$, donde $n = 13$ y $\alpha = 0.01$

9.101 Usted está probando la hipótesis $p = 0.7$ y ha decidido rechazar esta hipótesis si después de 15 intentos observa 14 o más éxitos.

a. Si la hipótesis nula es verdadera y usted observa 13 éxitos, ¿qué de lo siguiente haría? (1) Correctamente no rechazar H_o. (2) Correctamente rechazar H_o. (3) Cometer un error tipo I. (4) Cometer un error tipo II.

b. Encontrar el nivel de significación de su prueba.

c. Si la verdadera probabilidad de éxito es ½ y usted observa 13 éxitos, ¿qué de lo siguiente haría? (1) Correctamente no rechazar H_o. (2) Correctamente rechazar H_o. (3) Cometer un error tipo I. (4) Cometer un error tipo II.

d. Calcule el valor p para su prueba de hipótesis después que se observan 13 éxitos.

9.102 Usted está probando la hipótesis nula $p = 0.4$ y rechazará esta hipótesis si $z\bigstar$ es menor a -2.05.

a. Si la hipótesis nula es verdadera y usted observa $z\bigstar$ igual a -2.12, ¿qué de lo siguiente haría? (1) Correctamente no rechazar H_o. (2) Correctamente rechazar H_o. (3) Cometer un error tipo I. (4) Cometer un error tipo II.

b. ¿Cuál es el nivel de significación para esta prueba?

c. ¿Cuál es el valor p para $z\bigstar = -2.12$?

9.103 Una compañía de seguros dice que 90% de sus reclamaciones se resuelvan en no más de 30 días. Un grupo de consumidores seleccionó una muestra aleatoria de 75 de las reclamaciones de la compañía para probar esta afirmación. Si el grupo de consumidores encontró que 55 de las reclamaciones se resolvieron en no más de 30 días, ¿tienen suficiente razón para

apoyar su controversia de que menos del 90% de las reclamaciones se resuelvan en no más de 30 días? Use $\alpha = 0.05$.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.104 Un sondeo reciente llevado a cabo por ZOOM y Applied Research & Consulting LLC informó que los sucesos del 11 de septiembre han motivado a niños a ser voluntarios y que más de 80% se alistan como voluntarios. Un incrédulo de esta información tomó su propia muestra aleatoria de 500 niños en un intento por demostrar que el verdadero porcentaje de niños que se ofrecen como voluntarios es menor al 80%.

a. Encuentre el valor p si 384 de los niños encuestados dijo que realizan trabajo de voluntarios en la comunidad.

b. Explique por qué es importante se establezca el nivel de significación antes que se conozcan los resultados de la muestra.

9.105 Una política dice que ella recibirá 60% de los votos en una próxima elección. Los resultados de una muestra aleatoria correctamente diseñada de 100 electores mostró que 50 de los muestreados votará por ella. ¿Es probable que su afirmación sea correcta al nivel de significación de 0.05?

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.106 Un cuerpo estudiantil de tiempo completo de una universidad está compuesto de 50% de hombres y 50% de mujeres. Una muestra aleatoria de estudiantes (30 hombres, 20 mujeres) de un curso de introducción a la química, ¿muestra suficiente evidencia para rechazar la hipótesis de que la proporción de estudiantes hombres y de mujeres que toman este curso es igual que todo el cuerpo estudiantil? Use $\alpha = 0.05$.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.107 Si una encuesta dirigida por la American Association of Retired Persons (AARP) de 1706 adultos de 45 años o más, 72% estuvieron de acuerdo con la afirmación de que "a los adultos debe permitírseles legalmente usar marihuana para fines curativos si un médico lo recomienda".

Fuente: D&C, 19 de diciembre, 2004, "Most older Americans OK with medical pot"

Suponga que un estudio reciente de 200 adultos del oeste medio mostró que 134 estaban a favor del uso legal de marihuana para fines curativos. ¿Estos resultados muestran una menor proporción para el oeste medio con respecto al resto del país? Use un nivel de significación de 0.05.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.108 La popularidad de embarcaciones personales (también conocidas como esquís acuáticos) continúa en aumento, a pesar del aparente riesgo asociado con su uso. De hecho, una muestra de 54 accidentes ocurridos en estas embarcaciones y comunicados a la Comisión de Caza y Parques del estado de Nebraska reveló que 85% de ellos ocurrió en ellas, aun cuando sólo 8% de los botes de motor registrados en el estado son embarcaciones de este tipo.

Fuente: Nebraskaland, "Officer's Notebook: The Personal Problem"

Suponga que el promedio nacional de la proporción de accidentes en embarcaciones de este tipo fue 78%. El porcentaje de accidentes en embarcaciones personales en el estado de Nebraska, ¿rebasó al de la nación en su conjunto? Use un nivel de significación de 0.01.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.109 La USA Snapshot "Facing a crowd isn't easy" (Enfrentarse a una multitud no es fácil), (30 de mayo, 2002), informó que 35% de mujeres profesionistas norteamericanas temen hablar en público. Supongamos que usted dirige una encuesta de 1000 mujeres profesionistas seleccionadas al azar para probar que H_o: $p = 0.35$ contra H_a: $p < 0.35$, donde p representa la proporción que temen hablar en público. De las 1000 muestreadas, 324 temían hablar en público. Use $\alpha = 0.01$.

a. Calcule el valor de la estadística de prueba.

b. Resuelva usando el método del valor p.

c. Resuelva usando el método clásico.

9.110 Demuestre que la hipótesis de prueba completada como ejemplo 9.13 fue innecesaria porque el intervalo de confianza ya había sido completado en el ejemplo 9.8.

9.111 Consulte el caso práctico 9.14.

a. Encuentre la probabilidad de que 139 o más caras resulten cuando una moneda balanceada se lanza imparcialmente 250 veces al aire.

b. Encuentre la probabilidad de que 110 o menos caras resulten cuando una moneda balanceada se lanza imparcialmente al aire 250 veces.

c. Encuentre la probabilidad de que los resultados de lanzar imparcialmente al aire una moneda 250 veces es tan extremo como las 139 caras.

d. ¿Cuál es la reclamación de Romain Coenen? ¿Qué es lo que Barry Blight dice que es "sospechoso"? ¿Cómo es que estos dos enunciados forman los lados opuestos de una prueba de hipótesis? Exprese las hipótesis nula y alternativa.

e. El enunciado de Barry Blight, "Si la moneda no estuviera `cargada´, la probabilidad de obtener un resultado tan extremo como ese sería menos a 7%", es un enunciado de un valor p para una prueba de hipótesis de dos colas. Explique por qué.

f. Si los resultados muestrales se hubieran usado para estimar la probabilidad de que una moneda de un euro quedara con cara hacia arriba, ¿cuál hubiera sido el intervalo de confianza?

g. Si los medios hubieran informado estos resultados como 56% con ±6% de margen de error, ¿cómo sería esto semejante, o diferente, de una encuesta de opinión nacional que resultó en 56% ±6%?

9.112 Reliable Equipment (equipo confiable) fue perfeccionado como máquina, *The Flipper* (la aleta), que lanzará una moneda con resultados predecibles. Ellos dicen que una moneda lanzada por The Flipper quedará con la cara hacia arriba al menos en 88% de las veces. ¿Qué conclusión resultaría en una prueba de hipótesis, usando $\alpha = 0.05$, cuando 200 monedas se lanzan al aire y se obtienen los siguientes resultados?:

a. 181 caras
b. 172 caras
c. 168 caras
d. 153 caras

9.113 La siguiente salida de computadora se usó para completar una prueba de hipótesis.

Prueba para una proporción

```
Test of p = 0.225 vs p > 0.225
                                  95% Lower
Sample   X    N    Sample p   Bound        Z-Value   P-Value
1        61   200  0.305000   0.251451     2.71      0.003
```

a. Exprese las hipótesis nula y alternativa.

b. Si la prueba se completa usando $\alpha = 0.05$, ¿a qué decisión y conclusión se llega?

c. Verifique la "Muestra p".

9.114 Usando la salida de computadora e información del ejercicio 9.113, determine el valor de lo siguiente:

a. Valor hipotético de proporción poblacional

b. Proporción muestral

c. Estadística de prueba

9.4 Inferencias relacionadas con la varianza y la desviación estándar

Es frecuente que surjan problemas que nos exigen hacer inferencias acerca de la variabilidad. Por ejemplo, una compañía embotelladora de bebidas gaseosas tiene una máquina que llena botellas de 16 onzas. La compañía necesita controlar la desviación estándar σ (o varianza σ^2) en la cantidad de bebida gaseosa, x, puesta en cada botella. La cantidad media puesta en cada botella es importante, pero una cantidad media correcta no asegura que la máquina llenadora está funcionando correctamente. Si la varianza es demasiado grande, muchas botellas se derramarán y muchas quedarán sin llenar bien. Por tanto, la compañía embotelladora desea mantener la desviación estándar (o varianza) tan pequeña como sea posible.

Cuando se analizan inferencias acerca de la dispersión de datos, por lo general hablamos de varianza en lugar de desviación estándar porque las técnicas (las fórmulas empleadas) usan la varianza muestral en lugar de la desviación estándar.

No obstante, recuerde que la desviación estándar es la raíz cuadrada positiva de la varianza; por tanto, hablar de la varianza de una población es comparable a hablar de la desviación estándar.

Las inferencias acerca de la varianza de una población normalmente distribuida usan las distribuciones **ji cuadrada,** χ^2. Las distribuciones ji cuadrada, al igual que las distribuciones t de Student, son una familia de distribuciones de probabilidad, cada una de ellas identificada por el número de **parámetro** de **grados de libertad**. Para usar la distribución de ji cuadrada, debemos estar conscientes de sus propiedades (también, vea la figura 9.7).

Propiedades de la distribución ji cuadrada

1. χ^2 es no negativa en su valor; es cero o de valor positivo.
2. χ^2 no es simétrica; es sesgada a la derecha.
3. χ^2 está distribuida para formar una familia de distribuciones, una distribución separada para cada número diferente de grados de libertad.

FIGURA 9.7
Diversas distribuciones de ji cuadrada

Nota: cuando df > 2, el valor medio de la distribución ji cuadrada es df. La media está ubicada a la derecha del modo (el valor donde la curva llega a su punto más alto) y justo a la derecha de la mediana (el valor que divide la distribución, 50% a cada lado). Al localizar el cero a la extrema izquierda y el valor del df en su bosquejo de la distribución χ^2, el estudiante establecerá una escala aproximada de modo que los otros valores puedan ser ubicados en sus respectivas posiciones. Vea la figura 9.9.

FIGURA 9.9
Ubicación de la media, mediana y modo para la distribución χ^2

FIGURA 9.8
Distribución de ji cuadrada mostrando χ^2 (df, α)

Para valores de χ^2 en el lado izquierdo de la mediana, el área a la derecha será mayor a 0.50.

Los **valores críticos para ji cuadrada** se obtienen de la tabla 8 en el apéndice B. Cada valor crítico es identificado por dos piezas de información: df y área bajo la curva a la derecha del valor crítico que se busca. Por tanto, χ^2 (df, α) (se lee "*Ji cuadrada de df, alfa*") es el símbolo que se emplea para identificar el valor crítico de

ji cuadrada con df grados de libertad y con α área a la derecha, como se muestra en la figura 9.8. Como la distribución ji cuadrada no es simétrica, los valores críticos asociados con las colas derecha e izquierda se dan por separado en la tabla 8.

EJEMPLO 9.15 χ^2 asociada con la cola derecha

Encuentre $\chi^2(20, 0.05)$.

SOLUCIÓN Vea la figura. Use la tabla 8 del apéndice B para hallar el valor de $\chi^2(20, 0.05)$ en la intersección de la fila df = 20 y la columna $\alpha = 0.05$, como se ve en la parte de la tabla que sigue:

Parte de la tabla 8

	Área a la derecha		
df	...	0.05	...
⋮			
20		31.4	

$\chi^2(20, 0.05) = 31.4$

EJEMPLO 9.16 χ^2 asociada con la cola izquierda

Encuentre $\chi^2(14, 0.90)$.

SOLUCIÓN Vea la figura que sigue. Use la tabla 8 del apéndice B para hallar el valor de $\chi^2(14, 0.90)$ en la intersección de la fila df = 14 y la columna $\alpha = 0.90$, como se ve en la parte de la tabla que sigue:

Parte de la tabla 8

	Área a la derecha		
df	...	0.90	...
⋮			
14		7.79	

$\chi^2(14, 0.90) = 7.79$

Casi todos los paquetes de software de computadora o calculadoras estadísticas calcularán el área relacionada con un valor especificado de χ^2. La figura siguiente muestra la relación entre la probabilidad acumulativa y un valor χ^2 específica para una distribución χ^2 con df grados de libertad.

INSTRUCCIONES DE TECNOLOGÍA: PROBABILIDADES ACUMULATIVAS PARA χ^2

MINITAB (Versión 14)

Introduzca los datos en C1; luego continúe con:

```
Choose:     Calc > Probability Distributions > Chi-Square
Select:     Cumulative Probability
            Noncentrality Parameter: 0.0
```

```
Enter:      Degrees of freedom: df
Select:     Input constant*
Enter:      χ²-value (ex. 47.25) > OK
```

*Seleccione la columna Input si diversos valores de χ^2 están guardados en C1. Use C2 para guardado opcional.
Si es necesaria el área de la cola derecha, de uno reste la probabilidad calculada.

Excel

Si han de usarse diversos valores de χ^2, introduzca los valores en la columna A y active B1; luego continúe con:

```
Choose:     Insert function fx > Statistical > CHIDIST > OK
Enter:      X: individual χ²-value or (A1:A5 or select "χ²-value" cells)*
            Deg_freedom: df > OK
Drag*:      Bottom right corner of the B1 cell down to give other proba-
            bilities
```

TI-83/84 Plus

```
Choose:     2nd > DISTR > 7: χ²cdf(
Enter:      0, χ²-value, df)
```

Si es necesaria el área de la cola derecha, de uno reste la probabilidad calculada.

Ahora ya estamos listos para usar ji cuadrada para hacer inferencias acerca de la varianza poblacional o desviación estándar.

Suposiciones para inferencias acerca de la varianza σ^2 o desviación estándar σ: la población muestreada está normalmente distribuida.

Los procedimientos t para inferencias acerca de la media (vea sección 9.2) se basaron en la suposición de normalidad, pero los procedimientos t son generalmente útiles cuando la población muestreada no es normal, en especial para muestras más grandes. No obstante, lo mismo no es cierto acerca de los procedimientos de inferencia para la desviación estándar. Los procedimientos estadísticos para la desviación estándar son muy sensibles a distribuciones no normales (sesgo, en particular), y esto hace difícil determinar si un resultado significativo aparente es el resultado de la evidencia muestral o una violación de las suposiciones. Por tanto, el único procedimiento de inferencia a ser presentado aquí es la prueba de hipótesis para la desviación estándar de una población normal.

La **estadística de prueba** que se usará en probar hipótesis acerca de la varianza poblacional o desviación estándar se obtiene usando la fórmula:

Estadística de prueba para varianza y desviación estándar

$$\chi^2\bigstar = \frac{(n-1)s^2}{\sigma^2}, \quad \text{con df} = n - 1 \tag{9.10}$$

Cuando se sacan muestras aleatorias de una población normal con una varianza conocida σ^2, la cantidad $\frac{(n-1)s^2}{\sigma^2}$ posee una distribución de probabilidad que se conoce como la distribución ji cuadrada con $n - 1$ grados de libertad.

Procedimiento de prueba de hipótesis

Regresemos al ejemplo acerca de la compañía embotelladora que desea detectar cuando la variabilidad en la cantidad de bebida gaseosa que se pone en cada botella se sale de control. Una varianza de 0.0004 se considera aceptable, y la compañía desea ajustar la máquina llenadora de botellas cuando la varianza, σ^2, es más grande que este valor. La decisión se tomará usando el procedimiento de prueba de hipótesis.

EJEMPLO 9.17 Prueba de hipótesis de una cola para varianza, σ^2

La compañía embotelladora de bebidas gaseosas desea controlar la variabilidad en la cantidad de líquido, no permitiendo que la varianza exceda de 0.0004. Una muestra de tamaño 28 con una varianza de 0.0007, ¿indica que el proceso de embotellado está fuera de control (respecto a la varianza) al nivel de significación de 0.05?

SOLUCIÓN

Paso 1 El inicio:

a. Describir el parámetro poblacional de interés.
σ^2, varianza en la cantidad de líquido de una bebida gaseosa durante el proceso de embotellado.

b. Expresar la hipótesis nula (H_o) y la hipótesis alternativa (H_a).
H_o: $\sigma^2 = 0.0004$ ($\leq$) (varianza no es mayor a 0.0004)
H_a: $\sigma^2 > 0.0004$ (varianza es mayor a 0.0004)

Paso 2 Criterios de la prueba de hipótesis:

a. Comprobar las suposiciones.
La cantidad de líquido puesto en una botella está por lo general normalmente distribuida. Al comprobar la distribución de la muestra, podemos verificar esto.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.
La distribución ji cuadrada se usará y la fórmula (9.10), con df = $n - 1 = 28 - 1 = 27$.

c. Determinar el nivel de significación: $\alpha = 0.05$.

Paso 3 Evidencia muestral:

a. Recolectar la información muestral: $n = 28$ y $x^2 = 0.0007$.

b. Calcular el valor de la estadística de prueba.
Use la fórmula (9.10):

$$\chi^2\bigstar = \frac{(n-1)s^2}{\sigma^2}: \quad \chi^2\bigstar = \frac{(28-1)(0.0007)}{0.0004} = \frac{(27)(0.0007)}{0.0004} = 47.25$$

Paso 4 Distribución de probabilidad:

Usando el procedimiento del valor *p*

a. Calcular el valor *p* para la estadística de prueba.

Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor a".

O

Usando el procedimiento clásico:

a. Determinar la región crítica y valor(es) crítico(s)

La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "mayor a". El

P = $P(\chi^2\bigstar > 47.25$, con df = 27) como se muestra en la figura.

Para hallar el valor p, use uno de dos métodos:

1. Use la tabla 8 del apéndice B para poner límites al valor p: $0.005 < \mathbf{P} < 0.01$.
2. Use una computadora o calculadora para calcular el valor p: **P** = 0.0093.

Las instrucciones específicas siguen a esta ilustración.

b. Determinar si el valor *p* es o no es menor que *α*.

El valor p es menor que el nivel de significación, α (0.05).

El valor crítico se obtiene de la tabla 8, en la intersección de la fila df = 27 y la columna $\alpha = 0.05$: $\chi^2(27, 0.05) = 40.1$.

Para instrucciones específicas, vea la página 518.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

$\chi^2\bigstar$ está en la región crítica, como se muestra en **rojo** en la figura de líneas antes.

Paso 5 **Resultados:**

a. Expresar la decisión acerca de H_o: Rechazar H_o.

b. Expresar la conclusión acerca de H_a.

En el nivel de significación de 0.05, concluimos que el proceso de embotellado está fuera de control con respecto a la varianza.

Cálculo del valor *p* cuando se usa la distribución χ^2

Método 1: usar la tabla 8 del apéndice B para poner límites al valor p. Al inspeccionar la fila df = 27 de la tabla 8, se puede determinar un intervalo dentro del cual se encuentra el valor p. Localice $\chi^2\bigstar$ por toda la fila marcada df = 27. Si $\chi^2\bigstar$ no aparece, localice los dos valores entre los que $\chi^2\bigstar$ cae, y luego lea los límites al valor p desde la parte superior de la tabla. En este caso, $\chi^2\bigstar = 47.25$ está entre 47.0 y 49.6; por tanto, **P** está entre 0.005 y 0.01.

Parte de la tabla 8 Hallar $\mathbf{P} = P(\chi^2\bigstar > 47.25$, con df = 27)

	Área en cola derecha				
df	. . .	0.01	P	0.005	→ **0.005 < P < 0.01**
⋮		↑		↑	
27		47.0	*47.25*	49.6	

Método 2: usar computadora o calculadora. Use los comandos de distribución de probabilidad χ^2 de las páginas 518-519 para hallar el valor p asociado con $\chi^2\bigstar = 47.25$.

EJEMPLO 9.18 Prueba de hipótesis de valor *p* de una cola para varianza, σ^2

Encuentre el valor p para esta prueba de hipótesis:

H_o: $\sigma^2 = 12$
H_a: $\sigma^2 < 12$ con df = 15 y $\chi^2\bigstar = 7.88$

SOLUCIÓN Como el interés es por valores "menores que" (la hipótesis alternativa es "menor que"), el valor p es el área a la izquierda de $\chi^2\bigstar = 7.88$, como se muestra en la figura:

$$\mathbf{P} = P(\chi^2\bigstar < 7.88 \text{ con df} = 15)$$

Para hallar el valor p, use uno de dos métodos:

Método 1: *use la tabla 8 del apéndice B para poner límites al valor p*. Inspeccione la fila df = 15 para hallar $\chi^2\bigstar = 7.88$. El valor $\chi^2\bigstar$ está entre entradas, de modo que el intervalo que limita a **P** se lee del encabezado de la cola izquierda en la parte superior de la tabla.

Parte de la tabla 8 Hallar $\mathbf{P} = P(\chi^2\bigstar < 7.88$, con df $= 15)$

	Área en cola derecha				
df	. . .	0.05	P	0.10	$\longrightarrow$ **0.05 < P < 0.10**
⋮		↑		↑	
15		7.26	*7.88*	8.55	

Método 2: use computadora o calculadora. Use los comandos de distribución de probabilidad de las páginas 518-519 para hallar el valor p asociado con $\chi^2\bigstar = 7.88$.

EJEMPLO 9.19 Prueba de hipótesis de dos colas para desviación estándar, σ

El fabricante dice que una sustancia química fotográfica tiene una duración útil en depósito que está normalmente distribuida alrededor de una media de 180 días, con una desviación estándar de no más de 10 días. Como usuario de esta sustancia química, Fast Photo está interesado en que la desviación estándar pueda ser diferente de 10 días; de otro modo, comprará una cantidad mayor mientras la sustancia química sea parte de una promoción especial. Se seleccionaron y probaron 12 muestras aleatorias, resultando una desviación estándar de 14 días. Al nivel de significación de 0.05, ¿le parece que esta muestra presenta suficiente evidencia para demostrar que la desviación estándar es diferente de 10 días?

SOLUCIÓN

Paso 1 Inicio:

a. Describir el parámetro poblacional de interés.

σ, la desviación estándar para la duración útil en depósito de la sustancia química.

b. Exprese la hipótesis nula (H_o) y la hipótesis alternativa (H_a).

H_o: $\sigma = 10$ (desviación estándar es 10 días)

H_a: $\sigma \neq 10$ (desviación estándar es diferente de 10 días)

Paso 2 Criterios de hipótesis:

a. Comprobar las suposiciones.

El fabricante dice que la duración útil en depósito está normalmente distribuida; esto podría verificarse al comprobar la distribución de la muestra.

b. Identificar la distribución de probabilidad y la estadística de prueba a usar.

Se usará la distribución ji cuadrada y la fórmula (9.10), con df = $n - 1 = 12 - 1 = 11$.

c. Determine el nivel de significación: $\alpha = 0.05$.

Paso 3 **Evidencia muestral:**

a. Recolectar la información muestral: $n = 12$ y $s = 14$.

b. Calcular el valor de la estadística de prueba.

Use la fórmula (9.10):

$$\chi^2\bigstar = \frac{(n-1)s^2}{\sigma^2}: \quad \chi^2\bigstar = \frac{(12-1)(14)^2}{(10)^2} = \frac{2156}{100} = 21.56$$

Paso 4 **Distribución de probabilidad:**

Usando el procedimiento del valor *p*:

a. Calcular el valor *p* para la estadística de prueba.

Como el interés es para valores "diferentes de" 10, el valor p es el área de ambas colas. El área de cada cola representará ½**P.** Como $\chi^2\bigstar = 21.56$ está en la cola derecha, el área de la cola derecha es ½**P**:

½**P** = $P(\chi^2 > 21.56$, con df = 11) como se muestra en la figura.

Para hallar ½**P**, use uno de dos métodos:

1. Use la tabla 8 del apéndice B para poner límites a ½**P**: 0.025 < ½**P** < 0.05. Duplique ambos límites para hallar los límites para **P:** 2 × (0.025 < ½**P** < 0.05) se convierte en 0.05 < **P** < 0.10.
2. Use computadora o calculadora para hallar ½**P:** ½**P** = 0.0280; por tanto, **P** = 0.0560.

Instrucciones específicas siguen a esta ilustración.

b. Determinar si el valor *p* es o no es menor a *α*.

El valor p no es menor que el nivel de significación, (0.05).

O

Usando el procedimiento clásico:

a. Determinar la región crítica y valor(es) crítico(s).

La región crítica se divide en dos partes iguales porque H_a expresa interés por valores relacionados con "diferente de". Los valores críticos se obtienen de la tabla 8 en las intersecciones de la fila df = 11 con columnas $\alpha = 0.975$ y 0.025 (área a la derecha): $\chi^2(11, 0.975) = 3.82$ y $\chi^2(11, 0.025) = 21.9$

Para instrucciones específicas, vea la página 518.

b. Determinar si la estadística de prueba calculada está o no está en la región crítica.

$\chi^2\bigstar$ no está en la región crítica; vea la figura de líneas antes.

Paso 5 **Resultados:**

a. Exprese la decisión acerca de H_o: No rechazar H_o.

b. Expresar la conclusión acerca de H_a.

No hay suficiente evidencia en el nivel de significación de 0.05 para concluir que la duración útil en depósito de esta sustancia química tiene una desviación estándar diferente de 10 días. Por tanto, Fast Photo debería comprarla de conformidad.

Cálculo del valor *p* cuando se usa la distribución χ^2

Método 1: use la tabla 8 del apéndice B para poner límites al valor p. Inspeccione la fila df = 11 para localizar $\chi^2\bigstar = 21.56$. Observe que 21.56 está entre dos entradas de tabla. Los límites para ½**P** se leen del encabezado de la cola derecha en la parte superior de la tabla.

Parte de la tabla 8 hallar $\mathbf{P} = 2 \cdot P(\chi^2\bigstar > 21.56, \text{ con df} = 11)$

	Área en la cola derecha				
df	**. . .**	**0.05**	**$^1/_2$P**	**0.025**	→ **$0.025 < {}^1/_2\mathbf{P} < 0.05$**
⋮		↑		↑	**$0.05 < \mathbf{P} < 0.10$**
11		**19.7**	*21.56*	**21.9**	

Duplique ambos límites para hallar los límites para **P:** $2 < (0.025 < ½\mathbf{P} < 0.05)$ se convierte en **$0.05 < \mathbf{P} < 0.10$.**

Método 2: use computadora o calculadora. Use los comandos de distribución de probabilidad de las páginas 518-519 para hallar el valor *p* asociado con $\chi^2\bigstar = 21.56$. Recuerde duplicar la probabilidad.

Nota: cuando los datos muestrales son sesgados, sólo un resultado aislado puede afectar en gran medida la desviación estándar. Es muy importante, en especial cuando se usen muestras pequeñas, que la población muestreada sea normal porque de otro modo estos procedimientos no son confiables.

SECCIÓN 9.4 EJERCICIOS

9.115 a. Calcule la desviación estándar para cada conjunto.

A: 5, 6, 7, 7, 8, 10 B: 5, 6, 7, 7, 8, 15

b. ¿Qué efecto tuvo el cambio más grande en valor de 10 a 15 sobre la desviación estándar?

c. ¿Por qué piensa usted que 15 podría llamarse resultado aislado?

9.116 La varianza en medidas de zapatos para todos los fabricantes es 0.1024. ¿Cuál es la desviación estándar?

9.117 Encuentre:

a. $\chi^2(10, 0.01)$ b. $\chi^2(12, 0.025)$

c. $\chi^2(10, 0.95)$ d. $\chi^2(22, 0.995)$

9.118 Encuentre estos valores críticos usando la tabla 8 del apéndice B.

a. $\chi^2(18, 0.01)$ b. $\chi^2(16, 0.25)$ c. $\chi^2(8, 0.10)$

d. $\chi^2(28, 0.01)$ e. $\chi^2(22, 0.95)$ f. $\chi^2(10, 0.975)$

g. $\chi^2(50, 0.90)$ h. $\chi^2(24, 0.99)$

9.119 Usando la notación del ejercicio 9.118, mencione y encuentre los valores críticos de χ^2.

9.120 Usando la notación del ejercicio 9.118, mencione y encuentre los valores críticos de χ^2.

9.121 a. ¿Qué valor de ji cuadrada para 5 grados de libertad subdivide el área bajo la curva de distribución, tal que 5% está a la derecha y 95% a la izquierda?

b. ¿Cuál es el valor del 95avo percentil para la distribución ji cuadrada con 5 grados de libertad?

c. ¿Cuál es el valor del 90avo percentil para la distribución ji cuadrada con 5 grados de libertad?

9.122 a. El 90% central de la distribución ji cuadrada con 11 grados de libertad, ¿se encuentra entre qué valores?

b. El 95% central de la distribución ji cuadrada con 11 grados de libertad, ¿se encuentra entre qué valores?

c. El 99% central de la distribución ji cuadrada con 11 grados de libertad, ¿se encuentra entre qué valores?

9.123 Para una distribución ji cuadrada que tiene 12 grados de libertad, encuentre el área bajo la curva para valores de ji cuadrada que van de 3.57 a 21.0.

9.124 Para una distribución ji cuadrada que tiene 12 grados de libertad, encuentre el área bajo la curva entre $\chi^2(35, 0.96)$ y $\chi^2(35, 0.15)$.

9.125 Use computadora o calculadora para hallar el área (a) a la izquierda, y (b) a la derecha de $\chi^2\bigstar$ = 20.2 con df = 15.

9.126 Use computadora o calculadora para hallar el área (a) a la izquierda, y (b) a la derecha de $\chi^2\bigstar$ = 14.7 con df = 24.

9.127 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas afirmaciones:

a. La desviación estándar ha aumentado de su valor previo de 24.
b. La desviación estándar es no mayor a 0.5 onzas.
c. La desviación estándar no es igual a 10.
d. La varianza es no menor a 18.
e. La varianza es diferente del valor de 0.025, el valor pedido en las especificaciones.

9.128 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas afirmaciones:

a. La varianza ha disminuido de 34.5.
b. La desviación estándar de medida de zapatos es más de 0.32.
c. La desviación estándar es al menos 5.5.
d. La varianza es a lo sumo 35.
e. La varianza se ha reducido del valor de 0.34 porque las líneas de montaje se equiparon con nuevas máquinas.

9.129 Encuentre la estadística de prueba para la prueba de hipótesis:

a. H_o: $\sigma^2 = 532$ vs. H_a: $\sigma^2 > 532$ usando información muestral $n = 18$ y $s^2 = 785$.
b. H_o: $\sigma^2 = 52$ vs. H_a: $\sigma^2 \neq 52$ usando información muestral $n = 41$ y $s^2 = 78.2$.

9.130 Calcule el valor para la estadística de prueba, $\chi^2\bigstar$, para cada una de estas situaciones:

a. H_o: $\sigma^2 = 20$, $n = 15$, $s^2 = 17.8$
b. H_o: $\sigma^2 = 30$, $n = 18$, $s^2 = 5.7$
c. H_o: $\sigma = 42$, $n = 25$, $s = 37.8$
d. H_o: $\sigma^2 = 12$, $n = 37$, $s^2 = 163$

9.131 Calcule el valor p para cada una de las siguientes pruebas de hipótesis.

a. H_a: $\sigma^2 \neq 20$, $n = 15$, $\chi^2\bigstar = 27.8$
b. H_a: $\sigma^2 = 30$, $n = 18$, $\chi^2\bigstar = 33.4$
c. H_a: $\sigma^2 \neq 42$, df = 25, $\chi^2\bigstar = 37.9$
d. H_a: $\sigma^2 < 12$, df = 40, $\chi^2\bigstar = 26.3$

9.132 Determine la región crítica y valor(es) crítico(s) que se usarían para probar lo siguiente usando el método o enfoque clásico:

a. H_o: $\sigma = 0.5$ y H_a: $\sigma > 0.5$, con $n = 18$ y $\alpha = 0.05$
b. H_o: $\sigma^2 = 8.5$ y H_a: $\sigma^2 < 8.5$, con $n = 15$ y $\alpha = 0.01$
c. H_o: $\sigma = 20.3$ y H_a: $\sigma \neq 20.3$, con $n = 10$ y $\alpha = 0.10$
d. H_o: $\sigma^2 = 0.05$ y H_a: $\sigma^2 \neq 0.05$, con $n = 8$ y $\alpha = 0.02$
e. H_o: $\sigma = 0.5$ y H_a: $\sigma < 0.5$, con $n = 12$ y $\alpha = 0.10$

9.133 Complete la prueba de hipótesis del ejercicio 9.129a usando lo siguiente:

a. El método del valor p y $\alpha = 0.01$
b. El método clásico y $\alpha = 0.01$

9.134 Complete la prueba de hipótesis del ejercicio 9.129b usando lo siguiente:

a. El método del valor p y $\alpha = 0.05$
b. El método clásico y $\alpha = 0.05$

9.135 Una muestra aleatoria de 51 observaciones se seleccionó de una población normalmente distribuida. La media muestral fue $\bar{x} = 98.2$, y la varianza muestral fue $s^2 = 37.5$. ¿Esta muestra presenta suficiente razón para concluir que la desviación estándar poblacional no es igual a 8 al nivel de significación de 0.05?

a. Resuelva usando el método del valor p.
b. Resuelva usando el método clásico.

9.136 En el pasado, la desviación estándar de pesos de ciertos paquetes de 32.0 onzas llenados a máquina era de 0.25 onzas. Una muestra aleatoria de 20 paque-

tes mostró una desviación estándar de 0.35 onzas. El aparente aumento en variabilidad, ¿es significativo al nivel de significación de 0.10? Suponga que el peso de un paquete está normalmente distribuido.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.137 Quizá más importante que cuánto pesan, es que los discos que se usan en levantamiento de pesas sean del mismo peso. Cuando cada una de estas pesas cuelga de extremos opuestos de una barra, necesitan estar en equilibrio. Se seleccionó una muestra aleatoria de 24 pesas de 25 libras que se usan para levantar pesas, y se determinaron sus pesos (en libras):

25.3	22.1	25.7	24.2	25.7	23.9	23.1	21.9
24.7	26.3	26.5	22.2	25.9	23.5	25.8	27.1
25.4	22.0	25.2	21.1	27.9	22.9	27.3	25.7

Ha habido quejas respecto a la variabilidad excesiva en los pesos de estos discos de 25 libras. ¿La muestra presenta suficiente evidencia para concluir que la variabilidad en los pesos es mayor que la aceptable desviación estándar de 1 libra? Use $\alpha = 0.01$.

a. ¿Qué papel desempeña la suposición de normalidad en esta solución? Explique.

b. ¿Qué evidencia existe de que la suposición de normalidad es razonable? Explique.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.138 Un agricultor comercial cultiva todo su campo de una cosecha de legumbres a la vez. Por tanto, a él le gustaría plantar una variedad de judías verdes que maduren todas al mismo tiempo (pequeña desviación estándar entre tiempos de maduración de plantas individuales). Una empresa productora de semillas ha perfeccionado una nueva variedad híbrida de judías verdes que se cree es mejor para el agricultor comercial. El tiempo de maduración de la variedad estándar tiene un promedio de 50 días y una desviación estándar de 2.1 días. Una muestra aleatoria de 30 plantas de la nueva semilla híbrida mostró una desviación estándar de 1.65 días. ¿Esta muestra presenta una reducción significativa de la desviación estándar al nivel de significación de 0.05? Suponga que el tiempo de maduración está normalmente distribuido.

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.139. Los valores de bienes raíces de granjas en Estados Unidos fluctúan considerablemente de estado en estado y de condado a condado, haciendo así difícil que compradores adquieran tierras o a propietarios saber con precisión lo que valen en realidad sus propiedades. Por ejemplo, el valor promedio de tierras de cultivo en Missouri era de $548 el acre, mientras que el mismo promedio en tres estados cercanos (Kansas, Nebraska y Oklahoma) era más de $200 menos.

Fuente: Regional Economic Digest, "Survey of Agricultural Credit Conditions"

Esta discrepancia podría ser causada por una exagerada variabilidad en el valor de tierras de cultivo en Missouri. Suponga que la región combinada de los cuatro estados dio una desviación estándar de $85 por acre. Suponga que se tomó una muestra de 31 terratenientes de Missouri que en fechas recientes vendieron sus propiedades, y resultó una desviación muestral estándar de $125 por acre. La variabilidad en el valor de tierras de cultivo en Missouri, al nivel de significación de 0.05, ¿es mayor que la variabilidad para la región considerada en su conjunto?

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.140 Un fabricante de automóviles dice que las millas por galón para cierto modelo tiene una media igual a 40.5 millas, con una desviación estándar igual a 3.5 millas. Use los datos siguientes, obtenidos de una muestra aleatoria de 15 de esos autos, para probar la hipótesis de que la desviación estándar difiere de 3.5. Use $\alpha = 0.05$. Suponga normalidad.

37.0	38.0	42.5	45.0	34.0	32.0	36.0	35.5
38.0	42.5	40.0	42.5	35.0	30.0	37.5	

a. Resuelva usando el método del valor p.

b. Resuelva usando el método clásico.

9.141 El peso en seco de un corcho es otra cualidad que no afecta la capacidad del corcho para sellar la botella, pero es una variable que se vigila regularmente. Los pesos de corchos naturales Núm. 9 (24 mm de diá-

metro por 45 mm de longitud) tienen una distribución normal. Diez corchos seleccionados al azar se pesaron al centésimo de gramo más cercano.

Peso en seco (en gramos)

3.26	3.58	3.07	3.09	3.16	3.02	3.64	3.61	3.02	2.79

a. La muestra precedente, ¿presenta suficiente razón para indicar que la desviación estándar de los pesos en seco es diferente de 0.3275 gramos al nivel de significación de 0.02?

Se toma una muestra aleatoria diferente de 20 piezas del mismo lote.

Peso en seco (en gramos)

3.53	3.77	3.49	3.24	3.00	3.41	3.33	3.51	3.02	3.46
2.80	3.58	3.05	3.51	3.61	2.90	3.69	3.62	3.26	3.58

b. La muestra precedente, ¿presenta suficiente razón para indicar que la desviación estándar de los pesos en seco es diferente de 0.3275 gramos al nivel de significación de 0.02?

c. ¿Qué efecto tuvieron las dos desviaciones estándar muestrales diferentes sobre la estadística de prueba calculada en las partes a y b? ¿Qué efecto tuvieron en el valor p o valor crítico? Explique.

d. ¿Qué efecto tuvieron los dos tamaños muestrales diferentes sobre la estadística de prueba calculada en las partes a y b? ¿Qué efecto tuvieron en el valor p o valor crítico? Explique.

9.142 Use computadora o calculadora para hallar el valor p para la siguiente prueba de hipótesis: H_o: $\sigma^2 = 7$ vs. H_a: $\sigma^2 \neq 7$, si $\chi^2\bigstar = 6.87$ para una muestra de $n = 15$.

9.143 Use computadora o calculadora para hallar el valor p para la siguiente prueba de hipótesis: H_o: $\sigma = 12.4$ vs. H_a: $\sigma > 12.4$, si $\chi^2\bigstar = 36.59$ para una muestra de $n = 24$.

9.144 La distribución ji cuadrada se describió en la página 517 como una familia de distribuciones. Investiguemos estas distribuciones y observemos algunas de sus propiedades.

a. Use los comandos MINITAB que siguen y genere varias muestras aleatorias grandes de datos de diversas distribuciones ji cuadrada. Use valores df de 1, 2, 3, 5, 10, 20, y 80 (y otros si desea).

```
Choose:  Calc > Random Data > ChiSquare
Enter:   Generate: 1000 rows of data
         Store in column(s): C1
         Degrees of freedom: df
Use Stat > Basic Statistics > Display Descriptive Sta-
tistics to calculate the mean and median of the data in
C1. Use Graph > Histogram to construct a histogram of
the data in C1.
```

b. ¿Cuál parece ser la relación entre la media de la muestra y el número de grados de libertad?

c. ¿Cómo parecen estar relacionados los valores de la media, la mediana y la moda? ¿Sus resultados concuerdan con la información de la página 517?

d. Haga que la computadora genere muestras para dos grados de libertad adicionales df = 120 y 150. Describa la forma en que estas distribuciones parecen cambiar cuando df cambia.

9.145 ¿Qué tan importante es la suposición "la población muestreada está normalmente distribuida" para el uso de las distribuciones ji cuadrada? Use computadora y los dos conjuntos de comandos del MINITAB, para simular sacar 200 muestras de tamaño 10 de cada uno de dos tipos diferentes de distribuciones poblacionales. Los primeros comandos generarán 2000 valores de datos y construirán un histograma para que se pueda ver el aspecto de la población. Los siguientes comandos generarán 200 muestras de tamaño 10 de la misma población; cada una de las filas representa una muestra. Los siguientes comandos calcularán la desviación estándar y $\chi^2\bigstar$ para cada una de las 200 muestras. Los últimos comandos construirán histogramas de las 200 desviaciones estándar muestrales y los 200 valores de $\chi^2\bigstar$.

Para las muestras de la población normal:

a. ¿La distribución muestral de desviaciones estándar muestrales parece ser normal? Describa la distribución.

b. ¿La distribución χ^2 parece tener una distribución de ji cuadrada con df = 9? Encuentre porcentajes para intervalos (menores a 2, menores a 4, . . ., mayores a 15, mayores a 20, etc.), y compárelos con los porcentajes esperados como se estimó usando la tabla 8 del apéndice B.

Para las muestras de la población sesgada:

c. ¿La distribución muestral de desviaciones estándar muestrales parecen ser normales? Describa la distribución.

d. ¿La distribución χ^2 parece tener una distribución ji cuadrada con df = 9? Encuentre porcentajes para intervalos (menor que 2, menor que 4,..., mayor a 15, mayor a 20, etc.), y compárelos con los porcentajes esperados como se estimó usando la tabla 8.

En resumen:

e. ¿La condición de normalidad parece ser necesaria para que la estadística de prueba calculada χ^2★ tenga una distribución χ^2? Explique.

REPASO DEL CAPÍTULO

En retrospectiva

Hemos estado estudiando inferencias, intervalos de confianza y pruebas de hipótesis, para los tres parámetros poblacionales básicos (media μ, proporción p, y desviación estándar σ) de una población individual. Casi todas las inferencias acerca de una población individual se refieren a uno de estos tres parámetros. La figura 9.10 (p. 530) expone una organización visual de las técnicas presentadas en los capítulos 8 y 9 junto con las preguntas clave que el lector debe formular cuando decida cuál estadística de prueba y fórmula usar.

En esta capítulo también empleamos el máximo error de estimación, fórmula (9.7), para determinar el tamaño de la muestra requerida para hacer estimaciones acerca de la proporción poblacional con la precisión deseada. En el caso práctico 9.9 se describe el margen de error informado por los medios, y se estudia su relación con el máximo error de estimación, como se presenta en este capítulo. Al combinar la estimación puntual informada y el tamaño de muestra, podemos determinar el correspondiente máximo error de estimación de proporción binomial. Casi todas las encuestas y sondeos usan el nivel de confianza de 95% y luego usan el máximo error como estimación del margen de error y no informan un nivel de confianza, como explicó Humphrey Taylor.

En el siguiente capítulo estudiaremos inferencias acerca de dos poblaciones cuyas respectivas medias, proporciones y desviaciones estándar se comparan.

Vocabulario y conceptos clave

conclusión (pp. 483, 504, 521)
decisión (pp. 483, 504, 521)
error estándar (pp. 474, 498)
estadística de prueba (pp. 474, 481, 503, 519)
estadística muestral (p. 497)
estimador insesgado (p. 497)
experimento binomial (p. 496)
grados de libertad
inferencia (pp. 474, 496, 516)
intervalo de confianza (pp. 480, 498)
ji-cuadrada (p. 517)
máximo error de estimación (pp. 480, 498, 501)
nivel de confianza (pp. 480, 498, 506)
nivel de significación (pp. 482, 503, 520)
parámetro (pp. 475, 480, 496, 498, 503, 517, 520)
probabilidad binomial observada, p' (p. 496)
proporción (p. 497)
prueba de hipótesis (pp. 481, 502, 520)
región crítica (pp. 482, 504, 520)
regla práctica (p. 497)
suposiciones (pp. 479, 497, 519)
t de Student (p. 475)
tamaño muestral (pp. 474, 501)
valor calculado (pp. 482, 503, 520)
valor crítico (pp. 477, 504, 517)
valor p (pp. 482, 504, 520)
variable aleatoria (p. 496)
z, normal estándar (pp. 474, 497)
σ conocida (p. 475)
σ desconocida (p. 474)

FIGURA 9.10 Selección de la técnica correcta de inferencia

No** significa que se usa una técnica no paramétrica (distribución normal no requerida: vea el capítulo 14.

Objetivos de aprendizaje

✓ Entender que s, la desviación estándar muestral, es una estimación puntual para σ, la desviación estándar poblacional	pp. 474-475
✓ Entender que en casi todos los casos prácticos, σ es desconocida y s se usa como su mejor estimación.	pp. 474-475
✓ Comprender que cuando σ no se conoce, el estadístico z se reemplaza con el estadístico t de Student.	pp. 474-475
✓ Entender las propiedades de la distribución t, cómo es una serie de distribuciones basada en tamaño muestral (usando grados de libertad como índice), y cómo se aproxima a la distribución normal estándar cuando aumenta el tamaño austral.	pp. 475-476, EJ. 9.1, 9.2, 9.3, Ejer. 9.4, 9.19
✓ Entender que la suposición de inferencias acerca de la media μ cuando σ es desconocida es que la población muestreada está normalmente distribuida.	p. 479, Ejer. 9.27b
✓ Calcular, describir e interpretar un intervalo de confianza para la media poblacional, μ, usando la distribución t.	p. 480, EJ. 9.4, Ejer. 9.23, 9.27, 9.147
✓ Efectuar, describir e interpretar una prueba de hipótesis para la media poblacional, μ, usando la distribución t con el método del valor p y el método clásico.	pp. 481-482, EJ. 9.5, 9.6, Ejer. 9.47, 9.154
✓ Entender las propiedades fundamentales de un experimento binomial y el parámetro binomial, p.	pp. 496-497, Ejer. 9.63, 9.64
✓ Entender que p', la proporción muestral, es un estimador insesgado de la proporción poblacional, p.	pp. 496-498, Ejer. 9.65
✓ Entender que la distribución muestral de p' tiene una distribución aproximadamente normal si n es grande lo suficiente y por tanto la distribución normal estándar se puede usar para inferencias.	pp. 496-498, Ejer. 985
✓ Entender que la suposición de inferencias acerca del parámetro binomial, p, es que las observaciones aleatorias n que forman la muestra se seleccionan de manera independiente de una población que no está cambiando durante el muestreo.	p. 497
✓ Calcular, describir e interpretar un intervalo de confianza para la proporción poblacional, p, usando la distribución z.	EJ. 9.8, Ejer. 9.69, 9.73, 9.161
✓ Calcular y describir el tamaño muestral requerido para un intervalo de confianza de p, la proporción poblacional.	p. 501, EJ. 9.10, 9.11, Ejer. 9.86
✓ Efectuar, describir e interpretar una prueba de hipótesis para la proporción poblacional, p, usando la distribución z con el método del valor p y el método (o enfoque) clásico.	EJ. 9.12, 913, Ejer. 9.103, 9.108
✓ Entender las propiedades de la distribución ji cuadrada y cómo es una serie de distribuciones basada en un tamaño muestral (usando grados de libertad como el índice).	p. 517
✓ Entender que la suposición de inferencias acerca de la varianza, σ^2, o desviación estándar, σ, es que la población muestreada está normalmente distribuida.	p. 519, Ejer. 9.145
✓ Efectuar, describir e interpretar una prueba de hipótesis para la varianza poblacional, σ^2, o desviación estándar, σ, usando la distribución χ^2 con el método del valor p y el método clásico.	EJ. 9.17, 9.19, Ejer. 9.136, 9.177

Ejercicios del capítulo

9.146 Usted va de prisa al departamento local de emergencia con la esperanza de recibir atención urgente, sólo para encontrarse esperando lo que parecen horas. El gerente de un gran departamento de emergencias piensa que sus nuevos procedimientos han reducido en forma considerable el tiempo de espera para el paciente de atención urgente promedio. Él inicia un estudio para evaluar el tiempo de espera. Se revisan los registros de 18 pacientes seleccionados al azar desde que los nuevos procedimientos han sido implantados, y se ha observado el tiempo desde la entrada al departamento de emergencias hasta ser atendido por personal de atención urgente. El tiempo medio de espera fue 17.82 minutos con una desviación estándar de 5.68 minutos. Estime el tiempo medio de espera usando un intervalo de confianza de 99%. Suponga que los tiempos de espera están normalmente distribuidos.

9.147 Una empresa distribuidora de gas natural está considerando un contrato para comprar llantas para su flota de camiones de servicio. La decisión estará basada en millas de recorrido esperado. Para una muestra de 100 llantas probadas, la distancia en millas fue 36,000 y la desviación estándar fue de 2000 millas. Estime la distancia media en millas que la empresa debe esperar de estas llantas, usando un intervalo de 98%.

9.148 Uno de los objetivos de un importante estudio médico era estimar la tarifa media de un médico por remover cataratas. Para 25 casos seleccionados al azar, se encontró que la tarifa media era de \$1550 con una desviación estándar de \$125. Establezca un intervalo de confianza de 99% en μ, la tarifa media para todos los médicos. Suponga que las tarifas están normalmente distribuidas.

9.149 Al azar se seleccionan naranjas de un gran embarque que acaba de llegar. La muestra se tomó para estimar el tamaño (circunferencia, en pulgadas) de las naranjas. Los datos muestrales se resumen como sigue: $n = 100$, $\Sigma x = 878.2$, y $\Sigma(x-\bar{x})^2 = 49.91$.

a. Determine la media muestral y desviación estándar.

b. ¿Cuál es la estimación puntual para μ, la circunferencia media de todas las naranjas del embarque?

c. Encuentre el intervalo de confianza de 95% para μ.

9.150 Se usan moldes para la manufactura de lentes de contacto, de modo que el material del lente con una preparación y curado adecuados sean consistentes y satisfagan los criterios dimensionales diseñados. Se fabricaron moldes y una dimensión crítica se midió para 15 moldes seleccionados al azar. (Los datos han sido doblemente codificados para asegurar la patente.)

140	130	15	180	95	135	220	105
195	110	150	150	130	120	120	

Cortesía de Bausch & Lomb

a. Construya un histograma y encuentre la media y desviación estándar.

b. Demuestre la forma en que este conjunto de datos satisface las suposiciones para inferencia.

c. Encuentre el intervalo de confianza de 95% para μ.

d. Interprete el significado del intervalo de confianza.

9.151 Una compañía dice que su batería dura no menos de 42.5 horas en condiciones de uso continuo en un juguete especificado. Una muestra aleatoria simple de baterías da una duración media muestral de 41.89 horas, con una desviación estándar de 4.75 horas. Una computadora calcula una estadística de prueba de $t = -1.09$ y un valor p de 0.139. Si la prueba usa df = 71, ¿cuál es la mejor estimación del tamaño muestral?

9.152 Tener educación universitaria hoy en día es casi tan importante como respirar y es costosa. No es sólo la colegiatura, cuarto y comida, sino que los libros de texto también son caros. Es muy importante que los estudiantes, y sus padres, tengan una estimación precisa de los costos totales de libros de texto. Se recolectó el costo total de libros de texto requeridos para grupos de estudiantes de primer año o de segundo en 10

universidades públicas de Nueva York seleccionadas al azar:

582.19	806.40	913.44	915.75	932.35
957.45	960.92	996.24	1070.44	1223.44

a. Construya un histograma y encuentre la media y desviación estándar.
b. Demuestre la forma en que este conjunto de datos satisface las suposiciones de inferencia.
c. Encuentre el intervalo de confianza de 95% para μ, el costo medio total de libros de texto requeridos.
d. Interprete el significado del intervalo de confianza.

9.153 Se recolectó el costo total de libros de texto requeridos para nueve estudiantes de primer año o de segundo, en 10 universidades privadas de Nueva York seleccionadas al azar:

639.00	865.75	868.20	874.25	887.06
890.50	970.13	1013.22	1026.00	1048.96

a. Construya un histograma y encuentre la media y desviación estándar.
b. Demuestre la forma en que este conjunto de datos satisface las suposiciones de inferencia.
c. Encuentre el intervalo de confianza de 95% para μ, el costo medio total de libros de texto requeridos.
d. Interprete el significado del intervalo de confianza.
e. ¿Hay diferencia en el costo medio total de los nueve libros de texto requeridos, entre las universidades públicas del ejercicio 9.152 y de universidades privadas de este ejercicio? Explique.
f. Explique por qué el intervalo de confianza para universidades públicas es mucho más ancho que el intervalo correspondiente para universidades privadas. Sea exacto y detallado.

9.154 Un fabricante de televisores dice que los gastos de mantenimiento para su producto promediarán no más de $50 durante el primer año que sigue a la expiración de la garantía. Un grupo de consumidores le ha pedido a usted justificar o no dar crédito a lo dicho por el fabricante. Los resultados de una muestra aleatoria de 50 propietarios de estos televisores comprobaron que el gasto medio fue $61.60 con una desviación estándar de $32.46. Al nivel de significación de 0.01, ¿debe usted concluir que lo dicho por el fabricante es verdadero o no es probable que sea verdadero?

9.155 En una gran huerta de cerezas, el promedio de producción ha sido de 4.35 toneladas por acre durante los últimos varios años. Se ha probado un nuevo fertilizante en 15 terrenos de 1 acre seleccionados al azar. La producción (en toneladas) de estos lotes fue:

3.56	5.00	4.88	4.93	3.92	4.25	5.12	5.13
4.79	4.45	5.35	4.81	3.48	4.45	4.72	

Al nivel de significación de 0.05, ¿tenemos suficiente evidencia para decir que hubo aumento significativo en producción? Suponga que la producción por acre está normalmente distribuida.

9.156 Las lecturas de contaminación del agua en la playa del parque estatal parecen ser más bajas que las del año pasado. Se seleccionó una muestra de 12 lecturas (medidas en bacilos/100 mL) de los registros de lecturas diarias de este año:

3.5	3.9	2.8	3.1	3.1	3.4	4.8	3.2	2.5	3.5	4.4	3.1

Esta muestra, ¿da suficiente evidencia para concluir que la media de las lecturas de contaminación de este año es significativamente menor que la media del año pasado de 3.8, al nivel de 0.05? Suponga que todas esas lecturas tienen una distribución normal.

9.157 Se ha sugerido que el número de niños hombres anormales tiende a ser mayor cuando nacen de padres de más edad que el promedio. Se obtuvieron historias clínicas de 20 niños anormales, y las edades de las 20 madres fueron como sigue:

31	21	29	28	34	45	21	41	27	31
43	21	39	38	32	28	37	28	16	39

La edad media a la que las madres en la población general dan a luz es 28.0 años.

a. Calcule la media muestral y desviación estándar.
b. La muestra, ¿da suficiente evidencia para apoyar lo dicho de que niños hombres anormales tienen madres de más edad que el promedio? Use $\alpha = 0.05$. Suponga que las edades tienen una distribución normal.

9.158 Al azar se identificaron 24 condados productores de avena en Estados Unidos, con el fin de probar la frase de que "la producción media de avena es menos de 60 bushels (1 bushel = 35.237 litros) por acre". Se

obtuvo la producción, en bushels de avena por acre cultivado, para cada condado identificado:

44.0	65.0	78.5	50.0	52.5	51.0	47.5	67.5	76.7	33.3	73.6	20.0
57.0	52.0	68.6	42.9	63.0	80.0	37.0	70.0	43.0	30.0	60.0	67.5

Fuente: http://www.usda.gov/nass/graphics/county01/data/ot01.csv

a. ¿Se satisfacen las suposiciones de prueba? Explique.

b. Complete la prueba usando $\alpha = 0.05$.

9.159 A continuación se encuentran 100 mediciones de la velocidad de la luz en aire (km/s) registradas por Albert Michelson, físico norteamericano del 5 de junio al 2 de julio de 1879. A las mediciones se les había restado 299 000 de cada una y luego se ajustaron para correcciones usadas por Michelson. En esta forma, el verdadero valor constante para la velocidad de la luz en aire se convierte en 734.5 km/s. ¿Las mediciones de Michelson apoyan el verdadero valor que él estaba tratando de medir? Use un nivel de significación de 0.01.

850	740	900	1070	930	850	950	980	980	880	1000
980	930	650	760	810	1000	1000	960	960	960	940
960	940	880	800	850	880	900	840	830	790	810
880	880	830	800	790	760	800	880	880	880	860
720	720	620	860	970	950	880	910	850	870	840
840	850	840	840	840	890	810	810	820	800	770
760	740	750	760	910	920	890	860	880	720	840
850	850	780	890	840	780	810	760	810	790	810
820	850	870	870	810	740	810	940	950	800	810
870										

Fuente: http://lib.stat.cmu.edu/DASL/Stories/SpeedofLight.html

Nota: el "verdadero" valor aceptado en la actualidad es 299 792.5 km/s (sin ajustes.)

9.160 Incluso con una mayor información sobre la calidad de la carne de res, un 82% de norteamericanos indicó que su reciente gusto de comer hamburguesas ha seguido igual, según un sondeo reciente de los restaurantes T.G.I. Friday hecho a 1027 norteamericanos. De hecho, la mitad de la población en Estados Unidos come al menos una hamburguesa de carne de res por semana. Eso es un mínimo de 52 hamburguesas al año.

Fuente: Harris Interactive/Yankelovich Partners para los restaurantes T.G.I. Friday, http://www.knoxville3.com/fridays/News/burger.htm

a. ¿Cuál es la estimación puntual para la proporción de todos los norteamericanos que comen al menos una hamburguesa de carne de res por semana?

b. Encuentre el intervalo de confianza de 98% para la verdadera proporción p en la situación binomial donde $n = 1027$ y la proporción observada es la mitad.

c. Use los resultados de la parte b para estimar el porcentaje de todos los norteamericanos que comen al menos una hamburguesa de carne de res por semana.

9.161 El departamento de investigación de mercado de una empresa de café instantáneo llevó a cabo un sondeo de hombres casados, para determinar la proporción de hombres casados que prefieren su marca. De los 100 hombres de la muestra aleatoria, 20 de ellos preferían la marca de la compañía. Use un intervalo de confianza de 95% para estimar la proporción de todos los hombres casados que prefieren la marca de café instantáneo de esta compañía. Interprete su respuesta.

9.162 Una compañía está haciendo un proyecto de campaña publicitaria que hará necesarios los apoyos de notables atletas. Para que la compañía tenga éxito, quien la apoye debe ser una persona muy respetada y fácilmente reconocida. A una muestra aleatoria de 100 clientes en prospecto se le mostró fotografías de varios atletas; si el cliente reconoce un atleta, entonces al cliente se le pregunta si lo respeta. En el caso de una destacada golfista, 16 de los 100 que respondieron a la encuesta reconocieron la fotografía de ella e indicaron que también la respetaban. Al nivel de confianza de 95%, ¿cuál es la verdadera proporción con la que esta golfista es reconocida y respetada?

9.163 Un distribuidor de autos de la localidad anuncia que 90% de clientes cuyos autos fueron atendidos por su departamento de servicio están satisfechos con los resultados. Como investigador, usted se opone a esta aseveración porque está consciente de que muchas personas son reacias a expresar insatisfacción incluso cuando no están contentos con el servicio. Se inició un experimento de investigación en el que quienes estaban en la muestra habían recibido servicio por este distribuidor dentro de las últimas 2 semanas. Durante la entrevista, a las personas se les indujo a creer que el entrevistador era nuevo en la ciudad y estaba considerando llevar su auto a este

departamento de servicio. De los 60 muestreados, 14 dijeron que no estaban satisfechos y no recomendarían el departamento.

a. Estime la proporción de clientes insatisfechos usando un intervalo de confianza de 95%.

b. Dada su respuesta a la parte a, ¿qué puede concluirse acerca de lo dicho por el distribuidor?

9.164 Al obtener el tamaño muestral para estimar una proporción, se usa la fórmula $n = [z(\alpha/2)]^2\, pq/E^2$. Si no se dispone de una estimación razonable de p, se sugiere usar $p = 0.5$ porque esto dará el máximo valor de n. Calcule el valor de $pq = p(1 - p)$ para $p = 0.1$, 0.2, 0.3, . . . , 0.8, 0.9 para tener alguna idea acerca del comportamiento de la cantidad pq.

9.165 Un estudio de Pew Internet & American Life Project basado en 1100 llamadas telefónicas aleatorias, con niños de 12 a 17 años de edad y sus padres, fue realizado del 26 de octubre al 28 de noviembre de 2004. El estudio encontró que 13% de adolescentes norteamericanos no usan Internet. Hallazgos tienen un margen de error de muestreo de más o menos 3 puntos porcentuales.

Fuente: http://www.usatoday.com/news/bythenumbers/2005-03-18-teen-net-use_x.htm

Usted desea realizar un estudio para estimar el porcentaje de niños de 12 a 17 años de edad que no usan Internet y viven en su estado. Suponga que la proporción poblacional es 13% como lo informa el Pew Internet & American Life Project. ¿Qué tamaño muestral debe usar si usted desea que su estimación se encuentre dentro de:

a. 0.03 con 90% de confianza

b. 0.06 con 95% de confianza

c. 0.09 con 99% de confianza

9.166 La directora general de una pequeña empresa desea contratar los servicios de usted como consultor, para dirigir una muestra aleatoria sencilla de sus clientes. Ella desea determinar la proporción de sus clientes que consideran que su empresa es la fuente principal de sus productos. Ella pide que el margen de error en la proporción no sea mayor a 3% con 95% de confianza. Estudios anteriores han indicado que la proporción aproximada es 37%.

a. ¿Cuál es el tamaño mínimo de la muestra que usted recomendaría para satisfacer los requisitos de su cliente, si usa resultados anteriores?

b. ¿Cuál es el tamaño mínimo de la muestra que usted recomendaría para satisfacer los requisitos de su cliente, si pasa por alto los resultados anteriores?

c. ¿Es la proporción aproximada del valor cuando se realiza la encuesta? Explique.

9.167 El 25 de marzo de 2002, el artículo de *Newsweek* "Educando adultolescentes" citó una encuesta en línea hecha por Monster-TRAK.com. La encuesta encontró que 60% de los estudiantes universitarios planeaban vivir en casa después de graduarse. ¿Qué tan grande ha de ser el tamaño muestral que necesitaría para estimar la verdadera proporción de estudiantes que planean vivir en casa, después de graduarse, con variación de no más del 2% y confianza de 98%?

9.168 Se considera que una máquina está en operación de una manera aceptable si produce 0.5% o menos de piezas defectuosas; no está funcionando de modo aceptable si más de 0.5% de su producción es defectuosa. La hipótesis H_o: $p = 0.005$ se prueba contra la hipótesis H_a: $p > 0.005$ al tomar una muestra aleatoria de 50 piezas producidas por la máquina. La hipótesis nula es rechazada si dos o más piezas defectuosas se encuentran en la muestra. Encuentre la probabilidad del error tipo I.

9.169 Usted está interesado en comparar la hipótesis nula $p = 0.8$ contra la hipótesis alternativa $p < 0.8$. En 100 intentos observa 73 éxitos. Calcule el valor p asociado con este resultado.

9.170 La fundación de la familia Kaiser llevó a cabo una encuesta nacional en 2003 a 17 685 adultos mayores, con el fin de capturar información detallada acerca del uso, cobertura y experiencias de medicamentos para adultos mayores.

Fuente: http://www.kff.org/medicare/med041905nr.cfm

a. Si ésta fuera una muestra que satisficiera todos los requisitos para una inferencia acerca de p, ¿cuál sería el error estándar?

b. ¿Cuál sería el error máximo de estimación para un intervalo de confianza de 95%?

c. ¿Merece la pena una muestra de este tamaño? Dé razones para apoyar su respuesta.

9.171 La negociación Pizza Shack ha estado experimentando con diferentes recetas para su pasta de pizza, pensando que pudieran sustituir su actual receta. Están planeando muestrear pizzas hechas con la nueva pasta. Antes de hacer el muestreo, se necesita una estrategia para que después que estén los resultados de la prueba, Pizza Shack sepa cómo interpretar las preferencias de sus clientes. La decisión no se toma a la ligera porque hay mucho que ganar, o perder, dependiendo de si la decisión es o no es estimada. Se está planeando una prueba de hipótesis de una cola de $p = P$(prefiere nueva pasta) $= 0.50$.

a. Si se usa H_a: $p > 0.50$, explique el significado de cuatro posibles resultados y sus acciones resultantes.

b. Si se usa H_a: $p < 0.50$, explique el significado de cuatro posibles resultados y sus acciones resultantes.

c. ¿Cuál hipótesis alternativa recomienda usar, $p > 0.5$ o $p < 0.5$? Explique.

9.172 La Pizza Shack del ejercicio 9.171 ha completado su muestreo y los resultados ya están. Un martes en la tarde, muestrearon 15 clientes y 9 de ellos prefirieron la nueva pasta de pizza. El viernes por la noche, muestrearon 200 clientes y 120 prefirieron la nueva pasta de pizza. Ayude al gerente a interpretar el significado de estos resultados. Use una prueba de una cola con H_a: $p > 0.50$ y $\alpha = 0.02$. Use z como estadística de prueba.

a. ¿Hay suficiente evidencia para concluir una preferencia significativa para la nueva pasta, con base en los clientes del martes?

b. ¿Hay suficiente evidencia para concluir una preferencia significativa para la nueva pasta, con base en los clientes del viernes?

c. Como el porcentaje de clientes que prefieren la nueva pasta fue el mismo, $p' = 0.60$, en ambos muestreos, explique por qué las respuestas en las partes a y b no son iguales.

9.173 El propietario de Pizza Shack de los ejercicios 9.171 y 9.172 no entiende el uso de la distribución normal y z en el ejercicio 9.172. Ayude al gerente a interpretar el significado de los resultados volviendo a realizar ambas pruebas de hipótesis, usando x = número de clientes que prefieren la nueva pasta como la estadística de prueba y su distribución de probabilidad binomial. Use una prueba de una cola con H_a: $p > 0.50$ y $\alpha = 0.02$.

Los resultados fueron como sigue: el martes por la tarde, muestrearon 15 clientes y 9 de ellos prefirieron la nueva pasta de pizza; el viernes por la noche, muestrearon 200 clientes y hallaron que 120 prefirieron la nueva pasta de pizzas.

a. ¿Hay suficiente evidencia para concluir una preferencia significativa para la nueva pasta con base en los clientes del martes?

b. ¿Hay suficiente evidencia para concluir una preferencia significativa para la nueva pasta con base en los clientes del viernes?

c. Explique la relación entre las soluciones obtenidas en el ejercicio 9.172 y aquí.

9.174 Un profesor pide a cada uno de los 54 miembros de su grupo que escriban "al azar" uno de los números 1, 2, 3, . . . , 13, 14, 15. Como el profesor piensa que a los estudiantes les gustan los juegos de azar, considera que 7 y 11 son números de suerte. Él cuenta el número de estudiantes, x, que seleccionaron u 11. ¿Qué tan grande debe ser x antes que la hipótesis de aleatoriedad pueda ser rechazada al nivel de 0.05?

9.175 Los periódicos y revistas de hoy en día publican con frecuencia lo que se encuentra en encuestas acerca de varios aspectos de la vida. El Pew Internet & American Life Project (13 de enero-9 de febrero de 2005) encontró que "63% de usuarios de teléfono celular y edades entre 18 y 27 han usado mensajes de texto en el último mes". Otra información obtenida del proyecto incluyó "encuesta telefónica aleatoria de 1460 usuarios de teléfono celular" y "tiene un margen de error de muestreo de más o menos 3 puntos porcentuales". Relacione esta información con las inferencias estadísticas que ha estado estudiando en este capítulo.

a. ¿Un porcentaje de personas es un parámetro poblacional, y si es así, cómo está relacionado a cualquiera de los parámetros que hemos estado estudiando?

b. Con base en la información dada, encuentre el intervalo de confianza de 95% para la verdadera proporción de usuarios de teléfono celular que han usado mensajes de texto.

c. Explique la forma en que los términos "estimación puntual", "nivel de confianza", "máximo error de estimación," e "intervalo de confianza" se relacionan con los valores publicados en el artículo y con sus respuestas de la parte b.

9.176 Para probar la hipótesis de que la desviación estándar en una prueba estándar es 12, se probó una muestra de 40 estudiantes seleccionados al azar. Se encontró que la varianza muestral era de 155. ¿Esta muestra da suficiente evidencia para demostrar que la desviación estándar difiere de 12 al nivel de significación de 0.05?

9.177 Bright-Lite dice que su bombilla eléctrica de 60 watts se enciende con una duración que está normalmente distribuida en forma aproximada, con una desviación estándar de 81 horas. Una muestra de 101 bombillas tuvo una varianza de 8075. ¿Es esto suficiente evidencia para rechazar lo dicho por Bright-Lite a favor de la alternativa "la desviación estándar es mayor a 81 horas", al nivel de significación de 0.05?

9.178 Un proceso de producción está considerado fuera de control si las piezas producidas tienen una longitud media diferente de 27.5 mm o una desviación estándar mayor de 0.5 mm. Una muestra de 30 piezas da una media muestral de 27.63 mm y una desviación estándar muestral de 0.87 mm. Si suponemos que una parte de longitud es una variable normalmente distribuida, ¿esta muestra indica que el proceso debe ajustarse para corregir la desviación estándar del producto? Use $\alpha = 0.05$.

9.179 Julie Jackson opera un restaurante de franquicia que se especializa en conos blandos de nieve y helados de frutas. Recientemente ella recibió una carta de la casa matriz corporativa advirtiéndole que su restaurante estaba en peligro de perder su franquicia, porque los promedios de venta por cliente habían bajado "considerablemente por debajo del promedio del resto de la compañía". La frase puede ser verdadera, pero Julie está convencida de que no es válida para justificar una amenaza de cierre. La variación en ventas en su restaurante va camino de crecer más que muchos otros, principalmente porque les da servicio a niños, personas mayores y adultos solteros y no a familias grandes que corren a endeudarse en otros restaurantes. Por tanto, es probable que el promedio de sus comprobantes de venta sea menor y muestre mayor variabilidad. Para demostrar su punto de vista, Julie obtuvo los registros de ventas de toda la compañía y encontró que la desviación estándar era de \$2.45 por comprobante de venta. Ella entonces dirigió un estudio de los últimos 71 comprobantes de venta en su restaurante y encontró una desviación estándar de \$2.95 por comprobante. La variabilidad en ventas en la franquicia de Julie, al nivel de significación de 0.05, ¿es mayor que la variabilidad para la compañía?

9.180 Todos los tomates rojos que cierto supermercado compra de productores deben satisfacer las especificaciones de la tienda, con diámetro de 6.0 cm y desviación estándar de no más de 0.2 centímetros. El comprador del supermercado visita un potencial nuevo proveedor y selecciona una muestra aleatoria de 36 tomates rojos del invernadero del productor. Se mide el diámetro de cada tomate y se encuentra que la media es 5.94 y la desviación estándar es 0.24. ¿Estos tomates rojos satisfacen las especificaciones del supermercado?

a. Determine si es razonable una suposición de normalidad. Explique.

b. ¿La evidencia muestral presenta suficiente evidencia para concluir que los tomates no satisfacen las especificaciones respecto al diámetro medio? Use $\alpha = 0.05$.

c. ¿La evidencia muestral presenta suficiente evidencia para concluir que los tomates no satisfacen las especificaciones respecto a la desviación estándar? Use $\alpha = 0.05$.

d. Escriba un breve informe para el comprador, en donde resuma lo que encuentra y sus recomendaciones en cuanto a si usar este productor para proveer tomates para la venta en el supermercado.

9.181 La longitud uniforme de clavos es muy importante para un carpintero, es decir, la longitud de los clavos que se usan se iguala con los materiales que se unen con los clavos, haciendo con ello que una pequeña desviación estándar sea una propiedad importante de los clavos. Se selecciona al azar una muestra de 35 clavos de 2 pulgadas, tomándola de una gran cantidad de clavos de un lote de reciente producción de Nails,

Inc. Las mediciones de longitud resultantes tienen una longitud media de 2.025 pulgadas y una desviación estándar de 0.048 pulgadas.

a. Determine si es razonable una suposición de normalidad. Explique.

b. ¿La evidencia muestral presenta suficiente evidencia para rechazar la idea de que los clavos tienen una longitud media de 2 pulgadas? Use $\alpha = 0.05$.

c. ¿Hay suficiente evidencia, al nivel de 0.05, para indicar que la longitud de clavos de este lote de producción tiene una desviación estándar mayor a la anunciada de 0.040 pulgadas?

d. Escriba un breve informe que resuma lo que encuentre y sus recomendaciones en cuanto a que si el carpintero debe o no usar estos clavos para una aplicación que requiere clavos de 2 pulgadas.

9.182 Es importante que la fuerza necesaria para extraer un corcho de una botella de vino no tenga una desviación estándar grande. Años de producción y prueba indican que los corchos Núm. 9, del caso práctico 6.15 (p. 330), tienen una fuerza de extracción que está normalmente distribuida con una desviación estándar de 36 newtons. Se considera que cambios recientes en el proceso de manufactura han reducido la desviación estándar.

a. ¿Cuál sería el problema si la desviación estándar es relativamente grande? ¿Cuál sería la ventaja de una desviación estándar más pequeña?

Con fines de prueba se utiliza una muestra de 20 botellas seleccionadas al azar.

Fuerza de extracción en newtons

296	338	341	261	250	347	336	297	279	297
259	334	281	284	279	266	300	305	310	253

b. La muestra precedente, ¿indica suficiente razón para mostrar que la desviación estándar de fuerza de extracción es menor a 36.0 newtons, al nivel de significación de 0.02?

Durante una prueba diferente, se seleccionó y probó una muestra de ocho botellas seleccionadas al azar.

Fuerza de extracción en newtons

331.9	312.0	289.4	303.6	346.9	308.1	346.9	276.0

c. La muestra precedente, ¿indica suficiente razón para mostrar que la desviación estándar de fuerza de extracción es menor a 36.0 newtons, al nivel de significación de 0.02?

d. ¿Qué efecto tuvieron los dos tamaños muestrales diferentes en la estadística de prueba calculada en las partes b y c? ¿Qué efecto tuvieron en el valor p o valor crítico? Explique.

e. ¿Qué efecto tuvieron las dos desviaciones estándar diferentes en las respuestas a las partes b y c? ¿Qué efecto tuvieron en el valor p o valor crítico? Explique.

9.183 Una caja de Corn Flakes marcada "NET WT. 14 OZ." debe tener 14 onzas o más de cereal en su interior. Veinte de estas cajas se seleccionaron al azar y se determinó el peso de los contenidos (en onzas).

14.52	14.47	14.80	14.60	14.45	14.25	14.15	14.12	14.36	14.39
14.50	14.29	14.28	14.60	13.85	14.18	14.39	14.45	14.69	14.38

a. Trace un histograma del peso de cereal por caja.

b. Encuentre la media y desviación estándar de estadísticas muestrales.

c. ¿Qué porcentaje de la muestra está por debajo de 14.0 onzas de peso?

El supervisor de la planta está estudiando el proceso de llenado y necesita estimar el peso medio de todas las cajas que se llenan.

d. Determine si es razonable una suposición de normalidad. Explique.

e. Encuentre el intervalo de confianza de 95% para el peso medio.

f. Se piensa que el proceso de llenado opera con una desviación estándar de llenado de no más de 0.2 onzas. Pruebe esta hipótesis al nivel de 0.01.

9.184 El supervisor del ejercicio 9.183 piensa que la máquina llenadora de cereal que emplea Corn Flakes necesita ser cambiada, y que la nueva que él está considerando pagará la mejora dentro de corto tiempo, principalmente debido a menos variabilidad en la cantidad de llenado. La nueva máquina inicia su operación y se realiza un lote de prueba. Veinte de estas cajas se seleccionaron al azar del lote y se pesó el contenido (en onzas).

14.17	14.25	14.17	14.16	14.18	14.09	14.19	14.17	14.16	14.06
14.11	14.15	14.12	14.19	14.14	14.19	14.13	14.12	14.16	14.15

a. Trace un histograma del peso de cereal por caja.

b. Encuentre la media y desviación estándar de estadísticas muestrales.

c. ¿Qué porcentaje de la muestra de la nueva máquina está por debajo de 14.0 onzas de peso?

El supervisor necesita estimar el peso medio y probar la desviación estándar de todas las cajas que se llenan.

d. Determine si es razonable una suposición de normalidad. Explique.

e. Encuentre el intervalo de confianza de 95% para el peso medio.

f. Se dice que el proceso de llenado para la nueva máquina opera con una desviación estándar de llenado de menos de 0.1 onzas. Pruebe esta hipótesis al nivel de 0.01.

9.185 Las cajas de Corn Flakes de los ejercicios 9.183 y 9.184 que tengan más de 14.2 onzas de cereal están siendo consideradas "demasiado llenas". Como los pesos parecen estar normalmente distribuidos para ambas máquinas llenadoras, use la distribución normal y encuentre la siguiente información para el supervisor.

a. ¿Qué proporción de las cajas llenadas por la máquina actual llena las cajas con demasiado cereal?

b. ¿Qué proporción de las cajas llenadas por la nueva máquina llena las cajas con demasiado cereal?

c. Por cada 1000 cajas de cereal llenadas por la máquina actual, ¿cuántas cajas pueden ser llenadas por la nueva máquina usando la misma cantidad total de cereal?

d. Resuma lo que usted cree que deba ser la recomendación del supervisor para que la compañía adquiera la nueva máquina llenadora.

9.186 Un embotellamiento de tránsito no es un problema menor. En 2003, los viajeros demorados por congestiones de este tipo desperdiciaron 3700 millones de horas, y 2300 millones de galones de combustible para un costo total de $63 mil millones de dólares. Se estima que la mitad de todas las demoras en tránsito son causadas por colisiones entre autos. Una variable que con frecuencia se incluye en estudios de demoras de tránsito es el índice de tiempo de viaje (TTI). El TTI es la relación entre el tiempo en periodo pico y el tiempo en viaje sin atascos. Un valor de 1.30 indica que un viaje sin atascos de 20 minutos toma 26 minutos en el pico, como consecuencia de una fuerte demanda de tránsito e incidentes. El TTI medio nacional en ciudades grandes es 1.37.

Fuente: http://mobility.tamu.edu/ums/report/

En mayo de 2005, se identificaron 25 viajeros de cercanías que por lo general se transportan durante las horas pico de mayor movimiento. Fueron observados cuando viajaban en auto en su ruta regular al trabajo, tanto durante un periodo de viaje sin atascos como durante un periodo pico.

Sin atascos	Pico	Sin atascos	Pico	Sin atascos	Pico
50.7	64.9	21.0	30.5	24.1	34.7
27.4	32.1	51.5	66.4	48.3	64.2
47.9	69.5	50.1	66.6	34.2	40.7
22.9	30.0	48.9	72.4	27.5	35.8
29.0	45.8	36.6	42.1	44.6	57.1
29.7	38.3	35.9	43.1	33.5	44.9
34.4	40.6	38.1	54.9	22.2	34.4
34.4	46.4	31.2	44.9	26.1	32.6
38.7	55.3				

a. Calcule el TTI para cada uno de los viajeros de cercanías.

b. Construya un histograma y encuentre la media y desviación estándar de los TTI.

c. Demuestre la forma en que este conjunto de datos satisface las suposiciones de inferencia.

d. Encuentre el intervalo de confianza de 95% para μ el TTI medio.

e. Pruebe la hipótesis de que el TTI medio para esta zona urbana/suburbana es diferente de la media de 1.37 para grandes ciudades. Use $\alpha = 0.05$.

f. Compare las respuestas halladas en las partes d y e. ¿Piensa usted que estos viajeros de cercanías tienen un problema de tránsito? ¿Por qué? Explique por qué la parte e es una repetición en este caso.

Proyecto del capítulo

¿Se ejercita lo suficiente todos los días?

Numerosos estudios han demostrado que necesitamos hacer ejercicio para bajar diversos riesgos de salud, por ejemplo alta presión arterial, enfermedades cardiacas y colesterol alto. Pero, saber y hacer no son lo mismo. Los profesionales de la salud deberían estar más conscientes de la necesidad de hacer ejercicio. Los siguientes valores de datos son de la sección 9.1, "¿Se ejercita lo suficiente todos los días?" es un estudio que encuesta a técnicos cardiovasculares, personas que efectúan diversos procedimientos de diagnóstico cardiovascular, en cuanto a su propio ejercicio físico por semana (medido en minutos).

60	40	50	30	60	50	90	30	60	60
60	80	90	90	60	30	20	120	60	50
20	60	30	120	50	30	90	20	30	40
50	40	30	40	20	30	60	50	60	80

Trabajando en el contenido del capítulo 9

9.187 a. ¿Qué evidencia tiene usted para demostrar que la suposición de normalidad es razonable? Explique.

b. Estime la cantidad media de tiempo de ejercicio semanal para todos los técnicos cardiovasculares, usando una estimación puntual y un intervalo de confianza de 95%.

c. El artículo "Buenas noticias para mujeres" de la sección 9.1, "¿Se ejercita lo suficiente todos los días?" dice que las personas deben hacer ejercicio al menos 60 minutos por semana. Con base en los datos del estudio, determine si los técnicos se ejercitan al menos 60 minutos por semanas. Use un nivel de significación de 0.05.

Para su estudio

9.188 a. Defina la población en cuya cantidad de tiempo de ejercicio por semana usted estaría interesado en investigar.

b. Recolecte los "tiempos" de una muestra de 40 miembros de su población.

c. Encuentre la media y desviación estándar para la cantidad de tiempo de ejercicio por semana para los miembros de su muestra.

d. Construya una gráfica que muestre la distribución de sus datos.

e. Estime la cantidad media de tiempo de ejercicio por semana para su población usando una estimación puntual y un intervalo de confianza de 95%.

f. El artículo "Buenas noticias para mujeres" dice que las personas (¿pueden?) deben hacer ejercicio al menos 60 minutos por semana. ¿Le parece que los miembros de su muestra se ejercitan "al menos 60 minutos por semana"? Use un nivel de significación de 0.05. Justifique su respuesta.

g. ¿Sus datos satisficieron las suposiciones? Explique.

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

9.1 Las distribuciones t de Student tienen una distribución aproximadamente normal pero están **más** dispersas que la distribución normal estándar.

9.2 La distribución **ji cuadrada** se usa para inferencias acerca de la media cuando σ es desconocida.

9.3 La distribución ***t* de Student** se usa para todas las inferencias acerca de la varianza de una población.

9.4 Si la estadística de prueba cae en la región crítica, la hipótesis nula ha **sido probada verdadera**.

9.5 Cuando la estadística de prueba t y el número de grados de libertad se hacen muy grandes, el valor crítico de t es muy cercano al de la z **normal estándar**.

9.6 Cuando haga inferencias acerca de una media cuando el valor de σ no se conoce, el **puntaje** z es la estadística de prueba.

9.7 La distribución ji cuadrada es una distribución sesgada cuyo valor medio es **2** para df > 2.

9.8 Con frecuencia, el interés para probar la varianza (o desviación estándar) es mantener su tamaño bajo control o relativamente pequeño. Por tanto, muchas de las pruebas de hipótesis con ji cuadrada son **de una cola**.

9.9 $\sqrt{npq}$ es el error estándar de proporción.

9.10 La distribución muestral de p' está distribuida aproximadamente como una distribución ***t* de Student**.

SEGUNDA PARTE: Aplicación de conceptos

Conteste todas las preguntas, mostrando todas las fórmulas, sustituciones y trabajo.

9.11 Encuentre cada valor:

a. $z(0.02)$ b. $t(18, 0.95)$ c. $\chi^2(25, 0.95)$

9.12 Se seleccionó una muestra aleatoria de 25 datos de una población normalmente distribuida, con el fin de estimar la media poblacional, μ. Las estadísticas muestrales son $n = 25$, $\bar{x} = 28.6$, y $s = 3.50$.

a. Encuentre la estimación puntual para ·.

b. Encuentre el máximo error de estimación para la estimación de intervalo de confianza de 0.95.

c. Encuentre el límite inferior de confianza (LCL) y el límite superior de confianza (UCL) para la estimación del intervalo de confianza de 0.95 para μ.

9.13 A miles de estudiantes de escuela primaria de la zona se les aplicó recientemente un examen estandarizado en todo el país para probar sus conocimientos de composición. Si 64 de una muestra aleatoria de 100 estudiantes pasó este examen, construya una estimación de intervalo de confianza de 0.98 para la verdadera proporción de todos los estudiantes de la zona que pasaron el examen.

9.14 Exprese las hipótesis nula (H_o) y alternativa (H_a) que se usaría para probar cada uno de lo siguiente:

a. El peso medio de jugadores profesionales de baloncesto no es mayor a 225 libras.

b. Aproximadamente 40% de estudiantes de turno matutino tienen su propio auto.

c. La desviación estándar para las cantidades mensuales de lluvia en el condado de Monroe es menos de 3.7 pulgadas.

9.15 Determine el nivel de significación, estadística de prueba, región crítica y valor(es) crítico(s) que se usarían para completar cada prueba de hipótesis usando el método clásico con $\alpha = 0.05$.

a. H_o: $\mu = 43$ vs. H_a: $\mu < 43$, $\sigma = 6$

b. H_o: $\mu = 95$ vs. H_a: $\mu \neq 95$, σ desconocida, $n = 22$

c. H_o: $p = 0.80$ vs. H_a: $p > 0.80$

d. H_o: $\sigma = 12$ vs. H_a: $\sigma \neq 12$, $n = 28$

9.16 El fabricante de automóviles del Alero dice que éste promedia 32 millas por galón de gasolina. Un grupo independiente de consumidores está un tanto escéptico de esta frase y piensa que el rendimiento medio en millas por galón es menor a los 32 que dice el fabricante. Una muestra de 24 autos Alero seleccionados al azar produjo estas estadísticas muestrales: media 30.15 y desviación estándar 4.87. Al nivel de significación de 0.05, ¿el grupo de consumidores tiene suficiente evidencia para refutar lo dicho por el fabricante?

9.17 Se supone que una máquina para preparar café dosifica 6 onzas de café en una taza de papel. En realidad, la cantidad dosificada varía de una taza a la otra. No obstante, si la máquina funciona correctamente, la desviación estándar de las cantidades dosificadas debe ser 0.1 onzas o menos. Una muestra aleatoria de 15 tazas produjo una desviación estándar de 0.13 onzas. ¿Representa esto suficiente evidencia, a un nivel de significación de 0.10, para concluir que la máquina no está funcionando correctamente?

9.18 Un cliente nada contento está frustrado con el tiempo de espera en la oficina de correos cuando compra estampillas. Al registrar su queja, le in-

formaron, "Usted espera más de 1 minuto por un servicio y no más de la mitad del tiempo cuando sólo compre estampillas". No creyendo que éste fuera el caso, el cliente recolectó algunos datos de personas que sólo acababan de comprar estampillas. Las estadísticas muestrales son $n = 60$ y $x = n$ (esperar más de 1 minuto) $= 35$. Al nivel de significación de 0.02, ¿nuestro inconforme cliente tiene suficiente evidencia para refutar lo dicho por la oficina de corros?

TERCERA PARTE: Comprensión de conceptos

9.19 El estudiante B dice que el rango de un conjunto de datos puede usarse para obtener una estimación burda para la desviación estándar de una población. La estudiante A no está segura. ¿Cómo explicará el estudiante B correctamente en qué forma y bajo qué circunstancias es que su frase es correcta?

9.20 ¿Es la hipótesis nula o la hipótesis alternativa que el investigador suele creer que es verdadera? Explique.

9.21 Cuando usted rechaza una hipótesis nula, la estudiante A dice que usted está expresando incredulidad en el valor del parámetro como se dice en la hipótesis nula. El estudiante B dice que, en cambio, usted está expresando la creencia de que la estadística muestral provino de una población diferente de la relacionada con el parámetro dicho en la hipótesis nula. ¿Quién tiene razón? Explique.

9.22 "La distribución t de Student debe usarse cuando se hagan inferencias acerca de la media poblacional, μ, cuando la desviación estándar poblacional, σ, no es conocida" es un enunciado verdadero. La estudiante A expresa que el puntaje z a veces desempeña un papel cuando se usa la distribución t. Explique las condiciones que existen y el papel desempeñado por z que hace correcto el enunciado de la estudiante A.

9.23 La estudiante A dice que el porcentaje de las medias muestrales que caen fuera de los valores críticos de la distribución muestral, determinada por una hipótesis nula verdadera, es el valor p para la prueba. El estudiante B dice que el porcentaje que la estudiante A está describiendo es el nivel de significación. ¿Quién tiene razón? Explique.

9.24 La estudiante A lleva a cabo un estudio en el que ella está dispuesta a correr un 1% de riesgo de cometer un error tipo I. Ella rechaza la hipótesis nula y dice que su estadística es significativa al nivel de confianza de 99%. El estudiante B dice que lo dicho por la estudiante A no está correctamente expresado. ¿Quién tiene razón? Explique.

9.25 La estudiante A dice que cuando usted emplea un intervalo de confianza de 95% para determinar una estimación, usted no sabe con certeza si su inferencia es o no es correcta (el parámetro está contenido dentro del intervalo). El estudiante B dice que usted sí sabe; usted ha demostrado que el parámetro no puede ser menor que el límite inferior o mayor que el límite superior del intervalo. ¿Quién tiene razón? Explique.

9.26 La estudiante A dice que la mejor forma de mejorar la estimación de un intervalo de confianza es aumenta el nivel de confianza. El estudiante B dice que usar un alto nivel de confianza no mejora realmente la estimación del intervalo resultante. ¿Quién tiene razón? Explique.

CAPÍTULO

10 Inferencias que involucran a dos poblaciones

10.1 Estudiantes, tarjetas de crédito y débito

10.2 Muestras dependientes e independientes

10.3 Inferencias relacionadas con la diferencia de medias usando dos muestras dependientes

10.4 Inferencias relacionadas con las diferencias de medias usando dos muestras independientes

10.5 Inferencias relacionadas con las diferencias entre proporciones usando dos muestras independientes

10.6 Inferencias relacionadas con la razón de varianzas usando dos muestras independientes

© David Young-Wolff/Getty Images

10.1 Estudiantes, tarjetas de crédito y débito

Todos sabemos que "la universidad es costosa" y que "las tarjetas de crédito son fácilmente asequibles". También sabemos y pensamos que los jóvenes necesitan experiencia en el manejo de sus propias finanzas, pero es probable que dos viejos adagios "Compradores, cuidado" y "Conozca la información" desempeñen cada uno de ellos su función en el modo que tenga un estudiante universitario de manejar su tarjeta de crédito. A continuación veamos una parte del informe de Nellie Mae publicado en 2002.

TARJETAS DE CRÉDITO Y DÉBITO

No es de sorprender que la población de estudiantes de primer año tenga un porcentaje general más bajo de tarjetas de crédito, y niveles más bajos de deuda en sus tarjetas, que los estudiantes de clases superiores. Más de la mitad de todos los estudiantes de primer año (54%) tenía al menos una tarjeta de crédito, siendo 2.5 el número promedio de tarjetas; entre los que tienen tarjetas de crédito, 26% tienen cuatro o más. Los niveles de deuda también son más bajos que las cuentas generales en todas las categorías. El importe de su deuda mediana es $901, más baja que la mediana general de $1770; su saldo promedio es $1533 contra el general de $2327; los que tienen saldos superiores a $7000 son sólo 4% de estudiantes de primer año contra el 6% general; y los que tienen saldos de alto nivel entre $3000 y $7000 son 8% de estudiantes de primer año en comparación con el 21% general.

Uso de tarjeta de crédito por grado	1er. año	2o. año	3er. año	4o. año o más
Porcentaje que tienen tarjetas de crédito	54%	92%	87%	96%
Número promedio de tarjetas de crédito	2.5	3.67	4.5	6.13
Porcentaje que tienen 4 o más tarjetas	26%	44%	50%	66%
Deuda promedio de tarjeta de crédito	$1533	$1825	$2705	$3262
Deuda mediana de tarjeta de crédito	$901	$1564	$1872	$2185
Porcentaje con saldos entre $3000 y $7000	8%	18%	24%	31%
Porcentaje con saldos de más de $7000	4%	4%	7%	9%

A medida que los estudiantes avanzan en sus 4 años (o más) de universidad, hay un continuo aumento en el uso de tarjetas de crédito y saldos por año. Al graduarse, casi todos los estudiantes han más que duplicado su deuda promedio y casi triplicado el número de tarjetas que poseen. Lo más impresionante, sin embargo, es el salto de 70% que hay entre el

primero y segundo año en el porcentaje de estudiantes con al menos una tarjeta de 54% a 92% de la población total.

Una vez que el estudiante de primer año llega a las instalaciones de la universidad, hay muchos tentadores incentivos para firmar nuevas tarjetas de crédito y numerosas oportunidades para usarlas. El hecho de que el número promedio de tarjetas de crédito por estudiante continúa en aumento no es de sorprender. La proliferación de ofertas en el campus universitario, por correo e Internet para regalos gratis, bonos de millas de líneas aéreas y bajas tarifas de introducción para cada nueva tarjeta es difícil de resistir para estudiantes.

Fuente: Pasantes y tarjetas de crédito, publicado en abril de 2002, Nellie Mae, 50 Braintree Hill Park, Suite 3000, Braintree, MA 02184; 781-849-1325; http://www.nelliemae.com

Los datos que siguen son de muestras aleatorias de 200 estudiantes de primer año y 200 de segundo año a quienes se preguntó, "¿Tiene su propia tarjeta de crédito?" Un total de 97 estudiantes de primer año y 187 de segundo dijeron que tenían una o más tarjetas de crédito con su nombre en ellas. A los primeros 40 estudiantes de primer año, y los primeros 44 de segundo que contestaron "sí", se les preguntó entonces su "saldo actual total de deuda de tarjeta de crédito". A continuación veamos los saldos totales de deuda de tarjetas de crédito.

Saldo de deuda de estudiante de primer año, $n = 40$

1011.97	3998.72	2447.93	2457.39	855.63	1602.74	912.39	2478.49	1014.39	444.48	1293.36
1065.82	989.56	412.53	321.85	2578.39	2103.35	2917.65	3218.54	1384.34	4368.28	244.33
190.24	2778.17	1702.65	616.31	491.73	2205.95	1130.09	2402.92	767.42	657.83	1150.78
1102.28	154.11	1494.48	1324.01	2054.76	1762.31	644.31				

Saldo de deuda de estudiante de segundo año, $n = 44$

690.08	595.04	2983.50	1761.21	1020.91	2143.18	3048.87	1314.36	1378.99	1456.10	1893.37	1287.47
284.93	7135.64	3194.07	2565.71	3298.15	2747.14	839.57	393.20	1422.73	1652.03	2214.77	1126.76
3433.80	3962.25	1849.23	3037.52	328.29	3074.19	1194.87	889.40	1480.94	486.22	1688.81	1317.27
2624.01	2286.74	5341.94	633.37	873.18	3601.18	2023.29	4898.46				

Después de completar el capítulo 10, investigue más a fondo los datos precedentes y "Estudiantes, tarjetas de crédito y deuda" en el proyecto del capítulo en la página 613.

SECCIÓN 10.1 EJERCICIOS

10.1 Considere los datos muestrales citados líneas antes en la sección 10.1.

a. ¿Cuál es la población de interés?

b. ¿Qué porcentaje de cada grupo tiene su propia tarjeta de crédito? ¿Cómo se compara con lo encontrado e informado por la organización de Nellie Mae?

c. Describa la forma de la distribución que exhibirá la deuda total de tarjetas de crédito. Explique.

d. Construya un histograma de deuda total de tarjetas de crédito para cada clase. Use los mismos intervalos de clase para ambos histogramas. Compare lo que encuentre con la idea que tenga en la parte c.

e. Encuentre la media y desviación estándar para cada conjunto de datos. Compare las dos muestras usando estos hallazgos.

f. Compare la distribución de deudas de tarjetas de crédito para estudiantes de primer año con la de estudiantes de segundo año. Describa la forma en que parecen ser semejantes. Describa formas en las que parecen ser por completo diferentes.

10.2 a. Estime la deuda media de tarjeta de crédito para estudiantes de primer año, con un intervalo de confianza de 95%.

b. La población muestreada de estudiantes de segundo año, ¿tiene una deuda significativamente más alta en tarjeta de crédito que el promedio nacional informado por Nellie Mae? Use $\alpha = 0.05$.

FIGURA 10.1 "Mapa de caminos" para dos inferencias poblacionales

10.2 Muestras dependientes e independientes

En este capítulo vamos a estudiar los procedimientos para hacer inferencias acerca de dos poblaciones. Cuando comparemos dos poblaciones, necesitamos dos muestras, una de cada población. Pueden usarse dos clases básicas de muestra: independiente y dependiente. La dependencia o independencia de dos muestras está determinada por las fuentes de los datos. Una **fuente** puede ser una persona, un objeto, o cualquier cosa que produzca un valor de datos. Si el mismo conjunto de fuentes o conjuntos relacionados se usan para obtener los datos que representen ambas poblaciones, tenemos **muestras dependientes**. Si se usan dos conjuntos

no relacionados, un conjunto de cada población, tenemos **muestras independientes**. Los siguientes ejemplos deben aclarar estas ideas.

EJEMPLO 10.1 Muestras dependientes contra independientes

Se efectuará una prueba para ver si los participantes de un grupo de acondicionamiento físico en realidad mejoran en su nivel de condición física. Se anticipa que aproximadamente 500 personas se inscribirán en este curso. La instructora decide que dará a 50 de los participantes un conjunto de pruebas antes de iniciar el curso (examen previo), y luego aplicará otro conjunto de pruebas a 50 participantes al final del curso (después del examen). Se proponen dos procedimientos de muestreo:

Plan A: Seleccionar al azar 50 participantes de la lista de inscritos y aplicarles el examen previo. Al final del curso, hacer una segunda selección aleatoria de tamaño 50 y aplicarles el examen subsiguiente.

Plan B: Seleccionar al azar 50 participantes y aplicarles el examen previo; darles al mismo conjunto de 50 el examen subsiguiente cuando terminen el curso.

El Plan A ilustra muestreo independiente; las fuentes (los participantes del grupo) empleadas para cada muestra (examen previo y examen subsiguiente) se seleccionaron por separado. El Plan B ilustra muestreo dependiente; las fuentes empleadas para ambas muestras (examen previo y examen subsiguiente) son las mismas.

Típicamente, cuando se usan el examen previo y el subsiguiente, participan las mismas personas en el estudio. Así, los estudios de examen previo contra examen subsiguiente (antes contra después) usan muestras dependientes.

EJEMPLO 10.2 Muestras dependientes contra independientes

Se está diseñando una prueba para comparar la resistencia al desgaste de dos marcas de llantas para automóvil. Los automóviles se seleccionarán y equiparán con las nuevas llantas y luego se manejarán bajo condiciones "normales" durante un mes. A continuación se tomará una medición para determinar cuánto desgaste hubo. Se proponen dos planes:

Plan C: Una muestra de autos se seleccionará en forma aleatoria, se equipará con llantas de la marca A, y se manejará durante un mes. Se seleccionará otra muestra de autos, equipada con llantas de la marca B y se manejará durante un mes.

Plan D: Una muestra de autos se seleccionará en forma aleatoria, se equipará con una llanta de la marca A y una de la marca B (las otras dos llantas no son parte de la prueba), y se manejará durante un mes.

Sospechamos que muchos otros factores deben tomarse en cuenta cuando se prueban llantas para automóvil, por ejemplo edad, peso y condiciones mecánicas del auto; hábitos de manejo de conductores; ubicación de la llanta en el auto; así como dónde y cuánto se maneja el auto. No obstante, en este caso estamos tratando sólo de ilustrar muestras dependientes e independientes. El Plan C es independiente (fuentes no relacionadas), y el plan D es dependiente (fuentes comunes).

CASO PRÁCTICO 10.3 Exploración de facciones de gemelos

Los estudios donde participan gemelos idénticos son naturales para la técnica de muestreo dependiente estudiada en esta sección.

NUEVO ESTUDIO MUESTRA QUE PUEDEN HEREDARSE CARACTERÍSTICAS CLAVE

Al igual que muchos gemelos idénticos criados aparte, Jim Lewis y Jim Springer encontraron que habían estado llevando vidas extrañamente semejantes. Separados cuatro semanas después de su nacimiento en 1940, los gemelos Jim crecieron en lugares situados a 45 millas entre sí en Ohio y se volvieron a unir en 1979. A final de cuentas descubrieron que ambos manejaban el mismo modelo de Chevrolet azul, fumaban sin parar cigarrillos Salem, se mordían las uñas y tenían perros a los que llamaban Toy (Juguete). Cada uno pasaba buen tiempo de vacaciones en la misma playa de tres manzanas de largo en Florida. Lo que es más importante, cuando se sometieron a pruebas de características de personalidad como son la flexibilidad, autocontrol y sociabilidad, los gemelos respondieron casi exactamente iguales.

El proyecto es considerado el más completo de su clase. Investigadores de Minnesota informan los resultados de pruebas de seis días de sus sujetos, incluyendo 44 pares de gemelos idénticos que fueron criados aparte. Se encontró que vivir bien, no tener amigos, ser agresivos y apartarse de riesgos o peligros se debía tanto o más a la naturaleza que a la crianza. De once características clave de grupos de características analizadas en el estudio, los investigadores estimaron que un elevado 61% de lo que llaman "potencia de sociedad" (tendencia hacia liderazgo o dominio) es hereditario, mientras que la "cercanía social" (necesidad de intimar, confort y ayuda) fue más bajo, con un 33%.

Fuente: Time Inc. Todos los derechos reservados. Reimpreso con permiso de *TIME*.

Las muestras independientes y dependientes tienen cada una de ellas sus ventajas; aquí las veremos más adelante. Es frecuente que se usen ambos métodos.

SECCIÓN 10.2 EJERCICIOS

10.3 Explique por qué estudios que comprenden a gemelos idénticos, como en el caso práctico 10.3, resultan en muestras dependientes de datos.

10.4 a. Describa la forma en que podría usted seleccionar dos muestras independientes, de entre sus condiscípulos, para comparar las estaturas de estudiantes mujeres y hombres.

b. Describa la forma en que podría usted seleccionar dos muestras independientes, de entre sus condiscípulos, para comparar las estaturas de cuando entraron a preparatoria con sus estaturas y cuando entraron a universidad con sus estaturas.

10.5 A los estudiantes de una preparatoria local se les asignó hacer un proyecto para su clase de estadística. El proyecto comprendía hacer que estudiantes de primer año tomaran un examen de conceptos geométricos, tomándoles el tiempo para ello. Los estudiantes de estadística usarían entonces estos datos para determinar si había diferencia entre el rendimiento de hombres y mujeres. ¿Los conjuntos resultantes representarían muestras dependientes o independientes? Explique.

10.6 Al tratar de estimar la cantidad de crecimiento que tuvo lugar en los árboles recientemente plantados por la comisión de partes del condado, se seleccionaron al azar 6 árboles de los 4000 plantados. Las alturas de estos árboles se midieron y registraron. Un año después, otro conjunto de 42 árboles se seleccionó al azar y se midió. ¿Estos dos conjuntos de datos (36 alturas, 42 alturas) representan muestras dependientes o independientes? Explique.

10.7 Se seleccionaron 20 personas para que participaran en un experimento de psicología. Contestaron un breve cuestionario de opciones múltiples acerca de

sus actitudes en un tema en particular y luego vieron una película de 45 minutos. Al día siguiente se pidió a las mismas 20 personas contestaran un cuestionario de seguimiento acerca de sus actitudes. Al terminar el experimento, el investigador tendrá dos conjuntos de calificaciones. ¿Estas dos muestras representan muestras dependientes o independientes? Explique.

10.8 Un experimento está diseñado para estudiar el efecto que una dieta tiene en el nivel de ácido úrico. El estudio incluye 20 ratas blancas. Diez ratas se seleccionaron al azar y se les dieron dietas de comida chatarra; a las otras 10 ratas se les dio una dieta alta en fibras y baja en grasas. Se determinaron los niveles de ácido úrico de los dos grupos. ¿Los conjuntos de datos resultantes representan muestras dependientes o independientes? Explique.

10.9 Dos tipos diferentes de centrifugadoras de discos se usan para medir el tamaño de partículas en pintura de látex. Se selecciona al azar un galón de pintura, y de él se toman 10 especímenes para probar cada una de las centrifugadoras. Habrá dos conjuntos de datos, 10 valores de datos para cada una, como resultado de la prueba. ¿Los dos conjuntos de datos representan muestras dependientes e independientes? Explique.

10.10 Una compañía de seguros está preocupada de que el garaje A cobra más por reparaciones de lo que cobra el garaje B. Planea enviar 25 autos a cada garaje y obtener estimaciones separadas para las reparaciones necesarias para cada auto.

a. ¿Cómo puede hacer esto la compañía y obtener muestras independientes? Explique en detalle.

b. ¿Cómo puede hacer esto la compañía y obtener muestras dependientes? Explique en detalle.

10.11 Se está diseñando un estudio para determinar las razones por las que los adultos escogen seguir un plan de dieta sana. El estudio encuestará a 1000 hombres y 1000 mujeres. Al terminar el estudio, las razones por las que los hombres escogen una dieta sana se van a comparar con las razones por las que las mujeres escogen una dieta sana.

a. ¿Cómo pueden recolectarse datos si han de obtenerse muestras independientes? Explique en detalle.

b. ¿Cómo pueden recolectarse datos si han de obtenerse muestras dependientes? Explique en detalle.

10.12 Suponga que 400 estudiantes de cierta universidad están tomando clase de estadística elemental este semestre. Se necesitan dos muestras de tamaño 25 para probar algún conocimiento anterior al curso contra el mismo conocimiento después que los estudiantes terminen el curso.

a. Describa la forma en que usted podría obtener sus muestras si fuera a usar muestras dependientes.

b. Describa la forma en que usted podría obtener sus muestras si fuera a usar muestras independientes.

10.3 Inferencias relacionadas con la diferencia de medias usando dos muestras dependientes

Los procedimientos para comparar dos medias poblacionales están basados en la relación entre dos conjuntos de datos muestrales, una muestra de cada población. Cuando intervienen muestras dependientes, los datos se consideran como "datos pareados". Los datos pueden ser pareados como resultado de ser obtenidos de estudios de "antes" y "después"; de pares de gemelos idénticos como en el caso práctico 10.3; de una fuente "común", como con las cantidades de desgaste en llantas para cada marca en el plan D del ejemplo 10.2; o de comparar dos personas con características similares para formar "pares igualados". Los pares de valores de datos se comparan directamente entre sí usando la diferencia en sus valores numéricos. La diferencia resultante se denomina **diferencia pareada**.

Diferencia pareada

$$d = x_1 - x_2 \tag{10.1}$$

El uso de datos pareados en esta forma tiene una capacidad integrada para remover el efecto de factores que de otro modo no son controlados. El problema del desgaste de llantas del ejemplo 10.2 es un excelente ejemplo de estos factores adicionales. La resistencia al desgaste de una llanta resulta afectada en gran medida por una multitud de factores como son tamaño, peso, edad y condiciones del auto; los hábitos de manejo del conductor; el número de millas recorridas; la condición y tipos de carreteras sobre las que circula; la calidad del material empleado para hacer la llanta; etcétera. Creamos datos pareados al montar una llanta de cada marca en el mismo auto. Dado que se probará una llanta de cada marca bajo las mismas condiciones, usando el mismo auto, el mismo conductor, etcétera, se neutralizan las causas extrañas del desgaste.

Procedimientos y suposiciones de inferencia que comprenden datos pareados

Se llevó a cabo una prueba para comparar la resistencia al desgaste de las llantas producidas por dos empresas fabricantes de llantas, usando el plan D, como se describe en el ejemplo 10.2. Todos los factores citados líneas antes tendrán un efecto igual en ambas marcas de llantas, auto por auto. Una llanta de cada marca se puso en cada uno de los seis autos de prueba. La posición (lado izquierdo o derecho, delantera o trasera) fue determinada con ayuda de una tabla de números aleatorios. La tabla 10.1 detalla las cantidades de desgaste (en milésimos de pulgada) que resultaron de la prueba.

TABLA 10.1

Cantidad de desgaste en llanta

Auto	1	2	3	4	5	6
Marca A	125	64	94	38	90	106
Marca B	133	65	103	37	102	115

Como los diversos autos, conductores y condiciones son los mismos para cada llanta de un conjunto pareado de datos, tiene sentido usar una tercera variable, la diferencia pareada ***d***. Nuestras dos muestras dependientes de datos pueden combinarse en un conjunto de valores d, donde $d = \text{B} - \text{A}$.

Auto	1	2	3	4	5	6
$d = \text{B} - \text{A}$	8	1	9	−1	12	9

La diferencia entre las dos medias poblacionales, cuando se usan muestras dependientes (a veces llamadas **medias dependientes**), es equivalente a la **media de las diferencias pareadas**. Por tanto, cuando se ha de hacer una inferencia acerca de la diferencia de dos medias y se usan diferencias pareadas, la inferencia de

hecho será acerca de la media de las diferencias pareadas. La media muestral de las diferencias pareadas se usará como estimación puntual para estas inferencias.

Para hacer inferencias acerca de la media de todas las diferencias pareadas posibles μ_d, necesitamos conocer la *distribución muestral* de $\bar{d}$.

Cuando al azar se seleccionan observaciones pareadas de poblaciones normales, la diferencia pareada, $d = x_1 - x_2$, estará normalmente distribuida en forma aproximada alrededor de la media μ_d con una desviación estándar de σ_d.

Ésta es otra situación en la que se aplica la prueba t para una media; es decir, deseamos hacer inferencias acerca de una media desconocida (μ_d) donde la variable aleatoria (d) involucrada tiene una distribución aproximadamente normal con una desviación estándar desconocida (σ_d).

Las inferencias acerca de la media de todas las posibles diferencias pareadas μ_d están basadas en muestras de n pares dependientes de datos y la **distribución *t*** con $n - 1$ grados de libertad (df), bajo la siguiente suposición:

Suposición para inferencias acerca de la media de diferencias pareadas μ_d:
Los datos pareados se seleccionan al azar de poblaciones normalmente distribuidas.

Procedimiento de intervalo de confianza

El **intervalo de confianza** $1 - \alpha$ **para estimar la diferencia de media** μ_d se encuentra usando esta fórmula:

PARA SU INFORMACIÓN
La fórmula (10.2) es una adaptación de la fórmula (9.1).

Intervalo de confianza para diferencia de media (muestras dependientes)

$$\bar{d} - t(\text{df}, \alpha/2) \cdot \frac{s_d}{\sqrt{n}} \quad \text{a} \quad \bar{d} + t(\text{df}, \alpha/2) \cdot \frac{s_d}{\sqrt{n}}, \quad \text{donde df} = n - 1 \tag{10.2}$$

Donde $\bar{d}$ es la media de las diferencias muestrales:

$$\bar{d} = \frac{\sum d}{n} \tag{10.3}$$

PARA SU INFORMACIÓN
Las fórmulas (10.3) y (10.4) son adaptaciones de las fórmulas (2.1) y (2.10).

y s_d es la desviación estándar de las diferencias muestrales:

$$s_d = \sqrt{\frac{\sum d^2 - \left[\frac{(\sum d)^2}{n}\right]}{n - 1}} \tag{10.4}$$

EJEMPLO 10.4

Construcción de un intervalo de confianza para μ_d

Construya el intervalo de confianza de 95% para la diferencia de media de los datos pareados sobre desgaste de llantas, como se informa en la tabla 10.1. La informa-

ción muestral es $n = 6$ piezas de datos pareados $\bar{d} = 6.3$, y $s_d = 5.1$. Suponga que las cantidades de desgaste están normalmente distribuidas en forma aproximada para ambas marcas de llantas.

SOLUCIÓN

Paso 1 **Parámetro de interés:** μ_d, la diferencia media de las cantidades de desgaste entre las dos marcas de llantas.

Paso 2

a. **Suposiciones:** Ambas poblaciones muestreadas son aproximadamente normales.

b. **Distribución de probabilidad:** Se usarán la distribución t con df = 6 − 1 = 5 y la fórmula (10.2).

c. **Nivel de confianza:** $1 - \alpha = 0.95$.

Paso 3 **Información muestral:** $n = 6$, $\bar{d} = 6.3$, y $s_d = 5.1$.

La media:

$$\bar{d} = \frac{\sum d}{n} : \quad \bar{d} = \frac{38}{6} = 6.333 = \mathbf{6.3}$$

La desviación estándar:

$$s_d = \sqrt{\frac{\sum d^2 - \left[\frac{(\sum d)^2}{n}\right]}{n-1}} : \quad s_d = \sqrt{\frac{372 - \left[\frac{(38)^2}{6}\right]}{6-1}} = \sqrt{26.27} = 5.13 = \mathbf{5.1}$$

Paso 4

a. **Coeficiente de confianza:**

Ésta es una situación de dos colas con $\alpha/2 = 0.025$ en una cola. De la tabla 6 del apéndice B, $t(\text{df}, \alpha/2) = t(5, 0.025) = \mathbf{2.57}$.

b. **Máximo error de estimación:** Usando la parte del máximo error de la fórmula (10.2), tenemos

$$E = t(\text{df}, \alpha/2) \cdot \frac{s_d}{\sqrt{n}} : \quad E = 2.57 \cdot \left(\frac{5.1}{\sqrt{6}}\right) = (2.57)(2.082) = 5.351 = \mathbf{5.4}$$

c. **Límites inferior/superior de confianza:**

$$\bar{d} \pm E$$

$$6.3 \pm 5.4$$

$$6.3 - 5.4 = \mathbf{0.9} \quad \text{a} \quad 6.3 + 5.4 = \mathbf{11.7}$$

Paso 5

a. **Intervalo de confianza:** 0.9 a 11.7 es el intervalo de confianza de 95% para μ_d.

b. Esto es, con 95% de confianza podemos decir que la diferencia media en las cantidades de desgaste es entre 0.9 y 11.7 milésimos de pulgada. O bien, dicho de otro modo, el desgaste medio poblacional de llantas de la Marca B es entre 0.9 y 11.7 milésimos de pulgada más que el desgaste medio poblacional de llantas de la Marca A.

Nota: este intervalo de confianza es bastante amplio, en parte debido al pequeño tamaño muestral. Recuerde del teorema de límite central que cuando aumenta el tamaño muestral, el error estándar (estimado por $s_d/\sqrt{n}$) disminuye.

Para instrucciones específicas acerca de coeficientes de confianza y la tabla 6, vea las páginas 477-478.

INSTRUCCIONES DE TECNOLOGÍA: INTERVALO DE CONFIANZA $1 - \alpha$ PARA LA MEDIA μ_d CON DESVIACIÓN ESTÁNDAR DESCONOCIDA PARA DOS CONJUNTOS DEPENDIENTES DE DATOS MUESTRALES

MINITAB (Versión 14) Introduzca los datos pareados en C1 y C2; luego continúe con:

```
Choose:     Stat > Basic Statistics > Paired t
Select:     Samples in columns
Enter:      First sample: C1*
            Second sample: C2
Select:     Options
Enter:      Confidence level: 1 − α (ex. 0.95 or 95.0)
Select:     Alternative: not equal > OK > OK
```

*La *t* pareada evalúa la primera muestra menos la segunda muestra.

Excel Introduzca los datos pareados en las columnas A y B; active C1 y C2 (dependiendo de si se usan o no se usan encabezados de columna); luego continúe con:

```
Enter:      = A2 − B2* (if column headings are used)
Drag:       Bottom right corner of C2 down to give other differences
Choose:     Tools > Data Analysis Plus > t-Estimate: Mean
Enter:      Input range: (C2:C20 or select cells)
Select:     Labels (if necessary)
Enter:      Alpha: α (ex. 0.05) > OK
```

*Introduzca la expresión en el orden necesario: A2 − B2 y B2 − A2.

TI-83/84 Plus Introduzca los datos pareados en L1 y L2; luego continúe con lo siguiente, introduciendo los valores apropiados y resaltando Calculate:

```
Highlight:  L3
Enter:      L3 = L1 − L2*
Choose:     STAT > TESTS > 8:Tinterval
```

*Introduzca la expresión en el orden necesario: L1 − L2 o L2 − L1.

La solución al ejemplo 10.4 se asemeja a lo siguiente cuando se resuelva en MINITAB:

```
Paired T for Brand B − Brand A
             N      Mean      StDev      SE Mean
Brand B      6      92.5      35.2       14.4
Brand A      6      86.2      30.9       12.6
Difference   6      6.33      5.13       2.09
95% CI for mean difference: (0.95, 11.71)
```

Procedimiento de prueba de hipótesis

Cuando probamos una **hipótesis nula acerca de la diferencia media**, la estadística de prueba empleada será la diferencia entre la media muestral $\bar{d}$ y el valor hipotético de μ_d, dividido entre el **error estándar** estimado. Se supone que esta

estadística tiene una distribución t cuando la hipótesis nula es verdadera y las suposiciones para la prueba se satisfacen. El valor de la **estadística de prueba** $t\bigstar$ se calcula como sigue:

PARA SU INFORMACIÓN
La fórmula (10.5) es una adaptación de la fórmula (9.2).

Estadística de prueba para diferencia de media (muestras dependientes)

$$t\bigstar = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}, \quad \text{donde df} = n - 1 \tag{10.5}$$

Nota: una diferencia hipotética de media, μ_d, puede ser cualquier valor especificado. El valor especificado más común es cero pero la diferencia puede ser diferente de cero.

EJEMPLO 10.5 Prueba de hipótesis de una cola para μ_d

En un estudio de alta presión sanguínea y los medicamentos empleados para controlarla, el efecto de bloqueadores de canal de calcio en el ritmo cardiaco fue una de las inquietudes más específicas. Al azar se seleccionaron 26 pacientes de un numeroso grupo de potenciales sujetos de estudio, y se registraron sus ritmos cardiacos. Se aplicó un bloqueador de calcio a cada paciente durante un tiempo fijo, y luego se determinó de nuevo el ritmo cardiaco de cada paciente. Los dos conjuntos de datos resultantes parecieron tener distribuciones aproximadamente normales, y las estadísticas fueron $\bar{d} = 1.07$ y $s_d = 1.74$ (d = antes − después). ¿La información muestral da suficiente evidencia para demostrar que el ritmo cardiaco es menor después de tomar el medicamento? Use $\alpha = 0.05$.

PARA SU INFORMACIÓN
"Ritmo más bajo" significa "después" y es menor que "antes" y "antes − después" es positivo.

SOLUCIÓN

Paso 1
a. **Parámetro de interés:** μ_d, la diferencia media (reducción) en pulso cardiaco de antes a después, usando el bloqueador de canal de calcio durante el tiempo de la prueba.
b. **Enunciado de las hipótesis:**
H_o: $\mu_d = 0$ ($\leq$) (no bajó el ritmo cardiaco)
Recuerde: d = antes − después.
H_a: $\mu_d > 0$ (bajó el ritmo cardiaco)

Paso 2
a. **Suposiciones:** Como los datos de ambos conjuntos son aproximadamente normales, parece razonable suponer que las dos poblaciones están normalmente distribuidas en forma aproximada.
b. **Estadística de prueba:** La distribución t con df $= n - 1 = 25$, y la estadística de prueba es $t\bigstar$ de la fórmula (10.5).
c. **Nivel de significación:** $\alpha = 0.05$.

Paso 3
a. **Información muestral:** $n = 26$, $\bar{d} = 1.07$, y $s_d = 1.74$.
b. **Estadística calculada de prueba:**

$$t\bigstar = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}: \quad t\bigstar = \frac{1.07 - 0.0}{1.74/\sqrt{26}} = \frac{1.07}{0.34} = \mathbf{3.14}$$

Paso 4 **La distribución de probabilidad:**

Valor *p*:

a. Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor que". **P** = $P(t\bigstar > 3.14$, con df = 25) como se ve en la figura.

Para hallar el valor *p*, tiene tres opciones:

1. Usar la tabla 6 (apéndice B): **P < 0.005.**
2. Usar la tabla 7 (apéndice B) para leer el valor directamente: **P = 0.002.**
3. Usar computadora o calculadora para hallar el valor *p*: **P = 0.0022.**

Las instrucciones específicas están en las páginas 483-484.

b. El valor *p* es menor que el nivel de significación, α.

O

Clásico:

a. La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "mayor que". El valor crítico se obtiene de la tabla 6: $t(25, 0.05)$ = **1.71.**

Las instrucciones específicas están en páginas 477-478.

b. $t\bigstar$ está en la región crítica, como se muestra en el **área sombreada** en la figura.

Paso 5

a. Decisión: Rechazar H_o.

b. Conclusión: Al nivel de significación de 0.05, podemos concluir que el promedio de ritmo cardiaco es menor después de administrar el bloqueador de canal de calcio.

La *significación estadística* no siempre tiene el mismo significado cuando se considera la aplicación "práctica" de resultados. En la precedente prueba detallada de hipótesis, los resultados mostraron una significación estadística con un valor *p* de 0.002, es decir, 2 oportunidades en 1000. No obstante, una pregunta más práctica podría ser: ¿bajar el ritmo cardiaco en esta pequeña cantidad promedio, estimado en 1.07 pulsos por minuto, merece la pena correr los riesgos de posibles efectos colaterales de este medicamento? En realidad todo el problema es mucho más amplio que sólo este ritmo cardiaco.

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA LA MEDIA μ_d CON DESVIACIÓN ESTÁNDAR DESCONOCIDA PARA DOS CONJUNTOS DEPENDIENTES DE DATOS MUESTRALES.

MINITAB (Versión 14) Introduzca los datos pareados en C1 y C2; luego continúe con:

```
Choose:     Stat > Basic Statistics > Paired t
Select:     Samples in columns
Enter:      First sample: C1*
            Second sample: C2
Select:     Options
Enter:      Test mean: 0.0 or μd
Select:     Alternative: less than or not equal or greater than > OK > OK
```

*La *t* pareada evalúa la primera muestra menos la segunda muestra.

Excel Introduzca los datos pareados en las columnas A y B; luego continúe con:

```
Choose:      Tools > Data Analysis > t-Test: Paired Two Sample for Means
Enter:       Variable 1 Range: (A1:A20 or select cells)
             Variable 2 Range: (B1:B20 or select cells)
                               (subtracts: Var1 − Var2)
             Hypothesized Mean Difference: μd (usually 0)
Select:      Labels (if necessary)
Enter:       α (ex. 0.05)
Select:      Output Range
Enter:       (C1 or select cell) > OK
```

Use Format > Column > AutoFit Selection para hacer que la salida sea más fácil de leer. La salida muestra valores p y valores críticos para pruebas de una cola y dos colas. La prueba de hipótesis también se puede hacer si se restan primero las dos columnas, y luego se usan los comandos de inferencia alrededor de la media (sigma desconocida) de la página 487 sobre las diferencias.

TI-83/84 Plus Introduzca los datos pareados en L1 y L2; luego continúe con lo siguiente, introduciendo los valores apropiados, y resaltando Calculate:

```
Highlight:   L3
Enter:       L3 = L1 − L2*
Choose:      STAT > TESTS > 2:T-Test . . .
```

*Introduzca la expresión en el orden necesario: L1 − L2 o L2 − L1.

La solución al ejemplo 10.5 se parece a lo siguiente cuando se resuelve en MINITAB:

```
Paired T for Before − After
                N      Mean      StDev      SE Mean
Difference     26      1.07      1.74       0.34
T-Test of mean difference = 0 (vs > 0): T-Value = 3.14
P-Value = 0.002
```

EJEMPLO 10.6 Prueba de hipótesis de dos colas para μ_d

Suponga que los datos muestrales de la tabla 10.1 (p. 551) se recolectaron con la esperanza de mostrar que las dos marcas de llantas no se desgastan igual. ¿Los datos dan suficiente evidencia para concluir que las dos marcas muestran desgaste desigual, al nivel de significación de 0.05? Suponga que las cantidades de desgaste están normalmente distribuidas en forma aproximada para ambas marcas de llantas.

SOLUCIÓN

Paso 1 **a. Parámetro de interés:** μ_d, la diferencia media en las cantidades de desgaste entre las dos marcas.

b. Enunciado de las hipótesis:

H_o: $\mu_d = 0$ (no hay diferencia) Recuerde: $d = \text{B} - \text{A}$

H_a: $\mu_d \neq 0$ (hay diferencia)

Paso 2 **a. Suposiciones:** La suposición de normalidad está incluida en el enunciado de este problema.

b. Estadística de prueba: La distribución t con df $= n - 1 = 6 - 1 = 5$, y $t\bigstar = (\bar{d} - \mu_d)/(s_d/\sqrt{n})$.

c. Nivel de significación: $\alpha = 0.05$.

Paso 3 **a. Información muestral:** $n = 6$, $\bar{d} = 6.3$ y $s_d = 5.1$.

b. Estadística de prueba calculada:

$$t\bigstar = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}: \quad t\bigstar = \frac{6.3 - 0.0}{5.1/\sqrt{6}} = \frac{6.3}{2.08} = \mathbf{3.03}$$

Paso 4 **Distribución de probabilidad:**

Valor *p*

a. Use ambas colas porque H_a expresa interés por valores relacionados con "diferente de."

P = valor $p = P(t\bigstar < -3.03) + P(t\bigstar > 3.03)$

$= 2 \times P(|t\bigstar| > 3.03)$ como se muestra en la figura

Para hallar el valor p, tiene tres opciones:

1. Usar la tabla 6 (apéndice B): **0.02 < P < 0.05.**
2. Usar la tabla 7 (apéndice B) para poner límites en el valor p: **0.026** $<$ **P** $<$ **0.030**.
3. Usar computadora o calculadora para hallar el valor p: **P** $= 2 \times 0.0145 =$ **0.0290**.

Para instrucciones específicas, vea la página 486.

b. El valor p es menor que α.

O

Clásico:

a. La región crítica es de dos colas porque H_a expresa interés por valores relacionados con "diferente de". El valor crítico se obtiene de la tabla 6: $t(5, 0.025) =$ **2.57.**

Para instrucciones específicas, vea las páginas 477-478.

b. $t\bigstar$ está en la región crítica, como se ve en el **área sombreada** en la figura.

Paso 5 **a. Decisión:** Rechazar H_o.

b. Conclusión: Hay diferencia significativa en las cantidades medias de desgaste en el nivel de significación de 0.05.

CASO PRÁCTICO 10.7 Prueba de procedimientos de muestreo de asfalto

Esta aplicación es un extracto del informe de investigación del Departamento de Transportes de Florida.

COMPARACIÓN DE LOS MÉTODOS DE SACAR CON PALA Y DE CUARTEO, PARA OBTENER MUESTRAS DE MEZCLAS DE ASFALTO
Informe de investigación FL/DOT/SMO/00-441
Gregory A. Sholar, James A. Musselman, Gale C. Page
Oficina estatal de materiales

ABSTRACTO - El método estándar de una planta de cuarteo que producía una mezcla de asfalto para obtener muestras para máxima gravedad específica, clasificación y contenido de aglutinante asfáltico, ha sido empleado con gran éxito

y durante muchos años por el Departamento de Transportes de Florida (FDOT), contratistas y laboratorios independientes de pruebas. Este reporte examina un método alternativo para obtener muestras, que es un poco más fácil y que toma menos tiempo que el método tradicional de cuarteo. Este método, que en adelante denominamos "método de sacar con pala", sigue algunos de los mismos procedimientos y técnicas que se usan con el método de cuarteo. La diferencia en principio es que las muestras se sacan con pala de la pila de mezcla de asfalto hasta obtener el peso deseado de muestra, en lugar de dividir en cuartos la pila hasta obtener el peso deseado de la muestra. Se muestrearon doce mezclas diferentes para este estudio y se compararon las siguientes propiedades para los dos métodos diferentes de muestreo: densidad a granel, máxima gravedad específica, % de vacíos de aire, contenido de aglutinante de asfalto y clasificación. El análisis de los datos indica que los dos métodos de muestreo dan resultados estadísticamente equivalentes para las propiedades de mezcla citadas líneas antes. Incluida en este informe está una nueva versión del FM 1-T168, "Muestreo de mezclas bituminosas de pavimento", donde este nuevo método queda comprendido para muestrear mezclas de asfalto.

ANÁLISIS DE DATOS-Teóricamente, si los dos métodos de muestreo fueran idénticos, entonces la diferencia promedio entre valores obtenidos para cualquier propiedad del asfalto (por ejemplo el contenido de aglutinante del asfalto) sería cero. Se realizó un análisis pareado de diferencia para cada propiedad medida. Un análisis pareado de diferencia es una prueba t efectuada en las diferencias entre cada método de muestreo. Se utilizó un intervalo de confianza de 95%, es decir, $\alpha = 0.05$, para calcular el valor crítico t de dos lados. La hipótesis nula es que la diferencia promedio es cero. Si la t calculada es menor que la t crítica, entonces la hipótesis nula no puede ser rechazada. En los resúmenes de la prueba t, los valores importantes son la "t calculada" y la "t crítica". Para mayor sencillez, todos estos valores "t" se han recopilado en la tabla 14. Un examen de los resultados estadísticos indica que para todas las propiedades medidas, excepto por el % que pasa por el tamiz No. 4, la hipótesis nula no puede ser rechazada. Esto indica que los dos métodos son estadísticamente equivalentes. La excepción es para el % que pasa por el tamiz No. 4. Los valores de la t calculada y la t crítica fueron casi idénticos (2.224 vs. 2.228).

TABLA 14

Resumen de análisis pareado de diferencia

Propiedad de mezcla de asfalto	Valor absoluto		
	t calculada	t crítica	t calc. $<$ t crítica
Gmb (Nmáx)	1.442	2.306	SÍ
Gmm	0.802	2.201	SÍ
% vacíos de aire	1.719	2.306	SÍ
% de ignición)	0.534	2.201	SÍ
Medida de tamiz			
1/2″	0.672	2.228	SÍ
3/8″	0.783	2.228	SÍ
No. 4	2.224	2.228	Igual
No. 8	1.819	2.228	SÍ
No. 16	1.047	2.228	SÍ
No. 30	0.814	2.228	SÍ
No. 50	0.753	2.228	SÍ
No. 100	0.387	2.228	SÍ
No. 200	0.305	2.228	SÍ

CONCLUSIÓN: Con base en el análisis estadístico de los datos, los dos métodos de muestreo son equivalentes respecto al Gmb, Gmm, contenido de aglutinante asfáltico y clasificación. Como el método de sacar con pala es más fácil y rápido, se recomienda que el método modificado de la Florida para muestreo (FM 1-T 168) sea aceptado y puesto en práctica en todo el estado.

SECCIÓN 10.3 EJERCICIOS

10.13 Dado este conjunto de datos pareados:

Pares	1	2	3	4	5
Muestra A	3	6	1	4	7
Muestra B	2	5	1	2	8

Encuentre:

a. Las diferencias pareadas, $d = A - B$, para este conjunto de datos

b. La media $\bar{d}$ de las diferencias pareadas

c. La desviación estándar s_d de las diferencias pareadas

10.14 Encuentre $t(15, 0.025)$. Describa el papel que desempeña este número cuando forme un intervalo de confianza para la diferencia media.

10.15 a. Encuentre el intervalo de confianza de 95% para μ_d dadas $n = 26$, $\bar{d} = 6.3$, y $s_d = 5.1$. Suponga que los datos se seleccionan al azar de una población normal.

b. Compare su intervalo con el intervalo hallado en el ejemplo 10.4 (p. 552).

10.16 A todos los estudiantes inscritos en cierto curso para cultivar la memoria se les aplica un examen antes de iniciar el curso. Al terminarlo, 10 estudiantes son seleccionados al azar y se les aplica un examen después; sus calificaciones aparecen a continuación.

Estudiante	1	2	3	4	5	6	7	8	9	10
Antes	93	86	72	54	92	65	80	81	62	73
Después	98	92	80	62	91	78	89	78	71	80

Se utilizó el MINITAB para hallar el intervalo de confianza de 95% para el mejoramiento medio en memoria que resulta de tomar el curso, medido por la diferencia en calificaciones de la prueba (d = después − antes). Verifique los resultados mostrados en la salida al calcular usted mismo los valores. Suponga normalidad.

```
Confidence Intervals
Variable   N    Mean    StDev   SE Mean   95% C.I.
C3         10   6.10    4.79    1.52      (2.67, 9.53)
```

10.17 Diez personas con niveles de colesterol en el límite superior se seleccionaron al azar para un estudio que comprendía tomar una clase de educación sobre nutrición. Se tomaron lecturas de colesterol antes de la clase y 3 meses después de la clase.

Persona	1	2	3	4	5	6	7	8	9	10
Antes	295	279	250	235	255	290	310	260	275	240
Después	265	266	245	240	230	230	235	250	250	215

Sea d = colesterol antes de la clase – colesterol después de la clase. Se utilizó el Excel para hallar el intervalo de confianza de 95% para la cantidad media de reducción en lecturas de colesterol después de tomar la clase de educación sobre nutrición. Verifique los resultados mostrados en la salida al calcular usted mismo los valores. Suponga normalidad.

Estimación de *t*: media

	d = pre − post
Media	26.3
Desviación estándar	24.4997
Límite inferior de clase	8.773924024
Límite superior de clase	43.82607598

10.18 Use computadora o calculadora para hallar el intervalo de confianza de 95% para estimar μ_d con base en estos datos pareados y suponiendo normalidad:

Antes	75	68	40	30	43	65
Después	70	69	32	30	39	63

10.19 Es frecuente que se recomienden dietas sin sal a personas que tienen alta presión sanguínea. Los siguientes valores de datos se obtuvieron de un experimento diseñado para estimar la reducción en presión sanguínea diastólica, como resultado de consumir una dieta sin sal durante dos semanas. Suponga que las lecturas diastólicas están normalmente distribuidas.

Antes	93	106	87	92	102	95	88	110
Después	92	102	89	92	101	96	88	105

a. ¿Cuál es la estimación puntual para la reducción media en la lectura diastólica después de dos semanas con esta dieta?

b. Encuentre el intervalo de confianza de 98% para la reducción media.

10.20 Se diseñó un experimento para estimar la diferencia media en aumento de peso para cerdos alimentados con una ración A, comparado con aquellos alimentados con una ración B; se emplearon ocho pares de cerdos de la misma camada y las raciones se asignaron al azar a los dos animales de cada par. Los aumentos de peso (en libras) después de 45 días se muestran en la tabla siguiente.

Camada	1	2	3	4	5	6	7	8
Ración A	65	37	40	47	49	65	53	59
Ración B	58	39	31	45	47	55	59	51

Suponiendo que el aumento de peso es normal, encuentre la estimación del intervalo de confianza de 95% para la media de las diferencias μ_d, donde d = ración A – ración B.

10.21 Dos hombres, A y B, que viven en zonas suburbanas y por lo general van al trabajo juntos, deciden efectuar un experimento para ver si una ruta es más rápida que la otra. Los dos hombres piensan que sus hábitos de manejo son aproximadamente iguales, y por tanto deciden el siguiente procedimiento. Cada mañana durante dos semanas, A conducirá el auto al trabajo en una ruta y B usará la otra ruta. En la primera mañana, A lanzará al aire una moneda. Si sale cara, usará la ruta I; si sale cruz, usará la ruta II. A la mañana siguiente, B lanzará al aire la moneda; cara, ruta I; cruz, ruta II. Los tiempos, registrados al minuto más cercano, aparecen en la tabla siguiente. Suponga que los tiempos de viaje son normales y estime la diferencia media poblacional con un intervalo de confianza de 95%.

	Día									
Ruta	M	J	V	S	D	L	M	M	J	V
I	29	26	25	25	25	24	26	26	30	31
II	25	26	25	25	24	23	27	25	29	30

10.22 Al evaluar diferentes instrumentos de medición, primero hay que determinar si hay una diferencia sistemática entre los instrumentos. Con dos instrumentos diferentes, se midieron una sola vez lentes de varios poderes de aumento diferentes. Las diferencias en mediciones (Instrumento A – Instrumento B) se registraron. Las unidades de medida han sido codificadas por razones de patentes.

4	5	−2	−3	−7	10	11	−1	3	7	−5	3	−4
−5	−7	4	−1	−18	0	−17	12	9	4	17	−2	

¿Parece haber una diferencia sistemática entre los dos instrumentos?

a. Describa los datos usando un histograma y una gráfica.

b. Encuentre la media y la desviación estándar.

c. ¿Se necesitan suposiciones para que las inferencias queden satisfechas? Explique.

d. Usando un intervalo de confianza de 95%, estime la media poblacional de las diferencias.

e. ¿Hay cualquier evidencia de una diferencia? Explique.

10.23 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas frases:

a. Hay un aumento en la diferencia media entre calificaciones después de examen y antes de examen.

b. Siguiendo una sesión especial de capacitación, se piensa que la media de la diferencia en calificaciones de rendimiento no será cero.

c. En promedio, no hay diferencia entre las lecturas de dos inspectores en cada una de las partes seleccionadas.

d. La media de las diferencias entre calificaciones de autoestimas previas y de autoestimas subsiguientes mostró mejoría, después de participar en una comunidad de aprendizaje en una universidad.

10.24 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas frases:

a. La media de las diferencias entre las calificaciones de examen subsiguiente y previo es mayor que 15.

b. El aumento medio de peso, después del cambio de dieta para los animales de laboratorio, es al menos 10 onzas.

c. La pérdida media de peso que experimentan personas que siguen un nuevo plan de dieta no fue menor a 12 libras.

d. La diferencia media en reevaluaciones de casas, de dos asesores citadinos, no fue más de $200.

10.25 Determine el valor p para cada prueba de hipótesis para la diferencia media.

a. H_o: $\mu_d = 0$ y H_a: $\mu_d > 0$, con $n = 20$ y $t\bigstar = 1.86$

b. H_o: $\mu_d = 0$ y H_a: $\mu_d \neq 0$, con $n = 20$ y $t\bigstar = -1.86$

c. H_o: $\mu_d = 0$ y H_a: $\mu_d < 0$, con $n = 29$ y $t\bigstar = -2.63$

d. H_o: $\mu_d = 0.75$ y H_a: $\mu_d > 0.75$, con $n = 10$ y $t\bigstar = 3.57$.

10.26 Determine los criterios de prueba que se usarían con el método clásico para probar las siguientes hipótesis cuando se usa t como estadística de prueba.

a. H_o: $\mu_d = 0$ y H_a: $\mu_d > 0$, con $n = 15$ y $\alpha = 0.05$

b. H_o: $\mu_d = 0$ y H_a: $\mu_d \neq 0$, con $n = 25$ y $\alpha = 0.05$

c. H_o: $\mu_d = 0$ y H_a: $\mu_d < 0$, con $n = 12$ y $\alpha = 0.10$

d. H_o: $\mu_d = 0.75$ y H_a: $\mu_d > 0.75$, con $n = 18$ y $\alpha = 0.01$

10.27 Los efectos corrosivos de diversos suelos en tubos de acero con revestimiento y sin revestimiento se probaron usando un plan dependiente de muestreo. Los datos recolectados se resumen con $n = 40$, $\Sigma d = 220$, $\Sigma d^2 = 6222$, donde d es la cantidad de corrosión en la parte con recubrimiento y restada de la cantidad de corrosión de la parte sin recubrimiento. ¿Esta muestra aleatoria da suficiente razón para concluir que el recubrimiento es benéfico? Use $\alpha = 0.01$ y suponga normalidad.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.28 ¿El título del contenido ayuda a un lector a comprender un escrito? A 26 participantes se dio un artículo sin título para leerlo. Luego se clasificaron ellos mismos en comprensión de la información en una escala de 1 a 10, donde 10 era una comprensión completa. A los mismos 26 participantes se les dio otra vez el artículo, esta vez con un título apropiado, y se les pidió calificaran su comprensión. La información resumida resultante se dio como $\bar{d} = 4.76$ y $s_d = 2.33$, donde d = clasificación con título − calificación sin título. La comprensión es por lo general más alta en la segunda lectura que en la primera en un promedio de 3.2 en esta escala. ¿Esta muestra da suficiente evidencia de que el título de un contenido hace la diferencia respecto a la comprensión? Use $\alpha = 0.05$.

10.29 Se hizo una prueba a diez personas a quienes recientemente se les diagnosticó diabetes, para determinar si un programa educacional era eficiente para aumentar su conocimiento sobre la diabetes. Se les aplicó un examen, antes y después del programa educacional, respecto a su cuidado de la diabetes. Las calificaciones de la prueba fueron como sigue:

Paciente	1	2	3	4	5	6	7	8	9	10
Antes	75	62	67	70	55	59	60	64	72	59
Después	77	65	68	72	62	61	60	67	75	68

La siguiente salida de MINITAB se usó para determinar si la diferencia media poblacional es mayor a cero después del programa. Verifique los valores mostrados en la salida (diferencia media [MEAN], desviación estándar [STDEV], error estándar de la diferencia [SE MEAN], $t\bigstar$ [T], y valor p) al calcular usted mismo los valores.

```
TEST OF MU = 0.000 VS MU G.T. 0.000
     N     MEAN     STDEV    SE MEAN     T       P VALUE
C3   10    3.200    2.741    0.867       3.69    0.0025
```

10.30 Diez personas con niveles de colesterol en el límite superior fueron contratadas para un estudio que consistía en tomar un curso de educación nutricional. Se tomaron lecturas de colesterol antes del curso y 3 meses después del mismo.

Persona	1	2	3	4	5	6	7	8	9	10
Antes	295	279	250	235	255	290	310	260	275	240
Después	265	266	245	240	230	230	235	250	250	215

Sea d = colesterol antes − colesterol después. Use la siguiente salida de Excel para probar la hipótesis nula de que la diferencia media poblacional es igual a cero

contra la hipótesis alternativa de que la diferencia media poblacional es positiva con $\alpha = 0.05$. El rechazo de la hipótesis nula indicaría que el promedio de nivel de colesterol (población) después del curso es menor que el nivel promedio antes del curso. Suponga normalidad.

t-Test: Paired Two Sample for Means

	Pretest	Posttest
Mean	268.9	242.6
Variance	618.7666667	256.4888889
Observations	10	10
Hypothesized mean difference	0	
df	9	
t Stat	3.394655392	
$P(T \leq t)$ one-tail	0.003970146	
t Critical one-tail	1.833113856	

10.31 Complete la prueba de hipótesis con hipótesis alternativa $\mu_d > 0$ con base en los datos pareados que siguen y $d = B - A$. Use $\alpha = 0.05$. Suponga normalidad.

A	700	830	860	1080	930
B	720	820	890	1100	960

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

10.32 Use computadora o calculadora para completar la prueba de hipótesis, con hipótesis alternativa $\mu_d < 0$ con base en los datos pareados que siguen y $d = M - N$. Use $\alpha = 0.02$. Suponga normalidad.

M	58	78	45	38	49	62
N	62	86	42	39	47	68

10.33 Complete la prueba de hipótesis, con hipótesis alternativa $\mu_d \neq 0$ con base en los datos pareados que siguen y $d = 0 - Y$. Use $\alpha = 0.01$. Suponga normalidad.

Más viejo	199	162	174	159	173
Más joven	194	162	167	156	176

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

10.34 Diez estudiantes universitarios seleccionados al azar, que participaron en una comunidad de aprendizaje, fueron encuestados para medir su autoestima después de su aprendizaje. Una comunidad de este tipo es un grupo de estudiantes que toman dos o más cursos juntos. Por lo general, cada comunidad de aprendizaje tiene un tema, y el profesorado de que se trata coordina materias que enlazan los cursos. Una investigación ha demostrado que los beneficios de una más alta autoestima, promedios más altos de calificaciones, y más satisfacción en cursos, así como mejores grados de retención, resultan de participar en una comunidad de aprendizaje. Las puntuaciones en las encuestas son como sigue:

Estudiante	1	2	3	4	5	6	7	8	9	10
Antes	18	14	11	23	19	21	21	21	11	22
Después	17	17	10	25	20	10	24	22	10	24

¿Esta muestra de estudiantes presenta suficiente evidencia de que las calificaciones de autoestima fueron más altas después de la participación en una comunidad de aprendizaje? Las puntuaciones más bajas indican autoestima más alta. Use el nivel 0.05 de significación y suponga normalidad de puntuaciones.

10.35 Con referencia a los estudiantes universitarios que participaron en una comunidad de aprendizaje del ejercicio 10.34, también se formó un grupo de control de estudiantes para probar y comparar. A diez estudiantes universitarios seleccionados al azar, que no participaron en la comunidad de aprendizaje, se les practicó una encuesta de autoestima antes del curso y después del curso. Las puntuaciones en las encuestas para el grupo de control son como sigue:

Estudiante	1	2	3	4	5	6	7	8	9	10
Antes	19	23	12	20	26	20	15	10	22	12
Después	19	21	9	10	23	20	19	10	21	19

¿Esta muestra de estudiantes presenta suficiente evidencia de que las puntuaciones de autoestima fueron más altas después de participar en cursos universitarios? Las puntuaciones más bajas indican autoestima más alta. Use el nivel de significación 0.05 y suponga normalidad de calificaciones.

10.36 Para probar el efecto de un curso de acondicionamiento físico en la capacidad física de una persona, se registró el número de "sentadillas" que una persona puede hacer en 1 minuto, antes y después del curso. Diez participantes seleccionados al azar obtuvieron puntuaciones como se indica en la tabla siguiente. ¿Puede concluirse que tuvo lugar una cantidad significativa de mejoría? Use $\alpha = 0.01$ y suponga normalidad.

Antes	29	22	25	29	26	24	31	46	34	28
Después	30	26	25	35	33	36	32	54	50	43

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.37 Con referencia al caso práctico 10.7:

a. ¿Cuál hipótesis nula se está probando en cada una de estas 13 pruebas?

b. ¿Por qué los valores "t calculada" y "t crítica" son los valores importantes?

c. ¿Por qué es correcto informar sus valores absolutos para ambos valores t en la tabla 14?

d. ¿A qué decisión se llega para cada una de estas 13 pruebas de hipótesis?

e. ¿A que conclusión se llega como resultado de estas pruebas?

f. ¿Que acción se recomienda al estado de Florida como resultado de la conclusión?

10.38 Se emprendió un proyecto de investigación para evaluar dos focómetros. Cada uno de 20 lentes de diversos poderes fue leído una vez en cada focómetro. Las diferencias de medición se calcularon entonces, en donde cada diferencia es Focómetro A – Focómetro B. Suponga que las lecturas están normalmente distribuidas.

−0.016	0.013	0.009	0.000	−0.005	−0.015	−0.006	−0.016	−0.022	−0.006
−0.020	0.015	−0.017	−0.010	−0.003	0.011	−0.012	0.008	−0.005	−0.009

Cortesía: Bausch & Lomb

a. Usando una prueba t de estas diferencias pareadas y una $\alpha = 0.01$, determine si la correspondiente diferencia media poblacional es significativamente diferente de cero.

b. Construya un intervalo de confianza de 99% para la diferencia media en lecturas de focómetro.

c. Explique qué indican ambos procedimientos inferenciales acerca de las diferencias.

d. Si un experimentador con iniciativa efectúa esta misma prueba usando $\alpha = 10\%$, ¿cuál sería el resultado? Ofrezca comentarios acerca de continuar usando estas reglas de procedimiento.

10.4 Inferencias relacionadas con las diferencias de medias usando dos muestras independientes

Cuando se comparan las medias de dos poblaciones, por lo general consideramos la diferencia entre sus medias, $\mu_1 - \mu_2$ (a veces llamadas "**medias independientes**"). Las inferencias acerca de $\mu_1 - \mu_2$ estarán basadas en la diferencia entre las medias muestrales observadas, $\bar{x}_1 - \bar{x}_2$. Esta diferencia observada, $\bar{x}_1 - \bar{x}_2$, pertenece a una distribución muestral con las caractsaerísticas descritas en el siguiente enunciado.

PARA SU INFORMACIÓN
¿Por qué $\bar{x}_1 - \bar{x}_2$ es un estimador *insesgado* de $\mu_1 - \mu_2$?

Si unas muestras independientes de tamaños n_1 y n_2 se sacan al azar de poblaciones grandes con medias μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente, entonces la distribución muestral de $\bar{x}_1 - \bar{x}_2$, la diferencia entre las medias muestrales, tiene

1. media $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ y
2. error estándar $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)}$. (10.6)

Si ambas poblaciones tienen distribuciones normales, entonces la distribución muestral de $\bar{x}_1 - \bar{x}_2$ también será normalmente distribuida.

El enunciado precedente es verdadero para todos los tamaños muestrales dado que las poblaciones involucradas son normales y las varianzas poblacionales σ_1^2 y σ_2^2 son cantidades conocidas. No obstante, como en el caso de inferencias acerca de una media, la varianza de una población es generalmente una cantidad descono-

¿SABÍA USTED...?

La "distribución *t*"

Como director de la cervecería Guinness Brewing Company, William Gosset tenía ante sí numerosos y pequeños conjuntos de datos, pequeños por necesidad porque un periodo de 24 horas resultaba con frecuencia en sólo un valor de datos. Así, él perfeccionó la prueba *t* para manejar estas pequeñas muestras para el control de calidad en la elaboración de cerveza. En su artículo científico *El probable error de una media*, se propuso hallar la distribución de la cantidad de error en la media muestral $(x - \mu)$ dividida entre *s*, donde *s* era de una muestra de cualquier tamaño conocido. Encontró entonces el error probable de una media, $\bar{x}$, para una muestra de cualquier tamaño, usando la distribución de $(x - \mu)/(s/\sqrt{n})$. La distribución *t* de Student no ganó popularidad inmediata y, en 1922, incluso 14 años después de su publicación, Student escribió a Fisher: "Te envío un ejemplar de las tablas de Student porque eres el único que probablemente las uses alguna vez." En la actualidad, la distribución *t* de Student tiene uso generalizado y es respetada en investigaciones estadísticas.

cida. Por tanto, será necesario estimar el error estándar al sustituir las varianzas, σ_1^2 y σ_1^2, en la fórmula (10.6) con las mejores estimaciones disponibles, es decir, las varianzas muestrales, s_1^2 y s_2^2. El *error estándar estimado* se encontrará usando la fórmula siguiente:

$$\text{error estimado estándar} = \sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)} \tag{10.7}$$

Las inferencias acerca de la diferencia entre dos medias poblacionales, $\mu_1 - \mu_2$, estarán basadas en las suposiciones siguientes.

Suposiciones para inferencias acerca de la diferencia entre dos medias, $\mu_1 - \mu_2$: las muestras se seleccionan al azar de poblaciones normalmente distribuidas, y las muestras se seleccionan en forma independiente.

NO SE HACEN SUPOSICIONES ACERCA DE LAS VARIANZAS POBLACIONALES.

Como las muestras dan la información para determinar el error estándar, se usará la **distribución *t*** como la estadística de prueba. Las inferencias se dividen en dos casos.

Caso 1: Se usará la distribución *t*, y se calculará el número de grados de libertad.

Caso 2: Se usará la distribución *t*, y se aproximará el número de grados de libertad.

El caso 1 ocurrirá cuando esté completando la inferencia *usando computadora o calculadora de estadística y el software o programa estadístico calcula el número de grados de libertad* para el usuario. El valor calculado para los df es una función de tamaños muestrales y sus tamaños relativos, así como varianzas muestrales y sus tamaños relativos. El valor de df será un número entre el más pequeño de $df_1 = n_1 - 1$ o $df_2 = n_2 - 1$, y la suma de los grados de libertad, $df_1 + df_2 = [(n_1 - 1) + (n_2 - 1)] = n_1 + n_2 - 2$.

El caso 2 ocurrirá cuando esté completando la inferencia *sin ayuda de computadora o calculadora y su paquete de software estadístico*. El uso de la distribución *t* con el más pequeño de $df_1 = n_1 - 1$ o $df_2 = n_2 - 1$ dará resultados *conservadores*. Debido a esta aproximación, el verdadero nivel de confianza para una estimación de intervalo será ligeramente más alto que el nivel de confianza informado; o el verdadero valor *p* y el verdadero nivel de significación para una prueba de hipótesis serán ligeramente menores a lo informado. El espacio entre estos valores informados y los verdaderos valores será muy pequeño, a menos que los tamaños muestrales sean muy pequeños y desiguales o que las varianzas muestrales sean muy diferentes. El espacio disminuirá cuando las muestras aumenten en tamaño o cuando las varianzas muestrales sean más semejantes.

Como la única diferencia entre los dos casos es el número de grados de libertad empleados para identificar la distribución *t* de que se trata, estudiaremos primero el caso 2.

Nota: $A > B$ ("*A* es mayor que *B*") es equivalente a $B < A$ ("*B* es menor que *A*"). Cuando la diferencia entre *A* y *B* se discuta, es costumbre expresar la diferencia como "mayor − menor" de modo que la diferencia resultante sea positiva: $A - B > 0$.

PARA SU INFORMACIÓN
¿Se diría que la diferencia entre 5 y 8 es −3? ¿Cómo expresar la diferencia? Explique.

Expresar la diferencia como "menor − mayor" resulta en $B - A < 0$ (la diferencia es negativa) y suele ser innecesariamente confusa. Por tanto, se recomienda que la diferencia se exprese como "mayor − menor."

Procedimiento de intervalo de confianza

Usaremos la siguiente fórmula para calcular los puntos extremos del **intervalo de confianza $1 - \alpha$.**

Intervalo de confianza para la diferencia entre dos medias (muestras independientes)

$$(\bar{x}_1 - \bar{x}_2) - t(\text{df}, \alpha/2) \cdot \sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)} \quad \text{a} \quad (\bar{x}_1 - \bar{x}_2) + t(\text{df}, \alpha/2) \cdot \sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}$$

donde df es calculado o es el menor de df_1 o df_2 (ver p. 565)

(10.8)

EJEMPLO 10.8

Construcción de un intervalo de confianza para la diferencia entre dos medias

Las estaturas (en pulgadas) de 20 mujeres seleccionadas al azar y 30 hombres seleccionados al azar se obtuvieron de manera independiente, del alumnado de cierta universidad, para estimar la diferencia en sus estaturas medias. La información muestral se da en la tabla 10.2. Suponga que las estaturas están normalmente distribuidas en forma aproximada para ambas poblaciones.

TABLA 10.2

Información muestral sobre estaturas de estudiantes

Muestra	Número	Media	Desviación estándar
Mujer (f)	20	63.8	2.18
Hombre (m)	30	69.8	1.92

Encuentre el intervalo de confianza de 95% para la diferencia entre las estaturas medias, $\mu_m - \mu_f$.

SOLUCIÓN

Paso 1 **Parámetro de interés:** $\mu_m - \mu_f$, la diferencia entre la estatura media de estudiantes hombres y la estatura media de estudiantes mujeres.

Paso 2 **Suposiciones:** Ambas poblaciones son aproximadamente normales, y las muestras se seleccionaron al azar y de forma independiente.

b. Distribución de probabilidad: La distribución t con df = 19, el menor de $n_m - 1 = 30 - 1 = 29$ o $n_f - 1 = 20 - 1 = 19$, y la fórmula (10.8).

c. Nivel de confianza: $1 - \alpha = 0.95$.

Paso 3 **Información muestral:** Vea la tabla 10.2.

Paso 4 **Coeficiente de confianza:** Tenemos una situación de dos colas con $\alpha/2 = 0.025$ en una cola y df = 19. De la tabla 6 del apéndice B, $t(\text{df}, \alpha/2) = t(19, 0.025) = 2.09$. Vea la figura.

En páginas 477-479 vea instrucciones sobre el uso de la tabla 6.

b. Máximo error de estimación: Usando la parte del máximo error de la fórmula (10.8), tenemos

$$E = t\,(\text{df}, \alpha/2) \cdot \sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}: \qquad E = 2.09 \cdot \sqrt{\left(\frac{1.92^2}{30}\right) + \left(\frac{2.18^2}{20}\right)}$$

$$= (2.09)(0.60) = \mathbf{1.25}$$

c. Límites de confianza superior e inferior:

$$(\bar{x}_1 - \bar{x}_2) \pm E$$

$$6.00 \pm 1.25$$

$$6.00 - 1.25 = \mathbf{4.75} \quad \text{a} \quad 6.00 + 1.25 = \mathbf{7.25}$$

Paso 5 **a. Intervalo de confianza.**

4.75 a 7.25 es el intervalo de confianza de 95% para $\mu_m - \mu_f$.

b. Esto es, con 95% de confianza, podemos decir que la diferencia entre las estaturas medias de los estudiantes hombres y mujeres es entre 4.75 y 7.25 pulgadas; es decir, la estatura media de estudiantes hombres es entre 4.75 y 7.25 pulgadas más que la estatura media de estudiantes mujeres.

Procedimiento de prueba de hipótesis

Cuando probemos una **hipótesis acerca de la diferencia entre dos medias poblacionales**, la estadística de prueba empleada será la diferencia entre la diferencia observada y las medias muestrales y la diferencia hipotética de las medias poblacionales, dividida entre el error estándar estimado. Se supone que la estadística de prueba tiene aproximadamente una distribución t cuando la hipótesis nula es verdadera y se ha satisfecho la suposición de normalidad. El valor calculado de la **estadística de prueba** se encuentra usando la fórmula:

Estadística de prueba para la diferencia entre dos medias (muestras independientes)

$$t\bigstar = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} \tag{10.9}$$

donde df es calculado o es el menor de df_1 o df_2 (vea p. 565)

Nota: una diferencia hipotética entre las dos medias poblacionales $\mu_1 - \mu_2$, puede ser cualquier valor especificado. El valor más común especificado es cero, pero la diferencia puede ser diferente de cero.

EJEMPLO 10.9 Prueba de hipótesis de una cola para la diferencia entre dos medias

Suponga que estamos interesados en comparar el éxito académico de estudiantes universitarios, que pertenecen a organizaciones fraternales, con el éxito académico de quienes no pertenecen a organizaciones fraternales. La razón para la comparación es la reciente preocupación de que miembros de la fraternidad, en promedio, están obteniendo un nivel académico más bajo del que logran estudiantes que no son de la fraternidad. (El promedio de calificaciones se usa para medir éxito académico.) Se toman muestras aleatorias de tamaño 40 de cada población. Los resultados muestrales aparecen en la tabla 10.3.

TABLA 10.3

Información muestral de éxito académico

Muestra	Número	Media	Desviación estándar
Miembros de fraternidad (f)	40	2.03	0.68
No miembros de fraternidad (n)	40	2.21	0.59

Complete una prueba de hipótesis usando $\alpha = 0.05$. Suponga que los promedios de calificaciones para ambos grupos están normalmente distribuidos en forma aproximada.

SOLUCIÓN

Paso 1 **a. Parámetro de interés:** $\mu_n - \mu_f$, la diferencia entre los promedios de calificaciones para los miembros no de la fraternidad y los miembros de la fraternidad.

b. Enunciado de las hipótesis:

H_o: $\mu_n - \mu_f = 0$ ($\leq$) (promedios de la fraternidad no son más bajos)

H_a: $\mu_n - \mu_f > 0$ (promedios de la fraternidad son más bajos)

PARA SU INFORMACIÓN
Recuerde: "Mayor − menor" resulta en diferencia positiva.

Paso 2 **a. Suposiciones:** Ambas poblaciones son aproximadamente normales, y se seleccionaron muestras aleatorias. Como las dos poblaciones están separadas, las muestras son independientes.

b. Estadística de prueba: La distribución t con df = el menor de df_n o df_f; como ambas literales n son 40, df = 40 − 1 = **39;** y $t\bigstar$ se calcula usando la fórmula (10.9).

c. Nivel de significación: $\alpha = 0.05$.

Paso 3 **a. Información muestral:** Vea la tabla 10.3.

b. Estadística de prueba calculada:

PARA SU INFORMACIÓN
Cuando df no está en la tabla, use el siguiente valor df más pequeño.

$$t\bigstar = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} : \quad t\bigstar = \frac{(2.21 - 2.03) - (0.00)}{\sqrt{\left(\frac{0.59^2}{40}\right) + \left(\frac{0.68^2}{40}\right)}}$$

$$= \frac{0.18}{\sqrt{0.00870 + 0.01156}} = \frac{0.18}{0.1423} = \mathbf{1.26}$$

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor que". **P** = $P(t\bigstar > 1.26$, con df = 39) como se muestra en la figura.

Para hallar el valor *p*, use uno de tres métodos:

1. Use la tabla 6 (apéndice B) para poner límites al valor *p*: **0.10 < P < 0.25.**
2. Use la tabla 7 (apéndice B) para poner límites al valor *p*: **0.100 < P < 0.119.**
3. Use computadora o calculadora para hallar el valor *p*: **P = 0.1076.**

Detalles específicos siguen a esta ilustración.

b. El valor *p* no es menor que α.

O

Clásico

a. La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "mayor que". El valor crítico se obtiene de la tabla 6: $t(39, 0.05)$ = **1.69.**

Vea en páginas 477-478 información acerca de valores críticos.

b. $t\bigstar$ no está en la región crítica, como se muestra en el **área sombreada** en la figura.

Paso 5 **a. Decisión:** No rechazar H_o.

b. Conclusión: Al nivel de significación de 0.05, la frase de que miembros de la fraternidad logran un nivel más bajo que no miembros no está apoyada por los datos muestrales.

Para hallar el valor *p* para el ejemplo 10.9, use uno de tres métodos:

Método 1: use la tabla 6. Encuentre 1.26 entre dos entradas en la fila df = 39 (use df = 35) y lea los límites para **P** del encabezado de una cola en la parte superior de la tabla: **0.10 < P < 0.25.**

Método 2: use la tabla 7. Encuentre t = 1.26 entre dos filas y df = 39 entre dos columnas; lea los límites para $P(t\bigstar > 1.26|\text{df} = 39)$; **0.100 < P < 0.119.**

Método 3: si está haciendo la prueba de hipótesis con ayuda de una computadora o calculadora, lo más probable es que calculará el valor *p* para el usuario (vea pp. 486-487), o puede usar los comandos de distribución de probabilidad acumulativa descritos en el capítulo 9 (p. 479).

EJEMPLO 10.10 Hipótesis de dos colas para la diferencia entre dos medias

Muchos estudiantes se han quejado de que la máquina vendedora de bebidas gaseosas de la sala de recreación de estudiantes (A), expende una cantidad diferente de líquido que la máquina del salón de profesores (B). Para probar esta creencia, un estudiante al azar muestreó varias porciones de cada máquina y con todo cuidado las midió, con los resultados indicados en la tabla 10.4.

TABLA 10.4

Información muestral de máquinas vendedoras

Máquina	Número	Media	Desviación estándar
A	10	5.38	1.59
B	12	5.92	0.83

¿Esta evidencia apoya la hipótesis de que la cantidad media dosificada por la máquina A es diferente de la dosificada por la máquina B? Suponga que las cantidades dosificadas por ambas máquinas están normalmente distribuidas, y complete la prueba usando $\alpha = 0.10$.

SOLUCIÓN

Paso 1 **a.** **Parámetro de interés:** $\mu_B - \mu_A$, la diferencia entre la cantidad media dosificada por la máquina B y la cantidad media dosificada por la máquina A.

b. Enunciado de las hipótesis:

H_o: $\mu_B - \mu_A = 0$ (A dosifica la misma cantidad promedio que B)

H_a: $\mu_B - \mu_A \neq 0$ (A dosifica una cantidad promedio diferente que B)

Paso 2 **Suposiciones:** Se supone que ambas poblaciones son aproximadamente normales, y que las muestras se seleccionaron en forma aleatoria e independiente.

b. **Estadística de prueba:** La distribución t con df = al menor de $n_A - 1 = 10 - 1 = 9$ o $n_B - 1 = 12 - 1 = 11$, df = 9, y $t\bigstar$ calculada usando la fórmula (10.9).

c. **Nivel de significación:** $\alpha = 0.10$.

Paso 3 **a.** **Información muestral:** Vea la tabla 10.4.

b. **Estadística de prueba calculada:**

$$t\bigstar = \frac{(\bar{x}_B - \bar{x}_A) - (\mu_B - \mu_A)}{\sqrt{\left(\frac{s_B^2}{n_B}\right) + \left(\frac{s_A^2}{n_A}\right)}} : \quad t\bigstar = \frac{(5.92 - 5.38) - (0.00)}{\sqrt{\left(\frac{0.83^2}{12}\right) + \left(\frac{1.59^2}{10}\right)}}$$

$$= \frac{0.54}{\sqrt{0.0574 + 0.2528}} = \frac{0.54}{0.557} = \mathbf{0.97}$$

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Use ambas colas porque H_a expresa interés por valores relacionados con "diferente de".

P = valor $p = P(t\bigstar < -0.97) + P(t\bigstar > 0.97) = 2 \times P(|t\bigstar| > 0.97 | \text{df} = 9)$ como en la figura.

Para hallar el valor p, tiene tres opciones:

1. Usar la tabla 6 (apéndice B) **0.20 < P < 0.50**.
2. Usar la tabla 7 (apéndice B) para poner límites al valor p: **0.340 < P < 0.394**.
3. Usar computadora o calculadora para hallar el valor p: **P** = 2 × 0.1787 = **0.3574**.

Instrucciones específicas siguen a esta ilustración.

b. El valor p no es menor que α.

Clásico:

a. La región crítica es de dos colas porque H_a expresa interés por valores relacionados con "diferente de". El valor crítico derecho se obtiene de la tabla 6: $t(9, 0.05)$ = **1.83.** Vea la figura.

Para instrucciones específicas, vea las páginas 477-478.

b. $t\bigstar$ no está en la región crítica como se muestra en el **área sombreada** en la figura.

Paso 5 **a. Decisión:** No rechazar H_o.

b. Conclusión: La evidencia no es suficiente para demostrar que la máquina A dosifica una cantidad promedio diferente de bebida gaseosa que la máquina B, al nivel de significación de 0.10. Por tanto, la falta de evidencia proseguirá aun cuando las dos máquinas dosifiquen, en promedio, la misma cantidad.

Para hallar el valor p para el ejemplo 10.10, use uno de tres métodos:

Método 1: usar tabla 6. Encuentre 0.97 entre dos entradas en la fila df = 9 y lea los límites para **P** del encabezado de dos colas en la parte superior de la tabla: **0.20 < P < 0.50**.

Método 2: Usar la tabla 7. Encuentre $t\bigstar = 0.97$ entre dos filas y df = 9 entre dos columnas; lea los límites para $P(t\bigstar > 0.97|\text{df} = 9)$: $0.170 < \frac{1}{2}\mathbf{P} < 0.197$; por tanto, **0.340 < P < 0.394**.

Método 3: si está haciendo la prueba de hipótesis con ayuda de una computadora o calculadora, es muy probable que éstas calculen el valor p (no duplicar) por usted (vea pp. 486-487), o puede usar los comandos de distribución de probabilidad acumulativa descritos en el capítulo 9 (p. 479).

Casi todos los paquetes de estadística de computadora o calculadora completarán las inferencias para la diferencia entre dos medias, calculando el número de grados de libertad.

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA ENTRE DOS MEDIAS POBLACIONALES CON DESVIACIÓN ESTÁNDAR DESCONOCIDA DADOS DOS CONJUNTOS INDEPENDIENTES DE DATOS MUESTRALES

MINITAB (Versión 14) El comando de Muestra t (Prueba e Intervalo de Confianza) de MINITAB 2 efectúa al mismo tiempo el intervalo de confianza y la prueba de hipótesis.
Introduzca los dos conjuntos independientes de datos en C1 y C2; luego continúe con:

```
Choose:     Stat > Basic Statistics > 2-Sample t
Select:     Samples in different columns*
Enter:      First: C1  Second: C2
Select:     Assume equal variances (if known)
Select:     Options
Enter:      Confidence level:1 − α (ex. 0.95 or 95.0)
            Test mean: 0.0
Choose:     Alternative: less than or not equal or greater than > OK > OK
```

*Observe los otros posibles formatos de datos.

Excel Introduzca los dos conjuntos independientes de datos en las columnas A y B; luego continúe con:

```
Choose:     Tools > Data Analysis > t-Test: Two-Sample Assuming Unequal
            Variances
Enter:      Variable 1 Range: (A1:A20 or select cells)
            Variable 2 Range: (B1:B20 or select cells)
            Hypothesized Mean Difference: μB − μA (usually 0)
Select:     Labels (if necessary)
```

```
Enter:      α (ex. 0.05)
Select:     Output Range
Enter:      (C1 or select cell) > OK
```

Use Format > Column > AutoFit Selection para que la salida sea más fácil de leer. La salida muestra valores p y valores críticos para pruebas de una y dos colas.

TI-83/84 Plus Introduzca los dos conjuntos independientes de datos en L1 y L2.*
Para construir un intervalo de confianza de $1 - \alpha$ para la diferencia media, continúe con lo siguiente, introduciendo los valores apropiados y resaltando Calculate:

```
Choose:     STAT > TESTS > 0:2-SampTInt . . .
```

Para completar una prueba de hipótesis para la diferencia media, continúe con lo siguiente, introduciendo los valores apropiados y resaltando Calculate:

```
Choose:     STAT > TESTS > 4:2-SampTTest . . .
```

*Introduzca los datos en el orden que es necesario; el programa resta como L1–L2.

Resalte No for Pooled si no hay suposiciones acerca de la igualdad de varianzas.

El ejemplo 10.9 se resolvió usando MINITAB. Con promedios acumulativos de calificaciones para no miembros en C1 y 40 promedios para miembros de la fraternidad en C2, los comandos precedentes resultaron en la salida que se ilustra a continuación. Compare estos resultados con la solución del ejemplo 10.9. Observe la diferencia en **P** y valores df. Explique.

```
Two-Sample T-Test and CI
Sample      N       Mean      StDev      SE Mean
1           40      2.210     0.590      0.093
2           40      2.030     0.680      0.11
Difference = mu (1) − mu (2) Est. diff.: 0.180
95% CI for difference: (−0.10, 0.46)
T-Test diff. = 0 (vs >): T = 1.26 P = 0.105 DF = 76
```

CASO PRÁCTICO 10.11 Anatomía seccional: estrategia para dominio

Una estrategia para el dominio de anatomía seccional debe contener investigación para determinar si la metodología es o no es acertada. El método siguiente, basado en investigación, demuestra la efectividad de una estrategia específica. Una aplicación referente a la comprensión completa de las características anatómicas humanas, y sus estructuras adyacentes, es un método prescrito y de leyendas secuenciadas (PSLM, por sus siglas en inglés). El método hace contrastar la moderna convención de secciones anatómicas humanas marcadas al azar. El PSLM está basado en estudios que utilizan imágenes adquiridas del Proyecto Humano Visible, y fue realizado en el Triton College. Estos estudios han mostrado un significativo impacto en la rapidez de aprendizaje y comprensión de estructuras y relaciones anatómicas adyacentes.

El componente inicial de esta investigación estuvo formado por un grupo de 28 estudiantes que se dividieron en dos grupos: Grupo A: 15, Grupo B: 13. A ambos grupos se les aplicó un examen previo, seguido de un periodo de estudio y un singular examen después del curso. Ambas imágenes presentadas para estudio eran secciones transversales idénticas del cerebro. La imagen del Grupo A para estudio tenía leyendas de acuerdo con el protocolo PSLM. A la imagen del Grupo B para estudio se le aplicaron leyendas en forma aleatoria. Ambos grupos recibieron instrucciones de "hacer una lista y reconocer las partes/capas de una sección transversal del cerebro, de superficial a profunda", escribiendo sus respuestas en espacios provistos como único examen después del curso en este estudio preliminar.

Primer conjunto de estudios PSLM

Variable	Número de casos	Media	Desviación estándar	Error estándar de la media
Grupo A	15	9.6667	1.589	0.410
Grupo B	13	3.1538	3.023	0.839

Varianzas	Valor *t*	df	Sig.de 2 colas	Error estándar de la media	Intervalo de confianza
Desigual	6.98	17.57	.000	0.933	(4.548, 8.477)

La calificación promedio del Grupo A fue 9.6667 de una posible derecha total de 11 con una desviación estándar de 1.589. El Grupo B calificó un promedio de 3.1538 de 11 puntos posibles.

La diferencia media entre el Grupo A y el Grupo B es 6.5129, que es muy importante.

Fuente: Alexander Lane, Ph.D., Triton College, River Grove, Illinois, http://www.nlm.nih.gov/research/visible/vhpconf98/AUTHORS/LANE/LANE.HTM

SECCIÓN 10.4 EJERCICIOS

10.39 Dos muestras aleatorias independientes resultaron en lo siguiente:

Muestra 1: $n_1 = 12$, $s_1^2 = 190$

Muestra 2: $n_2 = 18$, $s_2^2 = 150$

Encuentre la estimación para el error estándar para la diferencia entre dos medias.

10.40 Dos muestras aleatorias independientes resultaron en lo siguiente:

Muestra A: $n_A = 24$, $s_A = 8.5$

Muestra B: $n_B = 21$, $s_B = 11.3$

Encuentre la estimación para el error estándar para la diferencia entre dos medias.

10.41 Se obtuvieron dos muestras aleatorias independientes de tamaños 18 y 24 para hacer inferencias acerca de la diferencia entre dos medias. ¿Cuál es el número de grados de libertad? Discuta ambos casos.

10.42 Encuentre el coeficiente de confianza, $t_{\left(\text{df}, \frac{\alpha}{2}\right)}$, que se usaría para hallar el máximo error para cada una de las siguientes situaciones, cuando estime la diferencia entre dos medias, $\mu_1 - \mu_2$.

a. $1 - \alpha = 0.95$, $n_1 = 25$, $n_2 = 15$

b. $1 - \alpha = 0.98$, $n_1 = 43$, $n_2 = 32$

c. $1 - \alpha = 0.99$, $n_1 = 19$, $n_2 = 45$

10.43 Encuentre el intervalo de confianza de 90% para la diferencia entre dos medias con base en esta información acerca de dos muestras. Suponga muestras independientes de poblaciones normales.

Muestra	Número	Media	Desv. est.
1	20	35	22
2	15	30	16

10.44 Se realizó un estudio que compara actitudes hacia la muerte, en el que donadores de órganos (personas que habían firmado tarjetas para donar órganos) se compararon contra no donadores. El estudio se publica en la revista *Death Studies.* A ambos grupos se administró la escala de ansiedad de muerte de Templer (DAS). En esta escala, las calificaciones altas indican elevada ansiedad respecto a la muerte. Los resultados se publicaron como sigue:

	n	Media	Desviación estándar
Donadores de órganos	25	5.36	2.91
No donadores de órganos	69	7.62	3.45

Construya un intervalo de confianza de 95% para la diferencia entre las medias, $\mu_{\text{no}} - \mu_{\text{donador}}$.

PARA SU INFORMACIÓN Los resultados obtenidos pueden ser notablemente diferentes, dependiendo del uso del grado de libertad (df) calculado o del uso del df para una muestra más pequeña.

10.45 "Quienes rentan autos encuentran tarifas más altas" (*USA Today,* 17 de mayo, 2005) informó que las tarifas por rentar autos están subiendo pero no han rebasado el valor pico promedio nacional de $35.98 por día de 2001. La tarifa promedio diaria nacional para el trimestre de abril a junio de 2005 fue de $30.15, pero, en algunas ciudades, la renta de autos puede costar más de $100 al día. Un estudio similar de dos ciudades importantes encontró los siguientes resultados:

Ciudad	n	Tarifa diaria promedio	Desviación estándar
Boston	10	128.25	7.50
Nueva York	15	116.60	8.90

Establezca un intervalo de confianza de 95% sobre la diferencia en tarifas diarias promedio, entre las dos ciudades principales de la costa del Atlántico, Boston y Nueva York. Suponga normalidad para las poblaciones muestreadas y que las muestras se seleccionaron al azar.

10.46 Parte de un programa de prueba sobre una familia de cohetes dobles de campana, diseñados para compensar el rendimiento a baja altitud, llevó a dos nuevos diseños ligeramente diferentes. Los dos conjuntos de magnitudes de flujo de prueba, medidos en libras por segundo, resultaron de probar estos dos nuevos diseños.

	n	$\sum x$	$\sum x^2$
Diseño 1	36	278.4	2163.76
Diseño 2	42	310.8	2332.26

Determine el intervalo de confianza de 99% para la diferencia entre las dos magnitudes medias de flujo de prueba. Suponga normalidad.

10.47 Se diseñó un estudio para estimar la diferencia en lecturas de presión sanguínea diastólica entre hombres y mujeres. Se utilizó el MINITAB para construir un intervalo de confianza de 99% para la diferencia entre las medias, con base en los siguientes datos muestrales.

Hombres	76	76	74	70	80	68	90	70
	90	72	76	80	68	72	96	80
Mujeres	76	70	82	90	68	60	62	68
	80	74	60	62	72			

```
Two-sample T for Males vs Females
         N        Mean     StDev    SE Mean
Males    16       77.37    8.35     2.1
Females  13       71.08    9.22     2.6
99% C.I. for mu males − mu females: ( −2.9, 15.5)
```

Verifique los resultados (las dos medias muestrales y desviaciones estándar, y límites de intervalo de confianza), calculando usted mismo los valores. Suponga normalidad de lecturas de la presión sanguínea.

10.48 "¿Es afectada la longitud de una barra de acero por la técnica de tratamiento térmico empleada?" Ésta fue la pregunta probada cuando se recolectaron los datos siguientes.

Tratamiento térmico	Longitudes (a la pulgada más cercana)										
1	156	159	151	153	157	159	155	155	151	152	158
	154	156	156	157	155	156	159	153	157	157	159
	158	155	159	152	150	154	156	156	157	160	
2	154	156	150	151	156	155	153	154	149	150	150
	151	154	155	155	154	154	156	150	151	156	154
	153	154	149	150	150	151	154	148	155	158	

a. Encuentre las medias y desviaciones estándar para los dos conjuntos de datos.

b. Encuentre evidencia acerca de los datos muestrales (gráfica y muestral) que apoye la suposición de normalidad para las dos poblaciones muestreadas.

c. Encuentre el intervalo de confianza de 95% para $\mu_1 - \mu_2$.

10.49 Aproximadamente 95% de los girasoles cosechados en Estados Unidos se producen en Dakota del Norte, Dakota del Sur y Minnesota. Para comparar las cantidades de producción entre Dakota del Norte y Dakota del Sur, se seleccionaron al azar 11 condados productores de girasoles de Dakota del Norte, así como 14 condados productores de girasoles de Dakota del Sur. Sus producciones en 2004 (en libras por acre) está a continuación.

Dakota del Norte										
1012	780	711	805	1294	666	779	1099	1314	1358	920

Dakota del Sur										
1101	1347	1520	1025	1462	1650	1903	1652	486	1144	1684
1530	1800	876								

Fuente: http://www.usda.gov/nass/graphics/county04/data/sf04.csv

Encuentre el intervalo de confianza de 95% para la diferencia entre la producción media de girasoles para todos los condados productores de girasoles de Dakota del Sur, y todos los condados productores de girasoles de Dakota del Norte. Suponga normalidad.

10.50 En una gran universidad se aplica un examen de conocimientos de matemáticas a todos los estudiantes. Se seleccionaron al azar muestras de 36 estudiantes hombres y 30 mujeres, del alumnado de este año y se registraron las siguientes calificaciones:

Hombres	72	68	75	82	81	60	75	85	80	70
	71	84	68	85	82	80	54	81	86	79
	99	90	68	82	60	63	67	72	77	51
	61	71	81	74	79	76				
Mujeres	81	76	94	89	83	78	85	91	83	83
	84	80	84	88	77	74	63	69	80	82
	89	69	74	97	73	79	55	76	78	81

a. Describa cada conjunto de datos con un histograma (use los mismos intervalos de clase en ambos histogramas), la media y desviación estándar.

b. Construya un intervalo de confianza de 95% para la calificación media para todos los estudiantes hombres. Haga lo mismo para todas las estudiantes mujeres.

c. Los resultados encontrados en la parte b, ¿muestran que las calificaciones medias para hombres y mujeres podrían ser iguales? Justifique su respuesta. ¡Tenga cuidado!

d. Construya el intervalo de confianza de 95% para la diferencia entre las calificaciones medias para estudiantes hombres y mujeres.

e. Los resultados encontrados en la parte d, ¿muestran que las calificaciones medias para hombres y mujeres podrían ser iguales? Explique.

f. Explique por qué los resultados de la parte b no pueden usarse para sacar conclusiones acerca de la diferencia entre las dos medias.

10.51 Exprese las hipótesis nula y alternativa que se usarían para probar lo siguiente:

a. Hay diferencia entre la edad media de los empleados de dos grandes empresas diferentes.

b. La media de la población 1 es mayor que la media de la población 2.

c. La producción media por condado de semillas de girasol en Dakota del Norte es menor que la producción media por acre por condado en Dakota del Sur.

d. No hay diferencia en el número medio de horas empleadas estudiando por semana entre estudiantes universitarios hombres y mujeres.

10.52 Exprese las hipótesis nula y alternativa que se usarían para probar lo siguiente:

a. La diferencia entre las medias de las dos poblaciones es más de 20 libras.

b. La media de la población A es menos de 50 más que la media de la población B.

c. La diferencia entre las dos poblaciones es al menos $500.

d. El patio de tamaño promedio para el vecindario A no es más de 30 yardas cuadradas mayor que el patio promedio del vecindario B.

10.53 Calcule la estimación para el error estándar de la diferencia entre dos medias independientes para cada uno de los casos siguientes:

a. $s_1^2 = 12$, $s_2^2 = 15$, $n_1 = 16$, y $n_2 = 21$

b. $s_1^2 = 0.054$, $s_2^2 = 0.087$, $n_1 = 8$, y $n_2 = 10$

c. $s_1 = 2.8$, $s_2 = 6.4$, $n_1 = 16$, y $n_2 = 21$

10.54 Encuentre el valor de $t\bigstar$ para la diferencia entre dos medias con base en una suposición de normalidad y esta información acerca de dos muestras:

Muestra	Número	Media	Desviación estándar
1	18	38.2	14.2
2	25	43.1	10.6

10.55 Encuentre el valor de $t\bigstar$ para la diferencia entre dos medias con base en una suposición de normalidad y esta información acerca de dos muestras:

Muestra	Número	Media	Desviación estándar
1	21	1.66	0.29
2	9	1.43	0.18

10.56 Determine el valor p para las siguientes pruebas de hipótesis, para la diferencia entre dos medias con varianzas poblacionales desconocidas.

a. H_a: $\mu_1 - \mu_2 > 0$, $n_1 = 6$, $n_2 = 10$, $t\bigstar = 1.3$

b. H_a: $\mu_1 - \mu_2 < 0$, $n_1 = 16$, $n_2 = 9$, $t\bigstar = -2.8$

c. H_a: $\mu_1 - \mu_2 \neq 0$, $n_1 = 26$, $n_2 = 16$, $t\bigstar = 1.8$

d. H_a: $\mu_1 - \mu_2 \neq 5$, $n_1 = 26$, $n_2 = 35$, $t\bigstar = -1.8$

10.57 Determine los valores críticos que se usarían para las siguientes pruebas de hipótesis (usando el método clásico), acerca de la diferencia entre dos medias con varianzas poblacionales desconocidas.

a. H_a: $\mu_1 - \mu_2 \neq 0$, $n_1 = 26$, $n_2 = 16$, $\alpha = 0.05$

b. H_a: $\mu_1 - \mu_2 < 0$, $n_1 = 36$, $n_2 = 27$, $\alpha = 0.01$

c. H_a: $\mu_1 - \mu_2 > 0$, $n_1 = 8$, $n_2 = 11$, $\alpha = 0.10$

d. H_a: $\mu_1 - \mu_2 \neq 10$, $n_1 = 14$, $n_2 = 15$, $\alpha = 0.05$

10.58 Para la prueba de hipótesis que comprende H_a: $\mu_B - \mu_A \neq 0$ con df = 18 y $t\bigstar = 1.3$.

a. Encuentre el valor p.

b. Encuentre los valores críticos dada $\alpha = 0.05$.

10.59 Suponga que la $t\bigstar$ calculada hubiera sido 1.80 en el ejemplo 10.10 (pp. 569-570). El uso de df = 9 o el de df = 20 resulta en respuestas diferentes. Explique la forma en que la palabra *conservador* (p. 565) se aplica aquí.

10.60 "En un mes típico, los hombres gastan $178 y las mujeres gastan $96 en su tiempo libre", según los resultados de una investigación internacional de comunicaciones (ICR) para una encuesta de American Express, como se informa en USA Snapshot en Internet, 25 de junio de 2005.

Suponga que se toman muestras aleatorias de la población de estudiantes universitarios hombres y mujeres. A cada estudiante se le pide determinar sus gastos en su tiempo libre en el mes anterior. Los resultados de datos muestrales tienen una desviación estándar de $75 para los hombres y de $50 para las mujeres.

a. Si ambas muestras son de tamaño 20, ¿cuál es el error estándar para la diferencia de dos medias?

b. Suponiendo normalidad en gastos en tiempo libre, ¿es significativa en $\alpha = 0.05$ la diferencia hallada en la encuesta ICR si se usan las muestras de la parte a? Explique.

10.61 Numerosos quesos se producen en forma de rueda. Debido a las diferencias en consistencia entre estos diferentes tipos de quesos, la cantidad de queso, medida en peso, varía de una rueda a otra. Heidi Cembert desea determinar si hay una diferencia significativa, al nivel de 10%, entre el peso por rueda de queso Gouda y el Brie. Ella muestrea al azar 16 ruedas de Gouda y encuentra que la media es 1.2 libras, con una desviación estándar de 0.32 libras; luego muestrea al azar 14 ruedas de queso Brie y encuentra una media

de 1.05 libras y una desviación estándar de 0.25 libras. ¿Cuál es el valor p para la hipótesis de igualdad de Heidi? Suponga normalidad.

10.62 Si una muestra aleatoria de 18 casas situadas al sur de la calle Centro en Provo tiene un precio medio de venta de $145 200 y una desviación estándar de $4700, y una muestra aleatoria de 18 casas situadas al norte de la calle Centro tiene un precio medio de venta de $148 600 y una desviación estándar de $5800, ¿puede concluirse que hay una diferencia significativa entre el precio de venta de casas en estas dos zonas de Provo al nivel de 0.05? Suponga normalidad.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.63 La era de las computadoras ha permitido a profesores usar material didáctico electrónico para motivar a sus alumnos a aprender. *Issues in Accounting Education* publicó los resultados de un estudio que mostró que un material didáctico electrónico, junto con la presión ejercida por un asesor intencionalmente inducida, fue efectiva para mejorar las preparaciones antes de clase y en mejorar la asistencia a clase, las calificaciones de exámenes, así como de evaluaciones del curso cuando son empleados por alumnos que estudian contabilidad de impuestos.

Suponga que se lleva a cabo un estudio semejante en su escuela, empleando una guía electrónica de estudio (ESG) como tutor para estudiantes de principios de contabilidad. Para una sección del curso, a los estudiantes se les pidió usar un nuevo programa ESG de cómputo que generaba y calificaba cuestionarios de repaso de capítulo y exámenes de práctica, presentaba repasos de capítulos del libro de texto y daba seguimiento al avance del estudiante. Los estudiantes podían usar computadora para elaborar, tomar y calificar sus propios exámenes simulados y revisar materiales a su propio ritmo, antes de tomar sus exámenes y cuestionarios formales en clase y formados por diferentes preguntas. El mismo profesor daba la clase a la otra sección del curso, usaba el mismo libro de texto y daba las mismas asignaturas a diario, pero no pedía a los estudiantes que usaran el ESG. Se aplicaron pruebas idénticas a ambas secciones, y las calificaciones medias de todos los exámenes y asignaturas al final del año se tabularon:

Sección	n	Calificación media	Desviación estándar
ESG (1)	38	79.6	6.9
No ESG (2)	36	72.8	7.6

¿Estos resultados muestran que las calificaciones medias de exámenes y asignaturas, para estudiantes que toman principios de contabilidad con un ESG para ayudarlos, son significativamente mejores que las de quienes no usan un ESG? Use un nivel de significación de 0.01.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.64 El departamento de compras de una cadena regional de supermercados está considerando dos fuentes de dónde comprar bolsas de patatas de 10 libras. Una muestra aleatoria tomada de cada fuente muestra los siguientes resultados.

	Idaho Supers	Idaho Best
Número de bolsas pesadas	100	100
Peso medio	10.2 lb	10.4 lb
Varianza muestral	0.36	0.25

Al nivel de significación de 0.05, ¿hay diferencia entre los pesos medios de las bolsas de patatas de 10 libras?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.65 Lauren, una morena, estaba cansada de oír "las rubias se divierten más". Se propuso "demostrar" que "las morenas son más inteligentes". Al azar, ella seleccionó (lo mejor que pudo) 40 rubias y 40 morenas en su preparatoria. Se calcularon las siguientes estadísticas de calificaciones generales:

Rubias	$n_{Bl} = 40$	$\overline{x}_{Bl} = 88.375$	$s_{Bl} = 6.134$
Morenas	$n_{Br} = 40$	$\overline{x}_{Br} = 87.600$	$s_{Br} = 6.640$

Al ver los resultados muestrales, ¿Lauren tiene apoyo para decir que "las morenas son más inteligentes que las rubias"? Explique. ¿Qué podría decir Lauren de la inteligencia de rubias y morenas?

10.66 Uno podría pensar que los estudiantes mayores de preparatoria parecerían tener más problemas monetarios que los estudiantes menores. Los mayores prevén gastos para la universidad, así como para viajes y bailes. Entonces, ¿esto significa que trabajan más que sus condiscípulos menores? Christine, estudiante mayor de la preparatoria HFL, seleccionó al azar los

siguientes datos (registrados en horas/semana) de estudiantes que trabajan:

Mayores	$n_S = 17$	$\bar{x}_S = 16.4$	$s_S = 10.48$
Menores	$n_J = 20$	$\bar{x}_J = 18.405$	$s_J = 9.69$

Suponiendo que las horas de trabajo están normalmente distribuidas, ¿estos datos sugieren que hay una diferencia significativa entre el número promedio de horas que mayores y menores de la HFL trabajan por semana? Use $\alpha = 0.10$.

10.67 Con referencia al caso práctico 10.11:

a. El término "Varianzas – Desigual" es equivalente a no hacer suposiciones acerca de las varianzas. Verifique el valor informado para el error estándar estimado, error estándar de diferencia.

b. Verifique el valor *t*.

c. ¿Cuál es el rango de posibles valores para el df para este estudio?

d. Explique cómo se obtuvo df = 17.57.

e. ¿Qué valor df usaría usted con base en el material presentado en la sección 10.4?

f. Explique por qué el valor *t* de 6.98 se dice que es "altamente significativo".

10.68 Se realizó un estudio para evaluar la seguridad y eficiencia de recibir nitroglicerina de un sistema transdérmico (es decir, un parche que se pone en la piel), que en forma intermitente administra el medicamento, contra el sistema de tomar medicamento oral (píldoras). Veinte pacientes que sufren de angina de pecho (dolor agudo) debido al esfuerzo físico se inscribieron en pruebas. A todos se les aplicaron parches, algunos ($n = 8$) contenían nitroglicerina; los otros ($n = 12$), un placebo. Suponga que los datos resultantes de "tiempo de angina" se resumieron como sigue:

Tiempo medio para una angina (segundos)

	Activo	Placebo	Diferencia	Error estándar	Valor *p*[a]
1 A.M. del día	320.00	287.00	33.00	9.68	0.0029
2 A.M. del día	314.00	285.25	28.75	13.74	0.0500

[a]Para diferencia de tratamiento.

a. Determine el valor de *t* para la diferencia entre dos medias independientes dada la diferencia y error estándar (SE) para los datos de la 1 A.M. del día. Suponga normalidad.

b. Verifique el valor *p*.

c. Determine el valor de *t* para la diferencia entre dos medias independientes, dada la diferencia y error estándar (SE) para los datos de las 7 P.M. del día.

d. Verifique el valor *p*.

10.69 Se utilizó el MINITAB para completar una prueba *t* de la diferencia entre las dos medias, usando las siguientes dos muestras independientes.

Muestra 1	33.7	21.6	32.1	38.2	33.2	35.9	34.1	39.8
	23.5	21.2	23.3	18.9	30.3			
Muestra 2	28.0	59.9	22.3	43.3	43.6	24.1	6.9	14.1
	30.2	3.1	13.9	19.7	16.6	13.8	62.1	28.1

```
Two-sample T for sample 1 vs sample 2
              N        Mean        StDev       SE Mean
sample1       13       29.68       7.07        2.0
sample2       16       26.9        17.4        4.4
T-Test mu sample1 = mu sample2 (vs not =): T=0.59
P=0.56 DF=20
```

a. Suponiendo normalidad, verifique los resultados (dos medias muestrales y desviaciones estándar, y la t★ calculada) al calcular los valores usted mismo.

b. Use la tabla 7 del apéndice B para verificar el valor *p* con base en el df calculado.

c. Encuentre el valor *p* usando el número más pequeño de grados de libertad. Compare los dos valores *p*.

10.70 De acuerdo con la Junta Directiva universitaria (http://www.collegeboard.com/press/article/0,3,8993,00.html), el costo promedio en el año escolar 2004-2005 (colegiatura, cuotas, alojamiento y comida) para una universidad pública fue \$11 354 contra \$26 057 de una universidad privada. ¿También hay diferencia en el costo promedio de libros de texto pedidos entre universidades públicas y privadas? Se tomaron las siguientes muestras de tamaño 10.

Pública	Privada	Pública	Privada	Pública	Privada
64.69	71.00	103.59	98.56	110.69	112.58
89.60	96.19	106.38	98.94	118.94	114.00
101.49	96.47	106.77	107.79	135.94	116.55
101.75	97.14				

Usando la salida Excel que sigue y $\alpha = 0.05$, determine si el costo promedio de libros de texto pedidos

por grupo es diferente entre universidades públicas y privadas. Suponga normalidad.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

```
t-Test: Two-Sample Assuming Unequal Variances
                              Public          Private
Mean                          103.984         100.922
Variance                      340.6249822     173.2995511
Observations                  10              10
Hypothesized Mean
  Difference                  0
df                            16
t Stat                        0.427125511
P(T ≤ t) two-tail             0.674980208
 t Critical two-tail          2.119904821
```

10.71 Veinte ratones de laboratorio se dividieron en forma aleatoria en dos grupos de 10; a cada grupo se le alimentó de acuerdo con una dieta prescrita. Al término de tres semanas se registró el aumento de peso de cada animal. Los datos de la siguiente tabla, ¿justifican la conclusión de que el aumento medio de peso en la dieta B fue mayor que el aumento de peso medio en la dieta A, al nivel de significación de $\alpha = 0.05$? Suponga normalidad.

Dieta A	5	14	7	9	11	7	13	14	12	8
Dieta B	5	21	16	23	4	16	13	19	9	21

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.72 Muchas de las personas que participan en el béisbol de ligas mayores (MLB) piensan que los partidos jugados por los Yanquis de Nueva York tienden a durar más que los partidos jugados por otros equipos. Para probar esta teoría, al azar se escogió un equipo de cada dos de la MLB, por ejemplo los Cardenales de San Luis. Se obtuvieron los tiempos de juego (en minutos) para siete partidos seleccionados al azar de los Cardenales en 2005, y ocho partidos seleccionados al azar de los Yanquis en 2005:

Cardenales	Yanquis	Cardenales	Yanquis
164	199	163	187
160	196	175	169
196	202	190	169
171	213		188

Fuente: MLB.com

a. Suponiendo normalidad, ¿estas muestras dan evidencia significativa para concluir que el tiempo medio, de todos los partidos de béisbol de los Yanquis, es significativamente mayor que el tiempo medio de los partidos de los Cardenales?

b. ¿Qué decisión resultará si $\alpha = 0.01$? Explique el cambio.

10.73 Penfield y Perinton son dos suburbios adyacentes del este de Rochester, Nueva York. En años anteriores, siempre habían sido considerados iguales respecto a la calidad de vida, vivienda y educación. Numerosas zonas residenciales se están construyendo en Penfield, y parece que el valor promedio de casas en Penfield es más alto que el de Perinton. Para probar esta teoría, se tomaron muestras aleatorias de transacciones de bienes raíces en cada uno de estos suburbios durante la semana del 12 de junio de 2005. ¿Estos datos apoyan la teoría para este marco de tiempo? Use $\alpha = 0.10$ y suponga normalidad.

Penfield	Perinton	Penfield	Perinton	Penfield	Perinton
$164 500	$127 500	$204 000	$300 000		$140 000
$134 900	$189 500		$212 500		$114 000
$295 000	$106 000		$210 000		$106 000
$235 000	$83 900				

10.74 Cuando se evalúan diferentes instrumentos de medición, primero se debe determinar si hay una diferencia sistemática entre los instrumentos. Se midieron lentes de dos grupos diferentes (1 y 2), uno a la vez, mediante dos instrumentos diferentes. Se registraron las diferencias en mediciones (Instrumento A – Instrumento B). Las unidades de medición han sido codificadas por razones de patente.

Grupo 1	4	5	−2	−3	−7	10	11	−1	3	7	−5	3	−4
	−5	−7	4	−1	−18	0	−17	12	9	4	17	−2	
Grupo 2	−13	−12	−5	11	15	7	−33	−10	−6	−2	−16	2	
	0	−19	6	−17	−4	−19	−22	−4	8	10	−6		

¿Parece haber una diferencia sistemática entre los dos instrumentos?

a. Describa cada conjunto de datos por separado usando un histograma y comparativamente usando una gráfica adyacente.

b. Encuentre la media y desviación estándar para cada conjunto de datos.

c. ¿Se satisfacen las suposiciones? Explique.

d. Pruebe la hipótesis de que no hay diferencia entre las medias de las dos diferencias. Use $\alpha = 0.05$.

e. ¿Hay alguna evidencia de una diferencia entre los dos instrumentos? Explique.

10.75 Use computadora para demostrar la verdad del enunciado que describe la distribución muestral de $\bar{x}_1 - \bar{x}_2$. Use dos poblaciones teóricas normales: $N_1(100, 20)$ y $N_2(120, 20)$.

a. Para conocer las dos poblaciones teóricas, seleccione al azar una muestra muy grande de cada una. Genere valores de datos de 2000, calcule la media y desviación estándar, y construya un histograma usando fronteras de clase que sean múltiplos de la mitad de una desviación estándar (10) comenzando en la media para cada población.

b. Si al azar se seleccionan muestras de tamaño 8 de cada población, ¿qué aspecto espera usted que tenga la distribución de $\bar{x}_1 - \bar{x}_2$ (forma de distribución, media, error estándar)?

c. En forma aleatoria saque una muestra de tamaño ocho de cada población, y encuentre la media de cada muestra. Encuentre la diferencia entre las medias muestrales. Repita 99 veces más.

d. El conjunto de 100 $(\bar{x}_1 - \bar{x}_2)$ valores forma una distribución muestral empírica de $\bar{x}_1 - \bar{x}_2$. Describa la distribución empírica: forma (histograma), media y error estándar. (Use fronteras de clase que sean múltiplos de error estándar de la media para fácil comparación contra lo esperado.)

e. Usando la información hallada en las partes a-d, verifique el enunciado acerca de la distribución muestral $\bar{x}_1 - \bar{x}_2$ hecha en la página 564.

f. Repita el experimento unas pocas veces y compare los resultados.

10.76 Una razón para ser conservador, cuando se determine el número de grados de libertad a usar con la distribución *t*, es la posibilidad de que las varianzas poblacionales puedan ser desiguales. Valores extremadamente diferentes ocasionan una reducción en el número de grados de libertad (df) empleados. Repita el ejercicio 10.75 usando distribuciones normales teóricas de $N(100,9)$ y $N(120, 27)$ y ambos tamaños muestrales de ocho. Compruebe las tres propiedades de la distribución muestral: normalidad, su valor medio y su error estándar. Describa en detalle lo que descubra. ¿Piensa usted que debemos preocuparnos por la selección del df? Explique.

10.77 Los tamaños muestrales no balanceados son un factor para determinar el número de grados de libertad para inferencias acerca de la diferencia entre dos medias. Repita el ejercicio 10.75 usando distribuciones normales teóricas de $N(100, 20)$ y $N(120, 20)$ y tamaños muestrales de = y 20. Compruebe las tres propiedades de la distribución muestral: normalidad, su valor medio y su error estándar. Describa en detalle lo que descubra. ¿Piensa usted que debemos preocuparnos al usar tamaños muestrales no balanceados? Explique.

10.78 Una suposición para la prueba *t* de dos muestras es que "las poblaciones muestreadas deben ser normalmente distribuidas". ¿Qué ocurre cuando no son normalmente distribuidas? Repita el ejercicio 10.75 usando dos poblaciones teóricas que no son normales y muestras de tamaño 10. La distribución exponencial utiliza una variable aleatoria continua, tiene una distribución en forma de J y su media y desviación estándar son del mismo valor. Use dos distribuciones exponenciales con medias de 50 y 80: Exp(50) y Exp(80). Compruebe las tres propiedades de la distribución muestral: normalidad, su valor medio y su error estándar. Describa en detalle lo que descubra. ¿Piensa usted que debemos preocuparnos al muestrear poblaciones no normales? Explique.

10.5 Inferencias relacionadas con las diferencias entre proporciones usando dos muestras independientes

PARA SU INFORMACIÓN
Las tres palabras con "p" (*proporción, porcentaje, probabilidad*) son todas del parámetro binomial p, P(éxito).

PARA SU INFORMACIÓN
En la página 287 se definen experimentos binomiales con más detalle.

Con frecuencia estamos interesados en hacer comparaciones estadísticas entre las **proporciones**, **porcentajes** o **probabilidades** asociados con dos poblaciones. Las siguientes preguntas hacen dichas comparaciones: ¿La proporción de propietarios de casas, que están a favor de cierta propuesta de impuestos, es diferente de la proporción de inquilinos que están a favor de ella? ¿Aprobó estadística un porcentaje más alto de alumnos de este semestre que los del semestre pasado? ¿La probabilidad de que un candidato demócrata gane en Nueva York es más grande que la probabilidad de que un candidato republicano gane en Texas? ¿Las opiniones de estudiantes acerca del nuevo código de conducta difieren de las que tiene el profesorado? Es probable que usted haya hecho preguntas similares.

Nota: éstas son las propiedades de un **experimento binomial**:

1. La probabilidad observada es $p' = x/n$, donde x es el número de éxitos observados en n intentos.
2. $q' = 1 - p'$.
3. p es la probabilidad de éxito en un intento individual en un experimento binomial de probabilidad de n intentos independientes repetidos.

En esta sección vamos a comparar dos proporciones poblacionales con el uso de la diferencia entre las proporciones observadas, $p'_1 - p'_2$, de dos muestras independientes. La diferencia observada, $p'_1 - p'_2$, pertenece a una distribución muestral con las características descritas en el siguiente enunciado.

Si muestras independientes de tamaños n_1 y n_2 se sacan al azar de poblaciones grandes con $p_1 = P_1$(éxito), respectivamente, entonces la distribución muestral de $p'_1 - p'_2$ tiene estas propiedades:

1. media $\mu_{p'_1 - p'_2} = p_1 - p_2$,
2. error estándar $$\sigma_{p'_1 - p'_2} = \sqrt{\left(\frac{p_1 q_1}{n_1}\right) + \left(\frac{p_2 q_2}{n_2}\right)} \qquad (10.10)$$
3. una distribución aproximadamente normal si n_1 y n_2 son lo suficiente grandes

En la práctica, usamos los siguientes *principios para asegurar normalidad:*

1. Ambos tamaños muestrales son mayores de 20.
2. Los productos n_1p_1, n_1q_1, n_2p_2 y n_2q_2 son todos mayores a 5.
3. Las muestras constan de menos de 10% de sus respectivas poblaciones.

Nota: p_1 y p_2 son incógnitas; por tanto, los productos mencionados en el principio serán estimados por $n_1p'_1$, $n_1q'_1$, $n_2p'_2$ y $n_2q'_2$.

Las inferencias acerca de la diferencia entre dos proporciones poblacionales, $p_1 - p_2$, estarán basadas en las siguientes suposiciones.

Suposiciones para inferencias acerca de la diferencia entre dos proporciones $p_1 - p_2$: las n_1 observaciones aleatorias y las n_2 observaciones aleatorias que forman las dos muestras se seleccionan de manera independiente de dos poblaciones que no están cambiando durante el muestreo.

Procedimiento de intervalo de confianza

Cuando estimamos las **diferencia entre dos proporciones,** $p_1 - p_2$, basaremos nuestras estimaciones en la **estadística muestral insesgada** $p'_1 - p'_2$. La estimación puntual $p'_1 - p'_2$ se convierte en el centro del intervalo de confianza y los límites del intervalo de confianza se encuentran usando la fórmula siguiente:

Intervalo de confianza para la diferencia entre dos proporciones

$$(p'_1 - p'_2) - z(\alpha/2) \cdot \sqrt{\left(\frac{p'_1 q'_1}{n_1}\right) + \left(\frac{p'_2 q'_2}{n_2}\right)} \quad \text{a} \quad (p'_1 - p'_2) + z(\alpha/2) \cdot \sqrt{\left(\frac{p'_1 q'_1}{n_1}\right) + \left(\frac{p'_2 q'_2}{n_2}\right)} \tag{10.11}$$

EJEMPLO 10.12

Construcción de un intervalo de confianza para la diferencia entre dos proporciones

Al estudiar sus planes de campaña, Mr. Morris desea estimar la diferencia entre los puntos de vista de hombres y mujeres respecto a su atractivo como candidato. Pide a su gerente de campaña que tome dos muestras aleatorias independientes y encuentra el intervalo de confianza de 99% para la diferencia entre las proporciones de electores mujeres y hombres que planean votar por Morris. Se tomó una muestra de 1000 electores de cada población, con 388 hombres y 459 mujeres que estaban a favor de Morris.

SOLUCIÓN

PARA SU INFORMACIÓN
Se acostumbra poner primero el valor más grande, de modo que la estimación puntual para la diferencia sea un valor positivo.

Paso 1 **Parámetro de interés:** $p_w - p_m$, la diferencia entre la proporción de electoras y la proporción de electores que planean votar por Morris.

Paso 2

a. **Suposiciones:** Las muestras se seleccionan en forma aleatoria e independiente.

b. **Distribución de probabilidad:** La distribución normal estándar. Las poblaciones son grandes (todos los electores); los tamaños muestrales son mayores a 20; y los valores estimados para $n_m p_m$, $n_m q_m$, $n_w p_w$, y $n_w q_w$ son todos mayores a 5. Por tanto, la distribución muestral de $p'_w - p'_m$ debe tener una distribución aproximadamente normal. $z\bigstar$ se calculará usando la fórmula (10.11).

c. **Nivel de confianza:** $1 - \alpha = 0.99$.

Paso 3 **Información muestral:**
Tenemos $n_m = 1000$, $x_m = 388$, $n_w = 1000$, y $x_w = 459$.

$$p'_m = \frac{x_m}{n_m} = \frac{388}{1000} = \mathbf{0.388} \qquad q'_m = 1 - 0.388 = \mathbf{0.612}$$

$$p'_w = \frac{x_w}{n_w} = \frac{459}{1000} = \mathbf{0.459} \qquad q'_w = 1 - 0.459 = \mathbf{0.541}$$

Paso 4 **a.** **Coeficiente de confianza:** Ésta es una situación de dos colas, con $\alpha/2$ en cada cola. De la tabla 4B, $z(\alpha/2) = z(0.005) = 2.58$.
Las instrucciones para usar la tabla 4B están en la página 405.

b. **Máximo error de estimación:** Usando la parte del máximo error de la fórmula (10.11), tenemos

$$E = z(\alpha/2) \cdot \sqrt{\left(\frac{p'_w q'_w}{n_w}\right) + \left(\frac{p'_m q'_m}{n_m}\right)}:$$

$$E = 2.58 \cdot \sqrt{\left(\frac{(0.459)(0.541)}{1000}\right) + \left(\frac{(0.388)(0.612)}{1000}\right)}$$

$$= 2.58\sqrt{0.000248 + 0.000237} = (2.58)(0.022) = \mathbf{0.057}$$

c. **Límites inferior/superior de confianza:**

$$(p'_w - p'_m) \pm E$$

$$0.071 \pm 0.057$$

$$0.071 - 0.057 = \mathbf{0.014} \quad \text{a} \quad 0.071 + 0.057 = \mathbf{0.128}$$

Paso 5 **a.** **Intervalo de confianza:** 0.014 a 0.128 es el intervalo de confianza de 99% para $p_w - p_m$. Con 99% de confianza, podemos decir que hay una diferencia de 1.4% a 12.8% en el atractivo para electores.

b. Esto es, una proporción más grande de mujeres que de hombres está a favor de Mr. Morris, y la diferencia en las proporciones es entre 1.4% y 12.8%.

A veces pueden intercambiarse intervalos de confianza y pruebas de hipótesis; esto es, un intervalo de confianza se puede usar en lugar de una prueba de hipótesis. Por ejemplo, el ejemplo 10.12 pidió un intervalo de confianza. Ahora supongamos que Mr. Morris pregunta, "¿Hay una diferencia en mi atractivo para electores hombres en comparación con electores mujeres?" Para contestar esta pregunta, no sería necesario completar una prueba de hipótesis si se escoge probar en $\alpha = 0.01$ usando una prueba de dos colas. "No hay diferencia" significaría una diferencia de cero, que no está incluida en el intervalo de 0.014 a 0.128 (el intervalo determinado en el ejemplo 10.12). Por tanto, una hipótesis nula de "no hay diferencia" sería rechazada, con lo cual se justifica la conclusión de que existe una diferencia significativa en atractivo para electores entre los dos grupos.

INSTRUCCIONES DE TECNOLOGÍA: INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS PROPORCIONES DADOS DOS CONJUNTOS INDEPENDIENTES DE DATOS MUESTRALES

MINITAB (Versión 14)

```
Choose:     Stat > Basic Statistics > 2 Proportions
Select:     Summarized data:
Enter:      First:    n (trials)     x (events)
            Second:   n (trials)     x (events)
Select:     Options
Enter:      Con?dence level: 1 − α (ex. 0.95 or 95.0)
Select:     Alternative: not equal > OK > OK
```

Excel Introduzca los datos para la primera muestra en la columna A usando números 0 por fallas (o palabras no) y números 1 por éxitos (o palabras sí); luego repita el mismo procedimiento para la segunda muestra en la columna B; luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Z-Estimate: Two Proportions
Enter:      Variable 1 Range: (A2:A20 or select cells)
            Variable 2 Range: (B1:B20 or select cells)
            Code for success: 1
Select:     Labels (if necessary)
Enter:      Alpha: α (ex. 0.05) > OK
```

TI-83/84 Plus

```
Choose:     STAT >TESTS > B:2-PropZint
```

Introduzca los valores apropiados y resalte Calculate.

Procedimiento de prueba de hipótesis

Cuando se pruebe la **hipótesis** nula, **no hay diferencia entre dos proporciones**, la **estadística de prueba** será la diferencia entre las proporciones observadas dividida entre el **error estándar**; se encuentra con la fórmula siguiente:

Estadística de prueba para la diferencia entre dos proporciones-proporción poblacional conocida

$$z\bigstar = \frac{p_1' - p_2'}{\sqrt{pq\left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}} \tag{10.12}$$

Notas:

1. La hipótesis nula es $p_1 = p_2$ o $p_1 - p_2 = 0$ (la diferencia es cero).
2. Las diferencias diferentes de cero entre proporciones no se estudian en esta sección.

3. El numerador de la fórmula (10.12) podría escribirse como $(p'_1 - p'_2) - (p_1 - p_2)$ pero como se supone que la hipótesis nula es verdadera durante la prueba, $p_1 - p_2 = 0$. Por sustitución, el numerador se convierte simplemente en $p'_1 - p'_2$.

4. Como la hipótesis nula es $p_1 = p_2$ el error estándar de $p'_1 - p'_2$, $\sqrt{\left(\frac{p_1 q_1}{n_1}\right) + \left(\frac{p_2 q_2}{n_2}\right)}$, se escribe como $\sqrt{pq\left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}$, donde $p = p_1 = p_2$ y $q = -p$.

5. Cuando la hipótesis nula establece $p_1 = p_2$ y no especifica el valor de p_1 o de p_2, los dos conjuntos de datos muestrales serán agrupados para obtener la estimación de p. Esta probabilidad agrupada (conocida como p'_p) es el número total de éxitos dividido entre el número total de observaciones con las dos muestras combinadas; se encuentra usando la fórmula siguiente:

$$p'_p = \frac{x_1 + x_2}{n_1 + n_2} \tag{10.13}$$

y p'_p es su complemento,

$$q'_p = 1 - p'_p \tag{10.14}$$

Cuando se utilice la estimación agrupada p'_p, la fórmula (10.12) se convierte en la fórmula (10.15):

Estadística de prueba para la diferencia entre dos proporciones-proporción poblacional desconocida

$$z\bigstar = \frac{p'_1 - p'_2}{\sqrt{(p'_p)(q'_p)\left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}} \tag{10.15}$$

EJEMPLO 10.13 Prueba de hipótesis de una cola para la diferencia entre dos proporciones

Una vendedora para un nuevo fabricante de teléfonos celulares dice no sólo que cuestan menos al detallista, sino que también el porcentaje de teléfonos celulares defectuosos hallados entre los productos de ella, no será más alto que el porcentaje de defectuosos hallados en la línea de un competidor. Para probar esta afirmación, el detallista tomó muestras aleatorias del producto de cada fabricante. Los resúmenes muestrales se dan en la tabla 10.5. ¿Podemos rechazar lo dicho por la vendedora al nivel de significación de 0.05?

TABLA 10.5

Información muestral de teléfonos celulares

Producto	Número defectuoso	Número verificado
De la vendedora	15	150
Del competidor	6	150

SOLUCIÓN

Paso 1 **a. Parámetro de interés:** $p_s - p_c$, la diferencia entre la proporción de defectuosos en el producto de la vendedora y la proporción de defectuosos en el producto del competidor.

b. Enunciado de las hipótesis: La preocupación del detallista es que el producto más barato de la vendedora pueda ser de más baja calidad, lo cual significa una mayor proporción de aparatos defectuosos. Si usamos la diferencia "proporción mayor sospechosa − proporción menor", entonces la hipótesis alternativa es "la diferencia es positiva (mayor a cero)".

H_o: $p_s - p_c = 0$ ($\leq$) (porcentaje defectuoso de la vendedora no es más alto que el del competidor)

H_a: $p_s - p_c \cdot 0$ (porcentaje defectuoso de la vendedora es más alto que el del competidor)

Paso 2 **a. Suposiciones:** Se seleccionaron muestras aleatorias de los productos de dos fabricantes diferentes.

b. La estadística de prueba a usar: La distribución normal estándar. Las poblaciones son muy grandes (todos los teléfonos celulares producidos); las muestras son mayores a 20; y los productos estimados $n_s p'_s$, $n_s q'_s$ y $n_c p'_c$, son todos mayores a 5. Por tanto, la distribución muestral debe tener una distribución aproximadamente normal. $z\bigstar$ será calculada usando la fórmula (10.15).

c. Nivel de significación: $\alpha = 0.05$.

Paso 3 **a. Información muestral:**

$$p'_s = \frac{x_s}{n_s} = \frac{15}{150} = \mathbf{0.10} \qquad p'_c = \frac{x_c}{n_c} = \frac{6}{150} = \mathbf{0.04}$$

$$p'_p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{15 + 6}{150 + 150} = \frac{21}{300} = \mathbf{0.07} \qquad q'_p = 1 - p'_p = 1 - 0.07 = \mathbf{0.93}$$

b. Estadística de prueba calculada:

$$z\bigstar = \frac{p'_s - p'_c}{\sqrt{(p'_p)(q'_p)\left[\left(\frac{1}{n_s}\right) + \left(\frac{1}{n_c}\right)\right]}}: \qquad z\bigstar = \frac{0.10 - 0.04}{\sqrt{(0.07)(0.93)\left[\left(\frac{1}{150}\right) + \left(\frac{1}{150}\right)\right]}}$$

$$= \frac{0.06}{\sqrt{0.000868}} = \frac{0.06}{0.02946} = \mathbf{2.04}$$

Paso 4 **Distribución de probabilidad**

Valor *p*:

a. Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor que". **P** = valor $p = P(z\bigstar > 2.04)$ como se ve en la figura.

Clásico:

a. La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "mayor que". El valor crítico se obtiene de la tabla 4A:$z(0.05)$ = **1.65.**

Para hallar el valor p, tiene tres opciones:

1. Usar la tabla 3 (apéndice B) para calcular el valor p: $\mathbf{P} = 0.5000 - 0.4793 = \mathbf{0.0207}$.
2. Usar la tabla 5 (apéndice B) para poner límites al valor p: $\mathbf{0.0202} < \mathbf{P} < \mathbf{0.0228}$.
3. Usar computadora o calculadora: $\mathbf{P} = \mathbf{0.0207}$.

Para instrucciones específicas, vea la página 432.

b. El valor p es menor que α.

Para instrucciones específicas, vea la página 451.

b. $z\bigstar$ está en la región crítica, como se muestra en **rojo** en la figura.

Paso 5

a. Decisión: Rechazar H_o.

b. Conclusión: Al nivel de significación de 0.05, hay suficiente evidencia para rechazar lo dicho por la vendedora; la proporción de los teléfonos celulares defectuosos de su compañía es más alta que la proporción de teléfonos celulares defectuosos de su competidor.

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA LA DIFERENCIA ENTRE DOS PROPORCIONES, $p_1 - p_2$, PARA DOS CONJUNTOS INDEPENDIENTES DE DATOS MUESTRALES

MINITAB (Versión 14)

```
Choose:     Stat > Basic Statistics > 2 Proportions
Select:     Summarized data:
Enter:      First:    n (trials)     x (events)
            Second:   n (trials)     x (events)
Select:     Options
Enter:      Test difference: 0.0
Select:     Alternative: less than or not equal or greater than
Select:     Use pooled estimate of p for test > OK > OK
```

Excel

Introduzca los datos para la primera muestra en la columna A usando números 0 por fracaso (o palabras no) y números 1 por éxitos (o palabras sí); luego repita el mismo procedimiento para la segunda muestra en la columna B, luego continúe con:

```
Choose:     Tools > Data Analysis Plus > Z-Test: Two Proportions
Enter:      Variable 1 Range: (A1:A20 or select cells)
            Variable 2 Range: (B1:B20 or select cells)
            Code for success: 1
            Hypothesized difference: 0
Select:     Labels (if necessary)
Enter:      Alpha: α (ex. 0.05)
```

TI-83/84 Plus

```
Choose:     STAT > TESTS > 6:2-PropZTest . . .
```

Introduzca los valores apropiados y resalte Calculate.

CASO PRÁCTICO 10.14 Los riñones que donan las personas fallecidas son buenos para trasplantes

En un descubrimiento que podría facilitar la grave escasez de órganos de donadores, investigadores suizos encontraron que los riñones trasplantados de cadáveres siguen funcionando tanto como los de un paciente cuyo corazón esté todavía latiendo. La mayor parte de los órganos de trasplantes se toman de pacientes con muerte cerebral pero cuyos corazones no se hayan detenido, porque durante mucho tiempo los médicos han pensado que si esperan hasta que el corazón se detenga los órganos se dañarán por falta de oxígeno.

No obstante, en el primer estudio de largo plazo que compara los dos métodos, los médicos del hospital universitario de Zurich hicieron seguimiento de hasta 15 años a casi 250 pacientes con trasplante, y encontraron porcentajes de supervivencia casi idénticos. Después de 10 años, 79% de pacientes cuyo riñón provenía de un donante cuyo corazón ya no latía estaban vivos, al igual que 77% de pacientes cuyo órgano provenía de un donador con muerte cerebral pero con su corazón todavía latiendo. El estudio, publicado en la revista *New England Journal of Medicine* del jueves, pudo resultar especialmente influyente porque fue una comparación directa de los dos métodos y fue el primero en dar seguimiento a los pacientes durante muchos años.

Los médicos piensan que pueden obtenerse resultados semejantes por trasplantes de hígado, páncreas y pulmones. Con el uso de órganos de donadores de "muerte cardiaca", el número de riñones disponible podría aumentar hasta 30%, lo que significa unos 1000 o más donadores extra al año en Estados Unidos, estiman expertos.

Fuente: Reimpreso con permiso de The Associated Press.

SECCIÓN 10.5 EJERCICIOS

10.79 Sólo 75 de las 250 personas entrevistadas pudieron dar el nombre del vicepresidente de Estados Unidos. Encuentre los valores para x, n, p' y q'.

10.80 Si $n_1 = 40$, $p'_1 = 0.9$, $n_2 = 50$ y $p'_2 = 0.9$:

a. Encuentre los valores estimados para ambas np y ambas nq.

b. ¿Esta situación satisfaría los principios para ser aproximadamente normal? Explique.

10.81 Calcule la estimación para el error estándar de la diferencia entre dos proporciones para cada uno de los casos siguientes:

a. $n_1 = 40$, $p'_1 = 0.8$, $n_2 = 50$ y $p'_2 = 0.8$

b. $n_1 = 33$, $p'_1 = 0.6$, $n_2 = 38$ y $p'_2 = 0.65$

10.82 Calcule el máximo error de estimación para un intervalo de confianza de 90% para la diferencia entre dos proporciones para los casos siguientes:

a. $n_1 = 40$, $p'_1 = 0.7$, $n_2 = 44$ y $p'_2 = 0.75$

b. $n_1 = 36$, $p'_1 = 0.33$, $n_2 = 38$ y $p'_2 = 0.42$

10.83 Un artículo de *Nursing Economics* titulado "Cambio de enfermeras ejecutivas" comparó dos grupos de enfermeras ejecutivas. Un grupo había participado en un programa de características únicas para enfermeras ejecutivas llamado Wharton Fellows, y el otro grupo no había participado en el programa. De 341 Wharton Fellows, 87 habían tenido un cambio en su posición; de 40 que no eran del Wharton Fellows, 9 habían tenido un cambio en su posición. Se usó el MINITAB para construir un intervalo de confianza de 99% para la diferencia en proporciones de población. Verifique los resultados que siguen calculándolos usted mismo.

```
Test and CI for Two Proportions
Sample   X        N        Sample p
1        87       341      0.255132
2        9        40       0.225000
Difference = p (1) - p (2)
Estimate for difference: 0.0301320
99% CI for difference: (-0.150483, 0.210747)
```

10.84 Encuentre el intervalo de confianza de 95% para $p_A - p_B$.

Muestra	n	x
A	125	45
B	150	48

10.85 "El atractivo del juego para adolescentes 'es igual que para adultos'", un artículo de *USA Today* (21 de diciembre de 2004), publicó una encuesta de 200 muchachos adolescentes y 199 muchachas adolescentes de entre 14 y 17 años. Los hallazgos publicados fueron que 66% de los muchachos "alguna vez habían apostado" y 37% de las muchachas "alguna vez habían apostado". La encuesta fue patrocinada por la encuesta Annenberg nacional de riesgos para la juventud efectuada en 2003.

Use un 95% de intervalo de confianza para estimar la diferencia en la proporción de adolescentes hombres y mujeres que alguna vez han apostado. Compare sus respuestas con los intervalos de confianza formados en el ejercicio 9.74 de la página 509. Comente sobre lo que dicen estos dos métodos. Incluya un repaso a los comentarios escritos en respuesta al ejercicio 9.74 parte c.

10.86 En una muestra aleatoria de 40 personas de cabello café, 22 de ellas indicaron que usaban tintes para el pelo. En otra muestra aleatoria de 40 rubias, 26 indicaron que habían usado tintes para el pelo. Use un intervalo de confianza de 92% para estimar la diferencia en las proporciones poblacionales de morenas y rubias que usan tintes para el pelo.

10.87 Se compararon las proporciones de piezas defectuosas producidas por dos máquinas, y se recolectaron los siguientes datos:

Máquina 1: $n = 150$; número de piezas defectuosas = 12

Máquina 2: $n = 150$; número de piezas defectuosas = 6

Determine un intervalo de confianza de 90% para $p_1 - p_2$.

10.88 En una encuesta hecha a 300 personas de la ciudad A, 128 de ellas preferían el jabón Nueva Primavera a todas las otras marcas de jabón desodorante. En la ciudad B, 149 de 400 personas preferían el jabón Nueva Primavera. Encuentre el intervalo de confianza de 98% para la diferencia en las proporciones de personas de las dos ciudades que prefieren el jabón Nueva Primavera.

10.89 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar estas frases:

a. No hay diferencia entre las proporciones de hombres y mujeres que votarán por el apoyado en la elección del próximo mes.

b. El porcentaje de muchachos que faltan a clases es mayor que el de muchachas que faltan a clases.

c. El porcentaje de estudiantes universitarios que manejan autos viejos es más alto que el porcentaje de personas no universitarias de la misma edad que manejan autos viejos.

10.90 Demuestre que el error estándar de $p'_1 - p'_2$, que es $\sqrt{\left(\frac{p_1q_1}{n_1}\right) + \left(\frac{p_2q_2}{n_2}\right)}$, se reduce a $\sqrt{pq\left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}$ cuando $p_1 = p_2 = p$.

10.91 Encuentre los valores de p'_p y q'_p para estas muestras:

Muestra	x	n
E	15	250
R	25	275

10.92 Encuentre el valor de $z\bigstar$ que se usaría para probar la diferencia entre las proporciones, dado lo siguiente:

Muestra	n	x
G	380	323
H	420	332

10.93 Encuentre el valor p para la prueba con hipótesis alternativa $p_E < p_R$ usando los datos del ejercicio 10.91.

10.94 Determine el valor p que se usaría para probar las siguientes hipótesis cuando se usa z como estadística de prueba.

a. H_o: $p_1 = p_2$ vs. H_a: $p_1 > p_2$, con $z\bigstar = 2.47$

b. H_o: $p_A = p_B$ vs. H_a: $p_A \neq p_B$, con $z\bigstar = -1.33$

c. H_o: $p_1 - p_2 = 0$ vs. H_a: $p_1 - p_2 < 0$, con $z\bigstar = -0.85$

d. H_o: $p_m - p_f = 0$ vs. H_a: $p_m - p_f > 0$, con $z\bigstar = 3.04$

10.95 Determine la región crítica y valor(es) crítico(s) que se usarían para probar (procedimiento clásico) las siguientes hipótesis cuando se usa z como estadística de prueba.

a. H_o: $p_1 = p_2$ vs. H_a: $p_1 > p_2$, con $\alpha = 0.05$

b. H_o: $p_A = p_B$ vs. H_a: $p_A \neq p_B$, con $\alpha = 0.05$

c. H_o: $p_1 - p_2 = 0$ vs. H_a: $p_1 - p_2 < 0$, con $\alpha = 0.04$

d. H_o: $p_m - p_f = 0$ vs. H_a: $p_m - p_f > 0$, con $\alpha = 0.01$

10.96 Es frecuente que usuarios de computadoras personales (PC) tengan problemas de hardware. Un estudio reveló que los problemas de hardware reportados a fabricantes podrían no ser reparados por uno de cada tres propietarios de computadoras personales. Les fue peor a usuarios de PC para el hogar que a quienes trabajan con estas máquinas, ya que esperan más tiempo para recibir servicio y se les resuelven menos problemas. Relativamente pocos propietarios calificaron bien a técnicos de servicio porque hayan tenido éstos experiencia adecuada, o por haberse esforzado sinceramente en ayudar a resolver problemas con el hardware.

Fuente: PC World, "Which PC Makers Can You Trust?".

Supongamos que se realiza un estudio para comparar el servicio proporcionado por fabricantes a propietarios de computadoras personales (PC) para el hogar y el trabajo. De 220 propietarios de PC para el hogar, 98 informaron que su problema no fue resuelto en forma satisfactoria. Cuando se hizo la misma pregunta de 180 propietarios de PC para el trabajo que tenían dificultades con sus máquinas, 52 informaron que su problema no había sido resuelto. ¿Los propietarios de PC para el hogar tenían una proporción mayor de problemas que no pudieron ser resueltos con ayuda del fabricante? Use el nivel de significación de 0.05 y la salida MINITAB siguiente para contestar la pregunta.

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

```
Test and CI for Two Proportions
Sample   X        N        Sample p
1        98       220      0.445455
2        52       180      0.288889
Difference = p (1) - p (2)
Estimate for difference: 0.156566
Test for difference = 0 (vs > 0): Z = 3.22
P-Value = 0.001
```

10.97 Dos grupos de ciudadanos seleccionados al azar fueron expuestos a diferentes campañas de medios de propaganda que se referían a la imagen de un candidato político. Una semana después, los grupos de ciudadanos fueron encuestados para ver si votarían por el candidato. Los resultados fueron como sigue:

	Expuesto a imagen conservadora	Expuesto a imagen moderada
Número de muestra	100	100
Proporción para el candidato	0.40	0.50

¿Hay suficiente evidencia para demostrar una diferencia en la efectividad de las dos campañas de imagen al nivel de significación de 0.05?

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

10.98 En una encuesta de familias en la que ambos padres trabajan, una de las preguntas formuladas fue, "Ha rechazado usted un trabajo, promoción o transferencia porque eso significaría menos tiempo con su familia?" A un total de 200 hombres y 200 mujeres se les hizo esta pregunta. "Sí" fue la respuesta dada por 29% de los hombres y 24% de las mujeres. Con base en esta encuesta, ¿podemos concluir que hay una diferencia en la proporción de hombres y mujeres que responden "sí" al nivel de significación de 0.05?

10.99 La Comisión de los 200, una organización profesional de prominentes empresarias y líderes corporativas, publicó lo siguiente: 60% de mujeres estudiantes de master en administración de empresa dice que "las empresas pagan a sus ejecutivos demasiado" y 50% de los estudiantes hombres de master en administración están de acuerdo.

a. ¿Parece haber diferencia en la proporción de mujeres y hombres que dicen que "A los ejecutivos se les paga demasiado"? Explique el significado de su respuesta.

b. Si los porcentajes precedentes resultaron de dos muestras de tamaño 20 cada una, ¿la diferencia es estadísticamente significativa al nivel de significación de 0.05? Justifique su respuesta.

c. Si los porcentajes precedentes resultaron de dos muestras de tamaño 500 cada una, ¿la diferencia es estadísticamente significativa al nivel de significación de 0.05? Justifique su respuesta.

d. Explique la forma en que las respuestas a las partes b y c afectan sus pensamientos acerca de su respuesta a la parte a.

10.100 Tanto los estudiantes como sus padres tienen muchas preocupaciones cuando hablan de universidades. Una de las tres principales preocupaciones, según el estudio College Partnership realizado el otoño de 2004, es "Escoger la mejor o más importante carrera". Diecinueve por ciento de los padres dijeron que "Escoger la mejor o más importante carrera" es la preocupación más importante, mientras que 15% de los estudiantes dijeron que es una preocupación importante.

Fuente: http://www.collegepartnership.com/pdf/Fall%20200 4%20Study%20Charts%20&%20 Analysis% 20_1_.pdf

Si el estudio fue realizado con una muestra de 1750 estudiantes y sus padres, pruebe la hipótesis de que "Escoger la mejor o más importante carrera" es una preocupación más grande para los padres, al nivel de significación de 0.05.

10.101 Se fabricaron y probaron 41 pequeños lotes de producto experimental para el caso de presentarse una indicación particular que es atributo en la naturaleza, pero que causa rechazo de la parte. Se hicieron 31 lotes usando un método particular de procesamiento, y se hicieron 10 lotes usando un segundo método de procesamiento. Cada lote fue igualmente muestreado ($n = 32$) para la presencia de esta indicación. En la práctica, las condiciones óptimas de procesamiento muestran poca o ninguna presencia de la indicación. El método 1, que comprende los diez lotes, se corrió antes que el método 2.

Métodos	n	Número de rechazos
Método 1	320	4
Método 2	992	26

Cortesía de Bausch & Lomb

Determine, al nivel de significación de 0.05, si hay diferencia en la proporción del rechazo de producto entre los dos métodos. (Guarde su respuesta para compararla con los ejercicios 11.44 de la p. 645.)

10.102 Los efectos colaterales adversos son siempre una preocupación cuando se prueban nuevos medicamentos. Se realizaron estudios clínicos controlados con placebo en pacientes de 12 años de edad y mayores, quienes recibieron una dosis de "una al día" de Allegra, un medicamento estacional para combatir alergias. Los siguientes resultados se publicaron en la edición de abril de 2005 de *Readers' Digest.*

	Allegra (dosis de una al día)	Placebo (dosis de una al día)
Efectos colaterales	$n = 283$	$n = 293$
Número de dolores de cabeza reportados	30	22

Determine, al nivel de significación de 0.05, si hay diferencia en la proporción de pacientes que informan de dolores de cabeza entre los dos grupos.

10.103 Una encuesta de Harris Interactive for Korbel, realizada en 2005, encontró que 63% de hombres y 55% de mujeres piensan que está bien que las mujeres propongan matrimonio a los hombres. Una diferencia del 8% puede o no puede ser estadísticamente significativa. ¿Qué tamaño muestral se requiere para que esta diferencia sea significativa?

Fuente: USA Snapshot en Internet, 25 de junio de 2005.

a. Si las estadísticas muestrales precedentes habían resultado de una muestra de 250 hombres y una muestra de 250 mujeres, ¿sería significativa la diferencia, usando $\alpha = 0.05$? Explique.

b. Si las muestras hubieran sido cada una de tamaño 500, ¿sería significativa la diferencia, usando $\alpha = 0.05$? Explique.

c. Determine el tamaño muestral que tendría la diferencia de 0.08 correspondiente a $p = 0.05$.

10.104 Los principios para asegurar que la distribución muestral de $P'_1 - P'_2$ sea normal incluyen varias condiciones acerca del tamaño de diversos valores. Las dos distribuciones binomiales $B(100, 0.3)$ y $B(100, 0.4)$ satisfacen todos estos principios.

a. Verifique que $B(100, 0.3)$ y $B(100, 0.4)$ satisfagan todos los principios.

b. Use computadora para generar al azar 200 muestras aleatorias de cada una de las poblaciones binomiales. Encuentre la proporción observada para cada muestra y el valor de las 200 diferencias entre dos proporciones.

c. Describa la distribución muestral observada usando estadísticas gráficas y numéricas.

d. ¿La distribución muestral empírica parece tener una distribución aproximadamente normal? Explique

10.6 Inferencias relacionadas con la razón de varianzas usando dos muestras independientes

Cuando se comparan dos poblaciones, naturalmente comparamos sus dos características de distribución más fundamentales, su "centro" y su "dispersión", al comparar sus medias y desviaciones estándar. Hemos aprendido, en dos secciones previas, cómo usar la distribución t para hacer inferencias para comparar dos medias poblacionales con muestras dependientes o independientes. Estos procedimientos tenían la intención de usarse con poblaciones normales, pero funcionan muy bien cuando las poblaciones no están distribuidas normalmente en forma exacta.

El siguiente paso lógico al comparar dos poblaciones es comparar sus desviaciones estándar, que es la medida de dispersión que se usa con más frecuencia. No obstante, las distribuciones muestrales que se refieren a desviaciones muestrales estándar (o varianzas) son muy sensibles a ligeras desviaciones a partir de las suposiciones. Por tanto, el único procedimiento de inferencia a presentarse aquí será la **prueba de hipótesis para la igualdad de desviaciones (o varianzas) estándar** para dos poblaciones normales.

La compañía embotelladora de bebidas gaseosas que vimos en la sección 9.4 (pp. 516, 520) está tratando de decidir si instalar una máquina embotelladora moderna y de alta velocidad. Hay, por supuesto, muchas preocupaciones para tomar esta decisión, y una de ellas es que la mayor velocidad puede resultar en una mayor variabilidad en la cantidad de líquido introducido en cada botella; este aumento podría no ser aceptable. Para esta preocupación, el fabricante del nuevo sistema respondió que la varianza en llenados no será mayor con la nueva máquina que con la anterior. (El nuevo sistema llenará varias botellas en el mismo tiempo en que el sistema anterior llena una botella; ésta es la razón por la que se está considerando el cambio.) Se establece una verificación para probar estadísticamente la preocupación de la compañía embotelladora, "La desviación estándar de la nueva máquina es mayor que la desviación estándar de la anterior", contra el dicho del fabricante de que "La desviación estándar de la nueva no es mayor que la desviación estándar de la anterior".

EJEMPLO 10.15 Escritura de hipótesis para la igualdad de varianzas

Exprese la hipótesis nula y alternativa a usar para comparar las varianzas de las dos máquinas embotelladoras de bebidas gaseosas.

SOLUCIÓN Hay varias formas equivalentes de expresar las hipótesis nula y alternativa, pero debido a que el procedimiento de prueba utiliza la razón entre varianzas, la convención recomendada es expresar las hipótesis nula y alternativa como razones entre las varianzas poblacionales. Además, se recomienda que varianza "más grande" o "que se espera más grande" sea el numerador. La preocupación de la compañía embotelladora de bebidas gaseosas es que la nueva máquina moderna (m) resulte en una desviación estándar más grande en las cantidades de llenado que su máquina actual (p); $\sigma_m > \sigma_p$ o, lo que es lo mismo, $\sigma_m^2 > \sigma_p^2$, que se

convierte en $\frac{\sigma_m^2}{\sigma_p^2} > 1$. Deseamos probar lo dicho por el fabricante (la hipótesis nula) contra la preocupación de la compañía (la hipótesis alternativa):

$$H_o: \frac{\sigma_m^2}{\sigma_p^2} = 1 \quad (m \text{ no es más variable})$$

$$H_a: \frac{\sigma_m^2}{\sigma_p^2} > 1 \quad (m \text{ es más variable})$$

Las inferencias acerca de la razón entre varianzas para dos poblaciones normalmente distribuidas usan la **distribución *F***. La distribución F, similar a la distribución t de Student y a la distribución χ^2, es una familia de distribuciones de probabilidad. Cada distribución F está identificada por dos números de grados de libertad, uno por cada una de las dos muestras involucradas.

Antes de continuar con los detalles del procedimiento de prueba de hipótesis, aprendamos acerca de la distribución F.

Propiedades de la distribución *F*:

1. F es no negativa; es cero o positiva.
2. F es no simétrica; es sesgada a la derecha.
3. F está distribuida de modo de formar una familia de distribuciones; hay una distribución separada por cada par de números de grados de libertad.

Para inferencias estudiadas en esta sección, el número de grados de libertad para cada muestra es $df_1 = n_1 - 1$ y $df_2 = n_2 - 1$. Cada combinación diferente de grados de libertad resulta en una distribución F diferente, y cada distribución F se ve aproximadamente como la distribución que se ilustra en la figura 10.2.

FIGURA 10.2
Distribución *F*

Los valores críticos par la distribución F se identifican usando tres valores:

df_n, los grados de libertad asociados con la muestra cuya varianza está en el numerador de la F calculada

df_d, los grados de libertad asociados con la muestra cuya varianza está en el denominador

α, el área bajo la curva de distribución a la derecha del valor crítico que se busca

Por tanto, el nombre simbólico para el valor crítico de F será $F(df_n, df_d, \alpha)$, como se ve en la figura 10.3 (vea p. 594).

Debido a que se necesitan tres valores para identificar un solo valor crítico de F, hacer tablas para F no es tan sencillo como con las distribuciones previamente estudiadas. Las tablas presentadas en este libro de texto están organizadas de modo que tengan una tabla diferente para cada valor diferente de α, el "área a la derecha".

FIGURA 10.3
Un valor crítico de *F*

La tabla 9A en el apéndice B muestra los valores críticos para $F(\mathrm{df}_n, \mathrm{df}_d, \alpha)$, cuando $\alpha = 0.05$; la tabla 9B da los valores críticos cuando $\alpha = 0.025$; la tabla 9C da los valores cuando $\alpha = 0.01$.

EJEMPLO 10.16 Búsqueda de valores *F* críticos

Encuentre $F(5, 8, 0.05)$, el valor crítico F para muestras de tamaño 6 y tamaño 9 con 5% del área en la cola derecha.

SOLUCIÓN Usando la tabla 9A ($\alpha = 0.05$), encuentre la intersección de la columna df = 5 (para el numerador) y la fila df = 8 (para el denominador) y lea el valor: $F(5, 8, 0.05) =$ **3.69**. Vea la siguiente tabla parcial.

Parte de la tabla 9A ($\alpha = 0.05$)

		df para numerador					
		...	5	...	8	...	
df	⋮						
para	5				**4.82**		$F(8, 5, 0.05) = 4.82$
denomi-	⋮						
nador	8		**3.69**				$F(5, 8, 0.05) = 3.69$
	⋮						

Observe que $F(5, 8, 0.05)$ es **4.82**. Los grados de libertad asociados con el numerador y con el denominador deben mantenerse en el orden correcto; 3.69 es diferente de 4.82. Compruebe otros pares para verificar que intercambiar los números de grados de libertad resultará en valores F diferentes.

INSTRUCCIONES DE TECNOLOGÍA: PROBABILIDAD ACUMULATIVA ASOCIADA CON UN VALOR ESPECIFICADO DE *F*

MINITAB (Versión 14)

```
Choose:     Calc > Probability Distributions > F
Select:     Cumulative Probability Noncentrality parameter: 0.0
Enter:      Numerator degrees of freedom: dfn
            Denominator degrees of freedom: dfd
Select:     Input constant*
Enter:      F-value (ex. 1.74) > OK
```

*Seleccione la columna Input si diversos valores *F* se guardan en C1. Use C2 para guardado opcional. Si es necesaria el área de la cola derecha, de uno reste la probabilidad calculada.

Excel Si han de usarse diversos valores F, introduzca los valores en la columna A y active B1; luego continúe con:

```
Choose:      Insert function fx > Statistical > FDIST > OK
Enter:       X: individual F-value or (A1:A5 or select 'F-value' cells)*
             Deg_freedom 1: dfn
             Deg_freedom 2: dfd > OK
*Drag:       Bottom right corner of the B1 cell down to give other proba-
             bilities
```

Para hallar la probabilidad para la cola izquierda (la probabilidad acumulativa hasta el valor F), de uno reste la probabilidad calculada.

TI-83/84 Plus

```
Choose:      2nd > DISTR > 9:Fcdf(
Enter:       0, F-value, dfn, dfd)
```

Nota: para hallar la probabilidad entre dos valores F, introduzca los dos valores en lugar de 0 y el valor F.

Si es necesaria el área en la cola derecha, de uno reste la probabilidad calculada.

El uso de la distribución F tiene una condición.

Suposiciones para inferencias acerca de la razón entre dos varianzas: las muestras se seleccionan al azar de poblaciones normalmente distribuidas, y las dos muestras se seleccionan de manera independiente.

Estadística de prueba para igualdad de varianzas

$$F\bigstar = \frac{s_m^2}{s_p^2}, \qquad \text{con } \mathrm{df}_m = n_m - 1 \text{ y } \mathrm{df}_p = n_p - 1 \tag{10.16}$$

Las varianzas muestrales se asignan al numerador y denominador en el orden establecido por las hipótesis nula y alternativa para pruebas de una cola. La razón calculada, $F\bigstar$, tendrá una distribución F con $\mathrm{df}_n = n_n - 1$ (numerador) y $\mathrm{df}_d = n_d - 1$ (denominador) cuando las suposiciones se satisfagan y la hipótesis nula sea verdadera.

Estamos listos para usar F para completar una prueba de hipótesis acerca de la razón entre dos varianzas poblacionales.

EJEMPLO 10.17 Prueba de hipótesis de una cola para la igualdad de varianzas

Recuerde que nuestra compañía embotelladora de bebidas gaseosas iba a tomar una decisión acerca de la igualdad de las varianzas de cantidades de llenado entre su máquina actual y el moderno equipo de alta velocidad. ¿La información muestral de la tabla 10.6 (p. 596) presenta suficiente evidencia para rechazar la hipótesis nula, lo dicho por el fabricante, de que la moderna máquina de alta velocidad llenadora de botellas llena botellas con una varianza no mayor a la de la máquina actual de la compañía? Suponga que las cantidades de llenado están normalmente distribuidas para ambas máquinas, y complete la prueba usando $\alpha = 0.01$.

TABLA 10.6

Información muestral sobre varianzas de llenado

Muestra	n	s^2
Máquina actual (p)	22	0.0008
Máquina moderna de alta velocidad (m)	25	0.0018

SOLUCIÓN

Paso 1 **a. Parámetro de interés:** $\frac{\sigma_m^2}{\sigma_p^2}$, la razón entre las varianzas en las cantidades de llenado puestas en botellas para la máquina moderna contra la máquina actual de la compañía.

b. Enunciado de las hipótesis: Las hipótesis se establecieron en el ejemplo 10.15 (pp. 592-593):

$$H_o: \frac{\sigma_m^2}{\sigma_p^2} = 1 \ (\leq) \ (m \text{ no es más variable})$$

$$H_a: \frac{\sigma_m^2}{\sigma_p^2} > 1 \qquad (m \text{ es más variable})$$

Nota: cuando la varianza "que se espera más grande" está en el numerador para una prueba de una cola, la hipótesis alternativa expresa que "La razón entre las varianzas es mayor a uno".

Paso 2 **a. Suposiciones:** Las poblaciones muestreadas están normalmente distribuidas (dadas en el enunciado del problema), y las muestras se seleccionan de manera independiente (sacadas de dos poblaciones separadas).

b. Estadística de prueba: La distribución F con la razón entre las varianzas muestrales y la fórmula (10.16):

c. Nivel de significación: $\alpha = 0.01$.

Paso 3 **a. Información muestral:** Vea la tabla 10.6.

b. Estadística de prueba calculada:
Usando la fórmula (10.16), tenemos

$$F\bigstar = \frac{s_m^2}{s_p^2}: \qquad F\bigstar = \frac{0.0018}{0.0008} = \mathbf{2.25}$$

El número de grados de libertad para el numerador es $df_n = 24$ (o $25 - 1$) porque la muestra de la máquina moderna de alta velocidad está asociada con el numerador, como especifica la hipótesis nula. También, $df_d = 21$ porque la muestra asociada con el denominador tiene tamaño 22.

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Use la cola derecha porque H_a expresa interés por valores relacionados con "mayor que." $\mathbf{P} = P(F\bigstar > 2.25$, con $df_n = 24$ y $df_d = 21)$ como se muestra en la figura.

Clásico:

a. La región crítica es la cola derecha porque H_a expresa interés por valores relacionados con "mayor que." $df_n = 24$ y $df_d = 21$. El valor crítico se obtiene de la tabla 9C: $F(24, 21, 0.01) = \mathbf{2.80}$.

Para hallar el valor p, tiene dos opciones:

1. Usar tablas 9A y 9B (apéndice B) para poner límites al valor p: $\mathbf{0.025 < P < 0.05}$.
2. Usar computadora o calculadora para hallar el valor p: $\mathbf{P = 0.0323}$.

Instrucciones específicas siguen a esta ilustración

b. El valor p no es menor que el nivel de significación, α (0.01).

Para instrucciones adicionales, vea página 594.

b. $F\bigstar$ no está en la región crítica, como se muestra en **rojo** en la figura.

Paso 5 **a. Decisión:** No rechazar H_o.

b. Conclusión: Al nivel de significación de 0.01, las muestras no presentan suficiente evidencia para indicar un aumento en varianza con la nueva máquina.

Cálculo del valor *p* cuando se usa la distribución *F*

Método 1: *use la tabla 9 del apéndice B para poner límites al valor p*. El uso de las tablas 9A, 9B y 9C del apéndice B para estimar el valor p es muy limitado. No obstante, para el ejemplo 10.17, el valor p se puede estimar. Por inspección de las tablas 9A y 9B, el lector encontrará que $F(24, 21, 0.025) = 2.37$ y $F(24, 21, 0.05) = 2.05$. $F\bigstar = 2.25$ está entre los valores 2.37 y 2.05; por tanto, el valor p está entre 0.025 y 0.05: $\mathbf{0.025 < P < 0.05}$. (Vea la figura al margen.)

Método 2: si está haciendo la prueba de hipótesis con ayuda de una computadora o calculadora, es muy probable que ésta calcule el valor p, o puede usted usar los comandos de distribución de probabilidad acumulativa descritos en las páginas 594-595.

Valores *F* críticos para pruebas de una cola y dos colas

PARA SU INFORMACIÓN

α todavía debe dividirse entre las dos colas para una H_a de dos colas.

Las tablas de valores críticos para la distribución F dan sólo los valores críticos derechos. Esto no será problema porque el valor crítico derecho es el único valor crítico que será necesario. Se puede ajustar el orden numerador-denominador de modo que toda la "actividad" esté en la cola derecha. Hay dos casos: pruebas de una cola y pruebas de dos colas.

Pruebas de una cola: arregle las hipótesis nula y alternativa de modo que la alternativa sea siempre "mayor que". El valor $F\bigstar$ se calcula usando el mismo orden como se especifica en la hipótesis nula (como en el ejemplo 10.17; también vea el ejemplo 10.18).

Pruebas de dos colas: cuando calcule el valor $F\bigstar$, siempre use la muestra con la varianza más grande para el numerador; esto hará a $F\bigstar$ mayor a uno y la pondrá en la cola derecha de la distribución. Así, necesitará solo el valor crítico para la cola derecha (vea ejemplo 10.19).

Todas las pruebas de hipótesis acerca de dos varianzas se pueden formular y completar en una forma que tanto el valor crítico de F como el valor calculado de $F\bigstar$ se encuentren en la cola derecha de la distribución. Como las tablas 9A, 9B y 9C

contienen sólo valores críticos para la cola derecha, esto será conveniente y nunca serán necesarios valores críticos para la cola izquierda. Los siguientes dos ejemplos demostrarán cómo se logra esto.

EJEMPLO 10.18 Formato para escribir hipótesis para la igualdad de varianzas

Reorganice la hipótesis alternativa de modo que la región crítica sea la cola derecha:

$$H_a\colon \sigma_1^2 < \sigma_2^2 \quad \text{o} \quad \frac{\sigma_1^2}{\sigma_2^2} < 1 \quad \text{(población 1 es más variable)}$$

SOLUCIÓN Invierta la dirección de la desigualdad, e invierta los papeles del numerador y denominador.

$$H_a\colon \sigma_2^2 > \sigma_1^2 \quad \text{o} \quad \frac{\sigma_2^2}{\sigma_1^2} > 1 \quad \text{(población 2 es más variable)}$$

La estadística de prueba calculada $F\bigstar$ será $\frac{s_2^2}{s_1^2}$.

EJEMPLO 10.19 Prueba de hipótesis de dos colas para la igualdad de varianzas

Encuentre $F\bigstar$ y valores críticos para la siguiente prueba de hipótesis, de modo que sólo sea necesario el valor crítico derecho. Use $\alpha = 0.05$ y la información muestral $n_1 = 10$, $n_2 = 8$, $s_1 = 5.4$, y $s_2 = 3.8$.

$$H_o\colon \sigma_2^2 = \sigma_1^2 \quad \text{o} \quad \frac{\sigma_2^2}{\sigma_1^2} = 1$$

$$H_a\colon \sigma_2^2 \neq \sigma_1^2 \quad \text{o} \quad \frac{\sigma_2^2}{\sigma_1^2} \neq 1$$

SOLUCIÓN Cuando la hipótesis alternativa sea de dos colas ($\neq$), la $F\bigstar$ calculada puede ser $F\bigstar = \frac{s_1^2}{s_2^2}$ o $F\bigstar = \frac{s_2^2}{s_1^2}$. La opción es nuestra; sólo necesitamos asegurarnos de mantener df_n y df_d en el orden correcto. Tomamos la opción al ver la información muestral y usar la muestra con la desviación estándar o varianza más grande como numerador. Por tanto, en esta ilustración,

$$F\bigstar = \frac{s_1^2}{s_2^2} = \frac{5.4^2}{3.8^2} = \frac{29.16}{14.44} = \mathbf{2.02}$$

Los valores críticos para esta prueba son la cola izquierda, $F(9, 7, 0.975)$, y la cola derecha, $F(9, 7, 0.025)$, como se muestra en la figura.

Como escogemos la muestra con la mayor desviación estándar (o varianza) para el numerador, el valor de $F\bigstar$ será mayor a 1 y estará en la cola derecha; por tanto, sólo es necesario el valor crítico derecho. (Todos los valores críticos para colas izquierdas serán valores entre 0 y 1.)

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS PARA LA RAZÓN ENTRE DOS VARIANZAS POBLACIONALES, σ_1^2/σ_2^2, PARA DOS CONJUNTOS INDEPENDIENTES DE DATOS MUESTRALES

MINITAB (Versión 14)

```
Choose:      Stat > Basic Statistics > 2 Variances
Select:      Samples in one column:
Enter:       Samples: C1        Subscripts: C2
Or           Select:    Samples in different columns:
             Enter:     First: C1        Second: C2
Or           Select:    Summarized data
             Enter:     Sample size and Variance for each sample:
Select:      Storage:   Standard Deviations > OK > OK
```

*El procedimiento 2 de varianzas evalúa la primera muestra dividida entre la segunda muestra.

Excel Introduzca los datos para el numerador (dispersión mayor) en la columna A y los datos para el denominador (dispersión menor) en la columna B; luego continúe con:

```
Choose:      Tools > Data Analysis > F-Test: Two-Sample for Variances
Enter:       Variable 1 Range: (A1:A20 or select cells)
             Variable 2 Range: (B1:B20 or select cells)
Select:      Labels (if necessary)
Enter:       α (ex. 0.05)
Select:      Output Range
Enter:       (C1 or select cell) > OK
```

Use Format > Column > AutoFit Selection para hacer que la salida sea más fácil de leer. La salida muestra el valor p y el valor crítico para una prueba de una cola.

TI-83/84 Plus Introduzca los datos para el numerador (dispersión mayor) en L1 y los datos para el denominador (dispersión menor) en L2; luego continúe con lo siguiente, introduciendo los valores apropiados y resaltando Calculate:

```
Choose:      STAT > TESTS > D:2-SampFTest . . .
```

CASO PRÁCTICO 10.20 Características de personalidad de aspirantes de la academia de policía

Bruce N. Carpenter y Susan M. Raza concluyeron que los "aspirantes de policía son un poco más semejantes a ellos que a la población normativa" cuando la prueba F de homogeneidad de varianza resultó en un valor p de menor de 0.005. *Homogeneidad* significa que las calificaciones del grupo son menos variables que las calificaciones para la población normativa.

COMPARACIONES ENTRE SUBGRUPOS Y CON OTRAS POBLACIONES

Para determinar si los aspirantes a policías son un grupo más homogéneo que la población normativa, se utilizó la prueba F de homogeneidad de varianza. Con la excepción de las escalas F, K y 6, donde las diferencias no son significativas, los resultados indican que los aspirantes a policías forman un grupo más homogéneo que la población normativa [$F(237, 305) = 1.36$, $p < 0.005$]. Así, los aspirantes a policías se asemejan más a ellos que a la población normativa.

Fuente: Reproducido de la *Journal of Police Science and Administration,* Vol.15, no.1, pp.10-17, con permiso de la International Association of Chiefs of Police, PO Box 6010, 13 Firstfield Road, Gaithersburg, MD 20878.

SECCIÓN 10.6 EJERCICIOS

10.105 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar lo siguiente:

a. Las varianzas de las poblaciones A y B no son iguales.

b. La desviación estándar de la población I es mayor que la desviación estándar de la población II.

c. La razón entre las varianzas para las poblaciones A y B es diferente de 1.

d. La variabilidad dentro de la población C es menor que la variabilidad dentro de la población D.

10.106 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar lo siguiente:

a. La desviación estándar de la población X es menor que la desviación estándar de la población Y.

b. La razón entre las varianzas de la población A sobre la población B es mayor a 1.

c. La desviación estándar de la población Q_1 es al menos la de la población Q_2.

d. La variabilidad dentro de la población I es más que la variabilidad dentro de la población II.

10.107 Explique por qué la desigualdad $\sigma_m^2 > \sigma_p^2$ es equivalente a $\dfrac{\sigma_m^2}{\sigma_p^2} > 1$.

10.108 Exprese la H_o y H_a del ejemplo 10.18 (p. 598) equivalentemente en términos de desviaciones estándar.

10.109 Usando la notación $F(\mathrm{df}_1, \mathrm{df}_2, \alpha)$, cite cada uno de los valores críticos mostrados en las siguientes figuras.

10.110 Encuentre los valores de $F(12, 24, 0.01)$ y $F(24, 12, 0.01)$.

10.111 Encuentre los siguientes valores críticos para F de las tablas 9A, 9B, y 9C del apéndice B.

a. $F(24, 12, 0.05)$

b. $F(30, 40, 0.01)$

c. $F(12, 10, 0.05)$

d. $F(5, 20, 0.01)$

e. $F(15, 18, 0.025)$

f. $F(15, 9, 0.025)$

g. $F(40, 30, 0.05)$

h. $F(8, 40, 0.01)$

10.112 Determine el valor p que se usaría para probar las siguientes hipótesis cuando F se usa como estadística de prueba:

a. H_o: $\sigma_1 = \sigma_2$ vs. H_a: $\sigma_1 > \sigma_2$, con $n_1 = 10$, $n_2 = 16$ y $F\bigstar = 2.47$

b. H_o: $\sigma_1^2 = \sigma_2^2$ vs. H_a: $\sigma_1^2 > \sigma_2^2$, con $n_1 = 25$, $n_2 = 21$ y $F\bigstar = 2.31$

c. H_o: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ vs. H_a: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$, con $n_1 = 41, n_2 = 61$ y $F\bigstar = 4.78$

d. H_o: $\sigma_1 = \sigma_2$ vs. H_a: $\sigma_1 < \sigma_2$, con $n_1 = 10$, $n_2 = 16$ y $F\bigstar = 2.47$

10.113 Encuentre el valor crítico para la prueba de hipótesis con H_a: $\sigma_1 > \sigma_2$, con $n_1 = 7$, $n_2 = 10$ y $\alpha = 0.05$.

10.114 Determine la región crítica y valor(es) crítico(s) que se usarían para probar las hipótesis siguientes usando el procedimiento clásico cuando $F\bigstar$ se use como estadística de prueba.

a. H_o: $\sigma_1^2 = \sigma_2^2$ vs. H_a: $\sigma_1^2 > \sigma_2^2$, con $n_1 = 10$, $n_2 = 16$ y $\alpha = 0.05$

b. H_o: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ vs. H_a: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$, con $n_1 = 25$, $n_2 = 31$ y $\alpha = 0.05$

c. H_o: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ vs. H_a: $\dfrac{\sigma_1^2}{\sigma_2^2} > 1$, con $n_1 = 10$, $n_2 = 10$, $\alpha = 0.01$

d. H_o: $\sigma_1 = \sigma_2$ vs. H_a: $\sigma_1 < \sigma_2$, con $n_1 = 25$, $n_2 = 16$ y $\alpha = 0.01$

10.115 Calcule $F\bigstar$ dadas $s_1 = 3.2$ y $s_2 = 2.6$.

10.116 Calcule $F\bigstar$ dadas $s_1^2 = 3.2$ y $s_2^2 = 2.6$.

10.117 ¿Cuál sería el valor de $F\bigstar$ en el ejemplo 10.20 si se usa $F\bigstar = \dfrac{s_2^2}{s_1^2}$? ¿Por qué es menor que 1?

10.118 a. Dos muestras independientes, cada una de tamaño 3, se sacan de una población normalmente distribuida. Encuentre la probabilidad de que una de las varianzas muestrales sea al menos 19 veces mayor que la otra.

b. Dos muestras independientes, cada una de tamaño 6, se sacan de una población normalmente distribuida. Encuentre la probabilidad de que una de las varianzas muestrales no sea más de 11 veces mayor que la otra.

10.119 Una panadería está considerando comprar uno de dos hornos de gas. La panadería requiere que la temperatura permanezca constante durante una operación de horneado. Se realizó un estudio para medir la varianza en temperatura de los hornos durante el proceso de horneado. La varianza en temperatura antes que el termostato reinicie la flama para el horno Monarca fue 2.4 para 16 mediciones. La varianza para el horno Kraft fue 3.2 por 12 mediciones. ¿Esta información da suficiente razón para concluir que hay una diferencia en las varianzas para los dos hornos? Suponga que las mediciones están normalmente distribuidas y use un nivel de significación de 0.02.

10.120 Un estudio en *Atención Pedriática de Emergencia* comparó la gravedad de lesiones entre niños de más corta edad y otros mayores. Una medida reportada fue la Puntuación de Gravedad de Lesiones (ISS). La desviación estándar de puntos de la ISS para 37 niños de 8 años o mayores fue 23.9, y la desviación estándar para 36 niños de más de 8 años de edad fue 6.8. Suponga que las puntuaciones de la ISS están normalmente distribuidas para ambos grupos de edades. Al nivel de significación de 0.01, ¿hay suficiente razón para concluir que la desviación estándar de puntuaciones de la ISS para niños menores es mayor que la desviación estándar de puntuaciones de la ISS para niños mayores?

10.121 La sacarosa, azúcar común de mesa, es probablemente el producto químico orgánico puro más abundante del mundo y el más ampliamente conocido por personas que no sean químicos. Ya sea de caña de azúcar (20% por peso) o de remolacha azucarera (15% por peso), cruda o refinada, el azúcar común es toda-

vía sacarosa. Quince condados productores de remolacha azucarera en Estados Unidos se seleccionaron al azar, y sus porcentajes de sacarosa en 2001 dieron una desviación estándar de 0.862. De manera semejante, 12 condados productores de caña de azúcar en Estados Unidos se seleccionaron al azar, y sus porcentajes de sacarosa en 2001 registraron una desviación estándar de 0.912. Al nivel de 0.05, ¿hay diferencia significativa entre las desviaciones estándar de porcentajes de remolacha azucarera y sacarosa de caña de azúcar?

10.122 Se llevó a cabo un estudio para determinar si había igual variabilidad en las lecturas de presión sistólica entre hombres y mujeres. Se emplearon muestras aleatorias de 16 hombres y 13 mujeres para probar lo dicho del experimentador de que las varianzas eran desiguales. Se usó el MINITAB para calcular las desviaciones estándar, $F\bigstar$, y el valor p. Suponga normalidad.

Hombres	120	120	118	112	120	114	130	114	124	125
	130	100	120	108	112	122				
Mujeres	122	102	118	126	108	130	104	116	102	122
	120	118	130							

```
Standard deviation of Men = 7.8864
Standard deviation of Women = 9.9176
F-Test (normal distribution)
Test Statistic: 1.581
P-Value: 0.398
```

Verifique estos resultados al calcular los valores usted mismo.

10.123 Cuando una prueba de hipótesis es de dos colas y se usa Excel para calcular el valor p, ¿qué paso adicional debe darse?

10.124 Con referencia al caso práctico 10.20 (p. 599):

a. ¿Qué hipótesis nula y alternativa probaron Carpenter y Raza?

b. ¿Qué significa "$p < 0.005$"?

c. Use una computadora o calculadora para calcular el valor p para $F(237, 305) = 1.36$.

10.125 La calidad del producto final está en cierto modo determinada por la calidad de los materiales empleados. Las fábricas textiles vigilan la resistencia a la tracción de las fibras empleadas en tejer sus artículos de taller. Las siguientes muestras aleatorias independientes son resistencias a la tracción de fibras de algodón provenientes de dos proveedores.

Proveedor A	78	82	85	83	77	84	90	82	93	82
	80	82	77	80	80					
Proveedor B	76	79	83	78	72	73	69	80	74	77
	78	78	73	76	78	79				

Calcule el valor observado de F, $F\bigstar$, para comparar las varianzas de estos dos conjuntos de datos.

10.126 Varios condados de Minnesota y Wisconsin se seleccionaron al azar, y se recolectó información acerca de la cosecha de 2001 de maíz dulce. Resultaron los siguientes porcentajes de producción en toneladas de maíz dulce por acre.

Producción, MN	5.90	6.50	6.20	5.90	6.20	5.70	5.91	5.90	6.00	5.60	6.51	6.30
Producción, WI	7.80	6.80	7.00	5.30	6.50	6.90	6.60	5.70	6.60	6.00	6.20	

Fuente: http://www.usda.gov/nass/graphis/county01/data/vsc01.csv

a. ¿Hay diferencia en la variabilidad de rendimientos de un condado, medida por la desviación estándar del porcentaje de rendimiento? Use $\alpha = 0.05$.

b. Suponiendo normalidad en porcentajes de rendimiento, ¿es significativamente mayor el porcentaje medio de rendimiento de Wisconsin que el porcentaje medio de rendimiento de Minnesota? Use $\alpha = 0.05$.

10.127 Un objetivo constante en la manufactura de lentes de contacto es mejorar el nivel y variación para aquellas características que afectan el poder de lentes y la agudeza visual. Una de esas características comprende el herramental del que finalmente se manufacturan los lentes.

Se examinaron los resultados de dos lotes de desarrollo de proceso inicial en busca de la Característica Crítica A. Se manufacturaron dos lotes de productos distintos con ligeras diferencias diseñadas para afectar la característica en cuestión. Cada lote se muestreó a continuación y resultó que el lote 1 tenía un tiempo de desarrollo más corto que el lote 2, y por tanto se tomaron más muestras del lote 2.

a. Calcule la media y desviación estándar de la Característica Crítica A para los lotes 1 y 2.

b. ¿Hay evidencia de una diferencia en la variabilidad de la Característica Crítica A entre los lotes 1 y 2? Use un valor alfa de 5% para hacer una determinación.

c. ¿Hay evidencia de una diferencia en los niveles medios de la Característica Crítica A entre los lotes

1 y 2. Use un valor alfa de 5% para hacer una determinación.

Muestra de lote 1	Característica Crítica A	Muestra de lote 2	Característica Crítica A	Muestra de lote 2	Característica Crítica A
1	0.017	1	0.026	14	0.041
2	0.021	2	0.027	15	0.021
3	0.006	3	0.024	16	0.022
4	0.009	4	0.023	17	0.027
5	0.018	5	0.034	18	0.032
6	0.021	6	0.035	19	0.023
7	0.013	7	0.035	20	0.023
8	0.017	8	0.033	21	0.024
		9	0.034	22	0.017
		10	0.033	23	0.023
		11	0.032	24	0.019
		12	0.038	25	0.027
		13	0.041		

10.128 Los norteamericanos dormitan los fines de semana, según una encuesta de 1506 adultos para la Fundación Nacional del Sueño y publicada en una USA Snapshot durante abril de 2005.

Horas de sueño	Días hábiles	Fines de semana
Menos de 6	0.1	0.10
6–6.9	0.24	0.15
7–7.9	0.31	0.24
8 o más	0.26	0.49

Se tomaron dos muestras aleatorias independientes en un gran complejo industrial. A los trabajadores seleccionados en una muestra se les preguntó, "¿Cuántas horas, a los 15 minutos más cercanos, durmió usted la noche del jueves de esta semana?". A los trabajadores seleccionados para la segunda muestra se les preguntó, "¿Cuántas horas, a los 15 minutos más cercano, durmió usted la noche del sábado el fin de semana pasado?".

Día hábil			Fin de semana			
5.00	7.75	7.25	9.00	7.25	8.75	7.50
9.25	7.25	8.75	6.25	5.25	9.25	9.25
7.00	7.75	6.75	7.50	8.50	8.75	6.50
9.25	7.00	7.75	8.00	8.75	9.50	8.00
9.25	9.25	6.00	8.75	7.75	8.75	7.50

a. Construya un histograma y encuentre la media y desviación estándar para cada conjunto de datos.

b. Las distribuciones de "horas de sueño en día hábil" y "horas de sueño en fin de semana" resultantes de la encuesta, ¿parecen ser similares en forma? ¿en el centro? ¿hay dispersión? Discuta sus respuestas.

c. ¿Es posible que ambas muestras se sacaran de poblaciones normales? Justifique su respuesta.

d. El número medio de horas de sueño en fin de semana, ¿es estadísticamente mayor que el número medio de horas de sueño en día hábil? Use $\alpha = 0.05$.

e. ¿Hay suficiente evidencia para demostrar que la desviación estándar de estas dos muestras son estadísticamente diferentes? Use $\alpha = 0.05$.

f Explique cómo las respuestas a las partes b-e ahora afectan en sus ideas acerca de su respuesta a la parte a.

10.129 ¿Qué cantidad en dólares debe gastar alguien en usted en un regalo del día del amor y la amistad? Los resultados encontrados por una encuesta de Greenfield Online de 653 personas que respondieron se muestran en la gráfica siguiente.

¿QUÉ CANTIDAD EN DÓLARES DEBE GASTAR ALGUIEN EN USTED EN UN REGALO DEL DÍA DEL AMOR Y LA AMISTAD?

Datos de Darryl Haralson y Karl Gelles, USA TODAY; Fuente: Greenfield Online survey of 653 respondents. Margen de error ±3 puntos de porcentaje.

Se seleccionaron muestras aleatorias en la región central del estado de Nueva York con los siguientes resultados.

Hombres	103	100	100	67	77	63	55	43	139	2	51	5	100	52	139
	86	23	56	40	84	15	32	157	35	4	24	102	52	43	75
	128	206	16	13	98										
Mujeres	36	5	77	97	25	62	91	170	108	112	198	161	54	40	111
	107	241	89	37	10	175	10	84	102	17	32	25	1	38	126
	121	30	147	135	45	230	29	88							

a. Construya un histograma y encuentre la media y desviación estándar para cada conjunto de datos de la región central del estado de Nueva York.

b. Las formas de "cantidad gastada" sugeridas por hombres y mujeres de la región central del estado de Nueva York ¿parecen se similares en forma? ¿en el centro? ¿hay dispersión? Discuta sus respuestas.

c. ¿Es posible que ambas muestras fueran sacadas de poblaciones normales? Justifique su respuesta.

d. ¿La cantidad media indicada por mujeres es estadísticamente mayor que la cantidad media indicada por hombres? Use $\alpha = 0.05$.

e. ¿Hay suficiente evidencia para demostrar que las desviaciones estándar de estas dos muestras son estadísticamente diferentes? Use $\alpha = 0.05$.

f. Explique la forma en que sus respuestas a las partes b-e ahora afectan sus pensamientos acerca de su respuesta a la parte a.

10.130 Use computadora para demostrar la verdad de la teoría presentada en esta sección.

a. Las suposiciones fundamentales son "las poblaciones están normalmente distribuidas", y mientras que se realiza una prueba de hipótesis para la igualdad de dos desviaciones estándar, se supone que las desviaciones estándar son iguales. Genere muestras muy grandes de dos poblaciones teóricas: $N(100, 20)$ y $N(120, 20)$. Encuentre evidencia gráfica y numérica de que las poblaciones satisfacen las suposiciones.

b. Al azar seleccione 100 muestras, cada una de tamaño ocho, de ambas poblaciones y encuentre la desviación estándar de cada muestra.

c. Usando la primera muestra sacada de cada población como par, calcule la estadística $F\bigstar$. Repita para todas las muestras. Describa la distribución muestral de los 100 valores de $F\bigstar$ usando estadísticas gráficas y numéricas.

d. Genere la distribución de probabilidad para $F(7, 7)$, y compárela con la distribución observada de $F\bigstar$. ¿Están de acuerdo las dos gráficas? Explique.

10.131 En esta sección dijimos que la prueba F es muy sensible a desviaciones menores respecto a las suposiciones. Repita el ejercicio 10.130 usando $N(100, 20)$ y $N(120, 30)$. Observe que el único cambio a partir del ejercicio 10.130 es el aparentemente ligero aumento en la desviación estándar de la segunda población. Conteste las mismas preguntas usando la misma clase de información y verá resultados muy diferentes.

REPASO DEL CAPÍTULO

En retrospectiva

En este capítulo iniciamos las comparaciones de dos poblaciones al distinguir entre muestras independientes y dependientes, que son procedimientos de muestreo estadísticamente importantes y útiles. A continuación procedimos a examinar las inferencias respecto a la comparación de medias, proporciones y varianzas para dos poblaciones.

Siempre estamos haciendo comparaciones entre dos grupos. Comparamos medias y comparamos proporciones. En este capítulo hemos aprendido a comparar estadísticamente dos poblaciones haciendo inferencias acerca de sus medias, proporciones o varianzas. Por comodidad, la tabla 10.7 identifica las fórmulas a usar cuando hagamos inferencias acerca de comparaciones entre dos poblaciones.

En los capítulos 8, 9 y 10 hemos aprendido a usar intervalos de confianza y pruebas de hipótesis para contestar preguntas acerca de medias, proporciones y desviaciones estándar para una o dos poblaciones. De aquí podemos ampliar nuestras técnicas para incluir inferencias acerca de más de dos poblaciones así como inferencias de tipos diferentes.

TABLA 10.7

Fórmulas a usar para inferencias que comprenden dos poblaciones

		Fórmula a usar	
Situaciones	Estadística de prueba	Intervalo de confianza	Prueba de hipótesis
Diferencia entre dos medias			
Muestras dependientes	t	Fórmula (10.2) (p. 552)	Fórmula (10.5) (p. 555)
Muestras independientes	t	Fórmula (10.8) (p. 566)	Fórmula (10.9) (p. 567)
Diferencia entre dos proporciones	z	Fórmula (10.11) (p. 582)	Fórmula (10.15) (p. 585)
Diferencia entre dos varianzas	F		Fórmula (10.16) (p. 595)

Vocabulario y conceptos clave

diferencia media (pp. 552, 554)
diferencia pareada (p. 550)
distribución F (p. 593)
distribución t (pp. 552, 565)
error estándar (pp. 555, 564, 581)
estadística de prueba (pp.555, 567, 584, 595)
estadística F (p. 595)
estadística t (pp. 555, 567)
estadística z (p. 584)
experimento binomial (p. 581)
fuente (de datos) (pp. 555, 564, 581)
intervalo de confianza (pp. 552, 566, 582)
medias dependientes (p. 551)
medias independientes (p. 564)
muestras dependientes (pp. 547, 550)
muestras independientes (pp. 548, 564, 581, 592)
p binomial (p. 581)
porcentaje (p. 581)
probabilidad (p. 581)
probabilidad agrupada observada (p. 585)
proporción (p. 581)
prueba de hipótesis (pp. 554, 567, 584, 592)
suposiciones (pp. 551, 565, 582, 595)
valor p (pp. 556, 569, 586, 596)

Objetivos de aprendizaje

✓ Entender la diferencia entre muestras dependientes e independientes	EJ. 10.1, 10.2 Ejer. 105, 107
✓ Entender que la diferencia media (media de las diferencias pareadas) debe usarse para analizar muestras dependientes.	pp. 550-551
✓ Calcular y/o entender cómo calcular la diferencia media y desviación estándar para datos pareados.	pp. 551-552, Ejer. 10.13
✓ Calcular, describir e interpretar un intervalo de confianza para la diferencia poblacional media.	EJ. 10.4, Ejer. 10.19
✓ Efectuar, describir e interpretar una prueba de hipótesis para la diferencia poblacional media, μ_d, usando el procedimiento del valor p y el procedimiento clásico.	EJ. 10.5, 10.6, Ejer. 10.35, 10.138
✓ Entender que la diferencia entre dos medias debe usarse para analizar muestras independientes.	pp. 564-565

✓ Entender cómo determinar los grados de libertad de la distribución t para la diferencia entre medias usando dos muestras independientes.	p. 565, Ejer. 10.14.
✓ Calcular, describir e interpretar un intervalo de confianza para la diferencia entre dos medias usando muestras independientes.	EJ. 10.8, Ejer. 10.43, 10.45
✓ Efectuar, describir e interpretar una prueba de hipótesis para la diferencia entre dos medias poblacionales, $\mu_1 - \mu_2$, usando el procedimiento del valor p y el procedimiento clásico.	EJ. 10.9, 10.10, Ejer. 10.62, 10.71
✓ Entender que la distribución z se usará para analizar la diferencia entre dos proporciones usando muestras independientes, siempre que se satisfagan los principios para asegurar normalidad.	pp. 581-582
✓ Calcular proporciones muestrales con base en tamaño muestral y número de sucesos.	EJ. 10.79
✓ Calcular, describir e interpretar un intervalo de confianza para la diferencia entre dos proporciones usando muestras independientes.	EJ. 10.12, Ejer. 10.87, 10.153
✓ Efectuar, describir e interpretar una prueba de hipótesis para la diferencia entre dos proporciones poblacionales, $p_1 - p_2$, usando el procedimiento del valor p y el procedimiento clásico.	EJ. 1013, Ejer. 10.101, 10.154
✓ Entender las propiedades de la distribución F y la forma en que es una serie de distribuciones basada en tamaños muestrales (usando pares de números de grados de libertad como el índice).	pp. 593–594, EJ. 10.16, 10.18, 10.19, Ejer. 10.110
✓ Entender que la suposición para inferencias acerca de la razón entre dos varianzas es que las poblaciones muestreadas están normalmente distribuidas, y las dos muestras se seleccionan de manera independiente.	p. 595
✓ Efectuar, describir e interpretar una prueba de hipótesis para la razón entre dos varianzas poblacionales, $\frac{\sigma_1^2}{\sigma_2^2}$, usando la distribución F con el procedimiento del valor p y el procedimiento clásico.	EJ. 10.17, Ejer. 10.119, 10.159

Ejercicios del capítulo

10.132 Una química está probando un método analítico recientemente propuesto y por comparación decide usar el método aceptado en la actualidad. Ella toma 12 especímenes de concentrado desconocido y determina la concentración de cada espécimen, usando tanto el método propuesto como el método actual. ¿Estas dos muestras representan muestras dependientes o independientes? Explique.

10.133 Usando un intervalo de 95% de confianza, estime la diferencia media en IQ entre los miembros mayores y menores (hermanos y hermanas) de una familia, con base en la siguiente muestra aleatoria de los IQ. Suponga normalidad.

Mayores	145	133	116	128	85	100	105	150	97	110	120	130
Menores	131	119	103	93	108	100	111	130	135	113	108	125

10.134 Las lecturas de presión sanguínea diastólica para 15 pacientes se determinaron usando dos técnicas: el método estándar empleado por personal médico y un método que utiliza un aparato electrónico con lectura digital. Los resultados fueron como sigue:

Paciente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Método estándar	72	80	88	80	80	75	92	77	80	65	69	96	77	75	60
Método digital	70	76	87	77	81	75	90	75	82	64	72	95	80	70	61

Suponiendo que la presión sanguínea está normalmente distribuida, determine el intervalo de confianza de 90% para la diferencia media en las dos lecturas, donde d = método estándar – lectura digital.

10.135 Deseamos saber cuál de dos tipos de filtro usar. Se diseñó una prueba en la que la fuerza de una señal podría hacerse variar de cero al punto donde el operador detecta primero la imagen. En este punto, se registra la intensidad del ajuste; los ajustes más bajos son mejores. A veinte operadores se pidió hacer una lectura por cada filtro.

Operador	1	2	3	4	5	6	7	8	9	10
Filtro 1	96	83	97	93	99	95	97	91	100	92
Filtro 2	92	84	92	90	93	91	92	90	93	90
Operador	11	12	13	14	15	16	17	18	19	20
Filtro 1	88	89	85	94	90	92	91	78	77	93
Filtro 2	88	89	86	91	89	90	90	80	80	90

Suponiendo que las lecturas de intensidad están normalmente distribuidas, estime la diferencia media entre las dos lecturas usando un intervalo de confianza de 90%.

10.136 Al final de su primer día en el campo de entrenamiento, 10 nuevos reclutas participaron en una competencia de tiro con fusil. Los mismos 10 compitieron de nuevo al final de toda una semana de entrenamiento y práctica. Sus calificaciones resultantes se muestran en la tabla siguiente.

Tiempo de competencia	Recluta									
	1	2	3	4	5	6	7	8	9	10
Primer día	72	29	62	60	68	59	61	73	38	48
Una semana después	75	43	63	63	61	72	73	82	47	43

Este conjunto de 10 pares de datos ¿muestra que hay una cantidad significativa de mejoría en la capacidad de disparo de reclutas durante la semana? Use $\alpha = 0.05$ y suponga normalidad.

10.137 La cantidad de anestésico general que un paciente debe recibir antes de una cirugía ha sido objeto de considerable atención pública. Según la American Society of Anesthesiologists, cada año, alrededor de 40 000 (algunos investigadores han puesto la cifra en cerca de 200 000) de los 28 millones de pacientes que son sometidos a anestesia general experimentan conciencia limitada durante la cirugía, debido a la resistencia al medicamento o por una dosis demasiado pequeña. Por lo general, los pacientes informan que oyen la conversación de médicos con enfermeras y asistentes durante operaciones.

Fuente: People, "Wake-Up Call"

Suponga que se realiza un estudio de 20 pacientes a quienes se practica cirugía en ambos ojos, con dos semanas de separación de tratamiento en cada ojo. A diez de los pacientes se les suministra una ligera dosis de anestésico general antes de la operación en el primer ojo, y a los otros diez pacientes se les aplica una dosis más fuerte. A la semana siguientes se invierte el procedimiento. Dos días después de cada cirugía, se pide a los pacientes que indiquen la magnitud de dolor e incomodidad que experimentan, en una escala de 0 (ninguno) a 10 (insoportable). A continuación veamos los resultados:

Sujeto	Dosis ligera	Dosis fuerte	Sujeto	Dosis ligera	Dosis fuerte
1	4	3	11	6	7
2	6	5	12	7	5
3	5	6	13	10	7
4	8	4	14	3	2
5	4	5	15	1	0
6	9	6	16	5	6
7	3	2	17	6	3
8	7	8	18	8	5
9	8	5	19	4	2
10	9	7	20	2	0

¿Puede concluirse que una dosis más fuerte de anestésico resultó en que los pacientes experimentan menos dolor e incomodidad después de una cirugía en ojos? Use el nivel de significación de 0.01.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.138 Los medicamentos de liberación inmediata sueltan rápidamente su contenido de droga, con la máxima concentración llegando en un corto tiempo; los medicamentos de liberación sostenida, por otra parte, toman más tiempo para llegar a máxima concentración. Como parte de un estudio, la codeína de liberación inmediata se comparó con una codeína de liberación sostenida, usando 13 pacientes sanos.

A los pacientes se les asignó al azar uno de los dos tipos de codeína y se les dio tratamiento durante 2.5 días; después del periodo de lavado de 7 días, a cada paciente se le dio el otro tipo de codeína con lo cual cada paciente recibió ambos tipos. La cantidad total (A) de droga disponible durante el tratamiento en (ng × mL)/hora es como sigue:

Paciente	1	2	3	4	5	6	7
Airc	1091.3	1064.5	1281.1	1921.4	1649.9	1423.6	1308.4
Asrc	1308.5	1494.2	1382.2	1978.3	2004.6	*	1211.1
Paciente	8	9	10	11	12	13	
Airc	1192.1	766.2	978.6	1618.9	582.9	972.1	
Asrc	1002.4	866.6	1345.8	979.2	576.3	999.1	

Fuente: http://exploringdata.cqu.edu.au/ws_coedn.htm

a. Explique por qué éste es un diseño de diferencia pareada.

b. ¿Qué ajuste es necesario dado que no hay el medicamento Asrc para el paciente 6?

¿Hay una diferencia significativa en la cantidad total de medicamento disponible durante el trata miento?

c. Compruebe las suposiciones de la prueba y describa lo que encuentre.

d. Pruebe el dicho usando $\alpha = 0.05$.

10.139 Una prueba que mide la ansiedad generada por las matemáticas se aplicó a 50 estudiantes hombres y 50 mujeres. Los resultados fueron como sigue:

Hombres: $\bar{x} = 70.5$, $s = 13.2$
Mujeres: $\bar{x} = 75.7$, $s = 13.6$

Construya un intervalo de 95% para la diferencia entre las calificaciones medias por la ansiedad.

10.140 La misma prueba de realización se aplica a soldados de dos unidades seleccionados al azar. Las calificaciones que alcanzaron se resumen como sigue:

Unidad 1: $n_1 = 70$, $\bar{x}_1 = 73.2$, $s_1 = 6.1$
Unidad 2: $n_2 = 60$, $\bar{x}_2 = 70.5$, $s_2 = 5.5$

Construya un intervalo de confianza de 90% para la diferencia en el nivel medio de las dos unidades.

10.141 Diez soldados fueron seleccionados al azar de cada una de dos compañías para participar en una competencia de tiro con fusil. Sus calificaciones se muestran en la tabla siguiente.

Compañía A	72	29	62	60	68	59	61	73	38	48
Compañía B	75	43	63	63	61	72	73	82	47	43

Construya un intervalo de confianza de 95% para la diferencia entre las calificaciones medias para las dos compañías. Suponga normalidad.

10.142 A dos grupos de estudiantes principiantes se aplicó una prueba para ver sus logros en un curso de ciencias de la computación. Uno de los grupos ya había tenido un curso previo de ciencias de la computación en secundaria; el otro grupo no lo había tenido. A continuación veamos los resultados de la prueba. Suponiendo que las calificaciones de la prueba son normales, construya un intervalo de confianza de 98% para la diferencia entre las dos medias poblacionales.

Grupo 1 (con curso en secundaria)	17	18	27	19	24	36	27	26	35	22	18
	29	29	26	33							
Grupo 2 (sin curso en secundaria)	19	25	28	27	21	24	18	14	28	21	22
	20	21	14	29	28	25	17	20	28	31	27

10.143 Se utilizaron dos métodos para estudiar el calor latente de fusión de hielo. El método A (eléctrico) y el método B (de mezclas) se realizaron con los especímenes enfriados a −0.72°C. Los datos de la tabla siguiente representan el cambio en calor total de 0.72°C al agua a 0°C en calorías por gramo de masa.

Método A	79.98	80.04	80.02	80.04	80.03	80.03	80.04
	79.97	80.05	80.03	80.02	80.00	80.02	
Método B	80.02	79.94	79.98	79.97	79.97	80.03	79.95
	79.97						

Suponiendo normalidad, construya un intervalo de confianza de 95% para la diferencia entre las medias.

10.144 La sacarosa, azúcar común de mesa, es probablemente el producto químico orgánico puro más abundante del mundo y el más ampliamente conocido por personas que no sean químicos. Ya sea de caña de azúcar (20% por peso) o de remolacha azucarera (15% por peso), cruda o refinada, el azúcar común es todavía sacarosa. Quince condados productores de remolacha azucarera en Estados Unidos se seleccionaron al azar y se registraron sus porcentajes de sacarosa en 2001. De manera semejante, 12 condados productores de caña de azúcar en Estados Unidos se seleccionaron al azar y se registraron sus porcentajes de sacarosa en 2001.

Sacarosa de remolacha	17.30	16.46	16.20	17.53	17.00	18.53	16.77	16.11
	15.30	17.90	15.98	17.30	17.94	17.30	16.60	
Sacarosa de caña de azúcar	14.1	13.5	15.2	15.0	13.6	13.6	11.7	14.3
	13.8	13.8	14.8	13.7				

Fuente: http://www.usda.gov/nass/graphics/county01/data/

Encuentre el intervalo de confianza de 95% para la diferencia entre el porcentaje medio de sacarosa, para todos los condados productores de remolacha de Estados Unidos y todos los condados productores de caña de azúcar de Estados Unidos. Suponga normalidad.

10.145 George Johnson es el director técnico de un equipo de futbol colegial y compite como local en un pasto artificial. George está preocupado porque el tiempo de sprint en 40 yardas, registrado por sus jugadores y otros, aumenta considerablemente cuando corren en pasto natural en comparación con un pasto artificial. Si es así, hay poca comparación entre la velocidad de sus jugadores y la de sus oponentes siempre que su equipo juegue en pasto natural. El siguiente oponente de George juega en pasto natural, de modo que hizo una encuesta a todos sus inicialistas para el siguiente juego y obtuvo sus mejores tiempos en el sprint de 40 yardas; a continuación los comparó con los mejores tiempos obtenidos por sus propios jugadores. Los resultados aparecen en la tabla siguiente:

Grupo de jugadores	n	Media (s)	Desviación estándar
Pasto artificial	22	4.85	0.31
Pasto natural	22	4.96	0.42

¿Los jugadores de Johnson tienen un tiempo medio de sprint más bajo? Suponiendo normalidad, pruebe al nivel de significación de 0.05 para asesorar a Johnson.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.146 Una prueba referente a algunos de los datos fundamentales acerca del síndrome de inmunodeficiencia adquirida (SIDA) se aplicó a dos grupos, uno de ellos formado por graduados universitarios y el otro por egresados de preparatoria. A continuación vea un resumen de los resultados de la prueba:

Graduados universitarios: $n = 75, \bar{x} = 77.5, s = 6.2$

Egresados de preparatoria: $n = 75, \bar{x} = 50.4, s = 9.4$

¿Estos datos muestran que los graduados universitarios, en promedio, calificaron más alto en la prueba? Use $\alpha = 0.05$.

10.147 Alrededor de 20 millones de norteamericanos visitan quiroprácticos cada año, y el número de practicantes en Estados Unidos es de 55,000, casi el doble de hace dos décadas, de acuerdo con la American Chiropractic Association. La revista *New England Journal of Medicine* publicó una información, que muestra los resultados de un estudio que comparaba la manipulación quiropráctica de la espina dorsal (CSM) con la terapia física para tratamiento de dolor agudo en la parte baja de la espalda. Después de dos años de tratamiento, se encontró que la CSM no era más eficaz para reducir ausencias laborales ni para prevenir una recaída.

Suponga que se hace un estudio similar de 60 pacientes, dividiendo la muestra en dos grupos. Durante 1 año, un grupo recibe la CSM y, el otro, terapia física. Durante el periodo de 1 año, se midió el número de días de ausencia en el trabajo como resultado de dolor en la parte baja de la espalda.

Grupo	n	Media	Desviación estándar
CSM (1)	32	10.6	4.8
Terapia (2)	28	12.5	6.3

¿Estos resultados muestran que el número medio de días de ausencia laboral, para personas que sufren de dolor agudo en la parte baja de la espalda es significativamente menor que quienes reciben terapia física? Suponga normalidad y use un nivel de significación de 0.01.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.148 Para comparar las cualidades de dos cohetes de corto alcance, 8 de primera clase y 10 de segunda clase se disparan a un objetivo. Si la primera clase tiene un error medio al objetivo de 36 pies y una desviación estándar de 15 pies, y la segunda clase tiene un error medio de objetivo de 52 pies y una desviación estándar de 18 pies, ¿indica esto que la segunda clase de cohete es menos precisa que la primera? Use $\alpha = 0.01$ y suponga distribución normal para error de objetivo.

10.149 El material empleado en la manufactura de piezas afecta no sólo el tiempo que dura la piezas sino también lo difícil que es su reparación. Las siguientes mediciones para la fuerza de torsión para remover un tornillo específico, después de varias operaciones de uso. La primera fila indica el número de pieza, la segunda contiene las mediciones de la fuerza de re-

moción para conjuntos hechos con el material A, y la tercera fila presenta las mediciones de la fuerza de torsión para conjuntos hechos con el material B. Suponga que las mediciones de torsión están normalmente distribuidas.

Torsión de remoción (NM, newton-metro)

Núm. de pieza	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Material A	16	14	13	17	18	15	17	16	14	16	15	17	14	16	15
Material B	11	14	13	13	10	15	14	12	11	14	13	12	11	13	12

Fuente: Datos proporcionados por AC Rochester Division, General Motors, Rochester, NY

a. Encuentre la media muestral, varianza y desviación estándar para los datos del material A.

b. Encuentre la media muestral, varianza y desviación estándar para los datos del material B.

c. Al nivel de 0.01, ¿estos datos muestran una diferencia significativa en la torsión media requerida para remover los tornillos de los dos diferentes materiales?

10.150 Un grupo de 17 estudiantes participó en una evaluación de una sesión especial de entrenamiento que decía mejoraba la memoria. Los estudiantes fueron asignados al azar en dos grupos: el grupo A, el grupo de prueba, y el grupo B, el grupo de control. Se hizo una prueba a los 17 estudiantes para conocer su capacidad de recordar cierto material. Al grupo A se le dio capacitación especial, no así al grupo B. Después de 1 mes, ambos grupos fueron sometidos a prueba de nuevo, con los resultados que se presentan en la tabla siguiente. ¿Estos datos apoyan la hipótesis alternativa de que la capacitación especial es eficaz al nivel de significación de $\alpha = 0.01$? Suponga normalidad.

	Grupo A de estudiantes									Grupo B de estudiantes							
Tiempo de prueba	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Antes	23	22	20	21	23	18	17	20	23	22	20	23	17	21	19	20	20
Después	28	29	26	23	31	25	22	26	26	23	25	26	18	21	17	18	20

10.151 En una gran universidad, se aplica un examen de conocimientos de matemáticas a todos los estudiantes. Este examen tiene una historia de producir calificaciones con una media de 77. Al azar se seleccionaron muestras de 36 estudiantes hombres y 30 mujeres, del alumnado de este año, y se registraron las siguientes calificaciones.

Hombres	72	68	75	82	81	60	75	85	80	70
	71	84	68	85	82	80	54	81	86	79
	99	90	68	82	60	63	67	72	77	51
	61	71	81	74	79	76				
Mujeres	81	76	94	89	83	78	85	91	83	83
	84	80	84	88	77	74	63	69	80	82
	89	69	74	97	73	79	55	76	78	81

a. Describa cada conjunto de datos con un histograma (use los mismos intervalos de clase en ambos histogramas), media y desviación estándar.

b. Pruebe las hipótesis, "La calificación media para todos los hombres es 77" y "la calificación media para todas las mujeres es 77", usando $\alpha = 0.05$.

c. ¿Los resultados precedentes muestran que las calificaciones medias para hombres y mujeres son iguales?

Justifique su respuesta. ¡Tenga cuidado!

d. Pruebe la hipótesis, "No hay diferencia entre las calificaciones medias para estudiantes hombres y mujeres", usando $\alpha = 0.05$.

e. Los resultados hallados en la parte d, ¿muestran que las calificaciones medias para hombres y mujeres son iguales? Explique.

f. Explique por qué los resultados hallados en la parte b no pueden usarse para concluir, "Las dos medias son iguales".

10.152 Se realizó un estudio para determinar la proporción de demócratas y republicanos que apoyan la política de "mano dura" en América del Sur. Los resultados del estudio fueron como sigue:

Demócratas: $n = 250$, número a favor = 120
Republicanos: $n = 200$, número a favor = 105

Construya un intervalo de confianza de 98% para la diferencia entre las proporciones de apoyo.

10.153 Un grupo de consumidores comparó la confiabilidad de dos microcomputadoras comparables hechas por dos fabricantes diferentes. Se determinó la proporción que requiere servicio antes de un año después de la compra para muestras de cada uno de los dos fabricantes.

Fabricante	Tamaño muestral	Proporción que necesita servicio
1	75	0.15
2	75	0.09

Encuentre un intervalo de confianza de 0.95 para $p_1 - p_2$.

10.154 Para determinar la "bondad" de una pregunta de examen, es frecuente que un profesor compare el porcentaje de los mejores estudiantes que la contesten bien contra los que no la contesten bien. Uno espera que la proporción de mejores estudiantes que contesten bien la pregunta es mayor que la de estudiantes malos. En el último examen, 35 de los estudiantes con las mejores 60 calificaciones y 27 de aquellos con las 60 calificaciones más bajas contestaron bien cierta pregunta. ¿Los estudiantes con las mejores calificaciones lo hacen mejor en esta pregunta? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.155 Según "Venus vs. Marte" en la edición de mayo/junio de 2005 de *Arthritis Today,* hombres y mujeres pueden ser más semejantes de lo que pensamos. El estudio de Boomers Wellness Lifestyle, de hombres y mujeres entre 35 y 65 años, encontró que 88% de las mujeres consideraban "manejar el estrés" como algo importante para mantener un bienestar general. El mismo estudio encontró que 75% de los hombres consideraban "manejar el estrés" como algo importante. Conteste lo siguiente y dé detalles para apoyar cada una de sus respuestas.

a. Si estas estadísticas provinieran de muestras de 100 hombres y 100 mujeres, ¿es significativa la diferencia?

b. Si estas estadísticas provinieran de muestras de 150 hombres y 150 mujeres, ¿es significativa la diferencia?

c. Si estas estadísticas provinieran de muestras de 200 hombres y 200 mujeres, ¿es significativa la diferencia?

d. ¿Qué efectos tuvo el aumento en el tamaño muestral en las soluciones en las partes a-c?

10.156 Un estudio del *New England Journal of Medicine* informó que con base en 987 muertes en el sur de California, los derechos murieron a un promedio de edad de 75 años y los zurdos murieron a un promedio de 66 años. Además, se encontró que 7.9% de los zurdos murieron por lesiones causadas en accidentes, excluyendo vehículos, contra 1.5% de los derechos; y 5.3% de los zurdos murieron cuando conducían sus vehículos contra 1.4% de los derechos.

Suponga que usted examina 1000 actas de defunción seleccionadas al azar, de las cuales 100 fueron de zurdos y 900 de derechos. Si encuentra que 5 de los zurdos y 18 de los derechos murieron cuando conducían un vehículo, ¿tendría evidencia para demostrar que la proporción de zurdos que murieron al volante es significativamente mayor que la de derechos que murieron al volante? Calcule el valor p e interprete su significado.

10.157 ¿Quién gana en casos de disputas siempre que hay un cambio en leyes impositivas, el contribuyente o el Servicio de Ingresos Internos (ISR)? La última tendencia indica que los gastos de pruebas en todos los casos judiciales ha pasado del contribuyente al IRS, cuyos expertos en impuestos predicen que podría desencadenar más preguntas. De los contadores, abogados y otros profesionales de impuestos encuestados por el grupo RIA, editor de información de impuestos, 55% espera al menos un ligero aumento en casos ganados a contribuyentes.

Fuente: Fortune, "Tax Reform?"

Suponga que a muestras de 175 contadores y 165 abogados se les pregunta, "Espera usted que los contribuyentes ganen más casos judiciales debido a las nuevas reglas de gastos de pruebas?". De los encuestados, 101 contadores contestaron afirmativamente al igual que 84 abogados. ¿Estos dos grupos de expertos difieren en sus opiniones? Use un nivel de significación de 0.01 para contestar la pregunta.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.158 "Es empate", según dos investigadores australianos. "A los 25 años, hasta 29% de todos los hombres y hasta 34% de todas las mujeres tienen canas, pero esta diferencia es tan pequeña que se considera insignificante."

Fuente: "Hilos de plata entre los de oro: ¿Quién los encontrará primero, un hombre o una mujer?" *Family Circle*

Si 1000 hombres y 1000 mujeres participaron en esta investigación, ¿sería significativo el 5% de diferencia al nivel de 0.01? Explique, incluyendo detalles para apoyar su respuesta.

10.159 Un fabricante diseñó un experimento para comparar la diferencia entre hombres y mujeres con respecto a los tiempos que necesitan para ensamblar un producto. Se hizo una prueba a un total de 15 hombres y 15 mujeres para determinar el tiempo que necesitaban, en promedio, para ensamblar el producto. El tiempo necesario para los hombres tuvo una desviación estándar de 4.5 minutos, y el tiempo necesario para las mujeres tuvo una desviación estándar de 2.8 minutos. ¿Estos datos muestran que el tiempo que necesitan los hombres es más variable que el tiempo que necesitan las mujeres? Use $\alpha = 0.05$ y suponga que los tiempos están normalmente distribuidos en forma aproximada.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.160 Un distribuidor de bebidas gaseosas está considerando dos nuevos modelos de máquinas dosificadoras. Tanto la máquina de la Harvard Company como la de Fizzit se pueden ajustar para llenar las tazas a cierta cantidad media. No obstante, la variación en la cantidad dosificada de una taza a otra es el principal interés. Diez tazas dosificadas de la máquina de Harvard mostraron una varianza de 0.065, mientras que 15 tazas dosificadas de la máquina Fizzit mostraron una varianza de 0.033. El representante de la fábrica de la Harvard Company sostiene que su máquina no tuvo más variabilidad que la máquina de Fizzit. Suponga que la cantidad dosificada está normalmente distribuida.

Al nivel de significación de 0.05, ¿la muestra refuta lo dicho por el representante?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.161 Mindy Fernández está a cargo de la producción en la nueva planta de ensamble de camionetas sport de uso general (SUV) que acaba de abrir en su ciudad. A últimas fechas ella ha estado preocupada porque los tornillos de lengüeta de las ruedas no se acoplan con las tuercas cromadas de lengüeta, para mantener en correcto funcionamiento el ensamble de las ruedas. Los trabajadores se quejan de que los daños a las cuerdas de tuercas y tornillos ocurren con tanta frecuencia, que las cuerdas son arrancadas por las máquinas neumáticas y que los ajustes de torsión tienen que bajarse para evitar dañar las cuerdas, incluso cuando las piezas ajustan bien. En un esfuerzo por determinar si la falla está en las tuercas de lengüeta o en los pernos, Mindy ha decidido pedir al departamento de control de calidad que pruebe una muestra aleatoria de 60 tuercas de lengüeta y 40 pernos para ver si las varianzas en cuerdas son iguales para ambas piezas. El reporte del técnico indicó que la varianza en las cuerdas de las tuercas de lengüeta muestreadas fue de 0.00213 y que la varianza de las cuerdas en los pernos muestreados fue de 0.00166. ¿Qué puede concluir Mindy acerca de la igualdad de las varianzas al nivel de significación de 0.05?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

10.162 Se seleccionaron muestras aleatorias de condados de Dakota del Norte y Dakota del Sur, del sitio web del USDA-NASS, con el fin de estimar la diferencia entre los porcentajes medios de producción de 2001 de avena para los dos estados.

Producción, ND	66.0	73.8	51.9	61.3	67.4	54.0	71.4	58.0	56.2	40.0
	66.7	64.4	75.7	74.4						
Producción, SD	65.0	62.5	30.0	90.0	70.0	62.7	42.4	47.1	62.2	76.2
	56.1	50.0	65.6	45.5	79.2	59.3				

Fuente: http://www.usda.gov/nass/graphics/county01/data/ot01.csv

a. ¿Se satisfacen las suposiciones? Explique.

b. ¿Hay evidencia suficiente para rechazar la hipótesis de varianzas iguales para los porcentajes de producción de avena para estos dos estados? Use $\alpha = 0.05$.

c. ¿Hay suficiente evidencia para rechazar la hipótesis de que no hay diferencia entre los porcentajes de producción de avena para estos dos estados? Use $\alpha = 0.05$.

PARA SU INFORMACIÓN

Cuando use la prueba t de dos muestras, seleccione "suponga varianzas iguales" de acuerdo con el resultado (b).

10.163 Se emprendió un proyecto de investigación para evaluar la cantidad de fuerza necesaria para producir una respuesta diseñada en equipo hecho de dos diseños distintos: el diseño existente y el diseño mejorado. La expectativa era que el equipo del nuevo diseño requeriría menos fuerza que el equipo actual. Se probaron 50 unidades de cada uno de los diseños y se registró la fuerza requerida. Un nivel menor de fuerza y una variabilidad reducida se consideran deseables.

Control o existencia	Prueba o nuevo diseño
0.003562	0.002477
0.005216	0.002725

a. Describa ambos conjuntos de datos usando medias, desviaciones estándar e histogramas.

b. Compruebe las suposiciones para comparar las varianzas y medias de dos muestras independientes. Describa lo que encuentre.

c. ¿Serán apropiadas unas pruebas de una cola o de dos colas para probar las expectativas para el nuevo diseño? ¿Por qué?

d. ¿Hay evidencia significativa para demostrar que el nuevo diseño ha reducido la variabilidad en la fuerza requerida? Use $\alpha = 0.05$.

e. ¿Hay evidencia significativa para demostrar que el nuevo diseño ha reducido la cantidad media de fuerza? Use $\alpha = 0.05$.

f. ¿El nuevo diseño cumplió con las expectativas? Explique.

Proyecto del capítulo

Estudiantes, tarjetas de crédito y débito

Como una forma de evaluar las técnicas estadísticas para dos poblaciones que hemos aprendido en este capítulo, regresemos a la sección 10.1, "Estudiantes, tarjetas de crédito y debito" (p. 545). Las compañías de tarjetas de crédito son muy conocidas por seducir a estudiantes universitarios a que firmen para tener nuevas tarjetas de crédito. La comodidad de las tarjetas de crédito, junto con la inexperiencia para manejar el dinero, lleva a numerosos estudiantes universitarios a tener deudas enormes. ¿Hay una diferencia significativa entre estudiantes de primer año y de segundo año con respecto a tarjetas de crédito y deuda? Investiguemos.

Trabajando en el contenido del capítulo 10

10.164 ¿Cómo se comparan las deudas de tarjetas de crédito para estudiantes de primer año y de segundo año? Con el uso de los dos conjuntos de datos muestrales de la página 546:

a. Encuentre la proporción de cada muestra que tenga al menos una tarjeta de crédito.

b. Encuentre la estimación puntual para la diferencia entre las dos proporciones.

c. ¿Se satisfacen las suposiciones para hacer inferencia acerca de la diferencia entre dos proporciones? Explique.

d. Encuentre el intervalo de confianza de 95% para la diferencia entre la proporción de estudiantes de primero y segundo años que tienen sus propias tarjetas de crédito.

e. Encuentre la deuda media de tarjetas de crédito para estudiantes de primero y segundo año en las muestras.

f. Trace gráficas de puntos para la cantidad de deuda para ambos grupos, usando una escala común. Interprete lo que muestran las gráficas de puntos, incluyendo la forma, centro y dispersión.

g. Encuentre la estimación puntual para la diferencia entre dos medias.

h. Compruebe las suposiciones para normalidad para ambos conjuntos de deuda de tarjeta de crédito. Verifique.

i. ¿Qué efecto tiene su respuesta a la parte h al contestar la pregunta: ¿La diferencia entre la deuda media de tarjeta de crédito para estudiantes de primer año y de segundo año de la parte g, es significativamente mayor que la diferencia de $292 indicada por Nellie Mae (páginas 545-546)?

j. Con base en lo que encuentre en las partes a-i, compare y contraste los hábitos de gasto de estudiantes universitarios de primer año y de segundo año.

Para su estudio

10.165 Diseñe su propio estudio que comprenda dos poblaciones.

a. Determine un conjunto de preguntas que compare las medias, proporciones o varianzas de dos poblaciones de interés para usted. Podría considerar dos diferentes niveles de clase como las dos poblaciones y preguntas similares a lo siguiente: ¿Hay diferencia entre sus costos medios de libros y víveres para un semestre? ¿Hay diferencia entre la proporción de quienes cuentan con sus propias tarjetas de crédito? ¿Hay diferencia entre la proporción de cuentan con cuatro o más tarjetas de crédito propias?

b. Defina dos poblaciones específicas que usted muestreará, describa su plan de muestreo, y recolecte los datos necesarios para contestar sus preguntas.

c. Analice cualesquiera diferencias y similitudes entre su estudio y el caso práctico del capítulo.

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

10.1 Cuando las medias de dos muestras no relacionadas se usan para comparar dos poblaciones, estamos hablando de **dos medias dependientes**.

10.2 Es frecuente que el uso de **datos pareados (medias dependientes)** permita el control de variables que no se pueden medir o sean confusas, porque cada par está sujeto por igual a estos efectos que confunden.

10.3 La **distribución ji cuadrada** se usa para hacer inferencias acerca de la razón entre las varianzas de dos poblaciones.

10.4 La **distribución** z se usa cuando dos medias dependientes hayan de ser comparadas.

10.5 Al comparar dos medias independientes cuando se desconocen las σ, necesitamos usar la distribución **normal estándar**.

10.6 El **puntaje normal estándar** se usa para todas las inferencias que traten de proporciones poblacionales.

10.7 La distribución F es una distribución **simétrica.**

10.8 El número de grados de libertad para el valor crítico de t es igual al **menor de** $n_1 - 1$ **o** $n_2 - 1$ cuando se hagan inferencias acerca de la diferencia entre dos medias independientes, en el caso que se estimen los grados de libertad.

10.9 En un intervalo de confianza para la diferencia media en datos pareados, el intervalo **aumenta** en ancho cuando el tamaño muestral se aumente.

10.10 Una **estimación agrupada**, para cualquier estadística en un problema que trate de dos poblaciones, es un valor al que se llega al combinar las dos estadísticas muestrales separadas de modo que se logre la mejor estimación puntual posible.

SEGUNDA PARTE: Aplicación de conceptos

Conteste todas las preguntas, mostrando todas las fórmulas, sustituciones y trabajo.

10.11 Exprese las hipótesis nula (H_o) y la alternativa (H_a) que se usarían para probar cada una de estas frases:

a. No hay diferencia significativa en los promedios medios de bateo para los jugadores de béisbol de las dos ligas mayores.

b. La desviación estándar para las cantidades mensuales de lluvia, en el condado de Monroe, es menor que la desviación estándar para las cantidades mensuales de lluvia en el condado de Orange.

c. Hay una diferencia significativa entre los porcentajes de estudiantes universitarios hombres y mujeres que poseen su propio automóvil.

10.12 Determine la estadística de prueba, región crítica y valor(es) crítico(s), que se usarían para completar cada prueba de hipótesis usando el procedimiento clásico con $\alpha = 0.05$.

a. H_o: $p_1 - p_2 = 0$
H_a: $p_1 - p_2 \neq 0$

b. H_o: $\mu_d = 12$
H_a: $\mu_d \neq 12$
$(n = 28)$

c. H_o: $\mu_1 - \mu_2 = 17$
H_a: $\mu_1 - \mu_2 > 17$
$(n_1 = 8, n_2 = 10)$

d. H_o: $\mu_1 - \mu_2 = 37$
H_a: $\mu_1 - \mu_2 < 37$
$(n_1 = 38, n_2 = 50)$

e. H_o: $\sigma_m^2 = \sigma_p^2$
H_a: $\sigma_m^2 > \sigma_p^2$
$(n_m = 16, n_p = 25)$

10.13 Encuentre cada uno de lo siguiente:

a. $z(0.02)$ b. $t\,(15, 0.025)$ c. $F(24,12, 0.05)$

d. $F(12, 24, 0.05)$ e. $z(0.04)$

f. $t(38, 0.05)$ g. $t(23, 0.99)$ h. $z(0.90)$

10.14 Al azar se dividieron, en dos grupos, 20 estudiantes universitarios de primer año. Los miembros de un grupo se asignaron a una sección de estadística que sólo utilizaba material programado, los del otro grupo se asignaron a una sección en la que el profesor daba conferencias. Al final del semestre, todos recibieron el mismo examen final. A continuación veamos los resultados.

Programado	76	60	85	58	91
	44	82	64	79	88
Conferencias	81	62	87	70	86
	77	90	63	85	83

Al nivel de significación de 5%, ¿estos datos dan suficiente evidencia para concluir que en promedio los estudiantes de las secciones de conferencias trabajaron significativamente mejor en el examen final? Suponga normalidad.

10.15 Los pesos de ocho personas antes que dejaran de fumar y = semanas después que dejaron de fumar son como sigue:

	1	2	3	4	5	6	7	8
Antes	148	176	153	116	129	128	120	132
Después	154	179	151	121	130	136	125	128

Al nivel de significación de 0.05, ¿esta muestra presenta suficiente evidencia para justificar la conclusión de que el peso aumenta si se deja de fumar? Suponga normalidad.

10.16 Una muestra a nivel nacional de 600 niños en edad escolar, y de 500 niñas de edad escolar, 288 niños y 175 niñas reconocieron haber cometido el delito de destrucción de propiedades. Use estos datos muestrales para construir un intervalo de confianza para la diferencia entre las proporciones de niños y niñas que han cometido este delito.

TERCERA PARTE: Entendiendo los conceptos

10.17 Para comparar la precisión de dos misiles de corto alcance, 8 de la primera clase y 10 de la segunda clase se disparan a un objetivo. Sea x la distancia por la que el proyectil no dio en el blanco. ¿Estos dos conjuntos de datos (8 distancias y 10 distancias) representan muestras dependientes o independientes? Explique.

10.18 Supongamos que 400 estudiantes de nuestra universidad están tomando estadística elemental este semestre. Describa cómo podría obtener dos muestras dependientes de tamaño 20 de estos estudiantes, para probar algún conocimiento que hayan tenido antes del curso y el mismo conocimiento después de terminar el curso. Sea específico.

10.19 El estudiante A dice, "No veo cuál es el alboroto por la diferencia entre medias independientes y dependientes; los resultados son casi iguales cualquiera que sea el método que se use". El profesor C sugiere que el estudiante A debe comparar los procedimientos con un poco de más cuidado. Ayude al estudiante A a descubrir que hay una diferencia significativa entre los procedimientos.

10.20 Suponga que usted está probando H_o: $\mu_d = 0$ contra H_a: $\mu_d < 0$ y las diferencias pareadas muestrales son todas negativas. ¿Significa esto que hay suficiente evidencia para rechazar la hipótesis nula? ¿Cómo puede no ser significativa? Explique.

10.21 La falta de asistencia a clase es muy perjudicial para el sistema de educación. Un grupo de profesores y asesores de preparatoria han creado un programa de asesoría en grupo, que esperan va a mejorar la situación de falta de asistencia a clase en su escuela. Han seleccionado los 80 estudiantes en su escuela con los peores registros de falta de asistencia y al azar han asignado la mitad de ellos al programa de asesoría en grupo. Al final del año escolar, los 80 estudiantes serán clasificados con respecto a su falta de asistencia. Cuando se hayan recolectado las calificaciones, éstas le serán entregadas a usted para su evaluación. Explique lo que hará para completar el estudio.

10.22 Usted desea estimar y comparar la proporción de familias católicas cuyos hijos asisten a una escuela privada, contra la proporción de familias no católicas cuyos hijos asisten a escuelas privadas. ¿Cómo estimará las dos proporciones y la diferencia entre ellas?

Trabajando con sus propios datos

La historia contiene muchos ejemplos de consumidores y los diversos productos que compran. Un objeto que se exhibe en el museo de ciencias de Boston habla del caso de un matemático panadero. Un hombre llamado Poincaré compraba una pieza de pan todos los días en la panadería local, pieza que se suponía pesaba 1 kilogramo. Después de un año de pesar y registrar el peso de cada pieza, Poincaré encontró una distribución normal con una media de 950 gramos. Llamó a la policía y al panadero le dijeron que se portara bien; no obstante, un año después Poincaré dijo que el panadero no se había reformado y otra vez la policía se presentó ante el panadero. El panadero fue interrogado, "¿Cómo pudo Poincaré saber que siempre le dábamos la pieza más grande?" Poincaré mostró entonces a la policía el segundo año de sus registros, una curva en forma de campana con una media de 950 gramos pero truncada a la izquierda.

Como consumidores, todos compramos infinidad de productos embotellados, en caja, enlatados y empacados. Raras veces, si acaso, alguno de nosotros preguntamos si el contenido tiene o no tiene realmente la cantidad indicada en el recipiente. Veamos a continuación algunas listas de contenido halladas en recipientes que compramos:

28 fl oz (1 pinta 12 onzas)
5 fl oz (148 ml)
Peso neto 10 oz
283 gramos
140 servilletas de una hoja
750 ml
32 fl oz (1 qt. 0.951
Peso neto 3¾ oz
48 bolsas de té de 106 g
77 pies cuadrados —92 hojas dobles— 11 × 11 pulgadas

¿Alguna vez se ha pregunta, "Obtengo la cantidad por la que pago"? Si esta idea cruzó por su mente, ¿trató de verificar la validez de lo dicho del contenido? El siguiente artículo apareció en el *Times Union* de Rochester, Nueva York, en 1972.

EMPRESA LECHERA ACUSADA DE MEDIR DE MENOS

El gerente de procesos de Dairylea Cooperative, Inc., ha sido citado en una orden judicial donde se acusa que la cooperativa está distribuyendo envases de leche en la zona de Rochester que contienen menos de la cantidad representada.

…un investigador encontró menos líquido en cuatro envases de leche Dairylea comprados el viernes.

El asistente del fiscal de distrito, Howard R. Relin, que emitió la orden judicial, dijo que las cantidades de menos iban de 1⅛ a 1¼ de onza por cuarto. Un cuarto de leche contiene 32 onzas de líquido.

…la ley de Agricultura y Mercados del estado…indica que un vendedor de una mercancía no debe vender o entregar menos de la cantidad representada a la venta.

…el propósito de la ley bajo la cual…la lechería es acusada es garantizar un trato honesto, preciso y justo con el público. No hay requerimiento de demostrar el intento de violar la ley, dijo.

*De *The Times-Union,* Rochester, NY, 16 de febrero de 1972.

Esta situación plantea un problema legal muy interesante: no hay necesidad de demostrar el intento de "dar de menos al cliente". Si son detenidos, los violadores a esta ley son multados automáticamente y los multas suelen ser muy fuertes.

A Una operación de llenado de alta velocidad

Una máquina tipo pistón de alta velocidad que se emplea para llenar latas de jugo de tomate con especias se vendió a una compañía empacadora. La garantía indicaba que la máquina llenaría latas de 48 onzas con una cantidad media de 49.5 onzas, una desviación estándar de 0.072 onzas, y una dispersión máxima de 0.282 onzas al operar a un ritmo de llenado de 150 a 170 latas por minuto. El 12 de agosto de 1994, se reunió

una muestra de 42 latas y se registraron los siguientes pesos. Éstos, medidos al $^1/_8$ de onza más cercano, se registraron como variaciones de 49.5 onzas.

$$-\frac{1}{8} \quad 0 \quad -\frac{1}{8} \quad 0 \quad 0 \quad 0 \quad -\frac{1}{8} \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad -\frac{1}{8} \quad 0$$

$$\frac{1}{8} \quad 0 \quad \frac{1}{8} \quad 0 \quad \frac{1}{8} \quad 0 \quad 0 \quad 0 \quad -\frac{1}{8} \quad 0 \quad 0 \quad 0 \quad 0 \quad -\frac{1}{8}$$

$$0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0$$

1. Calcule la media $\bar{x}$, la desviación estándar s, y el rango de los datos muestrales.
2. Construya un histograma que represente los datos muestrales.
3. ¿La cantidad de llenado difiere de las 49.5 onzas prescritas al nivel de $\alpha = 0.05$? Pruebe la hipótesis de que $\mu = 49.5$ contra una alternativa apropiada.
4. ¿La cantidad de variación, medida por el rango, satisface la garantía?
5. Suponiendo que la máquina llenadora continúa llenando latas con una cantidad de jugo de tomate distribuida normalmente, y que la media y desviación estándar son iguales a los valores hallados en la pregunta 1, ¿cuál es la probabilidad de que una lata seleccionada al azar contenga menos de las 48 onzas indicadas en la etiqueta?
6. Si la cantidad de llenado por lata está normalmente distribuida y la desviación estándar se puede mantener, encuentre el ajuste para el valor medio que permitiría que sólo una lata en cada 10 000 contenga menos de 48 onzas.

B Su propia investigación

Seleccione un producto empacado que tenga una cantidad de llenado por paquete que usted pueda y le gustaría investigar.

1. Describa su producto seleccionado, incluyendo la cantidad por paquete, y describa cómo piensa obtener sus datos.
2. Recolecte su muestra de datos. (Pida consejo a su profesor sobre el tamaño de muestra.)
3. Calcule la media $\bar{x}$ y la desviación estándar s para sus datos muestrales.
4. Construya un histograma o diagrama de tallo y hoja que represente los datos muestrales.
5. ¿La cantidad media de llenado está de acuerdo con la cantidad indicada en la etiqueta? Pruebe usando $\alpha = 0.05$.
6. Suponga que el artículo seleccionado está lleno continuamente. La cantidad de llenado está normalmente distribuida, y la media y desviación estándar son iguales a los valores hallados en la pregunta 3. ¿Cuál es la probabilidad de que un paquete seleccionado al azar contenga menos que la cantidad prescrita?

CAPÍTULO

11 Aplicaciones de *Ji* cuadrada

11.1 Algo dulce para contrarrestar el sabor picante

11.2 Estadístico *Ji* cuadrada

11.3 Inferencias relacionadas con experimentos multinomiales

11.4 Inferencias relacionadas con las tablas de contingencia

© Bonnie Kamin/Index Stock Imagery

11.1 Algo dulce para contrarrestar el sabor picante

Si le gustan los alimentos con picantes, es probable que prefiera algo dulce después de comer un delicioso platillo con especias. Algunos de los métodos más comunes que emplean las personas son beber agua, leche, refresco o comer pan u otro alimento. Hay personas incluso que prefieren no comer nada en esas ocasiones y así se quedan. La gráfica ilustra aquí las seis formas que prefieren personas adultas para contrarrestar el sabor picante.

PARA APAGAR EL FUEGO

Datos de Anne R. Carey y Suzy Parker, © 1995 USA Today.

A 200 personas que dicen les gustan mucho los alimentos con picante se les pidió citaran la forma en que contrarrestan el sabor picante después de comer con salsa. A continuación veamos el resumen del ejemplo resultante.

Método	Agua	Leche	Refresco	Cerveza	Pan	Otro	Nada
Número	73	35	20	19	29	11	13

Después de completar el capítulo 11, investigue con más detalle los datos precedentes y "Algo dulce para contrarrestar el sabor picante" en el proyecto del capítulo en la página 652.

SECCIÓN 11.1 EJERCICIOS

11.1 Con referencia a la muestra de 200 adultos recolectados en la sección 11.1:

a. ¿Qué información se recolectó de cada persona de la muestra?

b. Defina la población y la variable involucrada en la muestra.

c. Usando los datos muestrales, calcule porcentajes para los diversos métodos de contrarrestar el sabor picante.

d. ¿Cómo se comparan los porcentajes muestrales con los porcentajes gráficos?

11.2 Con referencia a la muestra de 200 adultos recolectados en la sección 11.1, ¿qué tan semejante es la distribución de la muestra y la distribución de porcentajes en la gráfica?

a. Construya una gráfica de barras horizontales de los 200 adultos, usando frecuencia relativa para la escala horizontal.

b. Sobreponga la gráfica de barras de "Para apagar el fuego" sobre la gráfica de barras de la parte a.

c. ¿Diría usted que la distribución de la muestra es "semejante a" o "muy diferente de" la distribución mostrada en la gráfica de "Para apagar el fuego"? Explique su respuesta.

11.2 Estadístico *Ji* cuadrada

Hay numerosos problemas por los que a los datos **enumerativos** se les asigna categoría y se muestran resultados por medio de cuentas. Por ejemplo, un conjunto de calificaciones de examen final pueden exhibirse como una distribución de frecuencia. Estos números de frecuencia son cuentas, el número de datos que caen en cada celda. Un estudio pregunta a electores si están registrados como republicanos, demócratas, o de otro partido, y si soportan o no a un candidato en particular. Los resultados suelen presentarse en una gráfica que muestra el número de electores en cada posible categoría. Numerosas ilustraciones de esta forma de presentar datos se han dado en los 10 capítulos previos.

Organización de datos

Suponga que tenemos varias **celdas** en las que n observaciones se han ordenado. (El término *celda* es sinónimo con el término *clase*; los términos *clase* y *frecuencia* se definieron y emplearon en capítulos anteriores. Antes de continuar, puede ser útil revisar las secciones 2.2, 2.3 y 3.2.) Las **frecuencias observadas** de cada celda están denotadas por $O_1, O_2, O_3, \ldots, O_k$ (vea la tabla 11.1). Nótese que la suma de todas las frecuencias observadas es

$$O_1 + O_2 + \cdots + O_k = n$$

donde n es el tamaño muestral. Lo que nos gustaría hacer es comparar las frecuencias observadas con algunas **frecuencias esperadas**, o teóricas, denotadas por $E_1, E_2, E_3, \ldots, E_k$ (vea la tabla 11.1), para cada una de estas celdas. De nueva cuenta, la suma de estas frecuencias esperadas debe ser exactamente n:

$$E_1 + E_2 + \cdots + E_k = n$$

TABLA 11.1

Frecuencias observadas

	Categorías k					
	1a	2a	3a	...	k-ésima	Total
Frecuencias observadas	O_1	O_2	O_3	...	O_k	n
Frecuencias esperadas	E_1	E_2	E_3	...	E_k	n

¿SABÍA USTED...?

Karl Pearson

Conocido como uno de los padres de la estadística moderna, Karl Pearson inventó la Ji-cuadrada (denotada por χ^2) en 1900. Es el procedimiento de inferencia más antiguo todavía en uso en su forma original y con frecuencia se usa en economía y negocios en la actualidad.

A continuación decidiremos si las frecuencias observadas parecen estar de acuerdo o en desacuerdo con las frecuencias esperadas. Haremos esto con el uso de una **prueba de hipótesis** con ***Ji*-cuadrada,** χ^2. χ es la letra minúscula del alfabeto griego que designa a la letra Ji.

Compendio del procedimiento de prueba

Estadística de prueba para *Ji*-cuadrada:

$$\chi^2\bigstar = \sum_{\text{todas las celdas}} \frac{(O-E)^2}{E} \qquad (11.1)$$

Este valor calculado para *Ji*-cuadrada es la suma de varios números no negativos, uno de cada celda (o categoría). El numerador de cada término de la fórmula para $\chi^2\bigstar$ es el cuadrado de la diferencia entre los valores de las frecuencias observada y esperada. Cuanto más cercanos entre sí sean estos valores, será más pequeño el valor de $(O-E)^2$; cuanto más lejanos entre sí, será mayor el valor de $(O-E)^2$. El denominador de cada celda pone el tamaño del numerador en perspectiva, es decir, una diferencia $(O-E)$ de 10 que resulte de frecuencias de 110 (O) y 100 (E) es muy distinta de una diferencia de 10 que resulte de 15 (O) y 5 (E).

Estas ideas sugieren que pequeños valores de *Ji*-cuadrada indican acuerdo entre los dos conjuntos de frecuencias, en tanto que valores más grandes indican desacuerdo. Por tanto, es usual que estas pruebas sean de una cola, con la región crítica a la derecha.

En muestreo repetido, el valor calculado de $\chi^2\bigstar$ en la fórmula (11.1) tendrá una distribución muestral que puede ser aproximada por la distribución de probabilidad de *Ji*-cuadrada cuando n es grande. Por lo general esta aproximación es considerada como adecuada cuando todas las frecuencias esperadas sean iguales o mayores a 5. Recuerde que las distribuciones de *Ji*-cuadrada, al igual que las distribuciones t de Student, son una familia de distribuciones de probabilidad, cada una siendo identificada por el número de parámetro de **grados de libertad**, df. El valor apropiado de df se describirá con cada prueba específica. Para usar la distribución *Ji*-cuadrada, debemos estar conscientes de sus propiedades, que se detallan en la sección 9.4 en la página 517. (También vea la figura 9.7.) Los valores críticos para *Ji*-cuadrada se obtienen de la tabla 8 en el apéndice B. (Se dan instrucciones específicas en la sección 9.4; vea las páginas 517-518.)

Suposición para usar *Ji*-cuadrada para hacer inferencias con base en datos enumerativos: la información muestral se obtiene usando una muestra aleatoria tomada de una población en la que cada individuo se clasifica de acuerdo con la(s) variable(s) categórica involucrada en la prueba.

Una *variable categórica* es aquella que clasifica o asigna categoría a cada individuo en exactamente una de varias celdas o clases; estas celdas o clases incluyen todo y son mutuamente exclusivas. El lado que queda hacia arriba en un dado que se lanza es una variable categórica: la lista de resultados {1, 2, 3, 4, 5, 6} es un conjunto de categorías que incluyen todo y son mutuamente exclusivas.

En este capítulo permitimos cierta cantidad de "liberalización" respecto a la hipótesis nula y su prueba. En capítulos previos la hipótesis nula fue siempre un enunciado acerca de un parámetro poblacional μ, σ, o p) No obstante, hay otros

tipos de hipótesis que se pueden probar, tales como "Este dado no está cargado" o "La estatura y peso de personas son independientes". Observe que estas hipótesis no son frases acerca de un parámetro, aun cuando a veces podrían indicarse con valores de parámetro especificados.

Suponga que digo "Este dado no está cargado", $p = P(\text{cualquier número}) = \frac{1}{6}$, y usted desea probar esta frase. ¿Qué haría? ¿Su respuesta fue algo como: tira muchas veces este dado y registra los resultados? Suponga que usted decide tirar el dado 60 veces. Si el dado no está cargado, ¿qué espera que ocurra? Cada número (1, 2, . . . , 6) debe aparecer aproximadamente $\frac{1}{6}$ del tiempo (es decir, 10 veces). Si aproximadamente ocurren 10 de cada número, es seguro que aceptará la frase de que el dado es bueno ($p = \frac{1}{6}$ por cada valor). Si el dado parece favorecer algunos números en particular, usted rechazará la frase. (La estadística de prueba χ^2★ tendrá un valor grande en este caso, como pronto veremos.)

SECCIÓN 11.2 EJERCICIOS

11.3 Usando la tabla 8 del apéndice B, encuentre lo siguiente:

a. $\chi^2(10, 0.01)$ b. $\chi^2(12, 0.025)$

c. $\chi^2(10, 0.95)$ d. $\chi^2(22, 0.995)$

11.4 Encuentre estos valores críticos usando la tabla 8 del apéndice B.

a. $\chi^2(18, 0.01)$ b. $\chi^2(16, 0.025)$

c. $\chi^2(40, 0.10)$ d. $\chi^2(45, 0.01)$

11.5 Usando la notación vista en el ejercicio 11.4, cite y encuentre los valores críticos de χ^2.

11.6 Usando la notación vista en el ejercicio 11.4, cite y encuentre los valores críticos de χ^2.

11.3 Inferencias relacionadas con experimentos multinomiales

El problema precedente de dados es una buena ilustración de un **experimento multinomial.** Consideremos este problema de nuevo. Suponga que deseamos probar este dado (a $\alpha = 0.05$) y decidir si no rechazar o rechazar la frase "Este dado no está cargado". (La probabilidad de cada número es $\frac{1}{6}$.) El dado se tira 60 veces desde un vaso sobre una superficie lisa y plana, con las siguientes frecuencias observadas:

Número	1	2	3	4	5	6
Frecuencia observada	7	12	10	12	8	11

Se supone que la hipótesis nula de que el dado no está cargado es verdadera, lo cual nos permite calcular las frecuencias esperadas. Si el dado no está cargado, ciertamente esperamos 10 sucesos de cada número.

Ahora calculemos un valor observado de χ^2. Estos cálculos se muestran en la tabla 11.2. El valor calculado es $\chi^2\bigstar = 2.2$.

TABLA 11.2

Cálculos para χ^2

Número	Observado (O)	Esperado (E)	$O - E$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
1	7	10	−3	9	0.9
2	12	10	2	4	0.4
3	10	10	0	0	0.0
4	12	10	2	4	0.4
5	8	10	−2	4	0.4
6	11	10	1	1	0.1
Total	60	60	0 ck		**2.2**

Nota: $\sum(O - E)$ debe ser igual a cero porque $\sum O = \sum E = n$. Se puede usar este dato como prueba, como se ve en la tabla 11.2.

Ahora usemos nuestro conocido formato de prueba de hipótesis.

Paso 1

a. **Parámetro de interés:** La probabilidad con la que cada lado queda hacia arriba: $P(1)$, $P(2)$, $P(3)$, $P(4)$, $P(5)$, $P(6)$.

b. **Enunciado de las hipótesis:**

H_o: El dado no está cargado (cada $p = \frac{1}{6}$).

H_a: El dado está cargado (al menos una p es diferente de las otras).

Paso 2

a. **Suposiciones:** Los datos se recolectaron de manera aleatoria, y cada resultado es uno de los seis números.

b. **Estadística de prueba:** La distribución *Ji*-cuadrada y fórmula (11.1), con df $= k - 1 = 6–1 = 5$.

En un experimento multinomial, df $= k - 1$, donde k es el número de celdas.

c. **Nivel de significación:** $\alpha = 0.05$.

Paso 3

a. **Información muestral:** Vea la tabla 11.2.

b. **Estadística calculada de prueba:** Usando la fórmula (11.1), tenemos

$$\chi^2\bigstar = \sum_{\text{todas las celdas}} \frac{(O - E)^2}{E}: \quad \chi^2\bigstar = \mathbf{2.2}$$

(los cálculos se muestran en la tabla 11.2)

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Use la cola derecha porque los valores "mayores" de *Ji*-cuadrada están en desacuerdo con la hipótesis nula:

$\mathbf{P = P(\chi^2\bigstar > 2.2 \mid df = 5)}$ como se ve en la figura.

Clásico:

a. La región crítica es la cola derecha porque los valores "mayores" de *Ji*-cuadrada están en desacuerdo con la hipótesis nula. El valor crítico se obtiene de la tabla 8, en la intersección de la fila df = 5 y columna = 0.05:

Para hallar el valor p, tiene dos opciones:

1. Usar la tabla 8 (apéndice B) para poner límite al valor p: **0.75 < P < 0.90.**
2. Usar computadora o calculadora para hallar el valor p: **P = 0.821.**

Para instrucciones específicas, vea la página 518.

b. El valor p no es menor que el nivel de significación, α.

Para instrucciones específicas, vea la página 518.

b. χ^2★ no está en la región crítica, como se muestra en **rojo** en la figura.

PARA SU INFORMACIÓN

Los comandos de computadora y calculadora, para hallar la probabilidad asociada con un valor especificado de *Ji*-cuadrada, se pueden hallar en el capítulo 9 (pp. 518-519).

Paso 5

a. Decisión: No rechazar H_o.

b. Conclusión: Al nivel de significación de 0.05, las frecuencias observadas no son significativamente diferentes de las esperadas de un dado no cargado.

Antes de ver otros ejemplos, debemos definir el término *experimento multinomial* y expresar los principios para completar la prueba *Ji*-cuadrada para ello.

Experimento multinomial: un experimento multinomial tiene las siguientes características:

1. Está formado por n intentos independientes idénticos.
2. El resultado de cada intento cabe en exactamente una de k posibles celdas.
3. Hay una probabilidad asociada con cada celda particular, y estas probabilidades individuales permanecen constantes durante el experimento. (Debe ser el caso que $p_1 + p_2 + \cdots + p_k = 1$.)
4. El experimento resultará en un conjunto de k frecuencias observadas, O_1, $O_2, \cdots, O_k$, donde cada O_i es el número de veces que el resultado de un intento cae en esa celda particular. (Debe ser el caso que $O_1 + O_2 + \cdots + O_k = n$.)

El ejemplo del dado satisface la definición de un experimento multinomial porque tiene las cuatro características descritas en la definición.

1. El dado se tiró n (60) veces de un modo idéntico, y estos intentos fueron independientes entre sí. (El resultado de cada intento no fue afectado por los resultados de otros intentos.)
2. Cada vez que se tiró el dado, resultó uno de seis números, y cada número estuvo asociado con una celda.
3. La probabilidad asociada con cada celda fue $\frac{1}{6}$, y ésta fue constante de un intento a otro. (Seis valores de $\frac{1}{6}$ dan 1.0.)
4. Cuando el experimento se completó, tenía una lista de seis frecuencias (7, 12, 10, 12, 8 y 11) que suman 60, lo que indica que cada uno de los resultados se tomó en cuenta.

El procedimiento de prueba para experimentos multinomiales es muy semejante al procedimiento de prueba descrito en capítulos previos. El cambio más grande

resulta con el enunciado de la hipótesis nula. Puede ser un enunciado verbal, como en el ejemplo del dado: "Este dado no está cargado". A veces, la alternativa de la hipótesis nula no se expresa. No obstante, en este libro la hipótesis alternativa se mostrará, porque ayuda a organizar y entender el problema. No se usará para determinar la ubicación de la región crítica, sin embargo, como fue el caso en capítulos previos. Para experimentos multinomiales siempre usaremos una región crítica de una cola, y será la cola derecha de la distribución χ^2 porque las desviaciones más grandes (positivas o negativas) a partir de los valores esperados llevan a un aumento en el valor calculado de $\chi^2\bigstar$.

El valor crítico se determinará por el nivel de significación asignado (α) y el número de grados de libertad. El número de grados de libertad (df) será 1 menos que el número de celdas (k) en las que se dividen los datos:

Grados de libertad para experimentos multinomiales

$$\text{df} = k - 1 \tag{11.2}$$

Cada frecuencia esperada, E_i, será determinada al multiplicar el número total de intentos n por la correspondiente probabilidad (p_i) para esa celda, es decir,

Valor esperado para experimentos multinomiales

$$E_i = n \cdot p_i \tag{11.3}$$

Un principio debe ser satisfecho para asegurar una buena aproximación a la distribución *Ji*-cuadrada: cada frecuencia esperada debe ser al menos 5 (es decir, cada $E_i \geq 5$). A veces es posible combinar celdas "menores" para satisfacer este principio. Si este principio no puede ser satisfecho, entonces deben usarse medidas correctivas para asegurar una buena aproximación. Estas medidas correctivas no se estudian en este libro pero se analizan en muchas otras fuentes.

EJEMPLO 11.1 Una prueba de hipótesis multinomial con iguales frecuencias esperadas

Los estudiantes universitarios han insistido regularmente en tener libertad de elección cuando se registran para cursos. Este semestre hubo siete secciones de un curso particular de matemáticas. Las secciones estuvieron programadas para satisfacerse en tiempos diversos con varios instructores. La tabla 11.3 muestra el número de estudiantes que seleccionaron cada una de las siete secciones. ¿Estos datos indican que los estudiantes tenían preferencia por ciertas secciones, o indican que cada sección era igualmente probable que la eligieran?

TABLA 11.3

Datos sobre inscripciones por sección

	Sección							
	1	2	3	4	5	6	7	Total
Número de estudiantes	18	12	25	23	8	19	14	119

SOLUCIÓN Si no se mostró preferencia en la selección de secciones, entonces esperaríamos que los 119 estudiantes estuvieran igualmente distribuidos entre los siete grupos. Esperaríamos que 17 de ellos se registraran en cada sección. La prueba de hipótesis se completa al nivel de significación de 5 por ciento.

Paso 1 **a. Parámetro de interés:** Preferencia por cada sección, la probabilidad de que una sección particular sea seleccionada al registrarse.

b. Enunciado de las hipótesis:

H_o: No se muestra preferencia (igualmente distribuida)

H_a: Se mostró preferencia (no igualmente distribuida)

Paso 2 **a. Suposiciones:** Los 119 estudiantes representan una muestra aleatoria de la población de todos los estudiantes que se inscriben para este curso particular. Como no se introdujeron regulaciones en la selección de cursos y la inscripción pareció continuar en su forma usual, no hay razón para pensar que ésta no es sino una muestra aleatoria.

b. Estadística de prueba: La distribución *Ji*-cuadrada y la fórmula (11.1), con df = 6.

c. Nivel de significación: $\alpha = 0.05$.

Paso 3 **a. Información muestral:** Vea la tabla 11.3 (p. 625).

b. Estadística de prueba calculada: Usando la fórmula (11.1), tenemos

$$\chi^2\bigstar = \sum_{\text{todas las celdas}} \frac{(O-E)^2}{E}: \quad \chi^2\bigstar = \frac{(18-17)^2}{17} + \frac{(12-17)^2}{17} + \frac{(25-17)^2}{17}$$

$$+ \frac{(23-17)^2}{17} + \frac{(8-17)^2}{17} + \frac{(19-17)^2}{17} + \frac{(14-17)^2}{17}$$

$$= \frac{(1)^2 + (-5)^2 + (8)^2 + (6)^2 + (-9)^2 + (2)^2 + (-3)^2}{17}$$

$$= \frac{1 + 25 + 64 + 36 + 81 + 4 + 9}{17} = \frac{220}{17} = 12.9411$$

$$= \mathbf{12.94}$$

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Use la cola derecha porque valores "mayores" de Ji-cuadrada están en desacuerdo con la hipótesis nula: $\mathbf{P = P(\chi^2\bigstar > 12.94 \mid df = 6)}$ como se ve en la figura:

Para hallar el valor *p*, tiene dos opciones:

1. Usar la tabla 8 (apéndice B) para poner límites al valor *p*: $\mathbf{0.025 < P < 0.05}$.
2. Use computadora o calculadora para hallar el valor *p*: $\mathbf{P = 0.044}$.

Para instrucciones específicas, vea la página 518.

b. El valor *p* es menor que el nivel de significación, α.

Clásico:

a. La región crítica es la cola derecha porque los valores "mayores" de Ji-cuadrada están en desacuerdo con la hipótesis nula. El valor crítico se obtiene de la tabla 8, en la intersección de la fila df=6 y la columna $\alpha = 0.05$:

$$\chi^2(6, 0.05) = \mathbf{12.6}$$

Para instrucciones específicas, vea la página 518.

b. $\chi^2\bigstar$ está en la región crítica, como se muestra en el **área sombreado** en la figura.

Paso 5

a. **Decisión:** Rechazar H_o.

b. **Conclusión:** Al nivel de significación de 0.05, parece haber una preferencia mostrada. No podemos determinar por la información cuál es la preferencia. Podría ser preferencia del profesor, preferencia del tiempo o un conflicto de horario.

Las conclusiones deben ser expresadas con todo cuidado para evitar sugerir conclusiones que los datos no puedan apoyar.

No todos los experimentos multinomiales resultan en iguales frecuencias esperadas, como veremos en el ejemplo 11.2.

EJEMPLO 11.2 Una prueba de hipótesis multinomial con frecuencias esperadas desiguales

La teoría de Mendel de la herencia dice que las frecuencias de guisantes (chícharos) redondos y amarillos, arrugados y amarillos, redondos y verdes y arrugados y verdes, se presentarán con una relación 9:3:3:1 cuando se crucen dos variedades específicas de chícharos. Al probar esta teoría, Mendel obtuvo frecuencias de 315, 101, 108, y 32, respectivamente. ¿Estos datos muestrales dan suficiente evidencia para rechazar la teoría al nivel de significación de 0.05?

SOLUCIÓN

Paso 1

a. **Parámetro de interés:** Las proporciones: P(redondo y amarillo), P(arrugado y amarillo), P(redondo y verde), P(arrugado y verde).

b. **Enunciado de las hipótesis:**

H_o: 9:3:3:1 es la razón de la herencia.

H_a: 9:3:3:1 no es la razón de la herencia.

Paso 2

a. **Suposiciones:** Supondremos que los resultados de Mendel forman una muestra aleatoria.

b. **Estadística de prueba:** La distribución *Ji*-cuadrada y fórmula (11.1), con df = 3.

c. **Nivel de significación:** $\alpha = 0.05$.

Paso 3

a. **Información muestral:** Las frecuencias observadas fueron: 315, 101, 108 y 32.

b. **Estadística de prueba calculada:** La razón 9:3:3:1 indica probabilidades de $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ y $\frac{1}{16}$.

Por tanto, las frecuencias esperadas son $\frac{9n}{16}$, $\frac{3n}{16}$, $\frac{3n}{16}$ y $\frac{1n}{16}$. Tenemos

$$n = \sum O_i = 315 + 101 + 108 + 32 = 556$$

Los cálculos para determinar $\chi^2\bigstar$ se muestran en la tabla 11.4.

TABLA 11.4

Cálculos necesarios para determinar $\chi^2\bigstar$

O	E	$O - E$	$\frac{(O-E)^2}{E}$
315	312.75	2.25	0.0162
101	104.25	−3.25	0.1013
108	104.25	3.75	0.1349
32	34.75	−2.75	0.2176
556	556.00	0 ck	**0.4700** $\longrightarrow \chi^2\bigstar = \sum_{\text{todas las celdas}} \frac{(O-E)^2}{E} = \mathbf{0.47}$

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Usar la cola derecha porque valores "mayores" de Ji-cuadrada están en desacuerdo con la hipótesis nula: $\mathbf{P = P\,(\chi^2\bigstar > 0.47 \mid df = 3)}$ como se muestra en la figura. Para hallar el valor *p*, tiene dos opciones:

1. Usar la tabla 8 (apéndice B) para poner límites al valor *p*: $\mathbf{0.90 < P < 0.95.}$
2. Usar computadora o calculadora para hallar el valor *p*: $\mathbf{P = 0.925.}$

Para instrucciones específicas, vea la página 518.

b. El valor *p* no es menor que el nivel de significación, α.

Clásico:

a. La región crítica es la cola derecha porque valores "mayores" de Ji-cuadrada están en desacuerdo con la hipótesis nula. El valor crítico se obtiene de la tabla 8, en la intersección de la fila df=3 y la columna $\alpha = 0.05$:

$$\chi^2(3, 0.05) = \mathbf{7.82}$$

Para instrucciones específicas, vea la página 518.

b. $\chi^2\bigstar$ no está en la región crítica, como se muestra en **el área sombreada** en la figura.

Paso 5 **a. Decisión:** No rechazar H_o.

b. Conclusión: Al nivel de significación de 0.05, no hay suficiente evidencia para rechazar la teoría de Mendel.

CASO PRÁCTICO 11.3 Cumpleaños

La Census Bureau recolecta datos para numerosas variables. La información exhibida por la siguiente gráfica está basada en el censo de 2003 de Estados Unidos, y se ajusta al formato de un experimento multinomial. Verifique que estos datos llenen los requisitos como experimento multinomial (vea ejercicio 11.7).

FECHAS ABUNDANTES DE CUMPLEAÑOS

Nacen más niños los martes que en cualquier otro día de la semana. En Estados Unidos, nacen un promedio de 13,000 niños cada martes. Día más lento para nacimientos: el domingo.

Datos de Anne R. Carey y Ron Coddington, © 2003 USA Today.

CASO PRÁCTICO 11.4 ¿Descargando qué?

La gráfica "Adolescentes y descargas" muestra los resultados de encuestar personas de 8 a 18 años de edad acerca de lo que descargan usando sus teléfonos celulares. Esta información no califica como experimento multinomial. ¿Qué propiedad se viola? (Vea ejercicio 11.8.)

ADOLESCENTES Y DESCARGAS

Datos de Justin Dickerson y Adrienne Lewis, © 2005 USA Today.

SECCIÓN 11.3 EJERCICIOS

11.7 Verifique que el caso práctico 11.3 (p. 628) sea un experimento multinomial. Sea específico.

a. ¿Qué es un intento?

b. ¿Cuál es la variable?

c. ¿Cuáles son los niveles posibles de resultados de cada intento?

11.8 ¿Por qué la información mostrada en el caso práctico 11.4 (arriba) no es la de un experimento multinomial? Sea específico.

11.9 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usaría para probar los siguientes enunciados:

a. Los cinco números 1, 2, 3, 4, y 5 tienen igual probabilidad de ser sacados.

b. La pregunta de opción múltiple tiene una historia de estudiantes que seleccionan respuestas en una razón de 2:3:2:1.

c. El sondeo mostrará una distribución de 16%, 38%, 41% y 5% para las posibles clasificaciones de excelente, malo, regular y malo en ese problema.

11.10 Determine el valor p para las siguientes pruebas de hipótesis que comprenden la distribución χ^2.

a. H_o: $P(1) = P(2) = P(3) = P(4) = 0.25$, con $\chi^2\bigstar = 12.25$

b. H_o: $P(\text{I}) = 0.25$, $P(\text{II}) = 0.40$, $P(\text{III}) = 0.35$, con $\chi^2\bigstar = 5.98$

11.11 Determine el valor crítico y región crítica que se usaría en el procedimiento clásico para probar la hipótesis nula para cada uno de los siguientes experimentos multinomiales.

a. H_o: $P(1) = P(2) = P(3) = P(4) = 0.25$, con $\alpha = 0.05$

b. H_o: $P(\text{I}) = 0.25$, $P(\text{II}) = 0.40$, $P(\text{III}) = 0.35$, con $\alpha = 0.01$

11.12 Explique la forma en que 9:3:3:1 se convierte en $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ y $\frac{1}{16}$ en el ejemplo 11.2 en la página 627.

11.13 Explique cómo se obtuvieron 312.75, 2.25 y 0.0162 en la primera fila de la tabla 11.4 en la página 628.

11.14 Un fabricante de cera para pisos realizó un experimento de preferencias de consumidores para determinar cuál de cinco marcas de cera para pisos era la más atrayente en apariencia. Una muestra de 100 consumidores vieron cinco superficies de piso que habían sido encerados con uno de los cinco tipos de cera. Cada uno de los consumidores indicó la superficie que prefería. La iluminación y fondo eran aproximadamente iguales para todas las superficies. Los resultados fueron como sigue:

Cera	A	B	C	D	E	Total
Frecuencia	27	17	15	22	19	100

Resuelva lo siguiente usando el procedimiento del valor p y resuelva usando el procedimiento clásico:

a. Exprese la hipótesis para "sin preferencia" en terminología estadística.

b. ¿Qué estadística de prueba se usará al probar esta hipótesis nula?

c. Complete la prueba de hipótesis usando $\alpha = 0.10$.

11.15 Cierto tipo de semilla de flores producirá flores magenta, amarillo verdoso y ocre con una proporción 6:3:1 (una flor por semilla). Se plantan un total de 100 semillas y todas germinan, dando los resultados siguientes.

Magenta	Amarillo verdoso	Ocre
52	36	12

Resuelva lo siguiente usando el procedimiento del valor p y resuelva usando el procedimiento clásico:

a. Si la hipótesis nula (6:3:1) es verdadera, ¿cuál es el número esperado de flores magenta?

b. ¿Cuántos grados de libertad están asociados con *Ji*-cuadrada?

c. Complete la prueba de hipótesis usando $\alpha = 0.10$.

11.16 Con los años, es más probable que actores afroamericanos que aparecen en estrenos cinematográficos importantes tengan papeles destacados en comedias que los actores de raza blanca. La siguiente tabla muestra el porcentaje de todos los papeles por tipo de película.

Tipo de película	Porcentaje de papeles	Tipo de película	Porcentaje de papeles
Acción y aventura	13.2	Horror y suspenso	12.5
Comedia	31.9	Comedia romántica	8.2
Drama	23.0	Otro	11.2

La tabla siguiente muestra el número de papeles principales desempeñados por afroamericanos para cada tipo de película en las últimas 89 películas estrenadas.

Tipo de película	Número de papeles	Tipo de película	Número de papeles
Acción y aventura	9	Horror y suspenso	11
Comedia	40	Comedia romántica	5
Drama	17	Otro	7

¿La distribución de papeles de afroamericanos difiere de la distribución general de papeles en estrenos cinematográficos principales?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.17 Un gran supermercado tiene en existencia cuatro calidades de carne molida de res. Se cree que sus clientes compran estas cuatro variedades con probabilidades de 0.10, 0.30, 0.35 y 0.25, respectivamente, de la variedad más barata a la más cara. Una muestra de 500 compras resultó en ventas de 46, 162, 191 y 101 de las calidades respectivas. ¿Esta muestra contradice las proporciones esperadas? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.18 Uno de los principales beneficios del correo electrónico es que hace posible comunicarse rápidamente sin obtener un tono de ocupado o que no haya contestación, dos críticas importantes de las llamadas telefónicas. Pero, ¿el e-mail es bueno para resolver problemas que tienen las personas cuando ejecutan un programa de computadora? Un estudio sondeó las opiniones de consumidores que trataban de usar e-mail para obtener ayuda, al pegar un mensaje en línea al fabricante de su PC o su representante autorizado. Los resultados se presentan en la tabla siguiente.

Resultado de pregunta en línea	Porcentaje
Nunca obtuvo respuesta	14
Obtuvo respuesta, pero no sirvió	30
La respuesta ayudó, pero no resolvió el problema	34
La respuesta resolvió el problema	22

Fuente: PC World, "PC World Reliability and Service Survey"

Como gerente de marketing para un gran fabricante de computadoras personales (PC), usted decide llevar a cabo un sondeo de sus clientes usando sus registros de e-mail para compararlos contra los resultados publicados. Para asegurar una comparación justa, elige usar el mismo cuestionario y examina las respuestas de 500 clientes que trataron de usar el e-mail para obtener ayuda de su personal de apoyo técnico. Los resultados son:

Resultado de pregunta en línea	Número que respondió
Nunca obtuvo respuesta	35
Obtuvo respuesta, pero no sirvió	102
La respuesta ayudó, pero no resolvió el problema	125
La respuesta resolvió el problema	238
Total	500

¿La distribución de respuestas difiere de la distribución obtenida del sondeo publicado? Pruebe al nivel de significación de 0.01.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.19 *Nursing Magazine* informó resultados de un sondeo de más de 1800 enfermeras de todo el país, respecto a la satisfacción y retención en el trabajo. Las enfermeras de hospitales "imanes" (hospitales que con éxito atraen y retienen enfermeras) describen la situación de empleo en sus unidades como sigue:

Situación de empleo	Porcentaje
1. Escasez desesperada de ayuda; mal cuidado a pacientes	12
2. Escaso, pero buen cuidado a pacientes	32
3. Adecuado	38
4. Más que adecuado	12
5. Excelente	6

Un sondeo de 500 enfermeras de hospitales no tan atractivos dio las siguientes respuestas a la situación de empleo.

Situación de empleo	1	2	3	4	5
Número	165	140	125	50	20

¿Los datos indican que las enfermeras de los hospitales no tan atractivos tienen una distribución diferente de opiniones? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.20 Se ha de probar un programa para generar números aleatorios en computadora; se le dan instrucciones para generar 100 números enteros de un dígito entre 0 y 9. Las frecuencias de los enteros fueron como sigue:

Entero	0	1	2	3	4	5	6	7	8	9
Frecuencia	11	8	7	7	10	10	8	11	14	14

Al nivel de significación de 0.05, ¿hay suficiente razón para pensar que los enteros no se están generando de manera uniforme?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.21 "Saliendo de una deuda, paso a paso," artículo del 29 de abril de 2005 de *USA Today*, publicó resultados de una encuesta de 260 miembros de la Financial Planning Association. Cada uno de los planificadores financieros informó que consideran ser uno de los pasos más valiosos que las personas pueden tomar para mejorar su vida financiera.

Paso más valioso	Porcentaje
1. Metas establecidas	30
2. Pagarse a sí mismo primero	21
3. Crear y apegarse a un presupuesto	17
4. Ahorrar en forma regular	12
5. Pagar deuda de tarjeta de crédito	7
6. Invertir el máximo en 401(k)	5
7. Otro	8

Un sondeo de 60 planificadores financieros de una zona metropolitana del interior del país dio las siguientes respuestas a la pregunta de "la meta más valiosa".

Respuesta a pregunta	1	2	3	4	5	6	7
Número	10	13	13	8	9	3	4

¿Los datos indican que los planeadores financieros de una zona metropolitana del interior del país tienen una distribución diferente de opiniones? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.22 El censo de 2003 de Estados Unidos encontró que bebés entraban al mundo en los días de la semana en las proporciones que siguen.

Día de la semana	*P*(día)	Día de la semana	*P*(día)
Domingo	0.098	Jueves	0.160
Lunes	0.149	Viernes	0.159
Martes	0.166	Sábado	0.111
Miércoles	0.157		

Fuente: U.S.Census Bureau

Una muestra aleatoria seleccionada de los registros de nacimientos para una gran zona metropolitana resultó en los siguientes datos:

Día	Dom	Lun	Mar	Mié	Jue	Vie	Sáb
Observado	10	6	9	13	9	17	11

a. ¿Estos datos dan suficiente evidencia para rechazar la frase de que "los nacimientos ocurren en esta zona metropolitana en las mismas proporciones diarias", como lo reporta la U.S. Census Bureau? Use $\alpha = 0.05$.

b. ¿Estos datos dan suficiente evidencia para rechazar la frase de que "los nacimientos ocurren en esta zona metropolitana todos los días con la misma posibilidad"? Use $\alpha = 0.05$.

c. Compare los resultados obtenidos en las partes a y b. Exprese sus conclusiones.

11.23 Los dulces Skittles Original Fruit, del tamaño de un bocadillo, son de muchos colores que se venden en bolsas y se puede "Probar el Arco Iris" con sus cinco colores y sabores: verde, lima; púrpura, uva; amarillo, limón; anaranjado, naranja; y rojo, fresa. A diferencia de algunos de los otros dulces multicolores que hay, Skittles dice que sus cinco colores son igualmente probables. En un intento por rechazar este dicho, se compró una bolsa de dulces Skittles de 4 onzas y se contaron los colores:

Rojo	Anaranjado	Amarillo	Verde	Púrpura
18	21	23	17	27

¿Esta muestra contradice el dicho de Skittles al nivel de 0.05?

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

11.24 Para demostrar o explorar el efecto que el tamaño muestral aumentado tiene en el valor calculado de *Ji*-cuadrada, consideremos los dulces Skittles del ejercicio 11.23 y tomemos muestras de algunas bolsas grandes de los dulces.

a. Suponga que compramos una bolsa de 16 onzas de Skittles, contamos los colores y observamos exactamente la misma proporción de colores hallados en el ejercicio 11.23:

Rojo	Anaranjado	Amarillo	Verde	Púrpura
72	84	92	68	108

Calcule el valor de *Ji*-cuadrada para estos datos. ¿Cómo está relacionado el nuevo valor de *Ji*-cuadrada con el hallado en el ejercicio 11.23? ¿Qué efecto tiene este nuevo valor en los resultados de la prueba? Explique.

b. Para continuar esta demostración o exploración, suponga que compramos una bolsa de 48 onzas, contamos los colores y observamos exactamente la misma proporción de colores del ejercicio 11.23 y la parte b de este ejercicio.

Rojo	Anaranjado	Amarillo	Verde	Púrpura
216	252	276	204	324

Calcule el valor de *Ji*-cuadrada para estos datos. ¿Cómo está relacionado el nuevo valor de *Ji*-cuadrada con el hallado en el ejercicio 11.23? Explique

c. ¿Qué efecto tiene el tamaño de la muestra en el valor calculado de *Ji*-cuadrada cuando la proporción de frecuencias observadas sigue igual cuando aumenta el tamaño muestral?

d. Explique en qué forma esto indica que si se toma una muestra grande lo suficiente, la prueba de hipótesis en última instancia resulta en un rechazo.

11.25 Según The Harris Poll, la proporción de todos los adultos que viven en familias con rifles (29%), escopetas (29%) o pistolas (23%) no ha cambiado de manera significativa desde 1996. No obstante, hoy en día más personas viven en familias sin armas de fuego (61%). Los 1014 adultos encuestados dieron los resultados siguientes.

	Todos adultos (%)	Todos dueños de arma (%)
Tienen rifle, escopeta y pistola (3 de 3)	16	41
Tienen 2 de 3 (rifle, escopeta o pistola)	11	27
Tienen 1 de 3 (rifle, escopeta o pistola)	11	29
Declina contestar/no está seguro	1	3
TOTAL	39%	100%

En un sondeo de 2000 adultos en Memphis que dijeron tener armas, 780 dijeron que tienen los 3 tipos, 550 dijeron que tenían 2 de 3, 560 dijeron que tenían 1 de 3, y 110 declinaron especificar qué tipos de armas poseían.

a. Pruebe la hipótesis nula de que la distribución del número de tipos en propiedad es igual en Memphis a como la publica The Harris Poll. Use un nivel de significación igual a 0.05.

b. ¿Qué hizo que el valor calculado de χ^2★ fuera tan grande? ¿Le parece bien que una celda tenga todo este efecto en los resultados? ¿Cómo podría completarse esta prueba de modo diferente (quizá de modo más significativo), para que los resultados no sean afectados como lo fueron en la parte a? Sea específico.

11.26 ¿Por qué la prueba de *Ji*-cuadrada es por lo general una prueba de una cola con la región crítica en la cola derecha?

a. ¿Qué clase de valor resultaría si las frecuencias observadas y las frecuencias esperadas estuvieran muy cercanas en valor? Explique cómo interpretaría esta situación.

b. Suponga que usted tuvo que lanzar un dado 60 veces como experimento para probar que no estaba cargado, como se dijo en el ejemplo de las páginas 622-624; pero en lugar de tirar el dado usted, le pagó a su hermanito $1 para que lo tirara 60 veces y llevara un total de los números. Él estuvo de acuerdo en hacer esto para usted y corrió a su cuarto con el dado, regresando a los pocos minutos con sus frecuencias resultantes. Él pidió el dólar y usted, por supuesto, le pagó antes que le diera los resultados, que son como sigue: 10, 10, 10, 10, 10 y 10. Los resultados observados son exactamente lo que usted había "esperado", ¿verdad? Explique sus reacciones. ¿Qué valor de χ^2★ resultaría? ¿Qué piensa usted que pasó? ¿Qué pide usted de su hermanito, y por qué? ¿Qué posible papel podría tener la cola izquierda en la prueba de hipótesis?

c. ¿Por qué la cola izquierda no es generalmente de interés?

11.4 Inferencias relacionadas con las tablas de contingencia

Una **tabla de contingencia** es un arreglo de datos en una clasificación de dos formas. Los datos están ordenados en celdas, y la cuenta para cada celda se informa. La tabla de contingencia comprende dos factores (o variables), y una pregunta común respecto a estas tablas es si los datos indican que las dos variables son independientes o dependientes (vea pp. 147-149, 243-245).

Dos pruebas diferentes usan el formato de tabla de contingencia. La primera que veremos es la *prueba de independencia.*

Prueba de independencia

Para ilustrar una prueba de independencia, consideremos una muestra aleatoria que muestra el género de estudiantes universitarios de artes liberales y su campo académico favorito.

EJEMPLO 11.5 Prueba de hipótesis para independencia

Cada persona de un grupo de 300 estudiantes se identificó como hombre o mujer y luego se le preguntó si prefería tomar cursos de artes liberales en el campo de ciencias matemáticas, ciencias sociales, o humanidades. La tabla 11.5 (p. 634) es una tabla de contingencia que muestra las frecuencias halladas para estas categorías. ¿Esta muestra presenta suficiente evidencia para rechazar la hipótesis nula:

"La preferencia por ciencias matemáticas, ciencias sociales, o humanidades es independiente del género de un estudiante universitario"? Complete la **prueba de hipótesis** usando el nivel de significación de 0.05.

TABLA 11.5

Resultados muestrales para preferencia de género y materia

	Materia favorita			
Género	Ciencias matemáticas (MS)	Ciencias sociales (SS)	Humanidades (H)	Total
Hombre (M)	37	41	44	122
Mujer (F)	35	72	71	178
Total	72	113	115	300

SOLUCIÓN

Paso 1

a. **Parámetro de interés:** Determinar la independencia de las variables "género" y "materia favorita" nos pide discutir la probabilidad de los diversos casos y el efecto que las respuestas acerca de una variable tiene en la probabilidad de respuestas acerca de la otra variable. La independencia, como se definió en el capítulo 4, requiere $P(\text{MS}|\text{M}) = P(\text{MS}|\text{F}) = P(\text{MS})$; esto es, el género no tiene efecto en la probabilidad en la elección de la materia que haga una persona.

b. **Enunciado de las hipótesis:**
H_o: La preferencia por ciencias matemáticas, ciencias sociales, o humanidades, es independiente del género de un estudiante universitario.
H_a: La preferencia de materia no es independiente del género del estudiante.

Paso 2

a. **Suposiciones:** La información muestral se obtiene usando una muestra aleatoria sacada de una población, con cada individuo clasificado de acuerdo con el género y materia favorita.

b. **Estadística de prueba.**

En el caso de tablas de contingencia, el número de grados de libertad es exactamente el mismo que el número de celdas de la tabla que se pueda llenar libremente cuando nos dan los *totales marginales.* Los totales en este ejemplo se muestran en la tabla siguiente:

			122
			178
72	113	115	300

Dados estos totales, se pueden llenar sólo dos celdas antes que las otras queden determinadas. (Por supuesto, los totales deben permanecer iguales.) Por ejemplo,

una vez que escojamos dos valores arbitrarios (por ejemplo 50 y 60) para las primeras dos celdas de la primera fila, los valores de las otras cuatro celdas son fijos (vea la tabla siguiente):

50	60	*C*	122
D	*E*	*F*	178
72	113	115	300

Los valores tienen que ser $C = 12$, $D = 22$, $E = 53$, y $F = 103$. De otro modo, los totales no serán correctos. Por tanto, para este problema hay dos opciones libres. Cada opción libre corresponde a 1 grado de libertad. En consecuencia, el número de grados de libertad para nuestro ejemplo es 2 (df = 2).

La distribución *Ji*-cuadrada se usará junto con la fórmula (11.1), con df = 2.

c. Nivel de significación: $\alpha = 0.05$.

Paso 3 **a. Información muestral:** Vea la tabla 11.5.

b. Estadística de prueba calculada.

Antes de calcular el valor de *Ji*-cuadrada, necesitamos determinar los valores esperados, *E*, para cada celda. Para hacer esto debemos recordar la hipótesis nula, que dice que estos factores son independientes. Por tanto, esperaríamos que los valores sean distribuidos en proporción a los totales marginales. Hay 122 hombres; esperaríamos que ellos se distribuyan entre MS, SS, y H proporcionalmente a 72, 113 y 115 totales. Así, las cuentas esperadas de celda para hombres son

$$\frac{72}{300} \cdot 122 \qquad \frac{113}{300} \cdot 122 \qquad \frac{115}{300} \cdot 122$$

Del mismo modo, esperaríamos para las mujeres

$$\frac{72}{300} \cdot 178 \qquad \frac{113}{300} \cdot 178 \qquad \frac{115}{300} \cdot 178$$

Entonces, los valores esperados son como se muestra en la tabla 11.6. Siempre verifique los totales marginales para los valores esperados contra los totales marginales para los valores observados.

TABLA 11.6

Valores esperados

	MS	SS	H	Total
Hombres	29.28	45.95	46.77	122.00
Mujeres	42.72	67.05	68.23	178.00
Total	72.00	113.00	115.00	300.00

Nota: podemos recordar el cálculo de los valores esperados de un segundo modo. Recuerde que supusimos que la hipótesis nula es verdadera hasta que haya evidencia para rechazarla. Habiendo hecho esta suposición en nuestro ejemplo, estamos

diciendo en efecto que el evento de que un estudiante escogido al azar sea hombre, y el evento de que un estudiante escogido al azar prefiera cursos de ciencias matemáticas, son independientes. Nuestra estimación puntual para la probabilidad de que un estudiante sea hombre es $\frac{122}{300}$, y la estimación puntual para la probabilidad de que el estudiante prefiera cursos de ciencias matemáticas es $\frac{72}{300}$. Por tanto, la probabilidad de que ambos eventos ocurran es el producto de las probabilidades. [Consulte la fórmula (4.7), p. 246.] Entonces, $\left(\frac{122}{300}\right)\left(\frac{72}{300}\right)$ es la probabilidad de que un estudiante seleccionado sea hombre y que prefiera ciencias matemáticas. El número de estudiantes, de entre 300 que se espera sean hombres y prefieran ciencias matemáticas, se encuentra al multiplicar la probabilidad (o proporción) por el número total de estudiantes (300). Así, el número esperado de hombres que prefieren ciencias matemáticas es $\left(\frac{122}{300}\right)\left(\frac{72}{300}\right)(300) = \left(\frac{122}{300}\right)(72) = 29.28$. Los otros valores esperados se pueden determinar en la misma forma.

En general, la tabla de contingencia se escribe de modo que contenga toda esta información (vea la tabla 11.7).

TABLA 11.7

Tabla de contingencia mostrando resultados muestrales y valores esperados

	Materia favorita			
Género	Mat	Sociales	Humanidades	Total
Hombre	37 (29.28)	41 (45.95)	44 (46.77)	122
Mujer	35 (42.72)	72 (67.05)	71 (68.23)	178
Total	72	113	115	300

El cálculo de Ji-cuadrada es

$$\chi^2\bigstar = \sum_{\text{Todas las celdas}} \frac{(O-E)^2}{E}: \chi^2\bigstar = \frac{(37-29.28)^2}{29.28} + \frac{(41-45.95)^2}{45.95} + \frac{(44-46.77)^2}{46.77}$$

$$+ \frac{(35-42.72)^2}{42.72} + \frac{(72-67.05)^2}{67.05} + \frac{(71-68.23)^2}{68.23}$$

$$= 2.035 + 0.533 + 0.164 + 1.395 + 0.365 + 0.112$$

$$= \mathbf{4.604}$$

Paso 4 **Distribución de probabilidad:**

Valor *p*:

a. Use la cola derecha porque los valores "mayores" de Ji-cuadrada están en desacuerdo con la hipótesis nula: $\mathbf{P = P\,(\chi^2\bigstar > 4.604 \mid df = 2)}$ como se muestra en la figura.

O

Clásico:

a. La región crítica es la cola derecha porque los valores "mayores" de Ji-cuadrada están en desacuerdo con la hipótesis nula. El valor crítico se obtiene de la tabla 8, en la intersección de la fila df = 2 y la columna $\alpha = 0.05$:

$$\chi^2(2,\ 0.05) = \mathbf{5.99}$$

Para hallar el valor p, tiene dos opciones:

1. Usar la tabla 8 (apéndice B) para poner límites al valor p: **0.10<P<0.25.**
2. Usar computadora o calculadora para hallar el valor p: **P = 0.1001.**

Para instrucciones específicas, vea la página 518.

b. El valor p no es menor que α.

Para instrucciones específicas, vea la página 518.

b. χ^2★ no está en la región crítica, como se muestra en el **área sombreada** en la figura.

Paso 5

a. Decisión: No rechazar H_o.

b. Conclusión: Al nivel de significación de 0.05, la evidencia nos permite rechazar independencia entre el género de un estudiante y la materia académica preferida del estudiante.

En general, la **tabla de contingencia $r \times c$** (r es el número de **filas;** c es el número de **columnas**) se usa para probar la independencia del factor de fila y el factor de columna. El número de **grados de libertad** está determinado por

Grados de libertad para tablas de contingencia

$$\text{df} = (r - 1) \cdot (c - 1) \tag{11.4}$$

donde r y c son ambos mayores a 1

(Este valor para df debe estar de acuerdo con el número de celdas contadas según la descripción general que aparece en las páginas 634-635.)

Las **frecuencias esperadas** para una tabla de contingencia $r \times c$ se encuentran por medio de las fórmulas dadas en cada celda en la tabla 11.8, donde n = gran total. En general, la frecuencia esperada en la intersección de la i-ésima fila y la j-ésima columna está dada por

Frecuencias esperadas para tablas de contingencia

$$E_{i,j} = \frac{\textit{total fila} \times \textit{total columna}}{\textit{gran total}} = \frac{R_i \times C_j}{n} \tag{11.5}$$

TABLA 11.8

Frecuencias esperadas para una tabla de contingencia $r \times c$

	Columna						
Fila	1	2	...	j-ésima columna	...	c	Total
1	$\frac{R_1 \times C_1}{n}$	$\frac{R_1 \times C_2}{n}$	...	$\frac{R_1 \times C_j}{n}$	...	$\frac{R_1 \times C_c}{n}$	R_1
2	$\frac{R_2 \times C_1}{n}$						R_2
⋮	⋮			⋮			⋮
i-ésima fila	$\frac{R_i \times C_1}{n}$			$\frac{R_i \times C_j}{n}$	...		R_i
⋮	⋮			⋮			⋮
r	$\frac{R_r \times C_1}{n}$						
Total	C_1	C_2	...	C_j	...	...	n

¿SABÍA USTED...?

¿Venus o Marte?

a. Exactamente 94 232 más niños que niñas nacieron en Estados Unidos durante 2004.

b. Desde 1940, han nacido un promedio de 91 685 más niños que niñas cada año, un total de 5 776 130 en un periodo de 63 años.

c. En 2003, había un total de 144 513 361 mujeres de todas las edades, en comparación con 138 396 524 hombres.

¿Cómo puede ser esto posible?

De nuevo debemos observar el principio antes mencionado: cada $E_{i,j}$ debe ser al menos 5.

Nota: la notación empleada en la tabla 11.8 y la fórmula (11.5) pueden ser no muy conocidas al estudiante. Para comodidad al referirnos a las celdas o entradas en una tabla, usamos $E_{i,j}$ para denotar la entrada en la i-ésima fila y la j-ésima columna. Es decir, la primera letra del subíndice corresponde al número de fila y la segunda letra corresponde al número de columna. Entonces, $E_{1,2}$ es la entrada en la primera fila, segunda columna, y $E_{2,1}$ es la entrada de la segunda fila, primera columna. En la tabla 11.6 (p. 635), $E_{1,2}$ es 45.95 y $E_{2,1}$ es 42.72. La notación empleada en la tabla 11.8 se interpreta de un modo semejante, es decir, R_1 corresponde al total de la fila 1, y C_1 corresponde al total de la columna 1.

Prueba de homogeneidad

El segundo tipo de problema de tabla de contingencia se denomina *prueba de homogeneidad*. Esta prueba se usa cuando una de las dos variables es controlada por el experimentador, de modo que los totales de fila o columna sean predeterminados.

Por ejemplo, suponga que deseamos obtener los electores registrados en torno a una pieza de legislación propuesta por el gobernador. En la votación, 200 residentes urbanos, 200 suburbanos y 100 rurales se seleccionan al azar y se les pregunta si están a favor o en contra de la propuesta del gobernador. Esto es, se toma una muestra aleatoria simple por cada uno de estos tres grupos. Se obtiene un total de 500 electores, pero observe que se ha predeterminado (antes de tomar la muestra) precisamente cuántos caen dentro de cada categoría de fila, como se ve en la tabla 11.9, y cada categoría se muestrea por separado.

TABLA 11.9

Votación registrada de elector con totales predeterminados de fila

	Propuesta del gobernador		
Residencia	Favor	En contra	Total
Urbano			200
Suburbano			200
Rural			100
Total			500

En una prueba de esta naturaleza, en realidad estamos probando la hipótesis: la distribución de proporciones dentro de las filas es igual para todas las filas. Esto es, la distribución de proporciones en la fila 1 es igual que en la fila 2, igual que en la fila 3, y así sucesivamente. La alternativa es: la distribución de proporciones dentro de las filas no es igual para todas las filas. Este tipo de ejemplo puede ser considerado como comparación de varios experimentos multinomiales.

Fuera de esta diferencia conceptual, la prueba real en cuanto a independencia y homogeneidad con tablas de contingencia es la misma. Demostremos esta **prueba de hipótesis** al completar la ilustración de la votación.

EJEMPLO 11.6 Prueba de hipótesis de homogeneidad

A cada una de las personas de una muestra aleatoria de 500 electores registrados (200 residentes urbanos, 200 suburbanos y 100 rurales) se le pidió su opinión acerca de la legislación propuesta por el gobernador. ¿La evidencia muestral de la tabla 11.10 apoya la hipótesis de que "Los votantes dentro de los diferentes grupos de residencia tienen diferentes opiniones acerca de la propuesta del gobernador"? Use $\alpha = 0.05$.

TABLA 11.10

Resultados muestrales por residencia y opinión

	Propuesta del gobernador		
Residencia	Favor	En contra	Total
Urbano	143	57	200
Suburbano	98	102	200
Rural	13	87	100
Total	254	246	500

SOLUCIÓN

Paso 1

a. **Parámetro de interés:** La proporción de electores a favor o en contra (es decir, la proporción de electores urbanos a favor, la proporción de electores suburbanos a favor, la proporción de electores rurales a favor y la proporción de los tres grupos, por separado, en contra).

b. **Enunciado de las hipótesis:**

H_o: La proporción de electores a favor de la legislación propuesta es igual en los tres grupos de residencia.

H_a: La proporción de electores a favor de la legislación propuesta no es igual en los tres grupos de residencia. (Esto es, en al menos un grupo la proporción es diferente de los otros.)

Paso 2

a. **Suposiciones**: La información muestral se obtiene usando tres muestras aleatorias tomadas de tres poblaciones separadas en las que cada persona se clasifica según su opinión.

b. **Estadística de prueba:** La distribución *Ji*-cuadrada y la fórmula (11.1), con df = $(r–1)(c–1) = (3–1)(2–1) = 2$

c. **Nivel de significación:** $\alpha = 0.05$.

Paso 3

a. **Información muestral:** Vea la tabla 11.10.

b. **Estadística de prueba calculada:** Los valores esperados se encuentran usando la fórmula (11.5) (p. 637) y se dan en la tabla 11.11.

TABLA 11.11

Resultados muestrales y valores esperados

Residencia	Propuesta del gobernador: Favor	Propuesta del gobernador: En contra	Total
Urbano	143 (101.6)	57 (98.4)	200
Suburbano	98 (101.6)	102 (98.4)	200
Rural	13 (50.8)	87 (49.2)	100
Total	254	246	500

Nota: cada uno de los valores esperados se usa dos veces en el cálculo de $\chi^2\bigstar$; por tanto, es buena idea conservar lugares decimales adicionales al hacer los cálculos.

La *Ji*-cuadrada calculada es

$$\chi^2\bigstar = \sum_{\text{todas las celdas}} \frac{(O-E)^2}{E}: \chi^2\bigstar = \frac{(143-101.6)^2}{101.6} + \frac{(57-98.4)^2}{98.4} + \frac{(98-101.6)^2}{101.6}$$

$$+ \frac{(102-98.4)^2}{98.4} + \frac{(13-50.8)^2}{50.8} + \frac{(87-49.2)^2}{49.2}$$

$$= 16.87 + 17.42 + 0.13 + 0.13 + 28.13 + 29.04$$

$$= \mathbf{91.72}$$

Paso 4 **Distribución de probabilidad**

Valor *p*:

a. Use la cola derecha porque los valores "mayores" de *Ji*-cuadrada están en desacuerdo con la hipótesis nula:

P = *P*($\chi^2\bigstar$ > 2.2|df = 5) como se ve en la figura.

Para hallar el valor *p*, tiene dos opciones:

1. Usar la tabla 8 (apéndice B) para poner límite al valor *p*: **P < 0.005.**
2. Usar computadora o calculadora para hallar el valor *p*: **P = 0.000+.**

Para instrucciones específicas, vea la página 518.

b. El valor *p* es menor que α.

O

Clásico:

a. La región crítica es la cola derecha porque los valores "mayores" de *Ji*-cuadrada están en desacuerdo con la hipótesis nula. El valor crítico se obtiene de la tabla 8, en la intersección de la fila df = 2 y la columna α = 0.05:

χ^2 (2, 0.05) = **5.99**

Para instrucciones específicas, vea la página 518.

b. $\chi^2\bigstar$ está en la región crítica, como se muestra en **rojo** en la figura.

Paso 5

a. Decisión: Rechazar H_o.

b. Conclusión: Los tres grupos de electores no tienen todos proporciones iguales a favor de la legislación propuesta.

INSTRUCCIONES DE TECNOLOGÍA: PRUEBA DE HIPÓTESIS DE INDEPENDENCIA U HOMOGENEIDAD

MINITAB (Versión 14) Introduzca, en C1, C2, . . . ,cada columna de frecuencias observada desde la tabla de contingencia; luego continúe con:

```
Choose:      Stat > Tables > Chi-Square Test (Table in Worksheet)
Enter:       Columns containing the table: C1 C2 > OK
```

```
COMPUTER SOLUTION MINITAB Printout for Example 11.6:
Chi-square Test: C1, C2
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts
                C1      C2      Total
             1  143     57      200
                101.60  98.40
                16.870  17.418
             2  98      102     200
                101.60  98.40
                0.128   0.132
             3  13      87      100
                50.80   49.20
                28.127  29.041
Total           254     246     500
Chi-Sq = 91.715, DF = 2, P-Value = 0.000
```

Excel Introduzca, en las columnas A, B, . . . ,cada columna de frecuencias observada desde la tabla de contingencia; luego continúe con:

```
Choose:      Tools > Data Analysis Plus > Contingency Table > OK
Enter:       Input range: (A1:B4 or select cells)
Select:      Labels (if necessary)
Enter:       Alpha: α (ex. 0.05)
```

TI-83/84 Plus Introduzca, en una matriz A de $r \times c$, las frecuencias observadas de la tabla de contingencia $r \times c$. Elabore la matriz B como una matriz $r \times c$ vacía para las frecuencias esperadas.

```
Choose:      MATRX > EDIT > 1:[A]
Enter:       r > ENTER > c > ENTER
             Each observed frequency with an ENTER afterward
```

Luego continúe con:

```
Choose:      MATRX > EDIT > 2[B]
Enter:       r > ENTER > c > ENTER
Choose:      STAT > TESTS > C:χ²-Test. . .
Enter:       Observed: [A] or wherever the contingency table is located
             Expected: [B] place for expected frequencies
Highlight:   Calculate > ENTER
```

CASO PRÁCTICO 11.7 Regla de papas horneadas para occidentales

La gráfica "Regla de papas horneadas para occidentales" informa del porcentaje de norteamericanos que prefieren comer papas horneadas por región, así como para todo el país. Si se da el número real de personas en cada categoría, tendríamos una tabla de contingencia y podríamos completar una prueba de hipótesis acerca de la homogeneidad de las cuatro regiones. (Vea ejercicios 11.38 y 11.45.)

REGLA DE PAPAS HORNEADAS PARA OCCIDENTALES

Los norteamericanos comen papas un promedio de tres veces a la semana y 47% las prefieren "horneadas" y no en puré (23%) o a la francesa (16%). Las prefirieron horneadas por región:

Datos de Anne R. Carey y Sam Ward, © 1998 USA Today.

SECCIÓN 11.4 EJERCICIOS

11.27 Exprese la hipótesis nula, H_o, y la hipótesis alternativa, H_a, que se usarían para probar los siguientes enunciados:

a. Los electores expresaron preferencias que no eran independientes de sus afiliaciones partidistas.

b. La distribución de opiniones es igual para las tres comunidades.

c. La proporción de respuestas afirmativas fue igual para todas las categorías encuestadas.

11.28 La "prueba de independencia" y la "prueba de homogeneidad" se completan en la misma forma, usando la tabla de contingencia para presentar y organizar los cálculos. Explique la forma en que difieren estas pruebas dos hipótesis.

11.29 Encuentre el valor esperado para la celda mostrada.

□ ...	50
⋮	
40	200

11.30 Identifique estos valores de la tabla 11.7:

a. C_2 b. R_1 c. n d. $E_{2,3}$

11.31 Se utilizó el MINITAB para completar una prueba de independencia de *Ji*-cuadrada entre el número de muertes de manatíes causadas por botes de recreo y dos condados de la Florida:

Condado	Muertes por botes	Muertes no por botes	Total muertes
Condado de Lee	23	25	48
Condado Colliers	8	23	31

```
Chi-Square Test: Boat-Related Deaths, Non-boat-Related
Deaths
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected
counts
        Boat-Related        Non-boat-Related
        Deaths              Deaths              Total
1       23                  25                  48
        18.84               29.16
        0.921               0.595

2       8                   23                  31
        12.16               18.84
        1.426               0.921
Total   31                  48                  79

Chi-Sq = 3.862, DF = 1, P-Value = 0.049
```

a. Verifique los resultados (valores esperados y la χ^2★ calculada) al calcular los valores usted mismo.

b. Use la tabla 8 para verificar el valor p con base en el df calculado.

c. La proporción de muertes relacionadas con botes ¿es independiente del condado? Use $\alpha = 0.05$.

11.32 Los resultados del uso del cinturón de seguridad, tomados del estudio de conducta de riesgo de la juventud en 2003, se publicaron en una USA Snapshot el 13 de enero de 2005. La siguiente tabla compendia los resultados de estudiantes de preparatoria encuestados en el estado de Nebraska. Se les preguntó si raras veces o nunca usaban cinturones de seguridad cuando viajaban en el auto de alguien.

	Mujeres	Hombres
Raras veces o nunca	208	324
Usa el cinturón de seguridad	1217	1184

Fuente: http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5302a1.htm#top

Usando $\alpha = 0.05$, ¿esta muestra presenta suficiente evidencia para rechazar la hipótesis de que el género es independiente del uso del cinturón de seguridad?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.33 Se tomó una muestra aleatoria de 500 hombres casados; cada persona fue clasificada en cruz en cuanto al tamaño de la comunidad en la que actualmente vive y el tamaño de la comunidad en la que se crió. Los resultados se presentan en la tabla siguiente.

	Tamaño de comunidad en la que reside			
Tamaño de comunidad en que se crió	Menos de 10 000	10 000 a 49 999	50 000 o más	Total
Menos de 10 000	24	45	45	114
10 000 a 49 999	18	64	70	152
50 000 o más	21	54	159	234
Total	63	163	274	500

¿Esta muestra contradice lo dicho de independencia, al nivel de significación de 0.01?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.34 Una encuesta de viajeros seleccionados al azar que entraban al sanitario de grandes gasolineras en Estados Unidos mostró los siguientes resultados:

	Calidad de instalaciones en sanitario			
Género de quien responde	Arriba de promedio	Promedio	Debajo de promedio	Totales
Mujeres	7	24	28	59
Hombres	8	26	7	41
Totales	15	50	35	100

Usando $\alpha = 0.05$, ¿esta muestra presenta suficiente evidencia para rechazar la hipótesis de que "La calidad de respuestas es independiente del género de quien responde"?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.35 Un sondeo de empleados de una compañía aseguradora se ocupaba de las relaciones entre trabajadores y supervisores. Una frase de evaluación era "No estoy seguro de lo que mi supervisor espera". Los resultados del sondeo se presentan en la siguiente tabla de contingencia.

	No estoy seguro de lo que mi supervisor espera		
Años de empleo	Verdadero	No es verdadero	Totales
Menos de 1 año	18	13	31
De 1 a 3 años	20	8	28
De 3 a 10 años	28	9	37
10 años o más	26	8	34
Totales	92	38	130

¿Podemos rechazar la hipótesis de que "Las respuestas a la frase y los años de empleo son independientes" al nivel de significación de 0.10?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.36 La tabla siguiente es de la publicación *Vital and Health Statistics* de los centros para control y prevención de enfermedades del Centro Nacional para Estadísticas de la Salud. Las personas de la tabla siguiente tienen una irritación en los ojos, una irritación en la nariz, o una irritación en la garganta. Tienen sólo una de las tres.

	Edad (años)			
Tipo de irritación	18–29	30–44	45–64	65 o mayor
Ojos	440	567	349	59
Nariz	924	1311	794	102
Garganta	253	311	157	19

¿Hay evidencia suficiente para rechazar la hipótesis de que el tipo de irritación en el oído, nariz o garganta es independiente del grupo de edades al nivel de significación igual a 0.05?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.37 La gerente de un proceso de ensamble desea determinar si el número de artículos defectuosos manufacturados depende o no del día de la semana en que se producen esos artículos. Ella recolectó la siguiente información.

Día de la semana	Lun	Mar	Miér	Jue	Vier
No defectuoso	85	90	95	95	90
Defectuoso	15	10	5	5	10

¿Hay suficiente evidencia para rechazar la hipótesis de que el número de artículos defectuosos es independiente del día de la semana en que se producen? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.38 Con referencia al caso práctico 11.7 (p. 642):

a. Exprese el porcentaje de norteamericanos que las "prefieren horneadas" a "otras" por región, como tabla de contingencia de 2×4.

b. Explique por qué la pregunta siguiente podría probarse usando la estadística *Ji*-cuadrada: "¿La preferencia por papas horneadas es igual en las cuatro regiones de Estados Unidos?"

c. Explique por qué es ésta una prueba de homogeneidad.

11.39 "Blogging" es un tema controvertido hoy día. Un "blog" es un registro en Internet. Los "blogs" son creados para uso personal o profesional. Según el sitio web Xtreme Recruiting (http.//www.xtremerecruiting.org/blog/archives/news/cat_news.html), hay un nuevo blog cada 7 segundos, y muy pocas personas leen estos blogs. La tabla que sigue muestra el número de nuevos lectores de blog por cada uno de los meses indicados. ¿Es igual la distribución de creadores y de lectores de blogs para los meses citados? Use $= 0.05$.

	Creadores de blog	Lectores de blog
Marzo 2003	74	205
Febrero 2004	93	316
Noviembre 2004	130	502

Fuente: USA Today, "Warning: Your clever little blog could get you fired", 15 de junio, 2005.

11.40 Los estudiantes emplean diversas clases de criterios cuando seleccionan cursos. "El maestro es calificador fácil" es con frecuencia un criterio. Tres maestros están programados para impartir estadística el próximo semestre. A continuación se presenta una muestra de distribuciones previas de calificación para estos tres maestros.

	Profesor		
Calificaciones	#1	#2	#3
A	12	11	27
B	16	29	25
C	35	30	15
Otras	27	40	23

Al nivel de significación de 0.01, ¿hay suficiente evidencia para concluir que "La distribución de calificaciones no es igual para los tres profesores"?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

c. ¿Cuál profesor es el calificador más fácil? Explique, citando evidencia específica de apoyo.

11.41 El miedo a la oscuridad es una emoción común. Se obtuvieron los siguientes datos al preguntar a 200 personas de cada grupo de edades si tenían mucho miedo a la oscuridad. Con $\alpha = 0.01$, ¿tenemos suficiente evidencia para rechazar la hipótesis de que "la misma proporción de cada grupo de edades tiene serios temores a la oscuridad"? (*Sugerencia:* la tabla de contingencia debe tomar en cuenta a las 1000 personas.)

Grupo de edad	Elemental	Menor	Mayor	Universidad	Adultos
Núm. que teme a oscuridad	83	72	49	36	114

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.42 El 21 de mayo de 2004, el Centro Nacional de Prevención de Enfermedades Crónicas y Promoción de Salud, y los Centros de Control y Prevención de Enfermedades, publicaron los resultados del sondeo de conductas de riesgo de la juventud, Estados Unidos, 2003. El informe dividió la muestra de 15,184 adolescentes norteamericanos en niveles de grados como se indica en la tabla siguiente. Los estudiantes reconocieron portar un arma en cualquiera de los últimos 30 días antes del sondeo, y estar en una pelea física durante el año pasado. La tabla siguiente resume dos partes de los resultados.

	Al menos una vez	Nunca	Total
Portaban arma			
Grados 9 y 10	1 436	7008	8 444
Grados 11 y 12	1 140	5 600	6 740
Total	2 576	12 608	15 184
En una pelea física			
Grados 9 y 10	3 057	5 387	8 444
Grados 11 y 12	1 942	4 798	6 740
Total	4 999	10 185	15 184

Fuente: Datos de http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5302a1.htm#top

¿La evidencia muestral indica que los estudiantes de grados 9 al 12 tienen diferente tendencia de portar armas en la escuela?, ¿de meterse en una pelea física? Use el nivel de significación 0.01 en cada caso.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.43 Todos los medicamentos nuevos deben pasar por un estudio antes de ser aprobados por la U.S. Food and Drug Administration (FDA). Un estudio de medicamentos por lo general incluye pruebas clínicas, donde los participantes se clasifican al azar para recibir dosis diferentes y placebo, pero no se les informa del grupo en el que se les registra. Para controlar tantos factores como sea posible, es mejor asignar participantes al azar pero de manera homogénea por todos los tratamientos. Considere la siguiente distribución de homogeneidad respecto al género y dosis.

Género	Droga de 10 mg	Droga de 20 mg	Placebo
Femenino	54	56	60
Masculino	32	27	26

Al nivel de significación de 0.01, ¿la distribución de medicamento es igual para ambos géneros?

Considerando el mismo estudio, la homogeneidad de edades también sería una característica importante. Al nivel de significación de 0.01, ¿la distribución de medicamento que sigue es igual para todos los grupos de edades?

Edad	Droga de 10 mg	Droga de 20 mg	Placebo
40–49	18	20	19
50–59	48	41	57
60–69	20	22	10

11.44 Se fabricaron 41 pequeños lotes de producto experimental, y se probó la existencia de una indicación particular que se atribuya a la naturaleza pero que causa el rechazo de la pieza. Se hicieron diez lotes usando un método particular de procesamiento, y se hicieron 31 lotes usando un segundo método de procesamiento. Se hizo muestreo por igual en cada lote ($n = 32$) en busca de esta indicación. En la práctica, las condiciones óptimas de procesamiento muestran poca o ninguna existencia de la indicación. El método 1, que comprende los diez lotes, se ejecutó antes que el método 2.

Fuente: Cortesía de Bausch & Lomb

Métodos	n	Número de rechazos
Método 1	320	4
Método 2	992	26

a. Determine, al nivel de significación de 0.05, si hay diferencia en la proporción de producto rechazado entre los dos métodos.

b. Compare lo que descubra en la parte a con lo que descubra en el ejercicio 10.101 (p. 591). Incluya todas las piezas de las pruebas de hipótesis en su comparación.

c. ¿Diría el lector que estos dos procedimientos diferentes de prueba son equivalentes? Dé evidencia específica que apoye su respuesta.

11.45 El caso práctico 11.7 (p. 642) presenta porcentajes que describen las preferencias de las personas respecto a cómo se elaboran las papas. ¿Cree usted que hay una diferencia significativa entre las cuatro regiones de Estados Unidos, respecto al porcentaje que prefieren papas horneadas? Nótese que el artículo no menciona el tamaño de la muestra.

a. Suponga que los porcentajes reportados estuvieron basados en cuatro muestras de tamaño 100 de cada región y calcule χ^2★ y su valor p.

b. Repita la parte a usando tamaños muestrales de 200 y 300.

c. ¿Son significativamente diferentes los cuatro porcentajes, indicados en la gráfica, quienes prefieren papas horneadas?

REPASO DEL CAPÍTULO

En retrospectiva

En este capítulo nos hemos ocupado de pruebas de hipótesis usando *Ji*-cuadrada, con las probabilidades de celda asociadas con el experimento multinomial, y con la tabla de contingencia simple. En cada caso, las suposiciones básicas son que se ha hecho un gran número de observaciones y que la estadística de prueba resultante, $\sum \frac{(O-E)^2}{E}$, está aproximadamente distribuida como *Ji*-cuadrada. En general, si n es grande y el tamaño mínimo permisible esperado de celda es 5, entonces esta suposición queda satisfecha.

La tabla de contingencia se puede usar para probar independencia y homogeneidad. La prueba de homogeneidad y la prueba de independencia parecen muy semejantes y, de hecho, se efectúan en exactamente la misma forma. Al probar los conceptos, no obstante, algunas distribuciones e independencia son bastante diferentes. Las dos pruebas se distinguen con facilidad porque la prueba de homogeneidad tiene márgenes totales predeterminados en una dirección de la tabla. Esto es, antes de recolectar los datos, el experimentador determina cuántos sujetos se observarán en cada categoría. El único número predeterminado de la prueba de independencia es el gran total.

Unas cuantas palabras de advertencia: el número correcto de grados de libertad es de importancia decisiva si es que los resultados de la prueba han de ser significativos. Los grados de libertad determinan, en parte, la región crítica, y su tamaño es importante. Al igual que en otras pruebas de hipótesis, no rechazar H_o no significa una aceptación directa de la hipótesis nula.

Vocabulario y conceptos clave

celda (p. 620)
columna (p. 637)
datos enumerativos (p. 620)
estadística de prueba (p. 621)
experimento multinomial (pp. 622, 624)
filas (p. 637)
frecuencia esperada (pp. 625, 637)
frecuencia observada (p. 620)
grados de libertad (pp. 625, 634)
homogeneidad (p. 638)
independencia (p. 633)
Ji-cuadrada (p. 621)
prueba de hipótesis (pp. 625, 627, 633, 639)
suposiciones (p. 621)
tabla de contingencia (pp. 633, 637)
tabla de contingencia $r \times c$ (p. 633)
totales marginales (p. 634)

Objetivos de aprendizaje

✓ Entender que los datos enumerativos son datos que se pueden contar y colocar en categorías.	pp. 620-622
✓ Entender que la distribución *Ji*-cuadrada se usará para probar hipótesis que contienen datos enumerativos.	pp. 620-621
✓ Entender que las propiedades de la distribución *Ji*-cuadrada y cómo está basada una serie de distribuciones en el tamaño muestral (usando grados de libertad como índice).	pp. 517,621, Ejer. 11.3, 11.5
✓ Entender los elementos clave de un experimento multinomial y ser capaz de definir n, k, O_i, y P_i.	pp. 622-624
✓ Saber y ser capaz de calcular valores esperados usando $E = np$.	EJ. 11.2, Ejer. 11.13
✓ Saber y ser capaz de calcular una estadística *Ji*-cuadrada: $\chi^2 = \sum_{\text{todas las celdas}} \frac{(O-E)^2}{E}$	EJ. 11.1
✓ Saber y ser capaz de calcular los grados de libertad para un experimento multinomial (df $= k-1$).	EJ. 11.1, 11.2
✓ Efectuar, describir e interpretar una prueba de hipótesis para un experimento multinomial, usando la distribución *Ji*-cuadrada con el procedimiento del valor p y el procedimiento clásico.	Ejer. 11.14, 11.15, 11.17
✓ Entender y saber la definición de independencia de dos eventos.	pp. 635-636
✓ Conocer y ser capaz de calcular valores esperados usando $E_{ij} = \frac{R_i \cdot C_j}{n}$.	pp. 635, 637-638, Ejer. 11.29
✓ Conocer y ser capaz de calcular los grados de libertad para una prueba de independencia u homogeneidad [df $= (r-1)(c-1)$].	p. 637
✓ Efectuar, describir e interpretar una prueba de hipótesis para una prueba de independencia u homogeneidad, usando la distribución *Ji*-cuadrada con el procedimiento del valor p y el procedimiento clásico.	EJ. 11.5, 11.6, Ejer. 11.34, 11.41
✓ Entender las diferencias y similitudes entre pruebas de independencia y pruebas de homogeneidad.	p. 646

Ejercicios del capítulo

11.46 El departamento de psicología de cierta universidad dice que las calificaciones en su curso de introducción están distribuidas como sigue: 10% A, 20% B, 40% C, 20% D y 10% F. En una encuesta de 200 estudiantes seleccionados al azar que habían terminado este curso, se encontró que 16 habían obtenido A; 43, B; 65, C; 48, D; y 28, F. ¿Esta muestra contradice lo dicho por el departamento al nivel de 0.05?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.47 Cuando se cruzan dos variedades de rosas, esperamos que la híbrida aparezca en tres clases genéticas con una proporción 1:3:4. Si los resultados de un experimento dan 80 híbridas del primer tipo, 340 del segundo tipo, y 380 del tercer tipo, ¿tenemos suficien-

te evidencia para rechazar la proporción genética hipotética al nivel de significación de 0.05?

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.48 Se prueba una muestra de 200 personas en cuanto a su tipo de sangre, y los resultados se usan para probar la distribución hipotética de tipos de sangre:

Tipo de sangre	A	B	O	AB
Porcentaje	0.41	0.09	0.46	0.04

Los resultados observados fueron como sigue:

Tipo de sangre	A	B	O	AB
Número	75	20	95	10

Al nivel de significación de 0.05, ¿hay suficiente evidencia para demostrar que la distribución expresada es incorrecta?

11.49 Como se publicó en *USA Today,* alrededor de 8.9 millones de familias enviaron estudiantes a la universidad este año, y más de la mitad viven lejos de casa. He aquí donde viven los estudiantes:

Casa de sus padres o un guardia	46%
Vivienda del campus	26%
Renta fuera de la universidad	18%
Poseen casa fuera del campus	9%
Otros arreglos	2%

Nota: rebasa 100% por error de redondeo

Una muestra aleatoria de 1000 estudiantes universitarios resultó en la siguiente información:

Casa de sus padres o un guardia	484
Vivienda del campus	230
Renta fuera de la universidad	168
Poseen casa fuera del campus	96
Otros arreglos	22

La distribución de esta muestra ¿es significativamente diferente de la distribución publicada en el periódico? Use $\alpha = 0.05$. (Para ajustar el error de redondeo, reste 2 de cada frecuencia esperada.)

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.50 ¿Con qué frecuencia revisa su talonario de pagos para comprobar que se retengan los impuestos correctos? *USA Today* informó que los norteamericanos hacen esta revisión como sigue:

Siempre	53%
Casi siempre	12%
Ocasionalmente	14%
Nunca	10%
No recibe cheques	10%
No está seguro	1%

Una muestra aleatoria de 650 trabajadores resultó en las siguientes frecuencias:

Siempre	342
Casi siempre	94
Ocasionalmente	68
Nunca	73
No recibe cheques	60
No está seguro	13

La distribución de esta muestra ¿es significativamente diferente de la distribución publicada en el periódico? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.51 Es probable que la mayoría de los golfistas se alegren de jugar 18 hoyos de golf siempre que tengan oportunidad de jugar. Ben Winter, golfista profesional, jugó 306 hoyos en 1 día en un maratón golfista de caridad en Stevens, Pennsylvania. Una encuesta nacional realizada por la revista *Golf* por Internet reveló la siguiente distribución de frecuencia del mayor número de hoyos jamás jugados por quienes respondieron en 1 día:

Más hoyos jugados en 1 día	Porcentaje	Más hoyos jugados en 1 día	Porcentaje
18	5	37 a 45	20
19 a 27	12	46 a 54	18
28 a 36	28	55 o más	17

Fuente: Golf, "18 Is Not Enough"

Suponga que uno de los campos de golf públicos de su localidad pide, a los siguientes 200 golfistas que dan el primer golpe, que contesten la misma pregunta. La tabla siguiente resume sus respuestas:

Más hoyos jugados en 1 día	Número	Más hoyos jugados en 1 día	Número
18	12	37 a 45	44
19 a 27	35	46 a 54	35
28 a 36	60	55 o más	14

La distribución del número mayor de hoyos jugados por "golfistas maratonistas" en su campo de golf ¿difiere de la distribución compilada por la revista *Golf*, usando respuestas obtenidas en Internet? Pruebe al nivel de significación de 0.01.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.52 El censo de 2003 de Estados Unidos encontró que entraban bebés a este mundo, durante los diversos meses, en las proporciones que siguen:

Mes	P(Mes)	Mes	P(Mes)	Mes	P(Mes)
Enero	0.082	Mayo	0.084	Septiembre	0.087
Febrero	0.076	Junio	0.081	Octubre	0.086
Marzo	0.082	Julio	0.089	Noviembre	0.079
Abril	0.081	Agosto	0.089	Diciembre	0.083

Fuente: U.S. Census Bureau

Una fuente aleatoria seleccionada de actas de nacimiento para una gran zona metropolitana resultó en los datos siguientes:

Mes	Ene	Feb	Mar	Abr	May	Jun	Jul	Agos	Sep	Oct	Nov	Dic
Observado	14	12	12	10	16	9	16	11	17	7	17	9

a. ¿Esta información da suficiente evidencia para rechazar la frase de que "Los nacimientos ocurren en esta zona metropolitana en las mismas proporciones mensuales", como informa la U.S. Census Bureau? $\alpha = 0.05$.

b. ¿Esta información da suficiente evidencia para rechazar la frase de que "Los nacimientos ocurren en esta zona metropolitana en todos los meses con la misma probabilidad"? Use $\alpha = 0.05$.

c. Compare los resultados obtenidos en las partes a y b. Exprese sus conclusiones.

11.53 Se determinaron y usaron los pesos (x) de 300 hombres adultos, para probar la hipótesis de que los pesos estaban normalmente distribuidos con una media de 160 libras y con una desviación estándar de 15 libras. Los datos se agruparon en las clases siguientes.

Peso (x)	Frecuencia observada	Peso (x)	Frecuencia observada
$x < 130$	7	$160 \leq x < 175$	102
$130 \leq x < 145$	38	$175 \leq x < 190$	40
$145 \leq x < 160$	100	190 y más	13

Usando las tablas normales, los porcentajes para las clases son 2.28%, 13.59%, 34.13%, 34.13%, 13.59%, y 2.28%, respectivamente. Los datos observados ¿muestran una razón importante para dudar de la hipótesis de que los pesos están normalmente distribuidos con una media de 160 libras, y una desviación estándar de 15 libras? Use $\alpha = 0.05$.

a. Verifique los porcentajes para las clases.

b. Resuelva usando el procedimiento del valor p.

c. Resuelva usando el procedimiento clásico.

11.54 ¿Tiene usted un "alimento cómodo" favorito? ¿Cómo lo obtiene? La gráfica siguiente presenta tres métodos de norteamericanos y el porcentaje en que se usa cada método. ¿A cuál categoría pertenece usted? A una muestra aleatoria de 120 norteamericanos que viven en la costa del Atlántico se les preguntó "¿Cómo obtiene su alimento cómodo favorito"?

OBTENIENDO NUESTRO ALIMENTO CÓMODO FAVORITO

Datos de Justin Dickerson y Suzy Parker, © 2005 USA Today.

Alimento cómodo	Lo compra	Lo hace	Pide a alguien que lo haga	No sabe
Costa Atlántica	57	44	12	7

Usando los porcentajes dados en la gráfica como el "estándar" nacional, ¿la evidencia indica que las respuestas de la costa del Atlántico son diferentes de las de toda la nación? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

11.55 A una muestra aleatoria de 120 norteamericanos que viven en el medio oeste se les preguntó "¿Cómo obtiene su alimento cómodo favorito"?

Alimento cómodo	Lo compra	Lo hace	Pide a alguien que lo haga	No sabe
Medio oeste	54	48	17	1

Usando los porcentajes dados en la gráfica (p. 649) como el "estándar" nacional, ¿la evidencia indica que las respuestas del medio oeste son diferentes de las de toda la nación? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

c. Explique las similitudes y diferencias halladas en los ejercicios 11.54 y 11.55.

11.56 A una muestra aleatoria de 120 norteamericanos que viven en la costa del Pacífico se les preguntó "¿Cómo obtiene su alimento cómodo favorito?"

Alimento cómodo	Lo compra	Lo hace	Pide a alguien que lo haga	No sabe
Costa del Pacífico	73	26	18	3

Usando los porcentajes dados en la gráfica (p. 649) como el "estándar" nacional, ¿la evidencia indica que las respuestas de la costa del Pacífico son diferentes de las de toda la nación? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

c. Explique las similitudes y diferencias halladas en los ejercicios 11.54, 11.55 y 11.56.

11.57 La tabla siguiente da la cantidad de colores para una muestra de 30 bolsas (tamaño de 47.9 gramos) de M&M's.

Caso	Rojo	Verde	Azul	Anaranjado	Amarillo	Café
1	15	9	3	3	9	19
2	9	17	19	3	3	8

Los datos para este ejercicio están en http://latinoamerica.cengage.com/johnson

Fuente: http://www.math.uah.edu/stat/,Christine Nickel and Jason York, ST 687 project, Fall 1988

Antes del voto mundial del color (GCV) de 2002, el porcentaje objetivo para cada color en la mezcla de seis colores era como sigue: café, 30%; rojo y amarillo, 20%; azul, verde y anaranjado, 10%.

a. ¿El caso 1 muestra que la bolsa 1 tiene una distribución significativamente diferente de colores que la distribución objetivo? Use $\alpha = 0.05$.

b. Combine los casos 1 y 2. ¿El total de las bolsas 1 y 2 muestra una distribución significativamente diferente de colores que la distribución objetivo?

c. Combine los resultados de los 30 casos. ¿El total de las 30 bolsas muestra una distribución significativamente diferente de colores que la distribución objetivo?

d. Discuta lo encontrado en las partes a-c.

11.58 En un artículo del 5 de enero de 2005, titulado "Hip-Hop rules radio", *USA Today* dio los siguientes resultados de Nielsen BDS y Arbitron de los principales 100 éxitos de 2003 y 2004.

	R&B/Hip-Hop	Pop	Rock	Country	Total
2003	53	25	8	14	100
2004	61	16	12	11	100

¿Esta información muestral indica que la distribución de los 100 más grandes éxitos de la radio cambió de 2003 a 2004? Use $\alpha = 0.05$.

11.59 Con base en datos de la U.S. Census Bureau, la Asociación Nacional de Constructores de Casas proyecta un aumento en los porcentajes de propiedades para esta década. Parte del proyecto consiste en pronosticar el inicio de nuevas construcciones por región. La tabla siguiente muestra lo que proyectan.

	Promedio de nuevas construcciones		
Region	1996–2000	2001–2005	2006–2010
Noroeste	145	161	170
Sur	710	687	688
Medio oeste	331	314	313
Oeste	382	385	373

¿Los datos presentan suficiente evidencia para rechazar la hipótesis de que la distribución de nuevas construcciones en todas las regiones es igual para todos los años?

a. Resuelva usando el procedimiento del valor *p*.

b. Resuelva usando el procedimiento clásico.

11.60 La tabla siguiente muestra el número de delitos cometidos el año pasado en la parte interior de una gran ciudad. Los delitos se clasificaron de acuerdo con

el tipo de delito y distrito de la ciudad interior donde ocurrieron. ¿Estos datos muestran suficiente evidencia para rechazar la hipótesis de que el tipo de delito y el distrito en el que ocurrió son independientes? Use $\alpha = 0.01$.

	Delito				
Distrito	Robo a mano armada	Asalto	Robo con allanamiento	Latrocinio	Robo de vehículo
1	54	331	227	1090	41
2	42	274	220	488	71
3	50	306	206	422	83
4	48	184	148	480	42
5	31	102	94	596	56
6	10	53	92	236	45

a. Resuelva usando el enfoque o método p.

b. Resuelva usando el método o enfoque clásico.

11.61 Con base en los resultados de un cuestionario de sondeo, 400 personas fueron clasificadas como conservadoras, moderadas, o liberales, todo desde el punto de vista político. Además, cada persona fue clasificada por edad, como se ve en la tabla siguiente.

	Grupo de edad			
	20–35	36–50	Mayor a 50	Totales
Conservador	20	40	20	80
Moderado	80	85	45	210
Liberal	40	25	45	110
Totales	140	150	110	400

¿Hay suficiente evidencia para rechazar la hipótesis de que "la preferencia política es independiente de la edad"? Use $\alpha = 0.01$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.62 "Espacio muy reducido" es queja común de pasajeros de líneas aéreas. A una muestra aleatoria de 150 pasajeros de negocios y 150 pasajeros de placer se les preguntó dónde les gustaría "más espacio". Las respuestas resultantes se resumen como sigue.

Lugar	Negocios	Placer
Espacio para equipaje de mano	15	9
Cuarto de hotel	29	49
Espacio para las piernas en el avión	91	66
Tamaño de auto de renta	10	20
Otro	5	6

Esta información muestral ¿presenta suficiente evidencia para concluir que el pasajero de negocios y el pasajero de placer difieren en cuanto al lugar en donde más les gustaría espacio adicional? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.63 Se probaron cuatro marcas de palomitas de maíz en cuanto al tiempo para estar listas. Se hicieron abrir 100 granos de cada marca, y el número de granos que no abrió se registró en cada prueba (vea la tabla siguiente). ¿Podemos rechazar la hipótesis nula de que las cuatro marcas abrieron igualmente? Pruebe a $\alpha = 0.05$.

Marca	A	B	C	D
No se abrieron	14	8	11	15

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.64 Un promedio de dos jugadores por equipo de baloncesto de muchachos y muchachas de preparatoria se lesiona durante una temporada. La tabla siguiente muestra la distribución de lesiones, para una muestra aleatoria de 1000 muchachas y 1000 muchachos tomados de los registros de temporada, de todas las lesiones registradas.

Lesión	Muchachas	Muchachos
Tobillo/pie	360	383
Cadera/muslo/pierna	166	147
Rodilla	130	103
Antebrazo/muñeca/mano	112	115
Rostro/cuero cabelludo	88	122
Todas las otras	144	130

Esta información muestral ¿presenta suficiente evidencia para concluir que la distribución de lesiones es diferente para muchachas que para muchachos? Use $\alpha = 0.05$.

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.65 "¿Ha usted designado en su licencia de manejo que es donador de órganos?" Todos hemos escuchado esta pregunta, y la respuesta debe salir de acuerdo con el artículo del 30 de marzo de 2005 "Trasplantes de Órganos Alcanzan Nuevo Récord de Casi 27 000 en 2004". Los resultados exactos para tipos de donaciones de órganos son como sigue.

	De donador fallecido	De donador en vida
2003	18,650	6,812
2004	20,018	6,966

Fuente: http://www.seniorjournal.com/NEWS/Features/5-03-30DonateLife.htm

a. ¿Qué porcentaje de donadores de órganos era de fallecidos para cada año? ¿Ve usted estos porcentajes como significativamente diferentes? Explique.

b. Al nivel de significación de 0.05, ¿cambiaron significativamente los porcentajes de donadores fallecidos respecto a los donadores en vida, entre 2003 y 2004?

c. Compare la decisión tomada en la parte b con su respuesta a la parte a. Describa cualesquiera diferencias y explique qué las ocasionó.

11.66 El registro de trabajo del año pasado por ausentismo, en cada una de cuatro categorías para 100 empleados seleccionados al azar, se compila en la tabla siguiente. ¿Estos datos dan suficiente evidencia para rechazar la hipótesis de que el porcentaje de ausentismo es igual para todas las categorías de empleados? Use $\alpha = 0.01$ y 240 días de trabajo para el año.

	Hombre casado	Hombre soltero	Mujer casada	Mujer soltera
Número de empleados	40	14	16	30
Días ausentismo	180	110	75	135

a. Resuelva usando el procedimiento del valor p.

b. Resuelva usando el procedimiento clásico.

11.67 Si usted fuera a tirar 600 veces un dado, ¿qué tan diferente de 100 podrían ser las frecuencias observadas para cada cara, antes que los resultados se hagan muy diferentes de los igualmente probables al nivel de 0.05?

11.68 Considere el siguiente conjunto de datos.

	Respuesta		
	Sí	No	Total
Grupo 1	75	25	100
Grupo 2	70	30	100
Total	145	55	200

a. Calcule el valor de la estadística de prueba $z\bigstar$ que se usaría para probar la hipótesis nula de que $p_1 = p_2$, donde p_1 y p_2 son las proporciones de respuestas "sí" en los grupos respectivos.

b. Calcule el valor de la estadística de prueba $\chi^2\bigstar$ que se usaría para probar la hipótesis de que "la respuesta es independiente del grupo".

c. Demuestre que $\chi^2\bigstar = (z\bigstar)^2$.

Proyecto del capítulo

Algo dulce para contrarrestar el sabor picante

La gráfica presentada en la sección 11.1, "Algo dulce para contrarrestar el sabor picante" (p. 619), muestra las principales seis formas en que los adultos dicen que contrarrestan el sabor picante después de tomar alimento con salsa picante. ¿Parecen razonables estas variables y porcentajes? Para unas personas puede que sí, pero no para otras. Aquí es donde entra en juego un experimento multinomial. Puede probarse una muestra contra las estadísticas indicadas, para determinar si la muestra proviene de la misma distribución dada o si contradice la distribución dada.

Suponga que 200 adultos, a los que les gusta tomar alimentos con salsas picantes, se les pide digan la forma en que contrarrestan el sabor después de comer con salsa picante. A continuación está el resumen de la muestra resultante.

Método	Agua	Leche	Soda	Cerveza	Pan	Otro	Nada
Número	73	35	20	19	29	11	13

Investiguemos esta muestra respecto a la distribución dada en la gráfica de la página 619.

Trabajando en el contenido del capítulo 11

11.69 a. ¿La muestra presenta una distribución que sea significativamente diferente de la distribución mostrada en la gráfica "Para apagar el fuego" (p. 619)? Use $\alpha = 0.05$.

b. Escriba un párrafo (más de 50 palabras) que describa por qué el método estadístico empleado en la parte a es apropiado para este conjunto de datos.

c. Escriba un párrafo (más de 50 palabras) que describa el significado de las suposiciones y los resultados del procedimiento estadístico escogido en la parte b.

Para su estudio

11.70 Diseñe su propio estudio para "la forma favorita en que las personas contrarrestan el sabor picante después de comer algo muy condimentado".

a. Defina una población específica que muestreará usted, describa su plan de muestreo y recolecte una muestra aleatoria de por lo menos 100 observaciones.

b. Haga una presentación descriptiva de sus datos muestrales, usando una tabla, al menos una gráfica, y un párrafo descriptivo.

c. ¿La muestra de usted presenta una distribución que sea significativamente diferente de la distribución presentada en la gráfica de la página 619? Use $\alpha = 0.05$.

d. Discuta las diferencias y similitudes entre su muestra y la distribución presentada en la gráfica "Para apagar el fuego".

Examen de práctica del capítulo

PRIMERA PARTE: Conociendo las definiciones

Conteste "Verdadero" si el enunciado es siempre verdadero. Si el enunciado no siempre es verdadero, cambie las palabras que aparecen en negritas con palabras que hagan que el enunciado sea siempre verdadero.

11.1 El número de grados de libertad para un examen de un experimento multinomial es **igual al** número de celdas en los datos del experimento.

11.2 La **frecuencia esperada** en una prueba *Ji*-cuadrada se encuentra multiplicando la probabilidad hipotética de una celda por el número total de observaciones en la muestra.

11.3 Debe permitirse que la frecuencia **observada** de una celda sea menor a 5 cuando se efectúe una prueba *Ji*-cuadrada.

11.4 En un **experimento multinomial** tenemos $(r-1)(c-1)$ grados de libertad (r es el número de filas, y c es el número de columnas).

11.5 Un experimento multinomial consta de ***n* intentos independientes idénticos.**

11.6 Un **experimento multinomial** distribuye los datos en una clasificación en dos formas, tal que los totales en una dirección están predeterminados.

11.7 Las gráficas para el experimento multinomial y la tabla de contingencia **deben** establecerse en forma tal que cada pieza de datos caerá exactamente en una de las categorías.

11.8 La estadística de prueba $\sum \frac{(O-E)^2}{E}$ tiene una distribución que es **aproximadamente normal**.

11.9 Los datos empleados en una prueba *Ji*-cuadrada multinomial son siempre **enumerativos**.

11.10 La hipótesis nula que se prueba por un examen de **homogeneidad** es que la distribución de proporciones es igual para cada una de las subpoblaciones.

SEGUNDA PARTE: Aplicación de conceptos

Conteste todas las preguntas. Presente fórmulas, sustituciones y trabajo.

11.11 Exprese las hipótesis nula y alternativa que se usarían para probar cada uno de estos enunciados.

a. Los números de un dígito generados por cierto generador de números aleatorios no fueron igualmente probables.

b. Los resultados de la última elección en nuestra ciudad sugieren que los votos emitidos no fueron independientes del partido registrado del electorado.

c. Las distribuciones de tipos de delitos cometidos contra la sociedad son iguales en las cuatro ciudades más grandes de Estados Unidos.

11.12 Encuentre cada valor:

a. $\chi^2(12, 0.975)$ b. $\chi^2(17, 0.005)$

11.13 Se pidió a 300 consumidores identificaran cuál, de tres artículos diferentes, encontraron más atractivo. La tabla muestra el número que prefirieron de cada uno.

Artículo	1	2	3
Número	85	103	112

¿Estos datos presentan suficiente evidencia al nivel de significación de 0.05 para indicar que los tres artículos no son igualmente preferidos?

11.14 Para estudiar el efecto del tipo de suelo sobre la cantidad de crecimiento alcanzado por una nueva planta híbrida, se plantaron arbolitos en tres tipos diferentes de suelo y se clasificaron sus cantidades subsiguientes de crecimiento en tres categorías:

	Tipo de suelo		
Crecimiento	Arcilla	Arena	Mantillo
Pobre	16	8	14
Promedio	31	16	21
Bueno	18	36	25
Total	65	60	60

¿La calidad del crecimiento parece estar distribuida de manera diferente para los tipos de suelo probados al nivel de 0.05?

a. Exprese las hipótesis nula y alternativa.

b. Encuentre el valor esperado para la celda que contenga 36.

c. Calcule el valor de *Ji*-cuadrada para estos datos.

d. Encuentre el valor p.

e. Encuentre los criterios de prueba [nivel de significación, estadística de prueba, su distribución, región crítica y valor(es) crítico(s)].

f. Exprese la decisión y conclusión para esta prueba de hipótesis.

TERCERA PARTE: Entendiendo los conceptos

11.15 Explique la forma en que un experimento multinomial y un experimento binomial son semejantes y también cómo son diferentes.

11.16 Explique la distinción entre una prueba para independencia y una prueba para homogeneidad.

11.17 El Estudiante A dice que las pruebas para independencia y homogeneidad son iguales, y el Estudiante B dice que para nada son iguales porque son pruebas de diferentes conceptos. Ambos estudiantes parcialmente tienen y no tienen la razón. Explique.

11.18 Usted está interpretando los resultados de una encuesta de opiniones sobre la función de reciclar en su ciudad. A una muestra aleatoria de 400 personas se les pidió contestaran firmemente a favor, ligeramente a favor, neutral, ligeramente en contra o firmemente a favor en cada una de varias preguntas. Hay cuatro preguntas que a usted le preocupan y piensa analizar sus resultados.

a. ¿Cómo calcula las probabilidades esperadas para cada respuesta?

b. ¿Cómo decidiría si las cuatro preguntas fueron contestadas iguales?

APÉNDICE

A Principios básicos de conteo

APÉNDICE

B Tablas

TABLA 1

Números aleatorios

10 09 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17	39 29 27 49 45
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02	00 82 29 16 65
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 43 64	35 08 03 36 06
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97	04 43 62 76 59
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77	12 17 17 68 33
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85	11 19 92 91 70
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39	23 40 30 97 32
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47	18 62 38 85 79
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09	83 49 12 56 24
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44	35 27 38 84 35
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33	50 50 07 39 98
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01	52 77 56 78 51
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10	68 71 17 78 17
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93	29 60 91 10 62
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68	23 47 83 41 13
65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86	40 21 81 65 44
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53	14 38 55 37 63
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37	96 28 60 26 55
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90	94 40 05 64 18
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22	54 38 21 45 98
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23	37 08 92 00 48
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	54 14 03 33 40	42 05 08 23 41
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81	22 22 20 64 13
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39	28 70 72 58 15
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82	07 20 73 17 90
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93	42 58 26 05 27
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18	33 21 15 94 66
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92	92 92 74 59 73
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59	25 70 14 66 70
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63	05 52 28 25 62
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35	65 33 71 24 72
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91	23 28 72 95 29
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24	90 10 33 93 33
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92	78 56 52 01 06
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47	70 61 74 29 41
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57	85 39 41 18 38
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23	97 11 89 63 38
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06	84 96 28 52 07
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33	20 82 66 95 41
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56	05 01 45 11 76
98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07	35 44 13 18 80
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94	37 54 87 30 43
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98	94 62 46 11 71
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39	00 38 75 95 79
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27	77 93 89 19 36

TABLA 1

Números aleatorios (continúa)

74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21	80 81 45 17 48
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55	36 04 09 03 24
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40	88 46 12 33 56
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46	15 02 00 99 94
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15	01 84 87 69 38
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74	64 27 85 80 44
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50	68 47 66 46 59
73 18 95 02 07	47 67 72 62 69	62 29 06 44 64	27 12 46 70 18	41 36 18 27 60
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71	93 82 34 31 78
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95	07 70 61 78 13
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53	31 22 30 84 20
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75	94 11 90 18 40
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95	77 76 22 07 91
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14	83 48 34 70 55
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06	94 54 13 74 08
77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80	72 89 35 55 07
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51	65 34 46 74 15
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46	31 85 33 84 52
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91	08 00 74 54 49
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42	43 86 07 28 34
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22	93 17 49 39 72
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00	71 14 84 36 43
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61	62 32 71 84 23
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60	81 60 41 88 80
65 13 85 68 06	87 60 88 52 61	34 31 36 58 61	45 87 52 10 69	85 64 44 72 77
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49	65 58 44 96 98
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05	40 03 03 74 38
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47	15 50 12 95 78
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41	03 85 65 45 52
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56	64 69 11 92 02
07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38	04 71 36 69 94
60 52 88 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39	61 21 20 64 55
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83	92 30 15 04 98
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 57 18 39	06 41 01 93 62
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51	31 02 47 31 67
59 58 00 64 78	75 56 97 88 00	88 83 55 44 86	23 76 80 61 56	04 11 10 84 08
38 50 80 73 41	23 79 34 87 63	90 82 29 70 22	17 71 90 42 07	95 95 44 99 53
30 69 27 06 68	94 68 81 61 27	56 19 68 00 91	82 06 76 34 00	05 46 26 92 00
65 44 39 56 59	18 28 82 74 37	49 63 22 40 41	08 33 76 56 76	96 29 99 08 36
27 26 75 02 64	13 19 27 22 94	07 47 74 46 06	17 98 54 89 11	97 34 13 03 58
91 30 70 69 91	19 07 22 42 10	36 69 95 37 28	28 82 53 57 93	28 97 66 62 52
68 43 49 46 88	84 47 31 36 22	62 12 69 84 08	12 84 38 25 90	09 81 59 31 46
48 90 81 58 77	54 74 52 45 91	35 70 00 47 54	83 82 45 26 92	54 13 05 51 60
06 91 34 51 97	42 67 27 86 01	11 88 30 95 28	63 01 19 89 01	14 97 44 03 44
10 45 51 60 19	14 21 03 37 12	91 34 23 78 21	88 32 58 08 51	43 66 77 08 83
12 88 39 73 43	65 02 76 11 84	04 28 50 13 92	17 97 41 50 77	90 71 22 67 69
21 77 83 09 76	38 80 73 69 61	31 64 94 20 96	63 28 10 20 23	08 81 64 74 49
19 52 35 95 15	65 12 25 96 59	86 28 36 82 58	69 57 21 37 98	16 43 59 15 29
67 24 55 26 70	35 58 31 65 63	79 24 68 66 86	76 46 33 42 22	26 65 59 08 02
60 58 44 73 77	07 50 03 79 92	45 13 42 65 29	26 76 08 36 37	41 32 64 43 44
53 85 34 13 77	36 06 69 48 50	58 83 87 38 59	49 36 47 33 31	96 24 04 36 42
24 63 73 97 36	74 38 48 93 42	52 62 30 79 92	12 36 91 86 01	03 74 28 38 73
83 08 01 24 51	38 99 22 28 15	07 75 95 17 77	97 37 72 75 85	51 97 23 78 67
16 44 42 43 34	36 15 19 90 73	27 49 37 09 39	85 13 03 25 52	54 84 65 47 59
60 79 01 81 57	57 17 86 57 62	11 16 17 85 76	45 81 95 29 79	65 13 00 48 60

De tablas de la RAND Corporation. Reimpresa de Wilfred J. Dixon y Frank J. Massey, Jr., *Introduction to Statistical Analysis.* 3rd ed. (New York: McGraw-Hill, 1969), pp. 446-447. Reimpreso con permiso de la RAND Corporation.

TABLA 2

Probabilidades binomiales $\left[\binom{n}{x} \cdot p^x \cdot q^{n-x}\right]$

		P													
n	*x*	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	*x*
2	0	.980	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002	0+	0
	1	.020	.095	.180	.320	.420	.480	.500	.480	.420	.320	.180	.095	.020	1
	2	0+	.002	.010	.040	.090	.160	.250	.360	.490	.640	.810	.902	.980	2
3	0	.970	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	0+	0+	0
	1	.029	.135	.243	.384	.441	.432	.375	.288	.189	.096	.027	.007	0+	1
	2	0+	.007	.027	.096	.189	.288	.375	.432	.441	.384	.243	.135	.029	2
	3	0+	0+	.001	.008	.027	.064	.125	.216	.343	.512	.729	.857	.970	3
4	0	.961	.815	.656	.410	.240	.130	.062	.026	.008	.002	0+	0+	0+	0
	1	.039	.171	.292	.410	.412	.346	.250	.154	.076	.026	.004	0+	0+	1
	2	.001	.014	.049	.154	.265	.346	.375	.346	.265	.154	.049	.014	.001	2
	3	0+	0+	.004	.026	.076	.154	.250	.346	.412	.410	.292	.171	.039	3
	4	0+	0+	0+	.002	.008	.026	.062	.130	.240	.410	.656	.815	.961	4
5	0	.951	.774	.590	.328	.168	.078	.031	.010	.002	0+	0+	0+	0+	0
	1	.048	.204	.328	.410	.360	.259	.156	.077	.028	.006	0+	0+	0+	1
	2	.001	.021	.073	.205	.309	.346	.312	.230	.132	.051	.008	.001	0+	2
	3	0+	.001	.008	.051	.132	.230	.312	.346	.309	.205	.073	.021	.001	3
	4	0+	0+	0+	.006	.028	.077	.156	.259	.360	.410	.328	.204	.048	4
	5	0+	0+	0+	0+	.002	.010	.031	.078	.168	.328	.590	.774	.951	5
6	0	.941	.735	.531	.262	.118	.047	.016	.004	.001	0+	0+	0+	0+	0
	1	.057	.232	.354	.393	.303	.187	.094	.037	.010	.002	0+	0+	0+	1
	2	.001	.031	.098	.246	.324	.311	.234	.138	.060	.015	.001	0+	0+	2
	3	0+	.002	.015	.082	.185	.276	.312	.276	.185	.082	.015	.002	0+	3
	4	0+	0+	.001	.015	.060	.138	.234	.311	.324	.246	.098	.031	.001	4
	5	0+	0+	0+	.002	.010	.037	.094	.187	.303	.393	.354	.232	.057	5
	6	0+	0+	0+	0+	.001	.004	.016	.047	.118	.262	.531	.735	.941	6
7	0	.932	.698	.478	.210	.082	.028	.008	.002	0+	0+	0+	0+	0+	0
	1	.066	.257	.372	.367	.247	.131	.055	.017	.004	0+	0+	0+	0+	1
	2	.002	.041	.124	.275	.318	.261	.164	.077	.025	.004	0+	0+	0+	2
	3	0+	.004	.023	.115	.227	.290	.273	.194	.097	.029	.003	0+	0+	3
	4	0+	0+	.003	.029	.097	.194	.273	.290	.227	.115	.023	.004	0+	4
	5	0+	0+	0+	.004	.025	.077	.164	.261	.318	.275	.124	.041	.002	5
	6	0+	0+	0+	0+	.004	.017	.055	.131	.247	.367	.372	.257	.066	6
	7	0+	0+	0+	0+	0+	.002	.008	.028	.082	.210	.478	.698	.932	7
8	0	.923	.663	.430	.168	.058	.017	.004	.001	0+	0+	0+	0+	0+	0
	1	.075	.279	.383	.336	.198	.090	.031	.008	.001	0+	0+	0+	0+	1
	2	.003	.051	.149	.294	.296	.209	.109	.041	.010	.001	0+	0+	0+	2
	3	0+	.005	.033	.147	.254	.279	.219	.124	.047	.009	0+	0+	0+	3
	4	0+	0+	.005	.046	.136	.232	.273	.232	.136	.046	.005	0+	0+	4
	5	0+	0+	0+	.009	.047	.124	.219	.279	.254	.147	.033	.005	0+	5
	6	0+	0+	0+	.001	.010	.041	.109	.209	.296	.294	.149	.051	.003	6
	7	0+	0+	0+	0+	.001	.008	.031	.090	.198	.336	.383	.279	.075	7
	8	0+	0+	0+	0+	0+	.001	.004	.017	.058	.168	.430	.663	.923	8

Para detalles específicos acerca del uso de esta tabla, vea páginas 291-292.

TABLA 2

Probabilidades binomiales $[\binom{n}{x} \cdot p^x \cdot q^{n-x}]$ (continúa)

		P													
n	*x*	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	*x*
9	0	.914	.630	.387	.134	.040	.010	.002	0+	0+	0+	0+	0+	0+	0
	1	.083	.299	.387	.302	.156	.060	.018	.004	0+	0+	0+	0+	0+	1
	2	.003	.063	.172	.302	.267	.161	.070	.021	.004	0+	0+	0+	0+	2
	3	0+	.008	.045	.176	.267	.251	.164	.074	.021	.003	0+	0+	0+	3
	4	0+	.001	.007	.066	.172	.251	.246	.167	.074	.017	.001	0+	0+	4
	5	0+	0+	.001	.017	.074	.167	.246	.251	.172	.066	.007	.001	0+	5
	6	0+	0+	0+	.003	.021	.074	.164	.251	.267	.176	.045	.008	0+	6
	7	0+	0+	0+	0+	.004	.021	.070	.161	.267	.302	.172	.063	.003	7
	8	0+	0+	0+	0+	0+	.004	.018	.060	.156	.302	.387	.299	.083	8
	9	0+	0+	0+	0+	0+	0+	.002	.010	.040	.134	.387	.630	.914	9
10	0	.904	.599	.349	.107	.028	.006	.001	0+	0+	0+	0+	0+	0+	0
	1	.091	.315	.387	.268	.121	.040	.010	.002	0+	0+	0+	0+	0+	1
	2	.004	.075	.194	.302	.233	.121	.044	.011	.001	0+	0+	0+	0+	2
	3	0+	.010	.057	.201	.267	.215	.117	.042	.009	.001	0+	0+	0+	3
	4	0+	.001	.011	.088	.200	.251	.205	.111	.037	.006	0+	0+	0+	4
	5	0+	0+	.001	.026	.103	.201	.246	.201	.103	.026	.001	0+	0+	5
	6	0+	0+	0+	.006	.037	.111	.205	.251	.200	.088	.011	.001	0+	6
	7	0+	0+	0+	.001	.009	.042	.117	.215	.267	.201	.057	.010	0+	7
	8	0+	0+	0+	0+	.001	.011	.044	.121	.233	.302	.194	.075	.004	8
	9	0+	0+	0+	0+	0+	.002	.010	.040	.121	.268	.387	.315	.091	9
	10	0+	0+	0+	0+	0+	0+	.001	.006	.028	.107	.349	.599	.904	10
11	0	.895	.569	.314	.086	.020	.004	0+	0+	0+	0+	0+	0+	0+	0
	1	.099	.329	.384	.236	.093	.027	.005	.001	0+	0+	0+	0+	0+	1
	2	.005	.087	.213	.295	.200	.089	.027	.005	.001	0+	0+	0+	0+	1
	3	0+	.014	.071	.221	.257	.177	.081	.023	.004	0+	0+	0+	0+	3
	4	0+	.001	.016	.111	.220	.236	.161	.070	.017	.002	0+	0+	0+	4
	5	0+	0+	.002	.039	.132	.221	.226	.147	.057	.010	0+	0+	0+	5
	6	0+	0+	0+	.010	.057	.147	.226	.221	.132	.039	.002	0+	0+	6
	7	0+	0+	0+	.002	.017	.070	.161	.236	.220	.111	.016	.001	0+	7
	8	0+	0+	0+	0+	.004	.023	.081	.177	.257	.221	.071	.014	0+	8
	9	0+	0+	0+	0+	.001	.005	.027	.089	.200	.295	.213	.087	.005	9
	10	0+	0+	0+	0+	0+	.001	.005	.027	.093	.236	.384	.329	.099	10
	11	0+	0+	0+	0+	0+	0+	0+	.004	.020	.086	.314	.569	.895	11
12	0	.886	.540	.282	.069	.014	.002	0+	0+	0+	0+	0+	0+	0+	0
	1	.107	.341	.377	.206	.071	.017	.003	0+	0+	0+	0+	0+	0+	1
	2	.006	.099	.230	.283	.168	.064	.016	.002	0+	0+	0+	0+	0+	2
	3	0+	.017	.085	.236	.240	.142	.054	.012	.001	0+	0+	0+	0+	3
	4	0+	.002	.021	.133	.231	.213	.121	.042	.008	.001	0+	0+	0+	4
	5	0+	0+	.004	.053	.158	.227	.193	.101	.029	.003	0+	0+	0+	5
	6	0+	0+	0+	.016	.079	.177	.226	.177	.079	.016	0+	0+	0+	6
	7	0+	0+	0+	.003	.029	.101	.193	.227	.158	.053	.004	0+	0+	7
	8	0+	0+	0+	.001	.008	.042	.121	.213	.231	.133	.021	.002	0+	8
	9	0+	0+	0+	0+	.001	.012	.054	.142	.240	.236	.085	.017	0+	9
	10	0+	0+	0+	0+	0+	.002	.016	.064	.168	.283	.230	.099	.006	10
	11	0+	0+	0+	0+	0+	0+	.003	.017	.071	.206	.377	.341	.107	11
	12	0+	0+	0+	0+	0+	0+	0+	.002	.014	.069	.282	.540	.886	12

TABLA 2

Probabilidades binomiales $[\binom{n}{x} \cdot p^x \cdot q^{n-x}]$ (continúa)

		p													
n	*x*	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	*x*
13	0	.878	.513	.254	.055	.010	.001	0+	0+	0+	0+	0+	0+	0+	0
	1	.115	.351	.367	.179	.054	.011	.002	0+	0+	0+	0+	0+	0+	1
	2	.007	.111	.245	.268	.139	.045	.010	.001	0+	0+	0+	0+	0+	2
	3	0+	.021	.100	.246	.218	.111	.035	.006	.001	0+	0+	0+	0+	3
	4	0+	.003	.028	.154	.234	.184	.087	.024	.003	0+	0+	0+	0+	4
	5	0+	0+	.006	.069	.180	.221	.157	.066	.014	.001	0+	0+	0+	5
	6	0+	0+	.001	.023	.103	.197	.209	.131	.044	.006	0+	0+	0+	6
	7	0+	0+	0+	.006	.044	.131	.209	.197	.103	.023	.001	0+	0+	7
	8	0+	0+	0+	.001	.014	.066	.157	.221	.180	.069	.006	0+	0+	8
	9	0+	0+	0+	0+	.003	.024	.087	.184	.234	.154	.028	.003	0+	9
	10	0+	0+	0+	0+	.001	.006	.035	.111	.218	.246	.100	.021	0+	10
	11	0+	0+	0+	0+	0+	.001	.010	.045	.139	.268	.245	.111	.007	11
	12	0+	0+	0+	0+	0+	0+	.002	.011	.054	.179	.367	.351	.115	12
	13	0+	0+	0+	0+	0+	0+	0+	.001	.010	.055	.254	.513	.878	13
14	0	.869	.488	.229	.044	.007	.001	0+	0+	0+	0+	0+	0+	0+	0
	1	.123	.359	.356	.154	.041	.007	.001	0+	0+	0+	0+	0+	0+	1
	2	.008	.123	.257	.250	.113	.032	.006	.001	0+	0+	0+	0+	0+	2
	3	0+	.026	.114	.250	.194	.085	.022	.003	0+	0+	0+	0+	0+	3
	4	0+	.004	.035	.172	.229	.155	.061	.014	.001	0+	0+	0+	0+	4
	5	0+	0+	.008	.086	.196	.207	.122	.041	.007	0+	0+	0+	0+	5
	6	0+	0+	.001	.032	.126	.207	.183	.092	.023	.002	0+	0+	0+	6
	7	0+	0+	0+	.009	.062	.157	.209	.157	.062	.009	0+	0+	0+	7
	8	0+	0+	0+	.002	.023	.092	.183	.207	.126	.032	.001	0+	0+	8
	9	0+	0+	0+	0+	.007	.041	.122	.207	.196	.086	.008	0+	0+	9
	10	0+	0+	0+	0+	.001	.014	.061	.155	.229	.172	.035	.004	0+	10
	11	0+	0+	0+	0+	0+	.003	.022	.085	.194	.250	.114	.026	0+	11
	12	0+	0+	0+	0+	0+	.001	.006	.032	.113	.250	.257	.123	.008	12
	13	0+	0+	0+	0+	0+	0+	.001	.007	.041	.154	.356	.359	.123	13
	14	0+	0+	0+	0+	0+	0+	0+	.001	.007	.044	.229	.488	.869	14
15	0	.860	.463	.206	.035	.005	0+	0+	0+	0+	0+	0+	0+	0+	0
	1	.130	.366	.343	.132	.031	.005	0+	0+	0+	0+	0+	0+	0+	1
	2	.009	.135	.267	.231	.092	.022	.003	0+	0+	0+	0+	0+	0+	2
	3	0+	.031	.129	.250	.170	.063	.014	.002	0+	0+	0+	0+	0+	3
	4	0+	.005	.043	.188	.219	.127	.042	.007	.001	0+	0+	0+	0+	4
	5	0+	.001	.010	.103	.206	.186	.092	.024	.003	0+	0+	0+	0+	5
	6	0+	0+	.002	.043	.147	.207	.153	.061	.012	.001	0+	0+	0+	6
	7	0+	0+	0+	.014	.081	.177	.196	.118	.035	.003	0+	0+	0+	7
	8	0+	0+	0+	.003	.035	.118	.196	.177	.081	.014	0+	0+	0+	8
	9	0+	0+	0+	.001	.012	.061	.153	.207	.147	.043	.002	0+	0+	9
	10	0+	0+	0+	0+	.003	.024	.092	.186	.206	.103	.010	.001	0+	10
	11	0+	0+	0+	0+	.001	.007	.042	.127	.219	.188	.043	.005	0+	11
	12	0+	0+	0+	0+	0+	.002	.014	.063	.170	.250	.129	.031	0+	12
	13	0+	0+	0+	0+	0+	0+	.003	.022	.092	.231	.267	.135	.009	13
	14	0+	0+	0+	0+	0+	0+	0+	.005	.031	.132	.343	.366	.130	14
	15	0+	0+	0+	0+	0+	0+	0+	0+	.005	.035	.206	.463	.860	15

TABLA 3

Áreas de la distribución normal estándar

Las entradas de esta tabla son las probabilidades de que una variable aleatoria, con una distribución normal estándar, tome un valor entre 0 y z; la probabilidad está representada por el área sombreada bajo la curva de la figura siguiente. Las áreas para valores negativos de z se obtienen por simetría.

Segundo lugar decimal en z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999									
4.0	0.49997									
4.5	0.499997									
5.0	0.4999997									

Para detalles específicos acerca del uso de esta tabla para hallar: probabilidades, vea páginas 317-320; coeficientes de confianza, páginas 338-339, 340-341; valores p, páginas 432-433, 435; valores críticos, páginas 317-320, 338-339.

TABLA 4

Valores críticos de distribución normal estándar

A SITUACIONES DE UNA COLA

Las entradas de esta tabla son los valores críticos para z para los que el área bajo la curva que representa α está en la cola derecha. Los valores críticos para la cola izquierda se encuentran por simetría.

Cantidad de α en una cola

α	0.25	0.10	0.05	0.025	0.02	0.01	0.005
$z(\alpha)$	0.67	1.28	1.65	1.96	2.05	2.33	2.58

Ejemplo de una cola:
$\alpha = 0.05$
$z(\alpha) = z(0.05) = 1.65$

B SITUACIONES DE DOS COLAS

Las entradas de esta tabla son los valores críticos para z para los que el área bajo la curva que representa α se divide igualmente entre las dos colas.

Cantidad de α en dos colas:

α	0.25	0.20	0.10	0.05	0.02	0.01
$z(\alpha/2)$	1.15	1.28	1.65	1.96	2.33	2.58
$1 - \alpha$	0.75	0.80	0.90	0.95	0.98	0.99

Área en el "centro"

Ejemplo de dos colas:
$\alpha = 0.05$ o $1 - \alpha = 0.95$
$\alpha/2 = 0.025$
$z(\alpha/2) = z(0.025) = 1.96$

Para detalles específicos acerca del uso de:

La tabla A para hallar: valores críticos, vea página 451.

La tabla B para hallar: coeficientes de confianza, vea páginas 403, 405; valores críticos, página 453.

TABLA 5

Valores *p* para distribución normal estándar

Las entradas de esta tabla son los valores *p* relacionados con la cola derecha para la $z^\star$ calculada para la distribución normal estándar.

$z^\star$	Valor *p*	$z^\star$	Valor *p*	$z^\star$	Valor *p*	$z^\star$	Valor *p*
0.00	0.5000	**1.00**	0.1587	**2.00**	0.0228	**3.00**	0.0013
0.05	0.4801	**1.05**	0.1469	**2.05**	0.0202	**3.05**	0.0011
0.10	0.4602	**1.10**	0.1357	**2.10**	0.0179	**3.10**	0.0010
0.15	0.4404	**1.15**	0.1251	**2.15**	0.0158	**3.15**	0.0008
0.20	0.4207	**1.20**	0.1151	**2.20**	0.0139	**3.20**	0.0007
0.25	0.4013	**1.25**	0.1056	**2.25**	0.0122	**3.25**	0.0006
0.30	0.3821	**1.30**	0.0968	**2.30**	0.0107	**3.30**	0.0005
0.35	0.3632	**1.35**	0.0885	**2.35**	0.0094	**3.35**	0.0004
0.40	0.3446	**1.40**	0.0808	**2.40**	0.0082	**3.40**	0.0003
0.45	0.3264	**1.45**	0.0735	**2.45**	0.0071	**3.45**	0.0003
0.50	0.3085	**1.50**	0.0668	**2.50**	0.0062	**3.50**	0.0002
0.55	0.2912	**1.55**	0.0606	**2.55**	0.0054	**3.55**	0.0002
0.60	0.2743	**1.60**	0.0548	**2.60**	0.0047	**3.60**	0.0002
0.65	0.2578	**1.65**	0.0495	**2.65**	0.0040	**3.65**	0.0001
0.70	0.2420	**1.70**	0.0446	**2.70**	0.0035	**3.70**	0.0001
0.75	0.2266	**1.75**	0.0401	**2.75**	0.0030	**3.75**	0.0001
0.80	0.2119	**1.80**	0.0359	**2.80**	0.0026	**3.80**	0.0001
0.85	0.1977	**1.85**	0.0322	**2.85**	0.0022	**3.85**	0.0001
0.90	0.1841	**1.90**	0.0287	**2.90**	0.0019	**3.90**	0+
0.95	0.1711	**1.95**	0.0256	**2.95**	0.0016	**3.95**	**0+**

Para detalles específicos acerca del uso de esta tabla para hallar valores *p*, vea páginas 432, 433, 435.

TABLA 6

Valores críticos de la distribución *t* de Student

Las entradas de esta tabla, $t(\text{df}, \alpha)$, son los valores críticos para la distribución t de Student para los que el área bajo la curva en la cola derecha es α. Los valores críticos para la cola izquierda se encuentran por simetría.

	Cantidad de α en una cola					
	0.25	0.10	0.05	0.025	0.01	0.005
	Cantidad de α en dos colas					
df	0.50	0.20	0.10	0.05	0.02	0.01
3	0.765	1.64	2.35	3.18	4.54	5.84
4	0.741	1.53	2.13	2.78	3.75	4.60
5	0.729	1.48	2.02	2.57	3.37	4.03
6	0.718	1.44	1.94	2.45	3.14	3.71
7	0.711	1.42	1.89	2.36	3.00	3.50
8	0.706	1.40	1.86	2.31	2.90	3.36
9	0.703	1.38	1.83	2.26	2.82	3.25
10	0.700	1.37	1.81	2.23	2.76	3.17
11	0.697	1.36	1.80	2.20	2.72	3.11
12	0.696	1.36	1.78	2.18	2.68	3.05
13	0.694	1.35	1.77	2.16	2.65	3.01
14	0.692	1.35	1.76	2.14	2.62	2.98
15	0.691	1.34	1.75	2.13	2.60	2.95
16	0.690	1.34	1.75	2.12	2.58	2.92
17	0.689	1.33	1.74	2.11	2.57	2.90
18	0.688	1.33	1.73	2.10	2.55	2.88
19	0.688	1.33	1.73	2.09	2.54	2.86
20	0.687	1.33	1.72	2.09	2.53	2.85
21	0.686	1.32	1.72	2.08	2.52	2.83
22	0.686	1.32	1.72	2.07	2.51	2.82
23	0.685	1.32	1.71	2.07	2.50	2.81
24	0.685	1.32	1.71	2.06	2.49	2.80
25	0.684	1.32	1.71	2.06	2.49	2.79
26	0.684	1.32	1.71	2.06	2.48	2.78
27	0.684	1.31	1.70	2.05	2.47	2.77
28	0.683	1.31	1.70	2.05	2.47	2.76
29	0.683	1.31	1.70	2.05	2.46	2.76
30	0.683	1.31	1.70	2.04	2.46	2.75
35	0.682	1.31	1.69	2.03	2.44	2.73
40	0.681	1.30	1.68	2.02	2.42	2.70
50	0.679	1.30	1.68	2.01	2.40	2.68
70	0.678	1.29	1.67	1.99	2.38	2.65
100	0.677	1.29	1.66	1.98	2.36	2.63
df > 100	0.675	1.28	1.65	1.96	2.33	2.58

Ejemplo de una cola:
df = 9 y = α = 0.10
$t(\text{df}, \alpha) = t(9, 0.10) = 1.38$

Ejemplo de dos colas:
df = 14, α = 0.02, 1 − α = 0.98
$t(\text{df}, \alpha/2) = t(14, 0.01) = 2.62$

Para detalles específicos acerca del uso de esta tabla para hallar: coeficientes de confianza, vea páginas 477-480; valores *p*, páginas 483, 486; valores críticos, páginas 477-478.

TABLA 7

Valores de probabilidad para distribución *t* de Student

Las entradas de esta tabla son los valores *p* relacionados con la cola derecha para el valor *t*★ calculado para la distribución *t* de df grados de libertad.

	Grados de libertad														
t★	3	4	5	6	7	8	10	12	15	18	21	25	29	35	df ≥ 45
0.0	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
0.1	0.463	0.463	0.462	0.462	0.462	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.461	0.460	0.460
0.2	0.427	0.426	0.425	0.424	0.424	0.423	0.423	0.422	0.422	0.422	0.422	0.422	0.421	0.421	0.421
0.3	0.392	0.390	0.388	0.387	0.386	0.386	0.385	0.385	0.384	0.384	0.384	0.383	0.383	0.383	0.383
0.4	0.358	0.355	0.353	0.352	0.351	0.350	0.349	0.348	0.347	0.347	0.347	0.346	0.346	0.346	0.346
0.5	0.326	0.322	0.319	0.317	0.316	0.315	0.314	0.313	0.312	0.312	0.311	0.311	0.310	0.310	0.310
0.6	0.295	0.290	0.287	0.285	0.284	0.283	0.281	0.280	0.279	0.278	0.277	0.277	0.277	0.276	0.276
0.7	0.267	0.261	0.258	0.255	0.253	0.252	0.250	0.249	0.247	0.246	0.246	0.245	0.245	0.244	0.244
0.8	0.241	0.234	0.230	0.227	0.225	0.223	0.221	0.220	0.218	0.217	0.216	0.216	0.215	0.215	0.214
0.9	0.217	0.210	0.205	0.201	0.199	0.197	0.195	0.193	0.191	0.190	0.189	0.188	0.188	0.187	0.186
1.0	0.196	0.187	0.182	0.178	0.175	0.173	0.170	0.169	0.167	0.165	0.164	0.163	0.163	0.162	0.161
1.1	0.176	0.167	0.161	0.157	0.154	0.152	0.149	0.146	0.144	0.143	0.142	0.141	0.140	0.139	0.139
1.2	0.158	0.148	0.142	0.138	0.135	0.132	0.129	0.127	0.124	0.123	0.122	0.121	0.120	0.119	0.118
1.3	0.142	0.132	0.125	0.121	0.117	0.115	0.111	0.109	0.107	0.105	0.104	0.103	0.102	0.101	0.100
1.4	0.128	0.117	0.110	0.106	0.102	0.100	0.096	0.093	0.091	0.089	0.088	0.087	0.086	0.085	0.084
1.5	0.115	0.104	0.097	0.092	0.089	0.086	0.082	0.080	0.077	0.075	0.074	0.073	0.072	0.071	0.070
1.6	0.104	0.092	0.085	0.080	0.077	0.074	0.070	0.068	0.065	0.064	0.062	0.061	0.060	0.059	0.058
1.7	0.094	0.082	0.075	0.070	0.066	0.064	0.060	0.057	0.055	0.053	0.052	0.051	0.050	0.049	0.048
1.8	0.085	0.073	0.066	0.061	0.057	0.055	0.051	0.049	0.046	0.044	0.043	0.042	0.041	0.040	0.039
1.9	0.077	0.065	0.058	0.053	0.050	0.047	0.043	0.041	0.038	0.037	0.036	0.035	0.034	0.033	0.032
2.0	0.070	0.058	0.051	0.046	0.043	0.040	0.037	0.034	0.032	0.030	0.029	0.028	0.027	0.027	0.026
2.1	0.063	0.052	0.045	0.040	0.037	0.034	0.031	0.029	0.027	0.025	0.024	0.023	0.022	0.022	0.021
2.2	0.058	0.046	0.040	0.035	0.032	0.029	0.026	0.024	0.022	0.021	0.020	0.019	0.018	0.017	0.016
2.3	0.052	0.041	0.035	0.031	0.027	0.025	0.022	0.020	0.018	0.017	0.016	0.015	0.014	0.014	0.013
2.4	0.048	0.037	0.031	0.027	0.024	0.022	0.019	0.017	0.015	0.014	0.013	0.012	0.012	0.011	0.010
2.5	0.044	0.033	0.027	0.023	0.020	0.018	0.016	0.014	0.012	0.011	0.010	0.010	0.009	0.009	0.008
2.6	0.040	0.030	0.024	0.020	0.018	0.016	0.013	0.012	0.010	0.009	0.008	0.008	0.007	0.007	0.006
2.7	0.037	0.027	0.021	0.018	0.015	0.014	0.011	0.010	0.008	0.007	0.007	0.006	0.006	0.005	0.005
2.8	0.034	0.024	0.019	0.016	0.013	0.012	0.009	0.008	0.007	0.006	0.005	0.005	0.005	0.004	0.004
2.9	0.031	0.022	0.017	0.014	0.011	0.010	0.008	0.007	0.005	0.005	0.004	0.004	0.004	0.003	0.003
3.0	0.029	0.020	0.015	0.012	0.010	0.009	0.007	0.006	0.004	0.004	0.003	0.003	0.003	0.002	0.002
3.1	0.027	0.018	0.013	0.011	0.009	0.007	0.006	0.005	0.004	0.003	0.003	0.002	0.002	0.002	0.002
3.2	0.025	0.016	0.012	0.009	0.008	0.006	0.005	0.004	0.003	0.002	0.002	0.002	0.002	0.001	0.001
3.3	0.023	0.015	0.011	0.008	0.007	0.005	0.004	0.003	0.002	0.002	0.002	0.001	0.001	0.001	0.001
3.4	0.021	0.014	0.010	0.007	0.006	0.005	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001
3.5	0.020	0.012	0.009	0.006	0.005	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001
3.6	0.018	0.011	0.008	0.006	0.004	0.004	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0+	0+
3.7	0.017	0.010	0.007	0.005	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0+	0+	0+
3.8	0.016	0.010	0.006	0.004	0.003	0.003	0.002	0.001	0.001	0.001	0.001	0+	0+	0+	0+
3.9	0.015	0.009	0.006	0.004	0.003	0.002	0.001	0.001	0.001	0.001	0+	0+	0+	0+	0+
4.0	0.014	0.008	0.005	0.004	0.003	0.002	0.001	0.001	0.001	0+	0+	0+	0+	0+	0+

Para detalles específicos acerca del uso de esta tabla para hallar valores *p*, vea páginas 484, 486.

TABLA 8

Valores críticos de la distribución χ^2 (Ji cuadrada)

Las entradas de esta tabla, χ^2 (df, α), son los valores críticos para la distribución χ^2 para la que el área bajo la curva a la derecha es α.

	Área a la derecha												
	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
	Área en cola izquierda						Mediana				Área en cola derecha		
df	0.005	0.01	0.025	0.0	0.10	0.25	0.50	0.25	0.10	0.05	0.025		
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.101	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.82	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.8
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.5	16.8	18.6
7	0.990	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.34	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.34	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.34	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.34	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.34	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.34	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.34	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.34	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.34	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.34	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.34	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.34	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.34	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.34	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.34	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.34	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.34	34.8	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	33.7	39.34	45.6	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	42.9	49.33	56.3	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	52.3	59.33	67.0	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	61.7	69.33	77.6	85.5	90.5	95.0	100.0	104.0
80	51.2	53.5	57.2	60.4	64.3	71.1	79.33	88.1	96.6	102.0	107.0	112.0	116.0
90	59.2	61.8	65.6	69.1	73.3	80.6	89.33	98.6	108.0	113.0	118.0	124.0	128.0
100	67.3	70.1	74.2	77.9	82.4	90.1	99.33	109.0	118.0	124.0	130.0	136.0	140.0

Ejemplo de cola izquierda:
Encontrar χ^2 con df = 28; área en cola izquierda = 0.10.

χ^2**(df, área a la derecha) =** χ^2**(28, 0.90) = 18.9**

Ejemplo de cola derecha:
Encontrar χ^2 con df = 23; área en cola derecha = 0.025

χ^2**(df, área a la derecha) =** χ^2**(23, 0.025) = 38.1**

Para detalles específicos acerca del uso de esta tabla para hallar: valores *p*, vea páginas 521, 522, 524; valores críticos, páginas 517-518, 521, 523.

TABLA 9A

Valores críticos de la distribución F ($\alpha = 0.05$)

Las entradas de esta tabla son valores críticos de F para los que el área bajo la curva a la derecha es igual a 0.05.

Grados de libertad para numerador

Grados de libertad para denominador	1	2	3	4	5	6	7	8	9	10
1	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Para detalles específicos acerca del uso de esta tabla para hallar: valores p, vea página 597; valores críticos, página 594.

TABLA 9A (CONTINÚA)

Grados de libertad para denominador	Grados de libertad para numerador								
	12	15	20	24	30	40	60	120	∞
1	244.	246.	248.	249.	250.	251.	252.	253.	254.
2	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

De E. S. Pearson y H. O. Hartley, *Biometrika Tables for Statisticians*, vol.1 (1958), pp. 159-163. Reimpreso con permiso de los Biometrika Trustees.

TABLA 9B

Valores críticos de la distribución $F(\alpha = 0.025)$

Las entradas de esta tabla son valores críticos de F para los que el área bajo la curva a la derecha es igual a 0.025.

Grados de libertad para numerador

Grados de libertad para denominador	1	2	3	4	5	6	7	8	9	10
1	648.	800.	864.	900.	922.	937.	948.	957.	963.	969.
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4
4	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.28	3.21	3.15
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05

Para detalles específicos acerca del uso de esta tabla para hallar: valores *p*, vea página 597; valores críticos, página 594.

TABLA 9B (CONTINÚA)

Grados de libertad para denominador	Grados de libertad para numerador								
	12	15	20	24	30	40	60	120	∞
1	977.	985.	993.	997.	1001.	1006.	1010.	1014.	1018.
2	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5
3	14.3	14.3	14.2	14.1	14.1	14.0	14.0	13.9	13.9
4	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
30	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	2.05	1.95	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

De E. S. Pearson y H. O. Hartley, *Biometrika Tables for Statisticians*, vol. I (1958), pp.159-163. Reimpreso con permiso de los Biometrika Trustees.

TABLA 9C

Valores críticos de la distribución F ($\alpha = 0.01$)

Las entradas de esta tabla son valores críticos de F para los que el área bajo la curva a la derecha es igual a 0.01.

Grados de libertad para numerador

Grados de libertad para denominador	1	2	3	4	5	6	7	8	9	10
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5982.	6024.	6056.
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Para detalles específicos acerca del uso de esta tabla para hallar: valores p, vea página 597; valores críticos, página 594.

TABLA 9C (CONTINÚA)

Grados de libertad para denominador	Grados de libertad para numerador								
	12	15	20	24	30	40	60	120	∞
1	6106.	6157.	6209.	6235.	6261.	6287.	6313.	6339.	6366.
2	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

De E. S. Pearson y H. O. Hartley, *Biometrika Tables for Statisticians*, vol. I (1958), pp.159-163. Reimpreso con permiso de los Biometrica Trustees.

TABLA 10

Bandas de confidencia para el coeficiente de correlación $(1 - \alpha) = 0.95$

Los números en las curvas son tamaños muestrales.

Para detalles específicos acerca del uso de esta tabla para hallar intervalos, vea páginas 704-705.

TABLA 11

Valores críticos de r cuando $\rho = 0$

Las entradas de esta tabla son los valores críticos de r para una prueba de dos colas en α. Para correlación simple, df $= n - 2$, donde n es el número de pares de datos de la muestra. Para una prueba de una cola, el valor de α mostrado en la parte superior de la tabla es el doble del valor de α que se usa en la prueba de hipótesis.

df \ α	0.10	0.05	0.02	0.01
1	0.988	0.997	1.000	1.000
2	0.900	0.950	0.980	0.990
3	0.805	0.878	0.934	0.959
4	0.729	0.811	0.882	0.917
5	0.669	0.754	0.833	0.874
6	0.621	0.707	0.789	0.834
7	0.582	0.666	0.750	0.798
8	0.549	0.632	0.716	0.765
9	0.521	0.602	0.685	0.735
10	0.497	0.576	0.658	0.708
11	0.476	0.553	0.634	0.684
12	0.458	0.532	0.612	0.661
13	0.441	0.514	0.592	0.641
14	0.426	0.497	0.574	0.623
15	0.412	0.482	0.558	0.606
16	0.400	0.468	0.542	0.590
17	0.389	0.456	0.528	0.575
18	0.378	0.444	0.516	0.561
19	0.369	0.433	0.503	0.549
20	0.360	0.423	0.492	0.537
25	0.323	0.381	0.445	0.487
30	0.296	0.349	0.409	0.449
35	0.275	0.325	0.381	0.418
40	0.257	0.304	0.358	0.393
45	0.243	0.288	0.338	0.372
50	0.231	0.273	0.322	0.354
60	0.211	0.250	0.295	0.325
70	0.195	0.232	0.274	0.302
80	0.183	0.217	0.256	0.283
90	0.173	0.205	0.242	0.267
100	0.164	0.195	0.230	0.254

De E. S. Pearson y H. O. Hartley, *Biometrika Tables for Statisticians*, vol.1 (1962), p. 138. Reimpreso con permiso de los Biometrika Trustees.

Para detalles específicos acerca del uso de esta tabla para hallar: valores p, vea páginas 706-707; valores críticos, página 707.

TABLA 12

Valores críticos de la prueba del signo

Las entradas de esta tabla son los valores críticos para el número del signo menos frecuente, para una prueba de dos colas en α para el binomio $p = 0.5$. Para una prueba de una cola, el valor de α mostrado en la parte superior de la tabla es el doble del valor de α que se use en la prueba de hipótesis

	α					α			
n	0.01	0.05	0.10	0.25	*n*	0.01	0.05	0.10	0.25
1					51	15	18	19	20
2					52	16	18	19	21
3				0	53	16	18	20	21
4				0	54	17	19	20	22
5			0	0	55	17	19	20	22
6		0	0	1	56	17	20	21	23
7		0	0	1	57	18	20	21	23
8	0	0	1	1	58	18	21	22	24
9	0	1	1	2	59	19	21	22	24
10	0	1	1	2	60	19	21	23	25
11	0	1	2	3	61	20	22	23	25
12	1	2	2	3	62	20	22	24	25
13	1	2	3	3	63	20	23	24	26
14	1	2	3	4	64	21	23	24	26
15	2	3	3	4	65	21	24	25	27
16	2	3	4	5	66	22	24	25	27
17	2	4	4	5	67	22	25	26	28
18	3	4	5	6	68	22	25	26	28
19	3	4	5	6	69	23	25	27	29
20	3	5	5	6	70	23	26	27	29
21	4	5	6	7	71	24	26	28	30
22	4	5	6	7	72	24	27	28	30
23	4	6	7	8	73	25	27	28	31
24	5	6	7	8	74	25	28	29	31
25	5	7	7	9	75	25	28	29	32
26	6	7	8	9	76	26	28	30	32
27	6	7	8	10	77	26	29	30	32
28	6	8	9	10	78	27	29	31	33
29	7	8	9	10	79	27	30	31	33
30	7	9	10	11	80	28	30	32	34
31	7	9	10	11	81	28	31	32	34
32	8	9	10	12	82	28	31	33	35
33	8	10	11	12	83	29	32	33	35
34	9	10	11	13	84	29	32	33	36
35	9	11	12	13	85	30	32	34	36
36	9	11	12	14	86	30	33	34	37
37	10	12	13	14	87	31	33	35	37
38	10	12	13	14	88	31	34	35	38
39	11	12	13	15	89	31	34	36	38
40	11	13	14	15	90	32	35	36	39
41	11	13	14	16	91	32	35	37	39
42	12	14	15	16	92	33	36	37	39
43	12	14	15	17	93	33	36	38	40
44	13	15	16	17	94	34	37	38	40
45	13	15	16	18	95	34	37	38	41
46	13	15	16	18	96	34	37	39	41
47	14	16	17	19	97	35	38	39	42
48	14	16	17	19	98	35	38	40	42
49	15	17	18	19	99	36	39	40	43
50	15	17	18	20	100	36	39	41	43

De Wilfred J. Dixon y Frank J. Massey, Jr., *Introduction to Statistical Analysis*, 3rd ed. (New York: McGraw-Hill, 1969), p. 509. Reimpreso con permiso.

Para detalles específicos acerca del uso de esta tabla para hallar: intervalos de confianza, vea página 753; valores p, páginas 754-755; valores críticos, páginas 753-755.

TABLA 13

Valores críticos de *U* en la prueba Mann-Whitney

A. Las entradas son los valores críticos de *U* para una prueba de una cola a 0.025 o para una prueba de dos colas a 0.05.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																				
2								0	0	0	0	1	1	1	1	1	2	2	2	2
3					0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4				0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5			0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6			1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7			1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8		0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9		0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10		0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11		0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12		1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13		1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14		1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15		1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16		1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17		2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18		2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19		2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20		2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

B. Las entradas son los valores críticos de *U* para una prueba de una cola a 0.05 o para una prueba de dos colas a 0.10.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																			0	0
2					0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3			0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4			0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5		0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6		0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7		0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8		1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9		1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10		1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11		1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12		2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13		2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14		2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
15		3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16		3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17		3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18		4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

Reproducida del *Bulletin of the Institute of Educational Research at Indiana University* vol. 1, No. 2; con el permiso del autor y el editor. Para detalles específicos acerca del uso de esta tabla para hallar: valores *p*, vea páginas 768-769; valores críticos, página 768.

TABLA 14

Valores críticos para el número total de corridas (V)

Las entradas de esta tabla son los valores críticos para una prueba de dos colas usando $\alpha = 0.05$. Para una prueba de una cola $\alpha = 0.025$, use sólo uno de los valores críticos: el valor crítico más pequeño para una región crítica izquierda, el más grande para una región crítica derecha.

El mayor de n_1 y n_2

El menor de n_1 y n_2	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2								2	2	2	2	2	2	2	2	2
								6	6	6	6	6	6	6	6	6
3		2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
		8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
4	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
	9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10
5	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
	10	10	11	11	12	12	12	12	12	12	12	12	12	12	12	12
6		3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
		11	12	12	13	13	13	13	14	14	14	14	14	14	14	14
7			3	4	4	5	5	5	5	5	6	6	6	6	6	6
			13	13	14	14	14	14	15	15	15	16	16	16	16	16
8				4	5	5	5	6	6	6	6	6	7	7	7	7
				14	14	15	15	16	16	16	16	17	17	17	17	17
9					5	5	6	6	6	7	7	7	7	8	8	8
					15	16	16	16	17	17	18	18	18	18	18	18
10						6	6	7	7	7	7	8	8	8	8	9
						16	17	17	18	18	18	19	19	19	20	20
11							7	7	7	8	8	8	9	9	9	9
							17	18	19	19	19	20	20	20	21	21
12								7	8	8	8	9	9	9	10	10
								19	19	20	20	21	21	21	22	22
13									8	9	9	9	10	10	10	10
									20	20	21	21	22	22	23	23
14										9	9	10	10	10	11	11
										21	22	22	23	23	23	24
15											10	10	11	11	11	12
											22	23	23	24	24	25
16												11	11	11	12	12
												23	24	25	25	25
17													11	12	12	13
													25	25	26	26
18														12	13	13
														26	26	27
19															13	13
															27	27
20																14
																28

De C. Eisenhart y F. Swed, "Tables for testing randomness of grouping in a sequence of alternatives", *Annals of Statistics*, vol. 14 (1943): 66-87. Reimpreso con permiso.
Para detalles específicos acerca del uso de esta tabla para hallar: valores p, vea página 778; valores críticos, página 778.

TABLA 15

Valores críticos del coeficiente de correlación de rango de Spearman

Las entradas de esta tabla son los valores críticos de r_s para una prueba de dos colas a α. Para una prueba de una cola, el valor de α que se ve en la parte superior de la tabla es el doble del valor de que se usa en la prueba de hipótesis.

n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$
5	0.900	–	–	–
6	0.829	0.886	0.943	–
7	0.714	0.786	0.893	–
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.536	0.618	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

De E. G. Olds, "Distribution of sums of squares of rank differences for small numbers of individuals". *Annals of Statistics* , vol. 9 (1938), pp. 138-148, y corregido, vol. 20 (1949), pp. 117-118. Reimpresa con permiso.

Para detalles específicos acerca del uso de esta tabla para hallar: valores *p*, vea páginas 788-789; valores críticos, página 788.

Respuestas a ejercicios seleccionados

Capítulo 1

1.1 a. norteamericanos
b. método de comunicación preferido por trabajadores
c. 63% de las personas encuestadas
e. 7.2 choques mortales por 100 millones de millas para jóvenes de 19 años

1.3 b. no aparece, los profesionales de Java trabajan una semana de 40 horas
c. sólo si se desean largas horas de trabajo

1.7 a. descriptivo
b. inferencial

1.9 a. jefes de familia norteamericanos
b. 1000
c. lugar más difícil de limpiar
d 120
e. 5% más abajo o 5% más arriba
f. Entre 30 y 40% piensan en persianas venecianas

1.11 a. norteamericanos que presentan impuestos
b. usuarios del TurboTax
c. acción tomada al recibir devolución de impuestos
d. plan de mayoría para pagar cuentas; parte más grande de un billete de dólar

1.13 a. sí, 50%
b. "salto de 50%" funciona para llamar la atención de personas

1.15 a. todos los norteamericanos adultos
b. 1200 seleccionados al azar
c. "estado de alergia" para cada adulto
d. 33.2% con base en los adultos muestreados
e. porcentaje de todos los norteamericanos adultos que tienen alergia, 36%

1.17 variable categórica

1.19 a. estado marital, código postal
b. nivel de educación, clasificación por primera impresión

1.21 a. los puntos se cuentan.
b. el tiempo se mide.

1.23 a. nivel de satisfacción
b. ordinal

1.25 a. todas las personas que tengan hipertensión
b. las 5000 personas
c. proporción de la población para la que la droga es eficaz
d. proporción de la muestra para la que la droga es eficaz, 80%
e. no

1.27 a. todas las piezas ensambladas desde la línea de montaje
b. infinito
c. las piezas revisadas
d. atributo, atributo, numérica

1.29 a. todas las personas que sufren de alergias estacionales
b. las 679 personas
c. estado de alivio y efectos colaterales
d. cualitativos

1.31 a. numérica
b. atributo
c. numérica
d. atributo
e. numérica
f. numérica

1.33 a. La población contiene todos los objetos de interés; la muestra contiene sólo los que se estudian en realidad.
b. comodidad, disponibilidad, sentido práctico

1.35 jugadores de futbol, rango más amplio

1.37 El precio por unidad estándar hace que el precio sea la única variable.

1.39 demasiado fácil, demasiado difícil, pueden distinguir entre el conocimiento de estudiantes

1.41 voluntario; sí

1.43 voluntario; sesgo

1.45 comodidad

1.49 muestras de probabilidad

1.51 Los métodos estadísticos suponen el uso de muestras aleatorias.

1.53 Seleccione al azar el primer artículo entre 1 y 25, de ahí en adelante seleccione cada 25avo.

1.55 Una muestra proporcional funcionaría mejor.

1.57 Serán consideradas sólo personas con teléfonos y números telefónicos de la guía

1.59 a. probabilidad
b. estadística

1.61 a. estadística
b. probabilidad
c. estadística
d. probabilidad

1.63 realizan muchos de los cálculos y pruebas rápida y fácilmente

1.65 Las calculadoras realizan sólo los cálculos que se les ordena que hagan.

1.67 a. color del pelo, edad, género, estado civil
b. número de cursos tomados, estatura, distancia de casa a la universidad

1.69 a. valor de datos
b. ¿Cuál es el promedio de la muestra?
c. ¿Cuál es el promedio para todas las personas?

1.71 a. adultos norteamericanos
b. estado de la vacuna para la gripe, lugar donde se obtuvo, precauciones, tipo de precaución
c. Todos son atributos

1.73 a. Todas las escuelas públicas en Estados Unidos
b. 1000 directores
c. muestra de probabilidad
d. estratificado

1.75 a. Ambas están aumentando al mismo ritmo; el número de no conductores sigue igual, uno aumenta a un ritmo mayor que el otro.
b. más conductores que vehículos antes de 1971, más vehículos que conductores después de 1973; número de conductores y vehículos fue igual

1.77 a. estudio de observación
b. porcentaje de uso de casco en niños
c. proporción de muestra que usaba casco, 41%
d. actividad, género, uso de casco —atributo; edad— numérico

1.79 a. todos los norteamericanos
b. actividades físicas y nutrición
c. edad —numérica; estado de actividad de descanso atributo; estado de nutrición— atributo

1.81 a. número de nuevas prescripciones
b. mujeres, por 1.5 de nuevas prescripciones
c. 75 y mayores, 13 nuevas prescripciones
d. sí

Capítulo 2

2.3 b. proporciones relativas en conjunto
c. proporciones relativas entre respuestas individuales

2.5 a. **El código actual de vestido en mi compañía es. . .**

b. **El código actual de vestido en mi compañía es . . .**

2.7 a. **Puntos anotados por equipos ganadores**
Noche de inauguración de la temporada 2004-2005 de la NBA

b. **Puntos anotados por equipos ganadores**
Noche de inauguración de la temporada 2004-2005 de la NBA

c. gráfica de barras en la parte a
d. inicia escala vertical en cero

2.9

2.11

Defecto	Talla incorrecta	Costura mala	Falta botón	Defecto en tela
Cantidad	258	153	67	22
Porcentaje	51.6	30.6	13.4	4.4
% Acum.	51.6	82.2	95.6	100.0

2.13 a.

Defecto	Limpieza	Lavandería	Otros	Cocinar	Lavar platos
Cantidad	53	18	12	9	8
Porcentaje	530	18.0	12.0	9.0	8.0
% Acum.	53.0	71.0	83.0	92.0	100.0

b. Es un conjunto de varias respuestas; es necesario desglosarlo.

2.15 a. 150 defectos
b. 0.30
c. 37.3 + 30.0 + 15.3 + 8.0
d. Defecto y arañazo, total 67.3

2.17 puntos anotados por juego por equipo de baloncesto

Points Scored per Game by Basketball Team

2.19 a. **Estaturas de seleccionados de la primera ronda de la NBA el 24 de junio, 2004**

b. 71 pulgadas, 89 pulgadas
c. 79 pulgadas, 5 jugadores
d. columna más alta

2.21

Overall Length of Commutators

2.23 **Puntos anotados por juego**

3	6
4	6
5	6 4 5 4 2 1
6	1 1 8 0 6 1 4
7	1

2.25 a. **Cargos de entrega inmediata**

2.	0
2.	9 8 8 9
3.	1 1
3.	5 8 8 5 8 6 6 8 7 7 8
4.	0 3 1 0 0
4.	5 5 9 6 8 6
5.	0 4 0 2 4
5.	6 7
6.	0 1
6.	8
7.	
7.	8

b. sesgado a la derecha

2.27 a. El valor de lugar de las hojas es centenas.
b. 16
c. 5.97, 6.01, 6.04, 6.08
d. Las frecuencias acumulativas se inician en la parte superior y en la inferior

2.29 a.

x	f
0	2
1	5
2	3
3	0
4	2

b. f es frecuencia; el valor de 1 se presenta 5 veces
c. 12
d. número de datos, o tamaño austral

2.31 a. gráfica de barras
c. histograma

2.33 b.

d. 0.667

2.35 a.

b. acampanada

c. centro en 5 cuartos; 3 a 7 cuartos en su mayoría

2.37 b.

2.39 a. 35–45

d.

2.41 a. 12 y 16

b. 2, 6, 10, 14, 18, 22, 26

c. 4.0

d. 0.08, 0.16, 0.16, 0.40, 0.12, 0.06, 0.02

e.

2.43 a. frecuencia: 1, 14, 22, 8, 5, 3, 2

b. 6

c. 27

d.

2.45 a.

b. frecuencia: 6, 10, 7, 6, 8, 11, 10, 6

c. frecuencia: 3, 10, 4, 9, 7, 11, 11, 7, 2

d. frecuencia: 3, 13, 13, 15, 17, 3

f. b y c, bimodal; d, sesgada a la izquierda; gráfica de puntos muestra que el modo es 9; el histograma indica dos clases modales de 4-7 y 16-22; modo; no en ninguna clase modal

2.47 a. 1, 9, 10, 12, 4
b. 1
c. 4.5, 5.5, 6.5, 7.5, 8.5

d.

2.51 a. frec. acumulada: 12, 49, 75, 94, 100
c.

2.53 a. frec. acum. relativa: 0.08, 0.24, 0.40, 0.80, 0.92, 0.98, 1.00
b.

2.55 a. frec.: 9, 15, 17, 8, 7, 11, 11, 2, 1, 1
b. frec. relativa: 0.110, 0.183, 0.207, 0.098, 0.085, 0.134, 0.134, 0.024, 0.012, 0.012

c. **Población que vive en vecindarios de alta pobreza**
Porcentajes en ciudades de E.U.

d. frec. acum. relativa: 0.110, 0.293, 0.500, 0.598, 0.683, 0.817, 0.951, 0.975, 0.987, 0.999

e. **Población que vive en vecindarios de alta pobreza en E.U.**

2.57 números variables cuantitativos con los que se puede trabajar en aritmética; con la variable cualitativa no se puede

2.59 a. 9
b. valor = 0

2.61 a. 3.4
b. 5.2
c. 5.9
d. 5.6
e. 4.95
f. 5.025

2.63 4.5avo; 4.55

2.65 2

2.67 a. 8.2, 8.5, 9.0, 8.0

2.69 a. 6.0
b. 3.5avo; 6.5
c. 7
d. 5.5

2.71 a. 40.2
b. 5.5avo; 41.5
c. 38.5
d. 48

2.73 a. 71.16%
b. 16avo; 72.66%

c. Porcentaje de tallo y hoja $N = 31$
Unidad de hoja = 1.0

2	5	89
5	6	244
10	6	57999
(17)	7	00111223333444444
4	7	578
1	8	0

d. la cola izquierda hace que la media sea menor a la mediana; 58.60 y 59.25; tienen un efecto reductor en el valor de la media

2.75 a. **Third Graders at Roth Elementary School**

b. 9

c. **Alumnos 3er. grado, prueba de resistencia física**

d. parece bimodal; 4-7, 16-19

f. no

g. el modo es valor de un solo dato; la clase modal, valores de grupo de datos

2.77 a. & b.

	Carreras local	Carreras visitante
Media	4.828	4.797
Mediana	4.870	4.860
Máximo	6.380	5.570
Mínimo	3.630	3.430
Rango medio	5.005	4.500

2.79 c. poca diferencia; más conductores; más conductoras

d. **Proporción entre conductores y conductoras por estado**

e. acampanado

f. 0.997

2.85 a. \$1465

b. 5.2%

2.87 $\Sigma(x - \bar{x}) = \Sigma x - n\bar{x} = \Sigma x - n \cdot (\Sigma x/n) = \Sigma x - \Sigma x = 0$

2.89 a. $\Sigma(x - \bar{x})^2 = 46$; 11.5

b. $\Sigma x^2 = 171$; 11.5

c. igual

2.91 a. 5

b. $\Sigma(x - \bar{x})^2 = 16$; 3.2

c. 1.8

2.93 a. $\Sigma(x - \bar{x})^2 = 42.95$; 3.1

b. $\Sigma x^2 = 764$; 3.1

c. 1.8

2.95 a. $n = 6$, $\Sigma x = 37{,}116$, $\Sigma x^2 = 229{,}710{,}344$; 22,153.6

b. $n = 6$, $\Sigma x = 1{,}116$, $\Sigma x^2 = 318{,}344$; 22,153.6

2.97 a.

b. 30.05

c. 9

d. $\Sigma x^2 = 18\,209$; 7.8

e. 2.8

g. Excepto por el valor $x = 30$, la distribución se ve rectangular.

2.99 a. 21.7; 5.242

b. Tallo y hoja de porcentaje $N = 31$
Unidad de hoja = 1.0

2	5	89	
5	6	24	
10	6	557999	
(17)	7	00111233334444444	s
4	7	578	
1	8	0	

rango

c. sesgado a la izquierda

2.101 incorrecto; la desviación estándar nunca es negativa; error en cálculos o error tipográfico

2.105 a. 44ava posición desde el valor bajo; 7ª posición desde el valor alto

b. 10.5avo; $P_{20} = 64$; 18avo $P_{35} = 70$

c. 10.5avo desde *H;* 88.5; 3º desde *H;* 95

2.107 a. **Salarios de maestros de escuela elemental**

b. 2º desde *L,* 17º desde *H*

c. 5º, \$36 700

d. 14avo, \$45 800

2.109 a. 3.8; 5.6

b. 4.7

c. 3.5º, 3.5; 7º, 4.0; 18.5º, 6.9

2.111

2.113 a.

Porcentajes de graduación de equipos hombres 2005
Torneo de baloncesto 1ª división de la NCAA

b.

Stem-and-leaf of Graduation Rates (%) $N = 64$
Leaf Unit = 1.0

3	0	008
11	1	11455779
20	2	055577799
27	3	0033368
(15)	4	000000334445557
22	5	003455577888
10	6	0477
6	7	135
3	8	
3	9	2
2	10	00

c. resumen de 5 números: 0, 27, 40, 55, 100

d. 4º, 11; 61avo, 75

e. sesgado a la derecha

2.115 a.

b.

```
Stem-and-leaf of % On Time N = 31
Leaf Unit = 0.10
1     | 69 | 9
1     | 70 |
2     | 71 | 9
2     | 72 |
4     | 73 | 58
5     | 74 | 6
5     | 75 |
7     | 76 | 04
10    | 77 | 689
13    | 78 | 239
15    | 79 | 45
(5)   | 80 | 55899
11    | 81 | 0012779
4     | 82 | 29
2     | 83 | 69
```

c. 69.96, 77.62, 80.50, 81.20, 83.93

d. 4º, 73.85; 7º, 76.45

e. sesgada a la izquierda

2.117 simétrica

2.119 1.67, −0.75

2.121 a. −1.76
b. −0.54
c. 0.42
d. 1.63

2.123 a. 120
b. 144.0
c. 92.0
d. 161.0

2.125 b. 0.03, 0.14, 0.20, 0.30, 0.55

c. 0.22; 0.16
d. −0.75, 1.66, −1.42, 0.22, 3.20

2.127 1.625, 1.2; A

2.129 175 a 225 palabras, inclusive

2.131 Casi todos los datos, 99.7% está dentro de 3 desviaciones estándar de la media.

2.133 a. 2.5%
b. 70.4 a 97.6 horas

2.135 a. 50%
b. 0.16
c. 0.84
d. 0.815

2.137 a. al menos 75%
b. al menos 89%

2.139 a. a lo sumo 11%
b. a lo sumo 6.25%

2.141 a. *f:* 6, 9, 8, 10, 6, 4, 4, 2, 1
b. 11.2, 6.2
c. −1.2 a 23.6; 96% (48/50)

2.143 a.

b. 74.4, 2.5
d. 71.9 a 76.9, 68%; 69.4 a 79.4, 97%; 66.9 a 81.9, 100%
e. 68%, 97% y 100% concuerdan
f. 97% y 100% satisfacen el teorema

2.147 sí

2.153 b. $\Sigma f = 20$; $\Sigma xf = 46$; $\Sigma x^2 f = 128$

2.155 a. 269/80 = 3.4
b. 3
c. 1.7277; 1.31

2.157 a. 13.1
b. 0.0532
c. 0.23

2.159 11.2; 12.225; 3.5

2.161 16.021; 0.026

2.163 24.2; 14.1

2.165 b.

c. sesgada a la derecha
d. 35.2
e. 38avo; 30
f. 90
g. 21.1
h.

	No agrupados	Agrupados	Porcentaje de error
Media	34.37	35.2	2.41%
Mediana	34	30	−11.76%
Rango	90	90	0%
Desv. estándar	20.95	21.08	0.62%

2.167 a. 125
b.

c. 2.5
d. 63avo; 1.5
e. 0

2.169 a. las clases primera y última no tienen mismo ancho
b. la última clase cambió a "85-94 años"

c. **Muestra aleatoria de 250 personas que viven en Nueva York**

d. & e. 38.7 años; 22.6 años

2.171 a. **¿Toma alimentos que hayan caído al piso?**

b. 30, 24, 12, 234

2.173 a. **¿Qué hará con su devolución extra de impuestos?**

2.175 a. **Principales causas de muerte en E.U. para 2002**

Cantidad	69.7	55.7	16.3	12.5	7.3	6.6	5.9	4.4	4.1	3.4
Porcentaje	37.5	30.0	8.8	6.7	3.9	3.6	3.2	2.4	2.2	1.8
% acum.	37.5	67.5	76.2	82.9	86.9	90.4	93.6	96.0	98.2	100.0

b. dos causas principales de muerte, enfermedades del corazón y neoplasmas malignos

2.177 a. numérico
b. atributo
c. numérico
d. atributo
e. numérico

2.179 a, d, e, f, y g aumentadas; b y c sin cambio

2.181 $n = 8$, $\Sigma x = 36.5$, $\Sigma x^2 = 179.11$
a. 4.56
b. 1.34
c. muy cercana a 4%

2.183 $n = 118$, $\Sigma x = 2364$
a. 20.0
b. 59.5ava, 17
c. 16
d. 30avo, 15; 89avo, 21
e. 12avo, 14; 113avo, 43

2.185 $n = 25$, $\Sigma x = 1997$, $\Sigma x^2 = 163\ 205$; 79.9; 12.4

2.187 b. datos
c. estadística
d. no

2.189 a. \$196 861, \$62 819
b. \$134 042, \$259 680
c. 34/50 = 68%
d. \$71 223, \$322 499
e. 48/50 = 96%
f. \$8404, \$385 318
g. 50/50 = 100%
h. concuerdan con las de Chebyshev
i. concuerdan con las empíricas
j. **Máximos Ganadores Torneo Nacional de 2004**

Media	196861
DesvEst	62819
N	50
AD	1.295
Valor *p*	<0.005

k. La prueba de normalidad sugiere que no es normal.

2.191 a. 13.15
b. 13.85
c. 15.0
d. 12.95

e. 5.7
f. 25.5avo, 10.95; 75.5avo, 14.9
g. 12.925
h. 35.5avo, 12.05; 64.5avo, 14.5
j. **Longitudes de 100 truchas cafés —criadero de peces**

k. mostrado en la parte i
l. **Longitudes de 100 truchas cafés —criadero de peces**

m. $n = 100$; $\Sigma xf = 1314$, $\Sigma x^2 f = 17\,635.26$; 13.14, 1.93

2.193 e. $n = 48$, $\Sigma x = 8503.88$; 177.2; 24.5th, 86.3; no hay modo; 539.425

Densidad de población estatal por milla cuadrada —E.U.

g. NJ, RI, MA, CT, MD; WY, MT, ND, SD, NM

2.195 b. **Porcentajes estatales de egresados de preparatoria que tomaron examen ACT en 2003**

c. bimodal
e. 0.4308
g. 0.2874
h. 53%;28/50 = 56%
i. bastante disperso

2.197 a. peso

30 bolsas de M&M's

30 bolsas de M&M

b. = 49.215, mediana = 49.07, $s = 1.522$, min = 46.22, max = 52.06
c. no
f. $\bar{x} = 57.1$, mediana = 58, $s = 2.383$, min = 50, max = 61
g. Una bolsa tiene "sólo 50" M&M's dentro.

2.199 a. ≈ -0.8 o -0.9
b. $\approx +1.6$ o $+1.7$

2.201 puntajes z deben cambiar a percentiles; P_{97}, P_{84}, P_{16}, P_{50}

2.203 $n = 8$, $\Sigma x = 31{,}825$, $\Sigma x^2 = 126{,}894{,}839$
a. 3978.1
b. 203.9
c. 3570.3 a 4385.9

2.205 a. calorías: 40, 111.88, 44.92
sodio (mg): 40, 566.3, 238.4
b. calorías: 22.04 y 201.72
sodio: 89.5 y 1043.1
c. sodio: 327.9 y 804.7, 67.5%

2.207 a.

b. 4248, 1624
c. 986, 3170, 3902, 6058, 650
d. 1000 y 7496
e. 2624 y 5872; 56.3; no

2.217 $n = 3570$; $\Sigma xf = 55{,}155$, $\Sigma x^2 f = 890\,655$

a.

b. 15.4; 16; 16; 11.5; 16
c. 15; 17
d. 14; 14
e. 17; 10.7967; 3.3

2.219 a.

b. $\Sigma f = 500$, $\Sigma xf = 994$, $\Sigma x^2 f = 5200$; 1.988, 1, 0, 4.5
c. 6.46; 2.5
d. 0, 4, 6
e. 2
f. 0, 0, 1, 4, 9

2.221 $n = 220$; $\Sigma xf = 219{,}100$, $\Sigma x^2 f = 224\,470\,000$

a.

b. 995.9
c. 169.2

2.223

b. \$2.00–\$3.00

Capítulo 3

3.1 a. sí
b. un poco

3.3 a.

	En avión	Cuarto de hotel	Todos los otros	Total marginal
Negocios	35.5%	9.5%	5.0%	50%
de placer	25.0%	16.5%	8.5%	50%

b.

	En avión	Cuarto de hotel	Todos los otros	Total marginal
Negocios	71.0%	19.0%	10.0%	100%
de placer	50.0%	33.0%	17.0%	100%

Negocios y placer son distribuciones separadas.

c.

	En avión	Cuarto de hotel	Todos los otros	Total marginal
Negocios	58.7%	36.5%	37.0%	50%
de placer	41.3%	63.5%	63.0%	50%

Cada categoría es una distribución separada.

3.5 a. dultos; género; a los viejos les gustaría quedarse el resto de su vida

b.

c. no

3.7 a. 3350

b. dos variables, afiliación política y cadena de televisión; ambas cualitativas

c. 880

d. 46.9%

e. 19.2%

f. 5.9%

3.9 Este: $\bar{x} = 4.438$, $\tilde{x} = 4.55$; Oeste: $\bar{x} = 4.838$, $\tilde{x} = 4.6$

3.11 a.

b. estaturas de la madre más dispersas

c.

d. Cuando las estaturas de madres aumentaron, las estaturas de hijas aumentaron.

3.13 estatura; el peso se pronostica con frecuencia

3.15 a.

b. Cuando aumentaron horas de estudio, aumentaron calificaciones en examen.

3.17 a. edad, estatura

b. edad = 3 años, estatura = 87 cm

c. El crecimiento es arriba o abajo del normal.

3.19

3.21

3.23 c.

d. no hay relación

3.25 b. fuerte y creciente patrón en los tres
c. Todos son más fuertes.
e. parece exhibir una relación lineal

3.27 a.

b.

c. El tipo 0 muestra un patrón diferente que el 1 y 2

d.

3.29 a. se acerca más a una recta con pendiente positiva
b. se acerca más a una recta con pendiente negativa

3.31 muy poca o ninguna correlación lineal

3.33 a. SS(x) = 10.8; SS(y) = 520; SS(xy) = 46
b. 0.61

3.35 a.

b. SS(x) = 2405.38822
c. SS(y) = 479.44078
d. SS(xy) = 1040.07546
e. 0.97

3.37 a. manatíes, botes de motor
b. número de registros, muertes de manatíes
c. Cuando una aumenta, la otra también.

3.39 a. cerca −0.75
b. SS(x) = 82.0; SS(y) = 67.6; SS(xy) = −60.0; −0.81

3.41 positivo vs. negativo; cercanía a recta, etcétera.

3.43 a.

b. sí
c. SS(x) = 3125.511; SS(y) = 48,505.6; SS(xy) = 12,264.84; r = 0.996
d. Fuerte, correlación positiva

e.

no hay relación lineal
SS(x) = 3794.1; SS(y) = 48 505.6;
SS(xy) = 2044.4; r = 0.15;
poca o ninguna correlación

3.45 a.

b. un poco
c. 0.79
d. sí
e. tiene más cruceros que millas de carretera
f.

más cercanamente sigue una recta
g. 0.892
h. valor positivo más fuerte

3.47 a.

b. lineal; muy diferente; Texas
c. r = 0.91
d.

r = 0.67

3.49 a. SS(x) = 700; SS(y) = 3286.7254;
SS(xy) = 1404.4; r = 0.93
b. **Primas mensuales para seguro de vida a no fumadores**

c. no
d. patrón alargado
e. La prima de un seguro aumenta (acelera) cuando aumenta la edad problema de la persona asegurada; de ahí el patrón "dobla hacia arriba"

3.51 Incendios más grandes exigen camiones más grandes

3.53

no

3.55 a. $\Sigma x^2 = 13\ 717$, SS(x) = 1396.9; $\Sigma y^2 = 15\ 298$, SS(y) = 858.0; $\Sigma xy = 14{,}257$, SS(xy) = 919.0

b. (Σ's) son sumas de datos, (SS( )) son partes de fórmulas

3.57 a. $\hat{y} = 28.1$; $\hat{y} = 47.9$

b. sí

3.59 a.

Lineal

b. SS(x) = 3125.511; SS(xy) = 12 264.84; $\hat{y} = 9.55 + 3.924x$

c. 166.51

d. 264.61

3.61 a. Por cada aumento de 1 pulgada en estatura, el peso aumenta en 4.71 libras.

b. La escala para el eje y empieza en $y = 95$, y la del eje x empieza en $x = 60$.

3.63 a. 121.14 o \$12 114

b. 56.58 o \$5658

c. 21.52(\$100) = \$2152

3.65 a. \$492 411 000

b. \$990 241 000

c. \$1 488 041 000

3.67 la escala vertical está en $x = 58$ y no es el eje y

3.69 5.83, −260.61

3.71 a.

b. sí

c. SS(x) = 17 500; SS(y) = 1222.8571; SS(xy) = −4600; $r = -0.994$

d. negativo muy fuerte

e. sí

f. $= 102.15 - 0.263x$

g.

h. 78.5%

3.73 c.

d. sí

e. $\hat{y} = 163 - 18.5x$

f. Juegos ganados disminuyen en 18.5 por cada punto de aumento en porcentaje de carreras limpias admitidas

g. sí

3.75 a. **Primas de seguro para un seguro de $250 000**

b. SS(x) = 914.117; SS(y) = 3286.725; SS(xy) = 1732.582; r = 0.9996

c. $\hat{y} = -5.83 + 1.90x$

d. prima para hombrees = $22.67

e. hombres pagaron prima más alta

3.77 a. **Quejas por mal manejo de equipaje por pasajero de líneas aéreas —octubre 2004**
The Office of Aviation Enforcement & Proceedings

b. sí

c. $\hat{y} = 1427 + 3.47x$

d. **Quejas por mal manejo de equipaje por pasajero de líneas aéreas —octubre 2004**
Quejas = 1427 + 3466 pasajeros

3.79 a. ambas aumentan al mismo ritmo, no paralelas, una aumenta a ritmo más rápido

b. más conductores que vehículos antes de 1971, luego más vehículos que conductores después conductores de 1973; 1971 a 1973, el número de conductores y vehículos fue casi igual

c. El número de conductores estuvo aumentando más rápido que el número de autos.

d. El número de vehículos de motor está aumentando más rápido que el número de conductores.

3.81 a. Temen: 138, no temen: 362

b.

	Elem.	Sec.	Prep.	Univ.	Adultos
Temen	7.4%	5.6%	5.0%	5.4%	4.2%
No temen	12.6%	14.4%	15.0%	14.6%	15.8%

d.

	Elem.	Sec.	Prep.	Univ.	Adultos
Temen	26.8%	20.3%	18.1%	19.6%	15.2%
No temen	17.4%	19.9%	20.7%	20.2%	21.8%

e. **Temen ir al dentista**

3.87 a. determinar si están linealmente relacionados; el resultado es r

b. determinar la ecuación de la recta de mejor ajuste; el resultado es la ecuación

3.89 a. **Diagrama de dispersión**

b. SS(x) = 4.0; SS(y) = 4.0; SS(xy) = 0.0; r = 0.00

c. $\hat{y} = 2.0 + 0.0x$

3.95 a. **Esperanza de vida en 2004**
The World Factbook

b. sí
c. $\hat{y} = -1.40 + 1.11x$
d. Por cada año adicional de esperanza de vida en hombres, la de mujeres aumenta 1.11 años

3.97 a.

b. $r = 0.3709$

3.99 b.

c. muy alargado; muestra una relación creciente
d. 167 689.2143; 477 586.4286; $\hat{y} = 12.3 + 2.85x$
e. 2.85, por cada 100 acres, 2850 tons. de caña de azúcar producidas

3.101 a.

b. lineal
c. SS(x) = 0.0967333; SS(y) = 181.333; SS(xy) = 2.71667; $r = 0.649$
d. $\hat{y} = 23.0 + 28.1x$
e. 28.62 mm

Capítulo 4

4.1 a. más: amarillo, azul y anaranjado; menos: café, rojo y verde
b. no exactamente, pero similar

4.3 5, 6, 6, 9, 8, 6, respectivamente

4.7 $P'(5) = 0.225$

4.9 a. 50%
b. 0.50
c. misma pregunta pero en dos formatos, porcentaje y formatos diferentes

4.11 a. 0.356
b. 0.389
c. 0.703
d. 0.18

4.13 a. 0.09; 0.95
b. 0.48; 0.27
c. San Diego

4.15 {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}

4.17 $P(5) = 4/36$; $P(6) = 5/36$; $P(7) = 6/36$; $P(8) = 5/36$; $P(9) = 4/36$; $P(10) = 3/36$; $P(11) = 2/36$; $P(12) = 1/36$

4.19 e. probabilidades teóricas

4.21 S = {JH, JC, JD, JS, QH, QC, QD, QS, KH, KC, KD, KS}

4.23 a. S = {\$1, \$5, \$10, \$20}
b.

4.25 a.

b. 0.55
c. 0.35

4.27 0.04; 4%

4.29 Todas son inapropiadas.

4.31 a. espere que ocurra un 1 aproximadamente 1/6 del tiempo cuando tire un solo dado
b. se espera que 50% de los tiros sean caras; el otro 50%, cruces

4.33

4.35 a.

b. {(H,1),(H,2),(H,3),(H,4),(H,5),(H,6), (T,1),(T,2),(T,3),(T,4),(T,5),(T,6)}

4.39 1/5

4.41 4/5

4.43 a. 1/13
b. 12:1

4.45 a. 0.00000000000629908
b. 0.00000153908
c. 6.29908×10^{-12}, 1.53908×10^{-6}

4.47 a. S = {HH, HT, TH, TT}; igualmente probables
b. S = {HH, HT, TH, TT}; no igualmente probables

4.49 a. 0.45
b. 0.40
c. 0.55

4.51 a. 0.35
b. 0.38
c. 0.64
d. 0.03
e. 0.91
f. 0.74
g. 0.98

4.53 a. 0.59
b. 0.41
c. 0.35
d. 0.27
e. 0.30
f. 0.60
g. 0.60
h. formas diferentes de hacer la misma pregunta

4.55 a. algunas categorías se contarían dos veces.
b. 0.08
c. 0.78
d. 0.09
e. 0.39

4.57 a. 0.3
b. 0.22

4.59 0.34

4.61 0.8

4.63 0.2

4.65 0.81

4.67 4%

4.69 0.28

4.71 0.5

4.73 0.098

4.75 0.90

4.77 a.

b. 2/5 o 1/5, dependiendo de la primera selección
c. 0.067
d. 0.067, misma probabilidad

4.79 0.62

4.81 0.133

4.83 a. 0.6
b. 0.7
c. 0.5

4.85 a. 0.4
b. 0.4
4.87 a. no mutuamente exclusivo
b. no mutuamente exclusivo
c. no mutuamente exclusivo
d. mutuamente exclusivo
4.89 No hay intersección.
4.91 a. 0.7
b. 0.6
c. 0.7
d. 0.0
4.93 a. sí
b. no
c. no
d. sí
e. no
f. sí
g. no
4.95 a. A y C y A y E son mutuamente exclusivos.
b. 12/36, 11/36, 10/36
4.97 a. sí
b. sí
c. no
d. 0.307
e. 0.587
f. 0.363
g. 0.559
h. 0.145
i. 0.658
j. 0.081
4.99 0.54
4.101 a. independiente
b. no independiente
c. independiente
d. independiente
e. no independiente
f. no independiente
4.103 0.28
4.105 0.5
4.107 a. 0.12
b. 0.4
c. 0.3
4.109 a. 0.5
b. 0.667
c. no
4.111 a. independiente
b. independiente
c. dependiente
4.113 a. 0.51
b. 0.15
c. 0.1326
4.115 a. 0.0289
b. 0.6889
c. 0.0008
4.117 0.0741
4.119 a. 0.36
b. 0.16
c. 0.48
4.121 a. 3/5
b. 0.16, 0.48, 0.36
4.123 a. de todos modos, de tiempo parcial y graduados son independientes
b. no
c. 0.4074
4.125 a. no puede ocurrir al mismo tiempo
b. El suceso de uno no tiene efecto en la probabilidad del otro.
c. mutuamente exclusivos —ya sea que compartan o no elementos comunes—; independencia —efecto que un evento tiene en la probabilidad de otro evento
4.127 a. 0.25
b. 0.2
c. 0.6
d. 0.8
e. 0.7
f. no
g. no
4.129 a. 0.0
b. 0.7
c. 0.6
d. 0.0
e. 0.5
f. no
4.131 a. 0.625
b. 0.25
c. $P(\text{satisfecho} \mid \text{mujer capacitada}) = 0.25$
$P(\text{satisfecho} \mid \text{mujer no capacitada}) = 0.667$
no independiente
4.133 a. 0.17
b. 0.17
c. 0.41
d. sin
4.135 0.300
4.137 a. $S = \{GGG, GGR, GRG, GRR, RGG, RGR, RRG, RRR\}$
b. 3/8
c. 7/8
4.139 7/8
4.141 a. 0.40
b. 0.49
c. 0.06
d. 0.82
e. 0.40
f. 0.45
4.143 a. Brasil, España, India, etcétera.
b. “con base en países incluidos”
c. 44%
d. 0.44
e. misma pregunta; respuesta en un formato diferente
4.145 $P(\text{A o B}) = P(\text{A}) + P(\text{B}) - P(\text{A y B}) =$
$P(\text{A}) + P(\text{B}) - P(\text{A}) = P(\text{B})$

4.147 a. 0.30
b. 0.60
c. 0.10
d. 0.60
e. 0.333
f. 0.25

4.149 a. 0.3168
b. 0.4659
c. no
d. no
e. "Candidato desea trabajo" y "RJB desea candidato" no podrían ocurrir ambos.

4.151 a. 0.429
b. 0.476
c. 0.905

4.153 a. 0.531
b. 0.262
c. 0.047

4.155 a. 0.5087
b. 0.2076
c. 0.3336
d. 0.56018
e. 0.1993
f. 0.4989
g. mutuamente exclusivo

4.157 a. falso
b. verdadero
c. falso
d. falso

4.159 8/30

4.161 a. 1/2, 1/4, 1/8
b. 9/16, 9/32, 9/64

4.163 0.592

4.165 a. 26/52
b. 26/52
c. 32/52
d. 20/52

4.167 a. 0.60
b. 0.648
c. 0.710
d. (a) 0.70 (b) 0.784 (c) 0.874
e. (a) 0.90 (b) 0.972 (c) 0.997
f. es probable que el "mejor" equipo gane más juegos; mayor diferencia entre equipos

4.169 a. 1/7
b. 1/7

Capítulo 5

5.1 a. 22%
b. más de 4 tazas o latas
c. número de tazas o latas diarias de bebida descafeinada
d. Sí, los eventos (ninguno, uno, dos, tres, 4 o más) no se traslapan.

5.3 número de hermanos $x = 0, 1, 2, 3, \ldots, n$; duración de conversación $x = 0$ a ? minutos

5.5 a. discretas, continuas; continuas, mensurables
b. discretas, cantidad
c. continuas, mensurables

5.7 a. número de nuevos trabajos
b. discreta, continua

5.9 distancia, $x = 0$ a n, n = radio del objetivo, continua

5.11 a. promedio de tiempo empleado en varias actividades
b. continua, mensurable

5.13

x	0	1
$P(x)$	1/2	1/2

5.15 a. Los eventos nunca se traslapan.
b. todos los eventos considerados

5.17 a. $P(x)$ es una función de probabilidad.

x	$P(x)$
1	0.12
2	0.18
3	0.28
4	0.42

b.

5.19

x	0	1	2	3
$P(x)$	0.20	0.30	0.40	0.10

5.21 a.

Número	Proporción	Número	Proporción
0	0.196	5	0.082
1	0.131	6	0.054
2	0.162	7 a 10	0.072
3	0.168	11 o más	0.014
4	0.121		

b. El estudiante escoge entre múltiples opciones.

5.23 los porcentajes no suman 1.00; no es variable aleatoria; variable de atributo

5.27 $\sigma^2 = \Sigma[(x - \mu)^2 P(x)]$
$= \Sigma[x^2 - 2x\mu + \mu^2)P(x)]$
$= \Sigma[x^2P(x) - 2x\mu P(x) + \mu^2 P(x)]$
$= \Sigma[x^2P(x)] - 2\mu\Sigma[xP(x)] + \mu^2[\Sigma P(x)]$
$= \Sigma[x^2P(x)] - 2\mu[\mu] + \mu^2[1]$
$= \Sigma[x^2P(x)] - 2\mu^2 + \mu^2$
$= \Sigma[x^2P(x)] - \mu^2$ o $\Sigma[x^2P(x)] - \{\Sigma[xP(x)]\}^2$

5.29 nada de ningún significado

5.31 2.0, 1.4

5.33 a. 0.54, 0.79
b.

5.35 11.3, 1.35

5.37 a. 2.0, 1.4
b. −0.8 a 4.8 comprende los números 1, 2, 3 y 4
c. 0.9

5.39 a. varía; cercano a −$0.20
b. −$0.20
c. semejantes; no; necesita media = 0

5.41 a. Cada pregunta es un intento separado.
b. cuatro formas diferentes, puede obtenerse una respuesta correcta y tres incorrectas
c. 1/3 es la probabilidad de éxito; 4 es el número de intentos independientes; número de preguntas

5.43 los artículos defectuosos deben ser más bien pequeños y fáciles de contar

5.45 a. 24
b. 5040
c. 1
d. 360
e. 10
f. 15
g. 0.0081
h. 35
i. 10
j. 1
k. 0.4096
l. 0.16807

5.47 $n = 100$ intentos (camisas), dos resultados (primera calidad o irregular), p = P(irregular), $x = n$ (irregular); cualquier valor entero de 0 a 100

5.49 a. Los intentos no son independientes.
b. $n = 4$; as, no as; $p = P(\text{as}) = 4/52$ y $q = P(\text{no as}) = 48/52$; $x = n(\text{ases})$, 0, 1, 2, 3, o 4

5.51 a.

e. $P(x) = \binom{3}{x} p^x q^{3-x}$, para $x = 0, 1, 2, 3$

5.53 $P(x) = \binom{3}{x} (0.5)^x (0.5)^{3-x}$, 0.125, 0.375, 0.125

5.55 a. 0.3585
b. 0.0159
c. 0.9245

5.57 a. 0.4116
b. 0.384
c. 0.5625
d. 0.329218
e. 0.375
f. 0.0046296

5.59 $n = 5$; $p = 1/2$, $q = 1/2$ $(p + q = 1)$; los exponentes suman 5; x = cualquier entero de 0 a $n = 5$; binomial binomial

5.61 0.930

5.63 a. 0.444
b. 0.238
c. 0.153

5.65 a. 0.590
b. 0.918

5.67 0.0011

5.69 0.984

5.71 a. 0.4596
b. 0.0457
c. 0.0042

5.73 0.007

5.75 a. 0.03132
b. 0.99962

5.79 a. 0.6858
b. 0.4458
c. 0.8300
d. 0.7414
e. Los valores de p son casi complementos.
f. Cuando p aumentó, la probabilidad del intervalo aumentó.

5.81 0.9292

5.83 a. $P(2 \text{ de } 2) = 0.308$; $P(9 \text{ de } 15) = 0.274$
b. $P(2 \text{ de } 2) = 0.687$; $P(8 \text{ de } 10) = 0.763$

5.85 18, 2.7

5.87 a. $\Sigma[xP(x)] = 0.55$, $\Sigma[x^2P(x)] = 0.819$, 0.549, 0.72
b. iguales
5.89 a. 25.0, 3.5
b. 4.4, 1.98
c. 24.0, 4.7
d. 44.0, 2.3
5.91 b. 0.4338
c. 4.92, 1.7
d. **En una muestra aleatoria de 12 pasajeros de línea aérea**
Probabilidad de número que se puso auriculares

5.93 $p = 0.5$, $n = 400$
5.95 $\mu = 6$, $\sigma = 1.9$; 0.03383
5.97 b. 0.886385
c. 0.99383
d. 0.12, 0.345
e. −0.225, 0.465; 0.88638
f. −0.57, 0.81; 0.88638
g. no concuerdan con la regla empírica; no concuerda con la de Chebyshev
5.99 1. Cada $P(x)$, es un valor entre 0 y 1 inclusive
2. La suma de todas las $P(x)$ es exactamente 1.
5.101 a. función de probabilidad
b. función de probabilidad
c. NO ES función de probabilidad
d. NO ES función de probabilidad
5.103 a. 0.1
b. 0.4
c. 0.6
5.105 a. 3.3
b. 1.187
5.107 a. 0.930
b. 0.264
5.109 0.103
5.111 a. 0.999
b. 0.206
c. 0.279
5.113 no, la variable es atributo
5.115 0.063
5.117 P(defectuosos) cambios; intentos no son independientes
5.119 a. 0.914
b. 0.625
c. Aun cuando P(defectuosos) cambie de intento en intento, si la población es muy grande las probabilidades son muy semejantes.
5.121 2600
5.123 a. 0.88
b. $B(50, 0.88)$
5.125 taller de herramientas: utilidad media = 7000.0; librería: utilidad media = 6000.0
5.127 $\mu = \Sigma[xP(x)]$
$= (1)(1/n) + (2)(1/n) + \ldots + (n)(1/n)$
$= (1/n)[1 + 2 + 3 + \ldots + n]$
$= (1/n)[(n)(n + 1)/2]$
$= (n + 1)/2$

Capítulo 6

6.1 a. Es un cociente.
b. IQ:100,16; SAT:500,100; calificación estándar: 0, 1
c. $z = (\text{IQ} - 100)/16$; $z = (\text{SAT} - 500)/100$
d. 2, 132, 700
e. iguales
6.3 a. proporción
b. porcentaje
c. probabilidad
6.5 a. media de 0 en forma de campana; desviación estándar de 1
b. referencia usada para determinar las probabilidades para todas las otras distribuciones normales
6.7 a. 0.4032
b. 0.3997
c. 0.4993
d. 0.4761
6.9 0.0212
6.11 0.4177
6.13 0.8571
6.15 a. 0.4394
b. 0.0606
c. 0.9394
d. 0.8788
6.17 a. 0.5000
b. 0.1469
c. 0.9893
d. 0.9452
e. 0.0548
6.19 a. 0.4906
b. 0.9725
c. 0.4483
d. 0.9306
6.21 a. 0.2704
b. 0.8528
c. 0.1056
d. 0.9599
6.23 0.2144
6.25 a. 0.7737
b. 0.8981
c. 0.8983
d. 0.3630

6.27 a. 1.14
b. 0.47
c. 1.66
d. 0.86
e. 1.74
f. 2.23
6.29 a. 1.65
b. 1.96
c. 2.33
6.31 −1.28 o +1.28
6.33 −0.67 y +0.67
6.35 a. 0.84
b. −1.15 y +1.15
6.37 a. 1.28
b. 1.65
c. 2.33
6.39 a. 0.7606
b. 0.0386
c. 0.2689
6.41 2.88
6.43 a. 0.5000
b. 0.3849
c. 0.6072
d. 0.2946
e. 0.9502
f. 0.0139
6.45 a. 0.3944
b. 0.8944
6.47 a. 0.6826 o 68.26%
b. 0.9544 o 95.44%
c. 0.9974 o 99.74%
d. 0.6826 ≈ 68%; 0.9544 ≈ 95%; 0.9974 ≈ 99.7%
6.49 a. 0.0397 o 4.0%
b. 0.1271 o 12.7%
c. 0.9147 o 91.5%
d. 0.3729 o 37.3%
e. 0.0375 o 3.8%
6.51 a. 0.0075
b. 0.3336
6.53 a. 0.3557
b. 100(0.3557) ≈ 36 botellas
6.55 a. de 23.6 mm a 24.6 mm
b. 0.998997 or 99.9%
c. 0.0001 o 0.01%
d. 0.9928 o 99.3%
6.57 a. 89.6
b. 79.2
c. 57.3
6.59 20.26
6.61 7.664
6.63 a. 20%
b. 50%
c. 50.375
d. 49.55
e. 51.395
f. para evitar multas
g. para evitar poner más M&M's en bolsas
6.65 a. 0.056241
b. 0.505544
c. 0.438215
d. 0.0559, 0.5077, 0.4364
e. Redondeo de errores en z
6.71 a. $z(0.03)$
b. $z(0.14)$
c. $z(0.75)$
d. $z(0.22)$
e. $z(0.87)$
f. $z(0.98)$
6.73 a. $z(0.01)$
b. $z(0.13)$
c. $z(0.975)$
d. $z(0.90)$
6.77 a. 1.96
b. 1.65
c. 2.33
6.79 a. 1.41
b. −1.41
6.81 a. 1.28, 1.65, 1.96, 2.05, 2.33, 2.58
b. −2.58, −2.33, −2.05, −1.96, −1.65, −1.28
6.83 a. área, 0.4602
b. puntaje z, 1.28
c. area, 0.5199
d. puntaje z 1.65
6.85 $np = 2$, $nq = 98$; no
6.87 binomial: 0.829; aprox. normal: 0.8133
6.89 0.1812; 0.183
6.91 0.6406; 0.655
6.93 $x = n$(sobreviven), $\mu = 225$, $\sigma = 4.74$; 0.9999997
6.95 $\mu = 25$, $\sigma = 3.54$
a. 0.5557
b. 0.0002
6.97 $\mu = 27.63$, $\sigma = 3.86$; 0.6844
6.99 $\mu = 420$, $\sigma = 16.52$
a. 0.0367
b. 0.034687
c. 0.032952
6.101 al menos $\frac{3}{4}$; 0.9544
6.103 a. 1.26
b. 2.16
c. 1.13
6.105 a. 0.0930
b. 0.9684
6.107 a. 1.175 o 1.18
b. 0.58
c. −1.04
d. −2.33
6.109 a. 0.1867
b. 0.4458
c. 0.1423
6.111 a. 0.0038
b. 0.5610
c. 0.0011

6.113 a. 0.3192
b. 0.1685
c. 156.0 a 184.2 minutos
d. 135.5 a 204.7 minutos

6.115 10.033

6.117 a. 0.0143
b. 619.4
c. 107.2
d. 755.4

6.119 a. $np = 7.5$, $nq = 17.5$; ambos mayores a 5
b. 7.5, 2.29

6.121 b. 0.77023
c. 0.751779

6.123 a. $P(0) + P(1) + \ldots + P(75)$
b. 0.9856
c. 0.9873

6.125 0.0087

6.127 $\mu = 8$, $\sigma = 2.6$
a. 0.0418
b. 0.4247
c. 0.7128

6.129 $\mu = 19.5$, $\sigma = 3.45$
a. 0.3859
b. 0.0735

6.131 $\mu = 130$, $\sigma = 11.1$
a. 0.0322
b. 0.3085

6.133

Nación	$\mu = np$	$\sigma = \sqrt{npq}$	$P(x \geq 70)$*
China	50	6.98	0.0026
Alemania	8	2.82	0.0000+
India	116	10.45	0.9999+
Japón	6	2.45	0.0000+
México	44	6.56	0.0001
Rusia	34	5.78	0.0000+
S. África	124	10.78	0.9999+
Estados Unidos	14	3.73	0.0000+

6.135 a. de −0.0392 a +0.0392
b. 0.8727 o 87.3%
c. 0.8664 o 86.6%

Capítulo 7

7.1 a. histograma

b. acampanado de 0 a 60 sesgado a la derecha
c. parece como población

7.3 a. no
b. variabilidad

7.5 a. distribución formada por medias de todas las muestras posibles de un tamaño fijo tomadas de una población
b. Es un elemento de la distribución muestral.

7.11 a. no todos sacados de la misma población; cada tipo de vehículo tiene un tamaño muestral diferente
b. vigilando las poblaciones de vehículos en tránsito, que continuamente están cambiando

7.17 b. muy cercana a $\mu = 65.15$
d. aproximadamente normal
e. tomó muchas muestras (1001) de tamaño 4 de una población aproximadamente normal; (1) media de las barra $x \approx \mu$, (2) $s_{\bar{x}} \approx \sigma/\sqrt{n}$, (3) distribución aproximadamente normal

7.19 a. 1.0
b. $\sigma_{\bar{x}} = \sigma/\sqrt{n}$; cuando n aumenta, el valor de esta fracción se hace más pequeño

7.21 a. 500
b. 5
c. aproximadamente normal

7.23 a. aproximadamente normal
b. 4.0 horas
c. 0.133

7.25 a. 17.71 libras/persona
b. 0.514
c. aproximadamente normal

7.29 2.69

7.31 0.4772

7.33 a. aproximadamente normal
b. 50
c. 1.667
d. 0.9974
e. 0.8849
f. 0.9282

7.35 a. aproximadamente normal, $\mu = 69$, $\sigma = 4$.
b. 0.4013
c. aproximadamente normal

d. $\mu_{\bar{x}} = 69$; $\sigma_{\bar{x}} = 1.0$
e. 0.1587
f. 0.0228

7.37 a. 0.3830
b. 0.9938
c. 0.3085
d. 0.0031

7.39 a. 0.2643
b. 0.0294
c. No, especialmente para la parte a; la velocidad del viento estará sesgada a la derecha, no normal.
d. Es factible que las probabilidades reales no sean tan altas como se encontraron.

7.41 a. 0.9821
b. 0.0001
c. La distribución normal debe tomar en cuenta estimaciones razonables porque $n > 30$.

7.43 38.73 pulgadas

7.45 a. computadora: 0.68269; tabla 3: 0.6826

7.47 6.067, 3.64, 2.6, 1.82

7.49 a. normalmente distribuida con una media = \$675 y una desviación estándar = \$85
b. 0.8900
c. normalmente distribuida en forma aproximada con una media =\$675 y error estándar = \$17
d. 0.8900

7.51 a. $e = 0.49$
b. $E = 0.049$

7.53 a. 0.0060
b. 0.1635
c. distribución sesgada

7.55 a. 0.1498
b. 0.0089

7.57 0.0228

7.59 0.0023

7.61 a. Σx peso total; aproximadamente 1.000
b. 0.9772

7.65 a. $\mu = 60$; $\sigma = 6.48$

Capítulo 8

8.1 a. mujeres profesionales de la salud
b. $\bar{x} = 64.7$, $s = 3.5$; acampanada hacia el centro, aproximadamente simétrica

8.3 La estimación puntual es un solo número; la estimación de intervalo es un intervalo de algún ancho

8.5 $n = 15$, $\Sigma x = 271$, $\Sigma x^2 = 5015$
a. 18.1 dólares
b. 8.5
c. 2.9 dólares

8.7 a. II tiene menos variabilidad.
b. II tiene un valor medio igual al parámetro.
c. Ninguna es una buena opción.

8.9 dificultad, fatiga escolar; costo de muestreo; destrucción de producto

8.11 3

8.13 a. 2(0.3997) = 0.7994
b. 2(0.4251) = 0.8502
c. 2(0.4750) = 0.9500
d. 2(0.4901) = 0.9802

8.15 19(\$17 320) + 6(\$20 200) = \$450 280 para 25 proyectos; 174(\$450 280/25)= \$3 133 948.80

8.17 a. entre 11:17 y 11:37 A.M.
b. Sí, 11.20:48 A.M. está dentro del intervalo 11:17 a 11:37 A.M.
c. 90% ocurre dentro del intervalo pronosticado

8.19 La distribución muestral de medias muestrales debe ser normal

8.21 a. $z(0.01) = 2.33$
b. $z(0.005) = 2.58$

8.23 a. 25.76 a 31.64
b. Sí, la población es normal.

8.25 a. 125.58 a 131.42
b. sí, Teorema del Límite Central

8.27 a. 128.5
b. $z(0.05) = 1.65$
c. 1.76845
d. 2.92
e. 125.58
f. 131.42

8.31 a. 15.9; ≈ 68%
b. 31.4; ≈ 95%
c. 41.2; ≈ 99%
d. el nivel más alto contribuye a un ancho más grande

8.33 a. 75.92
b. 0.368
c. 75.552 a 76.288

8.35 a. 14.01 a 14.59
b. 13.89 a 14.71

8.37 487.5 a 520.5

8.39 a. longitud media
b. 75.92
c. 75.512 a 76.328

8.41 a. 276.9 a 314.4
b. 284.69 a 344.01
c. Medias muestrales diferentes dieron puntos de centro diferentes.
d. Disminución en tamaño muestral; aumento en ancho
e. 310 está en ambos intervalos, ningún intervalo da razón para duda

8.43 Los números están calculados para muestras de nieve.
8.45 49
8.47 27
8.49 25
8.51 H_o: el sistema es confiable; H_a: el sistema no es confiable
8.53 a. H_a: el correo de entrega inmediata tarda demasiado.
b. H_a: El nuevo diseño es más cómodo.
c. H_a: El humo de cigarrillos tiene un efecto.
d. H_a: Un acondicionador del pelo es eficaz en "puntas abiertas".
8.55 A: La fiesta será un desastre; no fue.
B: La fiesta será un éxito, fue.
I: La fiesta será un desastre, fue.
II: La fiesta será un éxito, no fue.
8.57 a. H_a: La víctima no está viva
b. A: viva, tratada como si estuviera viva
I: viva, tratada como si estuviera viva
II: muerta, tratada como si estuviera viva
B: muerta, tratada como si estuviera muerta
c. I muy grave; la víctima puede morir pronto sin atención. II no es grave; la víctima está recibiendo atención que no es de valor.
8.59 se perdió de una gran fiesta
8.65 a. tipo I b. tipo II c. tipo I d. tipo II
8.67 a. El comercial no es eficaz.
b. El comercial es eficaz.
8.69 a. muy grave
b. un poco grave
c. nada grave
8.71 a. α
b. β
8.73 α es la probabilidad de rechazar una hipótesis nula VERDADERA; $1-\beta$ es la probabilidad de rechazar una hipótesis nula FALSA.
8.75 a. "Ves, te lo dije."
b. "Bien, no importa. Lo intentaré de nuevo mañana."
8.77 a. 0.1151
b. 0.2119
8.79 a. 29 corchos
b. lote rechazado; 3 corchos no satisfacen la especificación.
8.81 H_o: La resistencia media al corte es al menos 925 lb.
H_a: La resistencia media al corte es menor a 925 lb.
8.83 a. H_a: $\mu > 1.25$
b. H_a: $\mu < 335$
c. H_a: $\mu \neq 230{,}000$
d. H_a: $\mu > 210$
e. H_a: $\mu > 9.00$
8.85 II; comprar y usar remaches débiles
8.87 I: rechazar H_o se interpreta como "el cargo medio por hora es menor a \$60" cuando, de hecho, es al menos \$60
II: no rechazar H_o interpretado como "el cargo medio por hora es al menos \$60" cuando, de hecho, es menos de \$60.
8.89 a. 1.26
b. 1.35
c. 2.33
d. -0.74
8.91 a. rechazar H_o, no rechazar H_o
b. El valor p es menor o igual a α, rechazar H_o. El valor p es mayor a α, no rechazar H_o.
8.93 a. rechazar H_o, $\mathbf{P} < \alpha$
b. no rechazar H_o, $\mathbf{P} > \alpha$
c. rechazar H_o, $\mathbf{P} < \alpha$
d. rechazar H_o, $\mathbf{P} < \alpha$
8.95 a. no rechazar H_o
b. rechazar H_o
8.97 b. ≈ 0.0000
d. rechazar H_o
8.101 0.2714
8.103 a. 0.0694
b. 0.1977
c. 0.2420
d. 0.0174
e. 0.3524
8.105 a. 1.57
b. -2.13
c. -2.87, $+2.87$
8.107 6.67
8.109 a. H_a: $\mu < 525$
b. no rechazar H_o
c. 9.733
8.111 a. H_a: $\mu \neq 6.25$
b. rechazar H_o
c. 0.1585
d. 514.488, 3518.3437
8.113 a. precio medio para todas las portátiles
b. H_a: $\mu <$ \$1240
c. -2.57; $\mathbf{P} = 0.0051$
d. no rechazar H_o
8.115 H_a: $\mu < 12$; $z\bigstar = -2.53$; $\mathbf{P} = 0.0057$; rechazar H_o
8.117 H_a: $\mu <$ \$104.63; $z\bigstar = -2.72$; $\mathbf{P} = 0.0033$; rechazar H_o
8.119 a. precisión media de relojes de cuarzo
b. H_a: $\mu > 20$
c. normalidad supuesta, $n = 36$; $\sigma = 9.1$
d. $n = 36$, $\bar{x} = 22.7$
e. $z\bigstar = 1.78$; $\mathbf{P} = 0.0375$
f. $\mathbf{P} < \alpha$; rechazar H_o
8.123 H_o: la resistencia media al corte es al menos 925 lb.
H_a: la resistencia media al corte es menor a 925 lb.
8.125 a. H_a: $\mu < 1.25$
b. H_a: $\mu \neq 335$
c. H_a: $\mu > 230\ 000$
8.127 a. el contenido promedio de sal decidido es más de 350 mg cuando, de hecho, no lo es
b. el contenido promedio de sal decidido es menor o igual a 350 mg cuando, de hecho, es mayor

8.129 I: el mínimo medio decidido es mayor a \$85 cuando, de hecho, no lo es
II: el mínimo medio decidido es a lo sumo \$85 cuando, de hecho, es mayor

8.131 a. el conjunto de todos los valores de la estadística de prueba que nos hará rechazar H_o
b. El valor o valores crítico(s) es (son) valor(es) de la estadística de prueba que forman frontera entre la región crítica y la región no crítica; el valor crítico está en la región crítica.

8.133 Si uno se reduce, el otro se hace más grande.

8.135 $z \leq -2.33$

8.137 a. $z \leq -1.63$, $z \geq 1.65$
b. $z \geq 2.33$
c. $z \leq -1.65$
d. $z \leq -2.58$, $z \geq 2.58$

8.139 $\bar{x} = 247.1$; $\Sigma x = 21\ 004.133$

8.141 a. 3.0 errores estándar
b. La región crítica es $z \geq 2.33$, rechazar H_o.

8.143 a. rechazar H_o o no rechazar H_o
b. La estadística de prueba calculada cae en la región crítica; rechazar H_o. La estadística de prueba calculada cae en la región no crítica; no rechazar H_o.

8.145 a. H_o: $\mu = 15.0$ vs. H_a: $\mu \neq 15.0$
b. ± 2.58; rechazar H_o
c. 0.0913

8.147 a. H_o: $\mu = 72$ ($\leq$) vs. H_a: $\mu > 72$
b. no rechazar H_o
c. 2.0

8.149 H_a: $\mu > 79.68$; se supone normalidad, $n = 50$; $z\bigstar = 1.07$; $z(0.05) = 1.65$; no rechazar H_o

8.151 H_a: $\mu \neq 55$; se supone normalidad, $n = 35$; $z\bigstar = -1.70$; $z \leq -1.65$, $z \geq 1.65$; rechazar H_o

8.153 H_a: $\mu > 170.1$; normalidad indicada; $z\bigstar = 2.73$; $z \geq 1.65$; rechazar H_o

8.155 H_a: $\mu > 36.8$; se supone normalidad, $n = 42$; $z\bigstar = 1.45$; $z \geq 1.65$; no rechazar H_o

8.159 a. 32.0
b. 2.4
c. 64
d. 0.90
e. 1.65
f. 0.3
g. 0.495
h. 32.495
i. 31.505

8.161 43.3 a 46.7

8.163 a. 9.75 a 9.99
b. 9.71 a 10.03
c. amplió el intervalo

8.165 a. 69.89 a 75.31
b. sí

8.167 162.04 a 176.16

8.169 b. 2.00; $129.02 \pm (1.96)(2.00)$

8.171 92

8.173 60

8.175 a. "frontera" para decisión
b. ninguna

8.177 a. H_o: $\mu = 100$
b. H_a: $\mu \neq 100$
c. 0.01
d. 100
e. 96
f. 12
g. 1.70
h. -2.35
i. 0.0188
j. No rechazar H_o

8.179 H_a: $\mu > 45$; $z\bigstar = 2.47$; **P** = 0.0068; $z \geq 2.05$; rechazar H_o

8.181 a. H_a: $\mu \neq 0.50$
b. 0.2112
c. $z \leq -2.33$, $z \geq 2.33$

8.183 H_a: $\mu > 0.0113$; $z\bigstar = 2.47$; **P** = 0.0068; $z \geq 2.33$; rechazar H_o

8.185 H_a: $\mu > 9$; $z\bigstar = 3.14$; **P** = 0.0008; $z \geq 2.05$; rechazar H_o

8.187 a. H_a: $\mu < 129.2$; $z\bigstar = -2.02$; **P** = 0.0217; no rechazar H_o

8.189 a. H_a: $\mu \neq 18$; no rechazar H_o; la media poblacional no es significativamente diferente de 18
b. 0.756; $z\bigstar = -1.04$; valor $p = 0.2984$

8.191 a. 39.6 a 41.6
b. H_a: $\mu \neq 40$; $z\bigstar = 1.20$; **P** = 0.2302; no rechazar H_o
c. H_a: $\mu \neq 40$; $z\bigstar = 1.20$; $z \leq -1.96$, $z \geq 1.96$; no rechazar H_o

8.193 a. 39.9 a 41.9
b. H_a: $\mu > 40$; $z\bigstar = 1.80$; **P** = 0.0359; rechazar H_o
c. H_a: $\mu > 40$; $z\bigstar = 1.80$; $z \geq 1.65$; rechazar H_o

8.195 a. H_a: $r > A$, carga sobre medicina vieja
b. H_a: $r < A$, carga sobre medicina nueva

Capítulo 9

9.1 a. tiempo medio de ejercicio por semana para mujeres
b.

c. sesgado a la derecha o bimodal
d. sí

9.5 a. 2.68
b. 2.07
b. 1.30
d. 3.36

9.7 a. −1.33
b. −2.82
c. −2.03
d. computadora: −2.26; tabla 6: $-2.14 < t < -2.62$; interpolación: −2.30

9.9 a. 1.73
b. ±3.18
c. −2.55
d. 1.33

9.11 ±2.18

9.13 a. −2.49
b. 1.71
c. −0.685

9.15 df = 7

9.17 0.0241

9.19 a. Simétrica alrededor de la media: la media es 0.
b. La desviación estándar de t es mayor a 1; t tiene grado de libertad (df); la distribución t es una familia de distribuciones; una distribución z.

9.21 15.60 a 17.8

9.23 $82.63 a $87.37

9.25 a. $87.37
b. $15.78
c. $83.23 a $91.51

9.27 a. 88 oz; 4.68 lb
b. **Gráfica de probabilidad de peso (onzas)**

c. 74.10; 3.182
d. 71.95 a 76.25
e. 71 950 a 76 250 oz

9.31 (6.073, 9.427)

9.33 a. **Histograma (con curva normal) de millas/galón**

b. 18.07 a 18.83

9.35 a. H_o: $\mu = 11$ (≥) vs. H_a: $\mu < 11$
b. H_o: $\mu = 54$ (≤) vs. H_a: $\mu > 54$
c. H_o: $\mu = 75$ vs. H_a: $\mu \neq 75$

9.37 1.20

9.39 a. $0.025 < \mathbf{P} < 0.05$
b. $0.025 < \mathbf{P} < 0.05$
c. $0.05 < \mathbf{P} < 0.10$
d. $0.05 < \mathbf{P} < 0.10$

9.41 a. $0.10 < \mathbf{P} < 0.25$; no rechazar H_o
b. 1.75; no rechazar H_o

9.43 0.124

9.45 a. $0.10 < \mathbf{P} < 0.20$; $t = \pm 2.14$; no rechazar H_o
b. $0.01 < \mathbf{P} < 0.025$; $t \geq 1.71$; rechazar H_o
c. $0.025 < \mathbf{P} < 0.05$; $t \leq -1.68$; rechazar H_o
d. idénticos

9.47 H_a: $\mu < 25$; $t\bigstar = -3.25$; $\mathbf{P} < 0.005$; $t \leq -2.46$; rechazar H_o

9.49 H_a: $\mu > 130$; $t\bigstar = 1.53$; $0.05 < \mathbf{P} < 0.10$; $t \geq 2.50$; no rechazar H_o

9.53 no rechazar H_o

9.55 H_a: $\mu \neq 35$; $t\bigstar = 1.02$; $0.20 < \mathbf{P} < 0.50$; $\pm t(5, 0.025) = \pm 2.57$; no rechazar H_o

9.57 a. **Histograma de densidad, con curva normal**

b. H_a: $\mu < 5.517$; $t\bigstar = -1.68$; $0.05 < \mathbf{P} < 0.10$; $t \leq -1.70$; no rechazar H_o

9.59 a. sí
b. 593.93 a 598.67
c. La media es menor a 600 mg.
d. H_a: $\mu \neq 24.0$; $t\bigstar = 0.69$; $\mathbf{P} = 0.500$; $\pm t(17, 0.01) = \pm 2.57$; no rechazar H_o
e. La parte d fue mucho menor.

9.61 a. H_a: $\mu \neq 45.0$; $t\bigstar = -1.95$; $0.05 < \mathbf{P} < 0.10$; $\pm t(11, 0.01) = \pm 2.72$; no rechazar H_o
b. H_a: $\mu \neq 45.0$; $t\bigstar = 0.24$; $\mathbf{P} > 0.500$; $\pm t(17, 0.01) = \pm 2.57$; no rechazar H_o

9.63 a. número de éxitos, tamaño muestral
b. 0.30
c. 0.096
d. 0.312
e. 0.697

9.65 a. sí
b. La media de p' es p.

9.67 a. $z(\alpha/2) = 1.65$
b. $z(\alpha/2) = 1.96$
c. $z(\alpha/2) = 2.58$

9.69 a. 0.02104
b. 0.189 a 0.271

9.71 a. $p = P$(no sabían)
b. 0.20, estadística
c. 0.0143
e. 0.186 a 0.214

9.73 0.206 a 0.528

9.75 0.250 a 0.306

9.79 a. 0.028, 0.030, 0.022
b. productos diferentes de pq
c. sí
e. 0.5

9.83 a. 0.5005
b. 0.003227
c. 0.4942 a 0.5068

9.87 2401

9.89 a. 1048
b. 262
c. 2089
d. Aumentar máximo error disminuye tamaño muestral.
e. Aumentar nivel de confianza aumenta tamaño muestral.

9.91 a. H_a: $p > 0.60$
b. H_a: $p > 1/3$
c. H_a: $p > 0.50$
d. H_a: $p < 0.75$
e. H_a: $p \neq 0.50$

9.93 a. 1.78
b. −1.70
c. −0.49
d. 1.88

9.95 a. 0.1388
b. 0.0238
c. 0.1635
d. 0.0559

9.97 a. $z \geq 1.65$
b. $z \leq -1.96$, $z \geq 1.96$
c. $z \leq -1.28$
d. $z \geq 1.65$

9.99 a. 0.017
b. 0.085
c. 0.101
d. 0.004

9.101 a. correctamente no rechazar H_o
b. 0.036
c. cometer un error tipo II
d. 0.128

9.103 H_a: $p < 0.90$; $z\bigstar = -4.82$; $\mathbf{P} = 0.000003$; $z \leq -1.65$; rechazar H_o

9.105 H_a: $p < 0.60$; $z\bigstar = -2.04$; $\mathbf{P} = 0.0207$; $z \leq -1.65$; rechazar H_o

9.107 H_a: $p < 0.72$; $z\bigstar = -1.57$; $\mathbf{P} = 0.0582$; $z \leq -1.65$; no rechazar H_o

9.109 a. −1.72
b & c. $\mathbf{P} = 0.0427$; $z \leq -2.33$; no rechazar H_o

9.111 a. 0.0438
b. 0.0332
c. 0.04375
d. el de Coenen: "acuñado bien o balanceado" o H_o: $P(\text{H}) = 0.5$; el de Blight: "no balanceado" o H_a: $P(\text{H}) \neq 0.5$
e. La probabilidad de "obtener un resultado tan extremo" es la definición del valor p.
f. 0.498 a 0.622
g. experimento binomial —no hay probabilidad de sesgo debido a un proceso de entrevista

9.113 b. rechazar H_o
c. 0.305

9.115 a. A: 1.72; B: 3.58
b. aumentó
c. muy diferente del resto de los datos; tuvo un efecto grande en la desviación estándar

9.117 a. 23.2
b. 23.3
c. 3.94
d. 8.64

9.119 a. 30.1
b. 13.3
c. 7.56
d. 43.2
e. 11.6 y 32.7
f. 1.24 y 14.5

9.121 a. $\chi^2(5, 0.05) = 11.1$
b. $\chi^2(5, 0.05) = 11.1$
c. $\chi^2(5, 0.10) = 9.24$

9.123 0.94

9.125 a. 0.8356
b. 0.1644

9.127 a. H_a: $\sigma > 24$
b. H_a: $\sigma > 0.5$
c. H_a: $\sigma \neq 10$
d. H_a: $\sigma^2 < 18$
e. H_a: $\sigma^2 \neq 0.025$

9.129 a. 25.08
b. 60.15

9.131 a. $0.02 < \mathbf{P} < 0.05$
b. 0.01

c. $0.05 < \mathbf{P} < 0.10$
d. $0.025 < \mathbf{P} < 0.05$

9.133 a. $0.05 < \mathbf{P} < 0.10$; no rechazar H_o
b. $\chi^2 \geq 33.4$; no rechazar H_o

9.135 H_a: $\sigma \neq 8$; $\chi^2\bigstar = 29.3$; $0.01 < \mathbf{P} < 0.02$; $\chi^2 \leq 32.4$, $\chi^2 \geq 71.4$; rechazar H_o

9.137 b.

c & d. H_a: $\sigma > 1.0$; $\chi^2\bigstar = (23)(1.916^2)/(1.0^2) = 84.43$; $\mathbf{P} < 0.005$; $\chi^2 \geq 41.6$; rechazar H_o

9.139 H_a: $\sigma > 85$; $\chi^2\bigstar = 64.88$; $\mathbf{P} < 0.005$; $\chi^2 \geq 43.8$; rechazar H_o

9.141 a. H_a: $\sigma \neq 0.3275$; $\chi^2\bigstar = 7.15$; $0.50 < \mathbf{P} < 1.00$; $\chi^2 \leq 2.09$, $\chi^2 \geq 21.7$; no rechazar H_o
b. H_a: $\sigma \neq 0.3275$; $\chi^2\bigstar = 13.97$; $0.20 < \mathbf{P} < 0.50$; $\chi^2 \leq 2.09$, $\chi^2 \geq 21.7$; no rechazar H_o
c. Desviaciones muestrales estándar aumentan el valor de Ji cuadrada.

9.143 0.0359

9.147 35 524 a 36 476

9.149 a. 8.782, 0.710
b. 8.78
c. 8.64 a 8.92

9.151 72

9.153 a. $\bar{x} = \$908.30$, $s = \$118.50$

c. \$823.61 a \$992.99

9.155 H_a: $\mu > 4.35$; $t\bigstar = 1.63$; $0.05 < \mathbf{P} < 0.10$; $t \geq 1.76$; no rechazar H_o

9.157 a. 31.45, 8.049
b. H_a: $\mu > 28.0$; $t\bigstar = 1.92$; $0.025 < \mathbf{P} < 0.05$; $t \geq 1.73$; rechazar H_o

9.159

H_a: $\mu \neq 734.5$; $t\bigstar = 14.92$; $\mathbf{P} = 0.00+$; $t \leq -2.65$, $t \geq 2.65$; rechazar H_o

9.161 0.122 a 0.278

9.163 a. 0.126 a 0.340
b. porcentaje sobreestimado de clientes satisfechos

9.165 a. 343
b. 121
c. 93

9.167 3258

9.169 0.0401

9.175 a. parámetro; binomial p, P(éxito)
b. 0.60 a 0.66

9.177 H_a: $\sigma > 81$; $\chi^2\bigstar = 123.1$; $0.05 < \mathbf{P} < 0.10$; $\chi^2 \geq 124.0$; no rechazar H_o

9.179 H_a: $\sigma > \$2.45$; $\chi^2\bigstar = 101.5$; $0.005 < \mathbf{P} < 0.01$; $\chi^2 \geq 90.5$; rechazar H_o

9.181 b. H_a: $\mu \neq 2.0$; $t\bigstar = 3.08$; $\mathbf{P} < 0.010$; $t \leq -2.04$, $t \geq 2.04$; rechazar H_o
c. H_a: $\sigma > 0.040$; $\chi^2\bigstar = 48.96$; $0.01 < \mathbf{P} < 0.25$; $\chi^2 \geq 43.8$; rechazar H_o

9.183 a.

b. $\bar{x} = 14.386$, $s = 0.217$
c. 5%

d. **Gráfica de probabilidad de peso (onzas)**
Normal

e. $\bar{x} = 14.386$, $s = 0.217$; 14.285 a 14.487
f. H_a: $\sigma > 0.2$; $\chi^2\bigstar = 22.37$; $\mathbf{P} = 0.2662$; $\chi^2 \geq 36.2$; no rechazar H_o

9.185 a. 0.8051
b. 0.1271
c. 1016.46 cajas

Capítulo 10

10.1 a. estudiantes universitarios
b. F: 48.5%; 54%; S: 93.5%; 92%
d. **Deuda de tarjeta de crédito**
Alumnos 1er. año ($n = 40$)

Deuda de tarjeta de crédito
Alumnos 2o. año ($n = 44$)

Ambos acampanados y sesgados a la derecha; los alumnos de segundo año más dispersos
e. F: \$1519, \$1036; S: \$2079, \$1434

10.5 independientes; las muestras son conjuntos separados de estudiantes

10.7 independientes; cada persona proporciona un dato para cada muestra

10.9 independientes; muestras separadas

10.13 a. $d = A - B$: 1 1 0 2 −1
b. 0.6
c. 1.14

10.15 a. 4.24 a 8.36
b. intervalo de confianza más angosto

10.19 $n = 8$, $\Sigma d = 8$, $\Sigma d^2 = 48$
a. 1.0
b. −1.53 a 3.53

10.21 $d = \text{I} - \text{II}$; $n = 10$, $\bar{d} = 0.8$, $s_d = 1.32$; −0.143 to 1.743

10.23 a. H_a: $\mu_d > 0$; d = después de examen − antes de examen
b. H_a: $\mu_d \neq 0$; d = después − antes
c. H_a: $\mu_d \neq 0$; d = lectura1 − lectura2
d. H_a: $\mu_d > 0$; d = después de calificación − antes de calificación

10.25 a. $P(t > 1.86 \mid \text{df} = 19)$; $0.025 < \mathbf{P} < 0.05$
b. $2P(t < -1.86 \mid \text{df} = 19)$; $0.05 < \mathbf{P} < 0.10$
c. $P(t < -2.63 \mid \text{df} = 28)$; $0.005 < \mathbf{P} < 0.01$
d. $P(t > 3.57 \mid \text{df} = 9)$; $\mathbf{P} < 0.005$

10.27 H_a: $\mu_d > 0$ (beneficio); $t\bigstar = 3.067$; $\mathbf{P} < 0.005$; $t(39, 0.01) = 2.44$; rechazar H_o

10.31 $t\bigstar = 2.45$; $0.025 < \mathbf{P} < 0.05$; $t(4, 0.05) = 2.13$; rechazar H_o

10.33 $t\bigstar = 1.35$; $0.20 < \mathbf{P} < 0.50$; $\pm t(4, 0.005) = \pm 4.60$; no rechazar H_o

10.35 H_a: $\mu_d > 0$ (mejoría); $t\bigstar = 0.56$; $\mathbf{P} > 0.25$; $t(9, 0.05) = 1.83$; no rechazar H_o

10.37 a. La diferencia promedio es cero.
b. valores usados para tomar la decisión
c. La prueba es de dos colas; la distribución t es simétrica; la ausencia de números negativos la hace menos confusa.
d. no rechazar la hipótesis nula en 12 de ellas
e. Dos métodos son equivalentes.
f. método revisado de Florida para muestreo aceptado e implantado

10.39 4.92

10.41 caso I: entre 17 y 40; caso II: 17

10.43 −6.3 a 16.3

10.45 \$4.19 a \$19.11

10.49 N. Dakota: $n = 11$, $\bar{x} = 976.2$, $s = 255.7$; S. Dakota: $n = 14$, $\bar{x} = 1370$, $s = 397$; 101.3 to 686.3

10.51 a. H_a: $\mu_1 - \mu_2 \neq 0$
b. H_a: $\mu_1 - \mu_2 > 0$
c. H_a: $\mu_S - \mu_N > 0$
d. H_a: $\mu_M - \mu_F \neq 0$

10.53 a. 1.21
b. 0.1243
c. 1.56

10.55 2.64

10.57 a. ±2.13
b. −2.48

c. 1.42
d. ±2.16

10.61 $2P(t > 1.44 \mid \text{df} = 13)$; $0.10 < \mathbf{P} < 0.20$

10.63 H_a: $\mu_1 - \mu_2 > 0$; $t\bigstar = 4.02$; $\mathbf{P} < 0.005$; $t \geq 2.44$; rechazar H_o

10.65 no

10.67 c. 12 a 26
d. computadora
e. 12
f. el valor p es menor a 0.0005.

10.69 b. $0.554 < \mathbf{P} < 0.624$
c. $0.560 < \mathbf{P} < 0.626$

10.71 H_a: $\mu_B - \mu_A > 0$; $t\bigstar = 1.98$; $0.025 < \mathbf{P} < 0.05$; $t \geq 1.83$; rechazar H_o

10.73 H_a: $\mu_{PN} - \mu_{PR} > 0$; $t\bigstar = 1.36$; $0.10 < \mathbf{P} < 0.25$; $t \geq 1.53$; no rechazar H_o

10.79 75, 250, 0.30, 0.70

10.81 a. 0.085
b. 0.115

10.85 0.196 a 0.384; la diferencia positiva indica que la proporción de adolescentes hombres es significativamente mayor que la proporción de adolescentes mujeres

10.87 0.000 a 0.080

10.89 a. H_a: $p_m - p_w \neq 0$
b. H_a: $p_b - p_g > 0$
c. H_a: $p_c - p_{nc} > 0$

10.91 0.076; 0.924

10.93 1.34; 0.0901

10.95 a. $z \geq 1.65$
b. $z \leq -1.96$, $z \geq 1.96$
c. $z \leq -1.75$
d. $z \geq 2.33$

10.97 H_a: $p_m - p_c \neq 0$; $z\bigstar = 1.42$; $\mathbf{P} = 0.1556$; $z \leq -1.96$ y $z \geq 1.96$; no rechazar H_o

10.99 b. H_a: $p_w - p_m \neq 0$; $z\bigstar = 0.64$; $\mathbf{P} = 0.5222$; $z \leq -1.96$ y $z \geq 1.96$; no rechazar H_o
c. $z\bigstar = 3.18$; $\mathbf{P} = 0.0014$; $z \leq -1.96$ y $z \geq 1.96$; rechazar H_o
d. es necesario un tamaño muestral razonablemente grande para mostrar significación

10.101 H_a: $p_2 - p_1 \neq 0$; $z\bigstar = 1.43$; $\mathbf{P} = 0.1528$; $z \leq -1.96$ y $z \geq 1.96$; no rechazar H_o

10.103 a. H_a: $p_M - p_W \neq 0$; $z\bigstar = 1.82$; $\mathbf{P} = 0.0688$; $z \leq -1.96$ y $z \geq 1.96$; no rechazar H_o
b. $z\bigstar = 2.57$; $\mathbf{P} = 0.0102$; $z \leq -1.96$ y $z \geq 1.96$; rechazar H_o
c. 291

10.105 a. H_a: $\sigma_A^2 \neq \sigma_B^2$
b. H_a: $\sigma_{\text{I}}/\sigma_{\text{II}} > 1$
c. H_a: $\sigma_A^2/\sigma_B^2 \neq 1$
d. H_a: $\sigma_B^2/\sigma_C^2 > 1$

10.107 Dividir la desigualdad σ_p^2.

10.109 a. $F(9, 11, 0.025)$
b. $F(24, 19, 0.01)$
c. $F(8, 15, 0.01)$
d. $F(15, 9, 0.05)$

10.111 a. 2.51
b. 2.20
c. 2.91
d. 4.10
e. 2.67
f. 3.77
g. 1.79
h. 2.99

10.113 3.37

10.115 1.52

10.117 0.495; varianza menor en el numerador

10.119 H_a: $\sigma_k^2 \neq \sigma_m^2$; $F\bigstar = 1.33$; $\mathbf{P} > 0.10$; $F \geq 3.73$; no rechazar H_o

10.121 H_a: $\sigma_{sc} \neq \sigma_{sb}$; $F\bigstar = 1.12$; $\mathbf{P} > 0.10$; $F \geq 3.10$; no rechazar H_o

10.123 multiplicar por 2

10.125 $(4.43)^2/(3.50)^2 = 1.60$

10.127 a. $\bar{x}_1 = 0.01525$, $s_1 = 0.00547$; $\bar{x}_2 = 0.02856$, $s_2 = 0.00680$
b. H_a: $\sigma_1^2 \neq \sigma_2^2$; $F\bigstar = 1.55$; $\mathbf{P} > 0.10$; $F \geq 4.42$; no rechazar H_o
c. H_a: $\mu_2 - \mu_1 \neq 0$; $t\bigstar = 5.64$; $\mathbf{P} < 0.01$; $t \leq -2.36$, $t \geq 2.36$; rechazar H_o

10.129 a. hombres: $x_m = \$68.14$, $s_m = \$47.95$; mujeres: $\bar{x}_w = \$85.90$, $s_w = \$63.50$

d. H_a: $\mu_w - \mu_m > 0$; $t\bigstar = 1.36$; $0.05 < \mathbf{P} < 0.10$; $t \geq 1.70$; no rechazar H_o

e. H_a: $\sigma_w \neq \sigma_m$; $F\bigstar = 1.75$; $\mathbf{P} \approx 0.10$; $F \geq 2.72$; no rechazar H_o

f. La diferencia no fue significativa.

10.133 −8.85 a 16.02

10.135 0.95 a 3.05

10.137 H_a: $\mu_d > 0$; $t\bigstar = 3.70$; $\mathbf{P} < 0.005$; $t \geq 2.54$; rechazar H_o

10.139 −0.21 a 10.61

10.141 −9.14 a 19.54

10.143 0.012 a 0.072

10.145 H_a: $\mu_2 - \mu_1 > 0$; $t\bigstar = 0.988$; $0.10 < \mathbf{P} < 0.25$; $t \geq 1.72$; no rechazar H_o

10.147 H_a: $\mu_2 - \mu_1 > 0$; $t\bigstar = 1.30$; $0.10 < \mathbf{P} < 0.25$; $t \geq 2.47$; no rechazar H_o

10.149 c. H_a: $\mu_A - \mu_B \neq 0$; $t\bigstar = 5.84$; $\mathbf{P} < 0.01$; $t \leq -2.98$, $t \geq 2.98$; rechazar H_o

10.151 a. M: $\bar{x} = 74.69$, $s = 10.19$; F: $\bar{x} = 79.83$, $s = 8.80$

b. M: H_a: $\mu \neq 77$; $t\bigstar = -1.36$; $0.10 < \mathbf{P} < 0.20$; $t \leq -2.03$, $t \geq 2.03$; no rechazar H_o; F: H_a: $\mu \neq 77$; $t\bigstar = 1.76$; $0.05 < \mathbf{P} < 0.10$; $t \leq -2.05$, $t \geq 2.05$; no rechazar H_o

c. ambas no significativamente diferentes de 77

d. H_a: $\mu_F - \mu_M \neq 0$; $t\bigstar = 2.19$; $0.02 < \mathbf{P} < 0.05$; $t \leq -2.05$, $t \geq 2.05$; rechazar H_o

e y f. haciendo diferentes preguntas

10.153 −0.044 a 0.164

10.155 a. $z\bigstar = 2.37$; $\mathbf{P} = 0.0178$; diferencia significativa para $\alpha \geq 0.02$

b. $z\bigstar = 2.90$; $\mathbf{P} = 0.0038$; diferencia significativa para $\alpha \geq 0.01$

c. $z\bigstar = 3.35$; $\mathbf{P} = 0.0008$; diferencia significativa para $\alpha \geq 0.001$

d. El error estándar se hizo más pequeño.

10.157 H_a: $p_a - p_1 \neq 0$; $z\bigstar = 1.26$; $\mathbf{P} = 0.2076$; $z \leq -2.58$ y $z \geq 2.58$; no rechazar H_o

10.159 H_a: $\sigma_m^2 > \sigma_f^2$; $F\bigstar = 2.58$; $0.025 < \mathbf{P} < 0.05$; $F \geq 2.53$; rechazar H_o

10.161 H_a: $\sigma_n^2 \neq \sigma_s^2$; $F\bigstar = 1.28$; $\mathbf{P} > 0.10$; $F \geq 1.80$; no rechazar H_o

10.163 a.

	N	Media	DesvEst
Cont	50	0.005459	0.000763
Prueba	50	0.003507	0.000683

b. ambos aproximadamente normales

c. de una cola; busca una reducción

d. H_a: $\sigma_c^2 > \sigma_t^2$; $F\bigstar = 1.248$; $\mathbf{P} > 0.05$; $F \geq 1.69$; no rechazar H_o

e. H_a: $\mu_c - \mu_t > 0$; $t\bigstar = 13.48$; $\mathbf{P} = +0.000$; $t \leq -1.68$, $t \geq 1.68$; rechazar H_o

f. La fuerza media se ha reducido, pero no la variabilidad.

Capítulo 11

11.1 a. forma preferida de "enfriar" su boca después de comer salsa picante

b. Adultos norteamericanos que gustan de tomar alimentos con salsa picante, método de enfriar

c. 36.5%; 17.5%; 10%; 9.5%; 14.5%; 5.5%; 6.5%

d. muy semejante

11.3 a. 23.2

b. 23.3

c. 3.94

d. 8.64

11.5 a. χ^2 (14, 0.01) = 29.1

b. χ^2 (25, 0.05) = 40.1

11.7 a. preguntar a una persona

b. día de cumpleaños de la semana

c. los 7 días de la semana

11.9 a. H_o: $P(1) = P(2) = P(3) = P(4) = P(5) = 0.2$
H_a: no igualmente probable
b. H_o: $P(1) = 2/8$, $P(2) = 3/8$, $P(3) = 2/8$, $P(4) = 1/8$
H_a: al menos una es diferente
c. H_o: $P(\text{E}) = 0.16$, $P(\text{G}) = 0.38$, $P(\text{F}) = 0.41$, $P(\text{P}) = 0.05$
H_a: los porcentajes son diferentes a los especificados

11.11 a. $\chi^2 \geq 7.82$
b. $\chi^2 \geq 9.21$

11.15 a. 60
b. 2
c. H_a:la proporción es diferente de 6:3:1; $\chi^2\bigstar = 2.67$;
P = 0.263; $\chi^2 \geq 4.61$; no rechazar H_o

11.17 H_a: proporciones diferentes $\chi^2\bigstar = 7.35$; **P** = 0.062; $\chi^2 \geq 7.82$; no rechazar H_o

11.19 H_a: opiniones distribuidas de modo diferente $\chi^2\bigstar = 213.49$; **P** < 0.005; $\chi^2 \geq 9.49$; rechazar H_o

11.21 H_a: opiniones distribuidas de modo diferente; $\chi^2\bigstar = 10.05$; **P** = 0.123; $\chi^2 \geq 12.6$; no rechazar H_o

11.23 H_a: colores distribuidos de modo diferente; $\chi^2\bigstar = 3.057$; **P** = 0.548; $\chi^2 \geq 9.49$; no rechazar H_o

11.25 a. H_a: proporciones diferentes a la lista; $\chi^2\bigstar = 44.4928$; **P** < 0.005; $\chi^2 \geq 7.82$; rechazar H_o
b. 4ª celda

11.27 a. H_a: La preferencia de electores y afiliación de partido no son independientes.
b. H_a: La distribución no es igual para los tres
c. H_a: La proporción de respuestas positivas no es igual en todas las categorías.

11.29 10

11.33 H_a: el tamaño de comunidad de residencia no es independiente del tamaño de la comunidad en que se crió
$\chi^2\bigstar = 35.749$; **P** < 0.005; $\chi^2 \geq 13.3$; rechazar H_o

11.35 H_a: la respuesta no es independiente de los años $\chi^2\bigstar = 3.390$; **P** = 0.335; $\chi^2 \geq 6.25$; no rechazar H_o

11.37 H_a: el número de defectuosos no es independiente del día; $\chi^2\bigstar = 8.548$; **P** = 0.074; $\chi^2 \geq 9.49$; no rechazar H_o

11.39 H_a: los creadores y lectores de un blog no tienen igual proporción; $\chi^2\bigstar = 3.954$; **P** = 0.138; $\chi^2 \geq 5.99$; no rechazar H_o

11.41 H_a: El temor y no temor a la oscuridad no tienen igual proporción; $\chi^2\bigstar = 80.957$; **P** < 0.005; $\chi^2 \geq 13.3$; rechazar H_o

11.43 Género: H_a: mujeres y hombres no tienen igual proporción para cada dosis; $\chi^2\bigstar = 0.978$; **P** = 0.613; $\chi^2 \geq 9.21$; no rechazar H_o
Dosis: H_a: los grupos de edad no tienen igual proporción para cada dosis; $\chi^2\bigstar = 7.449$; **P** = 0.114; $\chi^2 \geq 13.3$; no rechazar H_o

11.45 a. $\chi^2\bigstar = 4.043$; **P** = 0.257
b. $\chi^2\bigstar = 8.083$, **P** = 0.044; $\chi^2\bigstar = 12.127$, **P** = 0.007
c. sí

11.47 H_a: las proporciones son diferentes a 1:3:4; $\chi^2\bigstar = 10.33$; **P** = 0.006; $\chi^2 \geq 5.99$; rechazar H_o

11.49 H_a: porcentajes diferentes a la lista; $\chi^2\bigstar = 6.693$; **P** = 0.153; $\chi^2 \geq 9.49$; no rechazar H_o

11.51 H_a: porcentajes diferentes a la lista; $\chi^2\bigstar = 17.92$; **P** = 0.003; $\chi^2 \geq 15.1$; rechazar H_o

11.53 $P(x < 130) = 0.0228$, $P(130 < x < 145) = 0.1359$, $P(145 < x < 160) = 0.3413$, $P(160 < x < 175) = 0.3413$, $P(175 < x < 190) =$
0.1359, $P(x > 190) = 0.0228$
H_a: los pesos no son $N(160, 15)$; $\chi^2\bigstar = 5.812$; **P** = 0.325; $\chi^2 \geq 11.1$; no rechazar H_o

11.55 H_a: porcentajes diferentes a la lista; $\chi^2\bigstar = 4.70$; **P** = 0.195; $\chi^2 \geq 7.82$; no rechazar H_o

11.57 a. H_a: las distribuciones son diferentes: $\chi^2\bigstar = 6.1954$; **P** = 0.2877; $\chi^2 \geq 11.1$; no rechazar H_o
b. $\chi^2\bigstar = 36.761$; **P** < 0.005; $\chi^2 \geq 11.1$; rechazar H_o
c. $\chi^2\bigstar = 92.93$; **P** < 0.005; $\chi^2 \geq 11.1$; rechazar H_o
d. Ji cuadrada se hace más sensible a variaciones cuando el tamaño muestral se hace más grande.

11.59 H_a: las distribuciones son diferentes:$\chi^2\bigstar = 3.123$; **P** = 0.793; $\chi^2 \geq 12.6$; no rechazar H_o

11.61 H_a: la preferencia política no es independiente de la edad; $\chi^2\bigstar = 23.339$; **P** < 0.005; $\chi^2 \geq 13.3$; rechazar H_o

11.63 H_a: la proporción de palomitas de maíz que revienta no es igual para todas las marcas; $\chi^2\bigstar = 2.839$; **P** = 0.417; $\chi^2 \geq 7.82$; no rechazar H_o

11.65 a. 2003: 73.2%, 2004: 74.2%
b. H_a: la proporción entre donadores de órganos no es igual; $\chi^2\bigstar = 5.955$; **P** = 0.015; $\chi^2 \geq 3.84$; rechazar H_o
c. Con tamaños muestrales muy grandes, las diferencias deben ser muy pequeñas para ser consideradas como inexistentes.

Respuestas a exámenes de práctica de los capítulos

Primera parte: sólo se da la sustitución de la(s) palabra(s) que aparecen en negritas. (Si el enunciado es verdadero, no se da respuesta; si es falso, se da una sustitución.)

Capítulo 1, página 36

Parte I

1.1 descriptiva
1.2 inferencial
1.4 muestra
1.5 población
1.6 atributo o cualitativa
1.7 cuantitativa
1.9 aleatoria

Parte II

1.11 **a.** Nominal **b.** Ordinal **c.** Continua
d. Discreta **e.** Nominal
1.12 c, g, h, b, e, a, d, f

Parte III

1.13 Vea definiciones; los ejemplos variarán. Nota: *población* es conjunto de TODO lo posible, en tanto que *muestra* es el conjunto real de subconjuntos estudiados.

1.14 Vea definiciones; los ejemplos variarán. Nota: *variable* es la idea de interés, en tanto que los *datos* son los valores reales obtenidos.

1.15 Vea definiciones; los ejemplos variarán. Nota: *valor de datos* es el valor que describe una fuente, la *estadística* es un valor (por lo general calculado) que describe todos los datos de la muestra, el *parámetro* es un valor que describe toda la población (por lo general desconocida).

1.16 Todo elemento de la población tiene igual probabilidad de ser seleccionado.

Capítulo 2, página 141

Parte I

2.1 mediana
2.2 dispersión
2.3 nunca
2.5 cero
2.6 más alta

Parte II

2.11 **a.** 30 **b.** 46 **c.** 91 **d.** 15 **e.** 1 **f.** 61
g. 75 **h.** 76 **i.** 91 **j.** 106 o 114
2.12 **a.** dos artículos comprados
b. Nueve personas compraron 3 artículos cada una.
c. 40 **d.** 120 **e.** 5 **f.** 2 **g.** 3 **h.** 3
i. 3.0 **j.** 1.795 **k.** 1.34
2.13 **a.** 6.7 **b.** 7 **c.** 8 **d.** 6.5 **e.** 5 **f.** 6
g. 3.0 **h.** 1.7 **i.** 5
2.14 **a.** −1.5 **b.** 153

Parte III

2.15 **a.** 98 **b.** 50 **c.** 121 **d.** 100
2.16 **a.** $32,000, $26,500, $20,000, $50,000
b.

c. Mr. VanCott, de rango medio; gerente de negocios, media; mayordomo, mediana; trabajador nuevo, modo

d. La distribución es de forma de J.

2.17 Hay más de una posible respuesta para éstos.
a. 12, 12, 12 **b.** 15, 20, 25
c. 12, 15, 15, 18 **d.** 12, 15, 16, 25 25
e. 12, 12, 15, 16, 17 **f.** 20, 25, 30, 32, 32, 80

2.18 A es correcta; B es errónea; la desviación estándar no cambiará.

2.19 B es correcta. Por ejemplo, si la desviación estándar es \$5, entonces la varianza (desviación estándar)^2, es "el cuadrado de 25 dólares." ¿Quién sabe qué son "cuadrado de dólares"?

Capítulo 3, página 200

Parte I

3.1 regresión
3.2 resistencia de la
3.3 +1 o −1
3.5 positiva
3.7 positivo
3.8 −1 y +1
3.9 valor de salida o pronosticado

Parte II

3.11 **a.** B, D, A, C **b.** 12 **c.** 10 **d.** 175 **e.** N **f.** (125, 13) **g.** N **h.** P
3.12 Alguien hizo un error aritmético; r debe ser entre −1 y +1.
3.13 **a.** 12 **b.** 10 **c.** 8 **d.** 0.73 **e.** 0.67 **f.** 4.33 **g.** $\hat{y} = 4.33 + 0.67x$

Parte III

3.14 Los niños de corta edad tienen pies pequeños y es probable que tiendan a tener menos capacidad en matemáticas, en tanto que los adultos tienen pies más grandes y tenderían a tener más capacidad.
3.15 El Estudiante B tiene razón. −1.78 se presenta sólo como resultado de una falla aritmética.
3.16 Estas respuestas variarán pero de algún modo deben incluir el pensamiento básico:
a. negativo fuerte **b.** positivo fuerte
c. sin correlación **d.** sin correlación
e. valor imposible, mala aritmética
3.17 Hay más de una posible respuesta para éstos.

a. (1, 1), (2, 1), (3, 1) **b.** (1, 1), (3, 3), (5, 5)
c. (1, 5), (3, 3), (5, 1) **d.** (1, 1), (5, 1), (1, 5), (5,5)

Capítulo 4, página 266

Parte I

4.1 cualquier valor numérico entre 0 y 1, inclusive
4.4 sencillo
4.5 raras veces
4.6 suman 1.0
4.7 dependiente
4.8 complementaria
4.9 mutuamente exclusivo o dependiente
4.10 regla de la multiplicación

Parte II

4.11 **a.** $\frac{4}{8}$ **b.** $\frac{4}{8}$ **c.** $\frac{2}{8}$ **d.** $\frac{6}{8}$ **e.** $\frac{2}{8}$

f. $\frac{6}{8}$ **g.** 0 **h.** $\frac{6}{8}$ **i.** $\frac{1}{8}$ **j.** $\frac{5}{8}$

k. $\frac{2}{4}$ **l.** 0 **m.** $\frac{1}{2}$ **n.** no (e)

o. sí (g) **p.** no (i) **q.** sí (a, k) **r.** no (b, 1)
s. sí (a, m)
4.12 **a.** 0 **b.** 0.7 **c.** 0 **d.** no (c)
4.13 **a.** 0.7 **b.** 0.5 **c.** no, $P(E \text{ y } F) = 0.2$
d. sí, $P(E) = P(E \mid F)$
4.14 0.51

Parte III

4.15 El Estudiante B tiene razón. *Mutuamente exclusivo* significa que no hay intersección, en tanto que *independencia* significa que un evento no afecta la probabilidad del otro.
4.16 Estas respuestas variarán pero de algún modo deben incluir las ideas básicas:
a. no hay suceso común
b. ningún evento tiene efecto sobre la probabilidad del otro
c. la frecuencia relativa con la que ocurre el evento.
d. probabilidad de que ocurra un evento aun cuando el evento condicional haya ocurrido previamente

Capítulo 5, página 311

Parte I

5.1 continuo
5.3 uno
5.5 exactamente dos
5.6 binomial
5.7 un éxito ocurre en 1 intento
5.8 población
5.9 parámetros poblacionales

Parte II

5.11 **a.** Cada $P(x)$ es entre 0 y 1, y la suma de todas las $P(x)$ es exactamente 1.
b. 0.2 **c.** 0 **d.** 0.8 **e.** 3.2 **f.** 1.25
5.12 **a.** 0.230 **b.** 0.085 **c.** 1.2 **d.** 1.04

Parte III

5.13 n intentos independientes repetidos de dos resultados; los dos resultados son "éxito" y "fracaso"; $p = P$(éxito) y $q = P$(fracaso) y $p + q = 1$; $x = n$(éxito) = 0, 1, 2, …, n
5.14 El Estudiante B tiene razón. La media muestral y desviación estándar son estadísticas halladas con el uso de fórmulas estudiadas en el capítulo 2. Las distribuciones de probabilidad estudiadas en el capítulo 5 son poblaciones teóricas y sus medias y desviaciones estándar son parámetros.
5.15 El Estudiante B tiene razón. No hay restricciones en los valores de la variable x.

Capítulo 6, página 357

Parte I

6.1 su media
6.4 1 desviación estándar
6.6 correcto
6.7 cero, 1
6.8 algunas (muchas)
6.9 eventos mutuamente exclusivos
6.10 normal

Parte II

6.11 **a.** 0.4922 **b.** 0.9162 **c.** 0.1020 **d.** 0.9082
6.12 **a.** 0.63 **b.** −0.95 **c.** 1.75
6.13 **a.** $z(0.8100)$ **b.** $z(0.2830)$
6.14 0.7910
6.15 28.03
6.16 **a.** 0.0569 **b.** 0.9890 **c.** 537 **d.** 417 **e.** 605

Parte III

6.17 Esta respuesta variará pero de algún modo debe incluir las propiedades básicas: forma de campana, media de 0, desviación estándar de 1.
6.18 Esta respuesta variará pero de algún modo debe incluir las ideas básicas: es un puntaje z, α representa el área bajo la curva y a la derecha de z.
6.19 Todas las distribuciones normales tienen la misma forma y propiedades relativas al puntaje z.

Capítulo 7, página 390

Parte I

7.1 no es
7.2 algunos (muchos)
7.3 población
7.4 dividido entre $\sqrt{n}$
7.5 disminuye
7.6 aproximadamente normal
7.7 muestreo
7.8 medias
7.9 aleatoria

Parte II

7.11 **a.** 0.4364 **b.** 0.2643
7.12 **a.** 0.0918 **b.** 0.9525
7.13 0.6247

Parte III

7.14 En este caso cada cabeza produjo una pieza de información, la longitud estimada de la recta. El teorema de límite central nos asegura que el valor medio de una muestra es mucho menos variable que valores individuales de la variable x.
7.15 Todas las muestras deben ser de un tamaño fijo.
7.16 El Estudiante A tiene razón. Una distribución poblacional es una distribución formada por todos los valores de x que conforman toda la población.
7.17 El Estudiante A tiene razón. El error estándar se encuentra al dividir la desviación estándar entre la raíz cuadrada del *tamaño muestral.*

Capítulo 8, página 470

Parte I

8.1 alfa
8.2 alfa
8.3 distribución muestral de la media
8.7 error tipo II
8.8 beta
8.9 decisión correcta
8.10 región crítica (de rechazo)

Parte II

8.11 4.72 o 5.88
8.12 **a.** H_o: $\mu = 245$, H_a: $\mu > 245$
b. H_o: $\mu = 4.5$, H_a: $\mu < 4.5$
c. H_o: $\mu = 35$, H_a: $\mu \neq 35$

8.13 **a.** 0.05, z, $z \leq -1.65$
b. 0.05, z, $z \geq +1.65$
c. 0.05, z, $z \leq -1.96$ o $z \geq +1.96$
8.14 **a.** 1.65 **b.** 2.33 **c.** 1.18 **d.** -1.65
e. -2.05 **f.** -0.67
8.15 **a.** $z\bigstar = 2.50$ **b.** 0.0062
8.16 H_o: $\mu = 1520$ vs. H_a: $\mu < 1520$, región crítica; $z \leq -2.33$, $z\bigstar = -1.61$,no rechazar H_o

Parte III

8.17 **a.** no hay efecto específico
b. lo reduce **c.** lo estrecha
d. no hay efecto
e. lo aumenta **f.** lo ensancha
8.18 **a.** H_o – (a), H_a – (b) **b.** 3 *c.* 2
d. P(error tipo I) es alfa, disminuye: P(error tipo II) aumenta
8.19 La hipótesis alternativa expresa el interés; la conclusión contesta el interés.

Capítulo 9, página 540

Parte I

9.2 t de Student
9.3 Ji cuadrada
9.4 para ser rechazada
9.5 calificación t
9.7 $n - 1$
9.9 $\sqrt{pq/n}$
9.10 z (normal)

Parte II

9.11 **a.** 2.05 **b.** -1.73 **c.** 14.6
9.12 **a.** 28.6 **b.** 1.44 **c.** 27.16 a 30.04
9.13 0.528 a 0.752
9.14 **a.** H_o: $\mu = 255$, H_a: $\mu > 225$
b. H_o: $p = 0.40$, H_a: $p \neq 0.40$
c. H_o: $\sigma = 3.7$, H_a: $\sigma < 3.7$
9.15 **a.** 0.05, z, $z \leq -1.65$
b. 0.05, t, $t \leq -2.08$ o $t \geq +2.08$
c. 0.05, z, $z \geq +1.65$
d. 0.05, χ^2, $\chi^2 \leq 14.6$ o $\chi^2 \geq 43.2$
9.16 H_o: $\mu = 26$ vs. H_a: $\mu < 26$, región crítica $t \leq -1.71$, $t\bigstar = -1.86$, rechazar H_o
9.17 H_o: $\sigma = 0.1$ vs. H_a: $\sigma > 0.1$, región crítica $\chi^2 \geq 21.1$, $\chi^2\bigstar = 23.66$, rechazar H_o
9.18 H_o: $p = 0.50$ vs. H_a: $p > 0.50$, región crítica. $z \geq 2.05$, $z\bigstar = 1.29$, no rechazar H_o

Parte III

9.19 Si la distribución es normal, 6 desviaciones estándar es aproximadamente igual al rango.
9.20 B
9.21 Ambos tienen razón.
9.22 Cuando el tamaño muestral, n, es grande, el valor crítico de t se estima usando el valor crítico de la distribución normal estándar de z.
9.23 El Estudiante A
9.24 El Estudiante B tiene razón. Es significativo al nivel de significación de 0.01.
9.25 El Estudiante A tiene razón.
9.26 Depende de lo que significa mejorar el intervalo de confianza. Para casi todos los fines, un tamaño muestral aumentado sería el mejor desarrollo.

Capítulo 10, página 614

Parte I

10.1 dos medias independientes
10.3 distribución F
10.4 distribución t de Student
10.5 t de Student
10.7 no simétrica (o sesgada)
10.9 disminuye

Parte II

10.11 **a.** H_o: $\mu_N - \mu_A = 0$, H_a: $\mu_N - \mu_A \neq 0$
b. H_o: $\sigma_o/\sigma_m = 1.0$, H_a: $\sigma_o/\sigma_m > 1.0$
c. H_o: $p_m - p_f = 0$, H_a: $p_m - p_f \neq 0$
d. H_o: $\mu_d = 0$, H_a: $\mu_d > 0$
10.12 **a.** z, $z \leq -1.96$ o $z \geq 1.96$
b. t, $t \leq -2.05$, $t \geq 2.05$
c. t, df = 7, $t \geq 1.89$
d. t, df = 37, $t \leq -1.69$
e. F, $F \geq 2.11$
10.13 **a.** 2.05 **b.** 2.13 **c.** 2.51 **d.** 2.18
e. 1.75 **f.** 1.69 **g.** -2.50 **h.** -1.28
10.14 H_o: $\mu_L - \mu_P = 0$ vs. H_a: $\mu_L - \mu_P > 0$, región crítica, $t \geq +1.83$, $t\bigstar = 0.979$, región crítica, H_o
10.15 H_o: $\mu_d = 0$ vs. H_a: $\mu_d > 0$, región crítica, $t \geq 1.89$, $t\bigstar = 1.88$, no rechazar, H_o
10.16 0.072 a 0.188

Parte III

10.17 Dependientes
10.18 Una posibilidad: examine a todos los estudiantes antes de empezar el curso, luego seleccione al azar 20 de los que terminen el curso y aplíqueles examen. Use las calificaciones previas de estos 20 como la muestra previa.

10.19 Para quienes se inician, si las dos muestras independientes son de tamaños diferentes, las técnicas para muestras dependientes no podrían completarse. Están probando conceptos muy diferentes, la "media de las diferencias de datos pareados" y la "diferencia entre dos valores medios".

10.20 Sólo es significativo si la calificación t calculada está en la región crítica. La variación entre los datos y su tamaño relativo harán su papel.

10.21 Las 80 calificaciones en realidad son dos muestras independientes de tamaño 40. Podría completarse una prueba para comparar las calificaciones medias de los dos grupos.

10.22 Habría que tomar una muestra más bien grande de familias católicas y no católicas, y sería necesario obtener el número de cada una cuyos hijos asistieran a escuelas privadas. Entonces podría estimarse la diferencia entre las dos proporciones.

Capítulo 11, página 653

Parte I

11.1 uno menos que
11.3 esperada
11.4 tabla de contingencia
11.6 prueba de homogeneidad
11.8 aproximada por Ji cuadrada

Parte II

11.11 **a.** H_o: Los dígitos generados se presentan con igual probabilidad.
H_a: Los dígitos no ocurren con igual probabilid
b. H_o: Los votos fueron emitidos independientemente de afiliación partidista.
H_a: Los votos no fueron emitidos independientemente de afiliación partidista.
c. H_o: Las distribuciones de delitos son iguales para las cuatro ciudades.
H_a: Las distribuciones de delitos no son iguales.

11.12 **a.** 4.40 **b.** 35.7

11.13 H_o: $P(1) = P(2) = P(3) = \frac{1}{3}$
H_a: preferencias no iguales, $\chi^{2\star} = 3.78$;
$0.10 < \mathbf{P} < 0.25$ o región crítica $\chi^2 \geq 5.99$;
no rechazar H_o

11.14 **a.** H_o: La distribución es igual para todos los tipos de suelo.
H_a: Las distribuciones no son iguales.
b. 25.622 **c.** 13.746
d. $0.005 < \mathbf{P} < 0.01$ **e.** $\chi^2 \geq 9.49$
f. Rechazar H_o: Hay suficiente evidencia para demostrar que la distribución de crecimiento es diferente para al menos uno de los tres tipos de suelo.

Parte III

11.15 Similares en que hay n intentos independientes repetidos. Diferentes en que el binomio tiene dos posibles resultados, mientras que el polinomio tiene varios. Cada posible resultado tiene una probabilidad y estas probabilidades suman 1 para cada experimento diferente, ambos para binomio y polinomio.

11.16 La prueba de homogeneidad compara varias distribuciones en una comparación paralela, mientras que la prueba para independencia prueba la independencia de los dos factores que crean las filas y columnas de la tabla de contingencia.

11.17 El Estudiante A tiene razón en que los cálculos se completan en la misma forma. El Estudiante B tiene razón en que la prueba de independencia se inicia con una muestra grande y la homogeneidad tiene varias muestras.

11.18 **a.** Si se usa una prueba de Ji cuadrada, los resultados de las cuatro preguntas se agruparían para estimar la probabilidad esperada.
b. Use una prueba de Ji cuadrada para homogeneidad.

Índice

Aleatoria, 22, 270
Aleatoriedad de datos, prueba para
Alfa, 403, 420, 452
Alfa, *P* (error tipo I), 420-1
"Al menos una…, ", 229
Análisis de correlación, 162
Análisis de regresión, 146, 174
Ancho de clase, 56
Aproximación normal a binomial, 343
  regla, 346
Área a la derecha, 338, 340, 403
Área sombreada, 315

Beta, intersección, 621
Beta, *P*(error tipo II), 420
Binomial *p*, 581
"Buenos datos, " 18

Calculadora TI-83-84 Plus, 28
Calculadoras, 28
Cálculo, 397, 402 (*ver también* Intervalo de confianza)
Cálculo de intervalo, 399
Cálculo del valor *p*
  χ^2, usando la Tabla 8, 521, 524
  F, usando la tabla 9, 597
  t, usando la tabla 6, 483, 486
  t, usando la tabla 7, 484, 486
  z, usando la tabla 3, 432
  z, usando la tabla 5, 432
Cálculos de intervalo (*ver* Intervalo de confianza)
Cálculos preliminares, 164, 175
Calificación estándar, *z*, 99, 316, 323
Calificaciones de desviación, 313
Calificaciones IQ, 313
Causa, 167
Causalidad, 167
Celdas, 620, 633
Censo, 12, 20
Centroide, 177
Clase, 56
Clase modal, 64
Clasificación de datos principios básicos, 57
  ancho de clase, 56
  en las dos direcciones, 633
  frecuencia de clase, 56
  fronteras de clase, 57
  marcas de clase, 58
Clasificación en dos direcciones, 633
Cociente de inteligencia, 313
Coeficiente binomial, 289
Coeficiente de confianza, 403, 498, 501
Coeficiente de correlación (*ver* Coeficiente de correlación lineal)
Coeficiente de correlación lineal, 163
  cálculo, 164
  comprendiéndolo, 166
  estimación, 166
  fórmula, 163, 164
Columnas, 637
Computadoras, 28
Conclusión, escritura, 422
Correlación causalidad, 167
  negativa, 163
  sin, 162, 163
  positiva, 163
Correlación lineal, 162, 168, causalidad, 167
Correlación negativa, 163
Correlación positiva, 163
Criterio de mínimos cuadrados, 174, 175, 176
Criterios de prueba, 427, 444
Cuartil central, 95
Cuartil, hallar, 93
Cuartiles, 93
Curva en forma de campana, 315, 317
Curva normal, 316

Dato, 12
Datos
  de dos variables, 146, 147
  de mejor ajuste, 174
  de Moivre, Abraham, 370
  definidos, 8
  enumerativos, 620, 621
  recolección de, 19
  ventana de, 153, 166
Datos (plural), 8
Datos clasificados, 74, 93
Datos cualitativos, 41
Datos cuantitativos, 44
Datos de dos variables (bivariados), 146, 147
  ambos cualitativos, 147
  ambos cuantitativos, 152
  mostrando, diagrama de dispersión, 153
  uno cualitativo, uno cuantitativo, 151
Datos enumerativos, 620, 621
Datos pareados, 153, 550
Datos totales, 58
Decisión
  escritura de, 422
  no rechazar H_0, 419, 420
  rechazar H_0, 419, 420
  regla para tomar, método clásico, 451
  regla para tomar, método de valor *p*, 433
Decisión correcta, tipo A, 419, 420
Decisión correcta, tipo B, 419, 420
Decisiones, tipos de, 419, 420
Dependencia, 547
Depósito, 56
Desviación de la media, 84
Desviación estándar, 84, 106
  cálculo de, 88, 89
  definida, 106
  distribución binomial, 300
  distribución de frecuencia, 117
  distribución de probabilidad, 117
  muestra, 87
  población (σ), 279
  unidad de medida, 88
Desviación, media, 85
Desviación, valor absoluto de, 85
Determinación de tamaño muestral, intervalo de confianza, 501
Diagrama de árbol, 209
Diagrama de caja y bigotes, 96, 151
  de escala común, 151
Diagrama de dispersión, 153
Diagrama de Pareto, 42
Diagrama de tallo y hoja, 46, 60, consecutivo, 47
Diagrama de Venn, 211
Diagramas de pastel, 41
Diagramas de puntos consecutivos, 47
Diferencia entre dos medias, 565
Diferencia entre dos proporciones, 582
Diferencia pareada, 550
Diseño muestral, 21
Dispersión, medida de, 84
Disraeli, Benjamin, 114
Dispersión, medidas de, 84
Distribución
  bimodal, 64
  de frecuencia, 55
  de muestreo, 363
  de probabilidad, 273
  definida, 44
  en forma de campana, 106
  en forma de J, 64
  estándar normal (ver Distribución normal estándar)
  normal, 63, 106, 315
  rectangular, 63
  sesgada, 64
  simétrica, 63
  uniforme, 63
Distribución bimodal, 64
Distribución binomial, 300, 343
Distribución de frecuencia acumulativa, 64
Distribución de frecuencia, 55, 59, 64, 364
  acumulativa, 64
  agrupada, 55, 56, 57
  cálculo usando, 117
  definida, 55
  desviación estándar, 117, 119
  media, 117, 118
  no agrupada, 55, 56
  relativa, 60, 207
  relativa acumulativa, 64, 65
  varianza, 119
Distribución de frecuencia agrupada, 55, 56, 57
Distribución de frecuencia no agrupada, 55, 56
Distribución de frecuencia relativa, 60, 270
Distribución de frecuencia relativa acumulativa, 64
Distribución de probabilidad, 364
  binomial, 284
  continua, 315
  definida, 273
  discreta, 279
  función para, 273
  propiedades de, 274

Distribución de probabilidad acumulativa,
Ji cuadrada, χ^2, 518
F, 595
t, 478
Distribución de probabilidad discreta, 279
Distribución de probabilidad normal estándar, 316
Distribución en forma de campana, 106
Distribución en forma de J, 64
Distribución *F*, 593, 597
Distribución muestral, 397, 497
Distribución muestral, $\bar{d}$, 552
diferencia entre medias, 564
diferencia entre proporciones, 581
dos medias dependientes, 552
estadística, 397
estadística muestral, 363
medias muestrales, 364, 369, 373, 377, 397, 444, 474
proporción, 497
rangos muestrales, 363
Distribución normal, 63, 106, 315
aplicaciones, 323
estándar (*ver* Distribución normal estándar)
representación de área, 315
Distribución normal de probabilidad, 315, 317, 343
función, 315
Distribución normal estándar, 476, 497
aplicación de, 323
desviación estándar, 316
media, 316
notación, 293
probabilidades, 317
propiedades, 316
Distribución rectangular, 63
Distribución sesgada, 63, 64, 370, 403, 517, 524, 593
Distribución simétrica, 63
Distribución *t*, 552, 564 (*ver también* Distribución *t* de Student)
Distribución *t* de Student, propiedades de, 475
Distribución uniforme, 63
σ desconocida, 474
Distribuciones de *Ji* cuadrada, 517
cálculo de estadística de prueba, 519, 621
propiedades, 517
pruebas de datos enumerativos, 620
pruebas de distribución estándar, 519
pruebas de varianza, 519
valor crítico, 517
Ecuación de recta
Ecuaciones de pronóstico, 174
Efecto, 167
Encuesta, 20
Engaño estadístico, 114
"Entra basura, sale basura", 29
Error estándar
definido, 370
diferencia media, 555
dos medias, 564
dos proporciones, 581, 584
media, 370, 403
media, estimada, 474
proporción, 498
Error estándar calculado, 565
Error máximo, 410, 501
Error máximo de estimación, 403, 410, 480, 498, 501
Error probable de media, 565
Error tipo I, 419
Error tipo II, 419
Errores, probabilidad de, 421
Errores, tipo I, 419
Errores, tipo II, 419
Espacio muestral, 208
definido, 208
diagrama de árbol, 210
diagrama de Venn, 211
lista de, 209
representación de gráfica, 210
Estadística
definida, 9
muestral, 9, 278, 397, 497
no sesgada, 397
sesgada, 397
Estadística, 3, 27
definida, 4
descriptiva, 4
inferencial, 4, 396
uso de, 7
Estadística de prueba
definida, 422
F★, dos varianzas, 595
notación de estrella, 431, 449
t★, diferencia media, 555
t★, dos medias, 567
t★, una media, 449
valor crítico de, 450
z★, dos proporciones, 584, 585
z★, una media, 449
z★, una proporción, 503
χ^2★, 621
χ^2★, una varianza, 519
Estadística muestral, 9, 278, 397, 497
cuartiles, 93
desviación estándar, 87, 279
distribución muestral, 363
media, 73, 279
mediana, 74
moda, 76
percentiles, 93
proporcional, 24
rango central, 76
rango, 84
varianza, 86, 279
Estadística muestral no sesgada, 497, 582
Estadística no sesgada, 397
Estadística sesgada, 397
Estadística *t* de Student, 475
Estadísticas descriptivas, 4, 7
Estadísticas inferenciales, 4, 396
Estimación agrupada, 585
Estimación puntual, 397, 498
Estimación puntual, σ, 475
Estimador no sesgado, 496-498
Estratos, 24
Estrella, uso de, 431, 449
Estudios observacionales, 19
Eventos
complementarios, 229
"de éxito", 287
"de fracaso", 287
dependientes, 243
igualmente probables, 208
independientes, 240, 243, 249
intersección de, 237
mutuamente exclusivos, 236, 239, 270
todo incluido, 213, 270
Eventos complementarios, 229
Eventos compuestos, 228
Eventos con sustitución, 290
Eventos dependientes, 243
Eventos igualmente probables, 208
Eventos mutuamente exclusivos, 236, 239, 270
n factorial, 289
n, tamaño muestral, 73
Eventos que incluyen todo, 213, 270
Eventos sin sustitución, 233
Exhibición de gráfica de puntos, 44
consecutiva, 47
Éxito, 287
Experimentación (*ver* Experimento estadístico)
Experimento, 8, 207
Experimento binomial, 496, 581
definición, 287
intento, 287
probabilidades, 287
propiedades, 287, 581
variable aleatoria, 287
Experimento estadístico, 8
Experimento multinomial, 622, 624
Extensiones, 118

F, 593, 597 *t*, 477
z, 338-340
F, estadística de prueba, dos varianzas, 595
Factor de corrección de continuidad, 344
variable continua (datos), 11
Filas, 637
Fisher, Sir Ronald A., 209, 422, 434, 450, 565
Fórmula breve (varianza), 88
Fórmulas múltiples, 90
Fortuito, vs. aleatorio, 22
Fracaso, 287
Frecuencia, 55, 620
Frecuencia acumulativa, 64
Frecuencia esperada, 620, 637
Frecuencia observada, 620
Frecuencia relativa, 60, 207
Frecuencia relativa observada, 207, 287, 438
Frecuencia teórica, 620
Frontera inferior de clase, 57
Fronteras de clase, 57
Fronteras superiores de clase, 57
Fuente, de datos, 547
Función constante, 274
Función, probabilidad, 273, 274

Galton, Sir Francis, 74, 179, 300
Gosset, William S., 475, 476, 565
Grados de libertad, 476, 517, 621, 637
de Moivre, Abraham, 370
F, 593
Ji cuadrada, multinomial, 621
Ji cuadrada, varianza, 517
número de, 476
t, medias dependientes, 552
t, medias independientes, 565
t, una media, 476
Gráfica de agujas, 275
Gráfica de barras, 41, 59
Gráficas de pastel, 41
Gran total, 148

Hipótesis, 417
Hipótesis alternativa, 417
de dos colas, 430
de una cola, 429, 446-7
expresión de, 428, 445
signo en, 452
Hipótesis alternativa de dos colas, 430
Hipótesis alternativa de una cola, 429, 446
Hipótesis, alternativas posibles, 428, 446
Hipótesis de investigación, 418
Hipótesis nula, 417, 428
Hipótesis, investigación, 418
Hipótesis, oponentes, 417, 428
Histograma, 59
de probabilidad, 275
en forma de J, 64
forma de, bimodal, 64
normal, 63
rectangular, 63
sesgado, 64
simétrico, 63
uniforme, 63

Histograma de frecuencia, 60
Histograma de frecuencia relativa, 60
Histograma de probabilidad, 275
Hoja, 46
Homogeneidad, 599, 638

Independencia, 243
Igual probabilidad, 22
Independencia probada por *Ji* cuadrada, 633
  eventos independientes, 240, 243, 249
  intentos independientes, 287, 624
  medias independientes, 565
  muestras independientes, 547
  variable independiente, 152, 162, 633
Inferencia
  desviación estándar, 516
  diferencia entre dos medias, 565
  diferencia entre proporciones, 581
  diferencia media, 550
  media poblacional, 474, 479
  proporción, 496
  razón entre dos varianzas, 592
  tablas de contingencia, 633
  tipos de (ver Intervalo de confianza; Prueba de hipótesis)
  variabilidad, 516
  varianza, 516
Intento (multinomial), 624
Intentos
  binomiales, 287
  independientes repetidos, 287
  número de (n), 287
Intersección de eventos, 237
Intervalo de confianza, 399, 402, 410, 506
  desconocida, 479
  dos medias dependientes, 552
  dos proporciones, 582
  error estándar, 403
  error máximo, 403
  estimación de punto, 403
  límite inferior, 403
  límite superior, 403
  nivel de confianza, 399
  procedimiento de cinco pasos, 404
  una media, sigma conocida, 402
  una media, suposición, 402
  una proporción, 497

Ji cuadrada, (χ^2), 179, 449
Ji cuadrada, estadística de prueba, datos enumerativos, 620
  una varianza, 519
Ji cuadrada, χ^2, 517
"Juego de la cuerda en tres direcciones", 410, 411, 421

Ley de números grandes, 215
Ley de tricotomía, 428, 445
Leyes de probabilidad, adición, 230, 239
Leyes de probabilidad, multiplicación, 246
Límite inferior de confianza, 403
Límite superior de confianza, 403
Límites, confianza, 403
Línea de mejor ajuste, 174, 712
  cálculo de ordenada en el origen, 175
  cálculo de pendiente, 175
  centroide, 177
  ecuación, 175
  en diagrama de dispersión, 181
  entender ecuación, 181
  estimación de ecuación, 181
  estimación de ordenada en el origen, 174, 177
  estimación de pendiente, 182
  ordenada en el origen, 175
  pendiente, 175
  predicciones, 180
  punto (x–, y–), 177, 727
Línea de regresión (*ver* Línea de mejor ajuste)

Mahalanobis, Prasanta Chandra, 21
Marca de clase, 58
Marco de muestreo representativo, 21
Marco muestral, 20
Media
  definida, 73
  desviación, 85
  desviación absoluta, 85
  diferencia, 551
  distribución binomial, 300
Media aritmética, 73
Media muestral, 73
Mediana, 74, 93, 96
Medias dependientes, 551
Medibilidad, 17
Medidas de dispersión, 84
Medidas de posición, 92
Medidas de tendencia central, 73
Método de mínimos cuadrados, 174
Método de muestreo no sesgado, 18
Método de muestreo sesgado, 18
Método del valor de probabilidad, prueba de hipótesis, 427, 433
Método del valor p, 427
Moda, 64, 76
Métodos estadísticos, uso de, 3
Modelo cuadrático, 174
Modelo de línea recta, 174
Modelo exponencial, 174
Modelo logarítmico, 174
Modelos, 174
Modelos de regresión, 174
Momento de producto r, de Pearson, 163
Mu (μ), media (*ver* Media)
Mu (M), mediana, 74
Muestra
  aleatoria sencilla, 22
  de conglomerados, 25
  de conveniencia, 19
  definida, 8
  encuesta de, 20
  estadística, 279
  estratificada, 24
  observacional, 19
  sistemática, 23
  voluntaria, 19
Muestra "al azar", 19
Muestra aleatoria, 22, 365
Muestra aleatoria sencilla, 22
Muestra de conveniencia, 19
Muestra de juicio, 21
Muestra de voluntarios, 19
Muestra estratificada, 24
Muestra no sesgada, 497
Muestra por conglomerados, 25
Muestra proporcional, 24
Muestra sistemática, 23
  diferencia de media, 555
  dos medias, 567
  t, estadística de prueba
  t, notación de valor crítico, 447
  una media, 481
Muestras
  aleatorias, 22
  de juicio, 21
  de probabilidad, 21
  dependientes, 547, 550, 552, 555
  independientes, 547
  repetidas, 363
Muestras de probabilidad, 21
Muestras dependientes, 547, 550, 552, 555
Muestreo aleatorio, 365
Muestreo repetido, 363

Neyman, Jerzy, 450
Nivel de confianza, 399, 407, 410, 480, 498, 501, 506
Nivel de significación, 422, 431, 449
No rechazar H_0, 419
Normalidad, directrices, 497
  prueba de, usando papel de probabilidad, 108
  prueba para, usando regla empírica, 106
Notación de sumatoria, 58, 73
Notación factorial, 289
Notación prima, 207
Notación, valor crítico, F, 593
Notación, valor crítico, *Ji* cuadrada, 517
Notación, valor crítico, t, 477
Notación, valor crítico, z, 338
Notación, z, 338-340
Número de clases, 56
Número de desviaciones no relacionadas, 476
Número de grados de libertad, 476
Número de posición, 74

Objetivo de estadística, 17
Objetivo estadístico, 17
Ojiva, 65, 108
Ordenar datos, 45

P(error tipo I), 421
P(error tipo II), 421
P(éxito), 287, 497, 498
Papel de probabilidad, 108
Parámetro, 397, 404, 427, 444, 476, 517
  definido, 8
Parámetro binomial, 496
Parámetro de población, 8, 279, 417
  desviación estándar, 279
  media, 73, 278
  mediana, 74
  proporción, 498
  varianza, 278
Parámetros, población, 279
Pares ordenados, 152, 162, 209
Pearson, Karl, 476, 621
Pendiente, 174, 175, 712, 718
Pendiente negativa, 174
Pendiente positiva, 174
Percentil, 93
Percentil k-ésimo, 93
Población, 7, 279
Población finita, 7, 22
Población infinita, 7, 22
Porcentaje, 316, 581
Porcentajes basados en gran total, 149
Porcentajes basados en totales de columna, 149
Porcentajes basados en totales de fila, 149
Posibilidades, 215
Posición, medida de, 92
Potencia de prueba estadística, 421
Presentación gráfica, 40
  datos cualitativos, 41
  datos cuantitativos, 44

Primer cuartil, 93, 96
Probabilidad, 27
binomial, 287, 288, 300, 343, 496, 581
condicional, 223
definida, 207
definida, espacio muestral, 208
empírica, 207
error tipo I, 421, 431, 449
esperada, 208
experimental, 207
función, 273
normal, 315
observada, 207, 496, 581
propiedades, 212, 274
regla especial de la adición, 239
regla especial de la multiplicación, 246
regla general de la adición, 230
regla general de la multiplicación, 231
subjetiva, 207, 212
teórica, 208
Probabilidad agrupada, 585
Probabilidad binomial, distribución de, 284, 288
media, 300
desviación estándar, 300
Probabilidad binomial por cálculo, 288
aproximación normal, 343
aproximación normal, regla de, 346
experimento, 287
función, 288
Probabilidad condicional, 223
Probabilidad de éxito, *p*, 287
Probabilidad de un evento, 207
Probabilidad empírica, 207
Probabilidad experimental, 207
Probabilidad observada, 207, 497, 581
Probabilidad subjetiva, 207, 212
Probabilidad teórica, 208
Probabilidad, de fracaso, *q*, 287
Procedimiento clásico de prueba de hipótesis, 444
Procedimiento de agrupación, 56
Procedimiento de cálculo, 404
Procedimiento de valor *p*, 427
Proceso estadístico, 7, 396
Profundidad, 74
Profundidad, mediana, 74
Programa para calculadora de la TI-83/84 Plus, introducción, 28
Promedio, 73, 77
Promedio a largo plazo, 215
Promedio, largo plazo, 215
Pronóstico para *y* dada *x*, 174, 179
Propiedades, distribución *F*, 593
Proporción, 316, 581
Prueba de dos colas, 429, 434, 436
Prueba de hipótesis, 417, 444
clásica de cinco pasos
conclusiones, 422
criterios de prueba, 430
de dos colas, 429, 434, 436, 447
de una cola, 429, 446, 447
estadística de prueba, 422, 431, 449
experimento multinomial, 622
hipótesis alternativa, 417, 428
hipótesis nula, 417, 428
homogeneidad (tabla de contingencia), 638
independencia (tabla de contingencia), 633
Ji cuadrada, 620
media, una, sigma conocida, 426, 444
media, una, sigma desconocida, 479
medias, dos dependientes, 554
medias, dos independientes, 567
método clásico, 444
método de valor p, 427
método de valor p, regla de decisión, 433
modelo de procedimiento, 444-445
naturaleza de, 416
nivel de significación, 422
nivel de significación, 430
procedimiento clásico, 444
proporción, dos, 584
proporción, una, 502
región crítica, 450
región de aceptación, 450
región no crítica, 450
resultados posibles, 419
valor calculado, 450
valor crítico, 451
valor *p*, 432
valor *p*, modelo de procedimiento de cinco pasos, 427
varianza (desviación estándar), dos, 592
varianza (desviación estándar), una, 520
Prueba de hipótesis estadística, 417
Prueba de homogeneidad, 638
Prueba de independencia, 240
Prueba de normalidad, 106, 108
Pruebas de hipótesis (*ver* Prueba de hipótesis)
Punto central de clase, 58
Punto de intersección, 175, 712
Puntos muestrales, 208
Rango, 55, 84
Rango central, 76
Rango de intercuartil, 96
Razón entre dos varianzas, 592
Rechazar H_o, 419
Recolección de datos, 19 (*ver* Muestra)
Recolección de datos, 19
Región de aceptación, 450
Región no crítica, 450
Región(es) crítica, 450, 451, 484, 504, 520, 621, 625
Regla de decisión, 422
Regla de decisión de valor *p*, 433
Regla de la adición, 230
Regla de multiplicación, 231
Regla de redondeo, 78
Regla empírica, 106, 315
Regla especial de la adición, 239-240
Regla especial de la multiplicación, 246
Regla general de la adición, 230
Regla general de la multiplicación, 231
Regla práctica, aproximación normal a binomial, 346, 497
σ conocida, 426, 444
σ desconocida, 474, 479
s, desviación estándar muestral, 279
s^2, varianza muestral, 279
Regresión curvilínea, 174
Regresión lineal, 173
Relación causa y efecto, 167
Relación, intervalo de confianza y prueba de hipótesis, 506
Representación de área para probabilidad, 316
Resultado, 208
Resultados conservativos, 565
Resumen de cinco números, 96
Riesgo, 452
Segundo cuartil, 93
Sentido común, 6
Sigma (σ), 279, 516
Sigma (Σ), 58, 73
Sigma conocida, 474
Sigma cuadrada (σ^2), 279
Significación estadística, 556
Sin correlación, 162, 163
Sin moda, 76
Sin patrón, 22
Software Excel, introducción, 28
Software MINITAB, introducción, 28
SS(*x*), 87
Student (*ver* Gosset, William S.)
Suma de cuadrados, 87, 164
fórmulas, 164
Suma de desviaciones, 86
Suma del cuadrado de desviaciones, 86
Sumatoria, 58
Suposiciones, 402
de *Ji* cuadrada, datos enumerativos, 621
dos medias dependientes, 552
dos medias independientes, 565
dos proporciones, 582
intervalo de confianza, una media, sigma conocida, 402
intervalo de confianza, una media, sigma desconocida, 479
para pruebas estadísticas, 402
prueba de hipótesis, una media, sigma conocida, 426, 444
prueba de hipótesis, una varianza, 519
razón entre dos varianzas, 595
una desviación estándar, 519
una proporción, 497
Sustitución, con, 290
Sustitución, sin, 233
Tabla de contingencia, 147, 633
columnas 637
con base en gran total, 148
con base en totales de columna, 149
con base en totales de fila, 149
filas, 637
frecuencias esperadas, 637
grados de libertad, 637
$r \times c$, 637
Tabla de contingencia $r \times c$, 637
Tabla de extensiones, 118, 165, 280, 281, 371, 372
Tabulación en cruz, 147
Tallo, 46
Tamaño muestral, 73, 86, 373, 410, 421, 422, 474, 480
determinación de, 403, 411, 412, 501
intervalo de confianza, media, sigma desconocida, 412
n, 86, 412
Técnica para ordenar, 46
Teorema de Chebyshev, 109
Teorema de límite central, 370, 373, 398, 402, 475, 482
aplicación, 402, 426, 444
ilustraciones de, 373
Tercer cuartil, 93, 95, 96
Toma de decisiones, 417 (*ver también* Prueba de hipótesis)
Totales de clase, 58
Totales marginales, 148, 634

Valor absoluto de desviación, 85
Valor crítico, 403, 451, 504
Valor de probabilidad, definido, 432 (*ver también* Valor *p*)
Valor esperado (E),
evento
multinomial, 620
Valor observado, 174
Valor *p* ventajas, 435
definido, 432
desventajas, 436
idea de, 435
procedimientos para hallar, 433
Valor pronosticado, $\hat{y}$, 174
Valores provisionales, 501
Variabilidad, 17
Variable (datos)
categóricos, 9, 147, 621
continuos, 11
cualitativa, 9, 10, 147, 151
cuantitativa, 10, 11, 59, 151, 152
de atributos, 9, 41
de entrada, 152, 179
dependientes, 152, 162
discretos, 11
independientes, 152, 162, 633
nominal, 10
oculta, 167
ordinal, 10
salida, 152
Variable aleatoria, 270, 496
binomial, 287
continua, 11, 271, 315, 327, 344
cuantitativa, 271
discreta, 271, 273, 279, 280, 315, 344
numérica, 271
Variable aleatoria continua, 11, 271, 315, 327, 344
Variable aleatoria cuantitativa, 271
Variable aleatoria discreta, 271, 273, 279, 280, 315, 344
Variable aleatoria discreta, 279
de desviaciones cuadradas, 86
de diferencias pareadas, 551
distribución de frecuencia, 118
distribución de probabilidad, 279
fórmula, 73
muestral, 73, 117, 278
poblacional, 278
representación física, 73
Variable binomial, 287
aleatoria, 270, 496
aleatoria continua, 11, 271, 315, 327, 344
aleatoria discreta, 271, 273, 279, 280, 315, 344
categórica, 9, 147, 621
definida, 8
dependiente, 152, 162
independiente, 633
numérica, 10, 11, 147, 271
respuesta, 8
Variable binomial aleatoria, 287
Variable categórica (datos), 9, 147, 621
Variable cualitativa (datos), 9, 10, 147, 151
Variable cuantitativa (datos), 10, 11, 59, 151, 152
Variable de atributos (datos), 9, 41
Variable de entrada, 152, 179
Variable de salida, 152
Variable dependiente, 152, 162
Variable discreta (datos), 11
Variable nominal (datos), 10
Variable numérica (datos), 10, 11, 47
Variable numérica aleatoria, 271
Variable oculta, 167
Variable ordinal (datos), 10
Varianza
cálculo usando definición, 86
cálculo usando método breve, 88
distribución de frecuencia, 119
distribución de probabilidad, 279
muestra (s^2), definida, 86
muestra (s^2), fórmula breve, 88
muestra (s^2), fórmula, 86
muestral, 279
población (σ^2), 278
razón entre dos, 595
σ^2, 516

"Ventana de datos", 153, 166

x barra, $\bar{x}$, 73
x barra, media muestral, 3, 279
x tilde, x', 74

y circunfleja, 174
ordenada en el origen, 175

$z(\alpha)$, 338, 499
z estrella, 431, 449
z, calificación estándar, 99, 316, 373
z, estadística de prueba
calificación *z*, fórmula básica, 99, 323
dos proporciones, 584, 585
una media, 449
una proporción, 503
z, notación, 338

Formulario para ESTADÍSTICA ELEMENTAL, de Johnson/Kuby, Décima edición

Media muestral:

$\bar{x} = \dfrac{\Sigma x}{n}$ **(2.1)** **o** $\dfrac{\Sigma xf}{\Sigma f}$ **(2.13)**

Profundidad de media muestral:

$d(\bar{x}) = (n + 1)/2$ **(2.2)**

Rango: $H - L$ **(2.4)**

Varianza muestral:

$$s^2 = \frac{\Sigma\left(x-\bar{x}\right)^2}{n-1} \quad \textbf{(2.6)}$$

o

$$s^2 = \frac{\Sigma x^2 - \dfrac{\left(\Sigma x\right)^2}{n}}{n-1} \quad \textbf{(2.10)}$$

o

$$s^2 = \frac{\Sigma x^2 f - \dfrac{\left(\Sigma xf\right)^2}{\Sigma f}}{\Sigma f - 1} \quad \textbf{(2.14)}$$

Desviación muestral estándar:

$s^2 = \sqrt{s^2}$ **(2.7)**

Teorema de Chebyshev: al menos $1 - (1/k^2)$ **(p. 109)**

Suma de cuadrados de *x*:

$SS(x) = \Sigma x^2 - ((\Sigma x)^2/n)$ **(2.9)**

Suma de cuadrados de *y*:

$SS(y) = \Sigma y^2 - ((\Sigma y)^2/n)$ **(3.3)**

Suma de cuadrados de *xy*:

$SS(xy) = \Sigma xy - ((\Sigma x \cdot \Sigma y)/n)$ **(3.4)**

Coeficiente de correlación de Pearson:

$r = SS(xy) / \sqrt{SS(x) \cdot SS(y)}$ **(3.2)**

Ecuación de la recta de mejor ajuste:

$\hat{y} = b_0 + b_1 x$ **(p. 174)**

Pendiente de la recta de mejor ajuste:

$b_1 = SS(xy)/SS(x)$ **(3.6)**

Intersección *y* para recta de mejor ajuste:

$b_0 = [\Sigma y - (b_1 \cdot \Sigma x)]/n$ **(3.7)**

Probabilidad empírica (observada):

$P'(A) = n(A)/n$ **(4.1)**

Probabilidad teórica para espacio muestral igualmente probable:

$P(A) = n(A)/n(S)$ **(4.2)**

Regla del complemento:

$P(\text{no A}) = P(\bar{A}) = 1 - P(A)$ **(4.3)**

Regla general de la adición:

$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$ **(4.4)**

Regla general de la multiplicación:

$P(A \text{ y } B) = P(A) \cdot P(B|A)$ **(4.5)**

Regla especial de la adición para eventos mutuamente exclusivos:

$P(A \text{ o } B \text{ o} \ldots \text{o } E) = P(A) + P(B) + \cdots + P(E)$ **(4.6)**

Regla especial de la multiplicación para eventos independientes:

$P(A \text{ y } B \text{ y} \ldots \text{y } E) = P(A) \cdot P(B) \ldots\ldots P(E)$ **(4.7)**

Media de variable aleatoria discreta:

$\mu = \Sigma[xP(x)]$ **(5.1)**

Varianza de variable aleatoria discreta:

$\sigma^2 = \Sigma[x^2P(x)] - \{\Sigma[xP(x)]\}^2$ **(5.3a)**

Desviación estándar de variable aleatoria discreta:

$\sigma = \sqrt{\sigma^2}$ **(5.4)**

Factorial: $n! = (n)(n-1)(n-2) \cdot \cdot \cdot \cdot \cdot 2 \cdot 1$ **(p. 289)**

Coeficiente del binomio:

$$\binom{n}{x} = \frac{n!}{x! \cdot (n-x)!} \quad \textbf{(5.6)}$$

Función de probabilidad del binomio

$$P(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x},\ x = 0, \ldots, n \quad \textbf{(5.5)}$$

Media de variable aleatoria del binomio:

$\mu = np$ **(5.7)**

Desviación estándar, variable aleatoria del binomio:

$\sigma = \sqrt{npq}$ **(5.8)**

Puntaje estándar: $z = (x - \mu)/\sigma$ **(6.3)**

Puntaje estándar para $\bar{x}$: $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ **(7.2)**

Intervalo de confianza para media, μ (σ conocida):

$\bar{x} \pm z(\alpha/2) \cdot (\sigma / \sqrt{n})$ **(8.1)**

Tamaño muestral para estimación de confianza $1 - \alpha$ para μ:

$n = [z(\alpha/2) \cdot \sigma/E]^2$ **(8.3)**

Estadístico de prueba calculado para H_o: $\mu = \mu_0$ (σ conocida):

$z\bigstar = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$ **(8.4)**

Estimación de intervalo de confianza para media, μ (σ desconocida):

$\bar{x} \pm t(\text{df}, \alpha/2) \cdot (s/\sqrt{n})$ con $\text{df} = n - 1$ **(9.1)**

Estadística de prueba calculada para H_o: $\mu = \mu_0$ (σ desconocida):

$t\bigstar = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$ con $\text{df} = n - 1$ **(9.2)**

Estimación de intervalo de confianza para proporción, *p*:

$p \pm z(\alpha/2) \cdot \sqrt{(p\,q)/n},\ p = x/n$ **(9.6)**

Estadístico de prueba calculado para H_o: $p = p_o$:

$z\bigstar = (p' - p_0) / \sqrt{(p_0 q_0/n)},\ p' = x/n$ **(9.9)**

Estadístico de prueba calculado para H_o: $\sigma^2 = \sigma_0^2$ o $\sigma = \sigma_0$:

$\chi^2\bigstar = (n-1)s^2/\sigma_0^2,\quad \text{df} = n - 1$ **(9.10)**

Diferencia media entre dos muestras dependientes:

Diferencia pareada: $d = x_1 - x_2$ **(10.1)**

Intervalo de confianza para media, μ_d:

$\bar{d} \pm t(g^l, \alpha/2) \cdot s_d/\sqrt{n}$ **(10.2)**

Media muestral de diferencias pareadas:

$\bar{d} = \Sigma d/n$ **(10.3)**

Desviación muestral estándar de diferencias pareadas:

$$s_d = \sqrt{\frac{\Sigma d^2 - \left[\frac{(\Sigma d)^2}{n}\right]}{n-1}} \quad \textbf{(10.4)}$$

Estadístico de prueba calculado para H_o: $\mu_d = \mu_o$

$$t\bigstar = (\bar{d} - \mu_o) / (s_d / \sqrt{n}), \text{ df} = n-1 \quad \textbf{(10.5)}$$

Diferencia entre medias de dos muestras independientes:

Grados de libertad:

df = menor de $(n_1 - 1)$ o $(n_2 - 1)$ **(p. 565)**

Estimación del intervalo de confianza para $\mu_1 - \mu_2$:

$$(\bar{x}_1 - \bar{x}_2) \pm t(\text{df}, \alpha/2) \cdot \sqrt{(s_1^2/n_1) + (s_2^2/n_2)} \quad \textbf{(10.8)}$$

Estadístico de prueba calculado para H_o: $\mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$:

$$t\bigstar = [(\bar{x}_1 - \bar{x}_2) - (\mu_1, \mu_2)_0] / \sqrt{(s_1^2/n_1) + (s_2^2/n_2)} \quad \textbf{(10.9)}$$

Diferencia entre proporciones de dos muestras independientes: $p'_1 q'_1 \backslash n_1$

Intervalo de confianza para $p_1 - p_2$:

$$(p'_1 - p'_2) \pm z(\alpha/2) \cdot \sqrt{\frac{p'_1 q'_1}{n_1} + \frac{p'_2 q'_2}{n_2}} \quad \textbf{(10.11)}$$

Probabilidad observada agrupada:

$$p'_p = (x_1 + x_2) / (n_1 + n_2) \quad \textbf{(10.13)}$$

$$q'_p = 1 - p'_p \quad \textbf{(10.14)}$$

Estadístico de prueba calculado para H_o: $p_1 - p_2 = 0$:

$$z\bigstar = \frac{p'_1 - p'_2}{\sqrt{(p'_p)(q'_p)\left[\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)\right]}} \quad \textbf{(10.15)}$$

Razón de varianzas entre dos muestras independientes:

Estadístico de prueba calculado para H_o: $\sigma_1^2/\sigma_2^2 = 1$:

$$F\bigstar = s_1^2 / s_2^2 \quad \textbf{(10.16)}$$

Estadístico de prueba calculado para datos enumerativos:

$$\chi^2\bigstar = \Sigma[(O - E)^2 / E] \quad \textbf{(11.1)}$$

Experimento multinomial:

Grados de libertad: df = $k - 1$ **(11.2)**

Frecuencia esperada: $E = n \cdot p$ **(11.3)**

Prueba para independencia o Prueba de homogeneidad:

Grados de libertad:

df = $(r - 1) \cdot (c - 1)$ **(11.4)**

Valor esperado: $E = (R \cdot C)/n$ **(11.5)**

Modelo matemático:

$$x_{c,k} = \mu + F_c + \epsilon_{k(c)} \quad \textbf{(12.13)}$$

Suma total de cuadrados:

$$\text{SS(total)} = \Sigma(x^2) - \frac{(\Sigma x)^2}{n} \quad \textbf{(12.2)}$$

Suma de cuadrados debido a factor:

$$\text{SS(factor)} = \left[\left(\frac{C_1^2}{k_1}\right) + \left(\frac{C_2^2}{k_2}\right) + \left(\frac{C_3^2}{k_3}\right) + \ldots\right] - \left[\frac{(\Sigma x)^2}{n}\right] \quad \textbf{(12.3)}$$

Suma de cuadrados debido a error:

$$\text{SS(error)} = \Sigma(x^2) - [(C_1^2/k_1) + (C_2^2/k_2) + (C_3^2/k_3) + \cdots] \quad \textbf{(12.4)}$$

Grados de libertad para total:

df(total) = $n - 1$ **(12.6)**

Grados de libertad por factor

df(factor) = $c - 1$ **(12.5)**

Grados de libertad por error

df(error) = $n - c$ **(12.7)**

Cuadrado medio por factor

MS(factor) = SS(factor)/df(factor) **(12.10)**

Cuadrado medio por error:

MS(error) = SS(error)/df(error) **(12.11)**

Estadístico de prueba calculado para H_o: el valor medio es igual en todos los niveles:

$F\bigstar$ = MS(factor)/MS(error) **(12.12)**

Covarianza de x y y:

$$\text{covar}(x, y) = \Sigma[(x - \bar{x})(y - \bar{y})]/(n - 1) \quad \textbf{(13.1)}$$

Coeficiente de correlación de Pearson:

$$r = \text{covar}(x, y)/(s_x \cdot s_y) \quad \textbf{(13.2)}$$

o

$$r = \text{SS}(xy) / \sqrt{SS(x)\ \text{SS}(y)} \quad \textbf{(3.2)} \text{ o } \textbf{(13.3)}$$

Error experimental: $e = y - \hat{y}$ **(13.5)**

Varianza de error ϵ: $s_e^2 = \Sigma(y - \hat{y})^2 / (n-2)$ **(13.6)**

o

$$s_e^2 = \frac{(\Sigma y^2) - (b_0)(\Sigma y) - (b_1)(\Sigma xy)}{n-2} \quad \textbf{(13.8)}$$

Desviación estándar alrededor de la recta de mejor ajuste:

$$s_e = \sqrt{s_e^2} \quad \textbf{(p. 715)}$$

Cuadrado de error estándar de regresión:

$$s_{b_1}^2 = \frac{s_e^2}{\text{SS}(x)} = \frac{s_e^2}{\Sigma x^2 - [(\Sigma x)^2/n]} \quad \textbf{(13.11)}$$

Intervalo de confianza para β_1:

$$b_1 \pm t(\text{df}, \alpha/2) \cdot s_{b_1} \quad \textbf{(13.12)}$$

Estadística de prueba calculada para H_o: $\beta_1 = 0$

$$t\bigstar = (b_1 - \beta_1)/s_{b_1} \text{ con df } = n-2 \quad \textbf{(13.13)}$$

Intervalo de confianza para valor medio de y en x_0:

$$\hat{y} \pm t(n-2, \alpha/2) \cdot s_e \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\text{SS}(x)}} \quad \textbf{(13.15)}$$

Intervalo de predicción para y en x_0:

$$\hat{y} \pm t(n-2, \alpha/2) \cdot s_e \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\text{SS}(x)}} \quad \textbf{(13.16)}$$

Prueba U de Mann-Whitney:

$$U_a = n_a \cdot n_b + [(n_b) \cdot (n_b + 1)/2] - R_b \quad \textbf{(14.3)}$$

$$U_b = n_a \cdot n_b + [(n_a) \cdot (n_a + 1)/2] - R_a \quad \textbf{(14.4)}$$

Coeficiente de correlación de rango de Spearman:

$$r_s = 1 - \left[\frac{6\Sigma d^2}{n(n^2 - 1)}\right] \quad \textbf{(14.11)}$$

Índice para instrucciones de computadora y calculadora

Técnica estadística	MINITAB	Excel	TI-83/84 Plus
Convenciones básicas	28	29	29
Gráfica de círculo	42	42	42
Diagrama de Pareto	43	43	44
Gráfica de puntos	45	45	45
Diagrama de tallo y hoja	47	47	47
Gráficas de puntos múltiples	49	49	49
Histograma	61	61	62
Ojiva	66	66	66
Media	74	74	74
Mediana	76	76	76
Desviación estándar	88	88	89
Estadística adicional	89	89	89
Percentiles	97	97	97
Resumen de 5º número	98	98	98
Diagrama caja y bigotes	98	98	99
Comandos adicionales	100	100	100
Seleccionar muestras aleatorias	101	101	—
Prueba para normalidad	109	109	109
Estadística de distribución de frecuencia	120	121	121
Instrucciones adicionales,			
ejercicio 2.212	137	137	—
ejercicio 2.216	138	—	—
Tablas de tabulación cruzada	150	150	150
Diagramas de caja lado a lado y gráficas de puntos	152	152	152
Diagrama de dispersión	155	155	155
Coeficiente de correlación	166	166	166
Instrucciones adicionales, ejercicio 3.27	162	—	162
Recta de mejor ajuste	180	180	180
Simular dados, ejercicio 4.20	219	219	220
Instrucciones adicionales, simular M&Ms, ejercicio 4.171	266	266	—
Instrucciones adicionales, discretas, ejercicio 5.25	278	278	—
Probabilidades binomiales	292	292	293
Probabilidad acumulativa, binomial	292	292	293
Instrucciones adicionales,			
ejercicio 5.77	298	298	298
ejercicio 5.78	298	298	298
ejercicio 5.97	303	304	304
Ordenada para curva de distribución normal	328	328	329
Probabilidad acumulativa, distribución normal	329	329	330
Instrucciones adicionales,			
ejercicio 6.65	336	336	336
ejercicio 6.66	336	336	—
ejercicio 6.67	337	337	—
ejercicio 6.69	337	337	338
ejercicio 6.86	347	347	—
ejercicio 6.118	352	352	353
Instrucciones adicionales,			
ejercicio 7.13	368	368	368
ejercicio 7.15	369	369	—
ejercicio 7.27	376	377	377
ejercicio 7.45	383	383	383
ejercicio 7.63	387	388	—
ejercicio 7.64	388	388	—
Intervalo de confianza, media, sigma conocida	409	409	410
Prueba de hipótesis, media, sigma conocida	438	438	439
Probabilidad acumulativa, distribución *t*	479	479	479
Intervalo de confianza, media, sigma desconocida	481	481	481
Prueba de hipótesis, media, sigma desconocida	486	487	487
Intervalo de confianza, proporción	499	500	500
Prueba de hipótesis, proporción	506	506	506
Instrucciones adicionales,			
ejercicio 9.83	511	511	511
ejercicio 9.85	512	512	512
Probabilidad acumulativa, distribución *Ji* cuadrada	518	519	519
Instrucciones adicionales, ejercicio 9.144	528	—	—
Intervalo de confianza, diferencia media	554	554	554
Prueba de hipótesis, media diferencia	556	557	557
Prueba de hipótesis, diferencia de dos medias	571	571	572
Intervalo de confianza, diferencia de dos proporciones	584	584	584
Prueba de hipótesis, diferencia de dos proporciones	587	587	587
Probabilidad acumulativa, distribución *F*	594	595	595
Prueba de hipótesis, razón entre dos varianzas	599	599	599
Prueba de *Ji* cuadrada	641	641	641
Análisis de una vía de varianza	674	674	674
Análisis de regresión (*line of best fit*)	722	722	723
Análisis de regresión (bandas de confianza, predicción)	732	—	—
Prueba de hipótesis, mediana, prueba de signo de muestra I	755	755	756
Prueba de hipótesis, diferencia mediana, prueba de signo	758	758	758
Prueba de hipótesis, Mann-Whitney	771	772	772
Prueba de hipótesis, corre prueba	780	781	781
Prueba de hipótesis, correlación de rango de Spearman	789	789	789

Áreas de la distribución normal estándar

Las entradas de esta tabla son las probabilidades de que una variable aleatoria, con una distribución normal estándar, tome un valor entre 0 y *z*; la probabilidad está representada por el área sombreada bajo la curva de la figura siguiente. Las áreas para valores negativos de *z* se obtienen por simetría.

Segundo lugar decimal en *z*

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999									
4.0	0.49997									
4.5	0.499997									
5.0	0.4999997									

Para detalles específicos acerca del uso de esta tabla para hallar: probabilidades, vea páginas 317-320; coeficientes de confianza, páginas 338-339, 340-341; valores *p*, páginas 432-433, 435; valores críticos, páginas 317-320, 338-339.

Valores críticos de la distribución *t* de Student

Las entradas de esta tabla, $t(\text{df}, \alpha)$, son los valores críticos para la distribución t de Student para los que el área bajo la curva en la cola derecha es α. Los valores críticos para la cola izquierda se encuentran por simetría.

	Cantidad de α en una cola					
	0.25	0.10	0.05	0.025	0.01	0.005
	Cantidad de α en dos colas					
df	0.50	0.20	0.10	0.05	0.02	0.01
3	0.765	1.64	2.35	3.18	4.54	5.84
4	0.741	1.53	2.13	2.78	3.75	4.60
5	0.729	1.48	2.02	2.57	3.37	4.03
6	0.718	1.44	1.94	2.45	3.14	3.71
7	0.711	1.42	1.89	2.36	3.00	3.50
8	0.706	1.40	1.86	2.31	2.90	3.36
9	0.703	1.38	1.83	2.26	2.82	3.25
10	0.700	1.37	1.81	2.23	2.76	3.17
11	0.697	1.36	1.80	2.20	2.72	3.11
12	0.696	1.36	1.78	2.18	2.68	3.05
13	0.694	1.35	1.77	2.16	2.65	3.01
14	0.692	1.35	1.76	2.14	2.62	2.98
15	0.691	1.34	1.75	2.13	2.60	2.95
16	0.690	1.34	1.75	2.12	2.58	2.92
17	0.689	1.33	1.74	2.11	2.57	2.90
18	0.688	1.33	1.73	2.10	2.55	2.88
19	0.688	1.33	1.73	2.09	2.54	2.86
20	0.687	1.33	1.72	2.09	2.53	2.85
21	0.686	1.32	1.72	2.08	2.52	2.83
22	0.686	1.32	1.72	2.07	2.51	2.82
23	0.685	1.32	1.71	2.07	2.50	2.81
24	0.685	1.32	1.71	2.06	2.49	2.80
25	0.684	1.32	1.71	2.06	2.49	2.79
26	0.684	1.32	1.71	2.06	2.48	2.78
27	0.684	1.31	1.70	2.05	2.47	2.77
28	0.683	1.31	1.70	2.05	2.47	2.76
29	0.683	1.31	1.70	2.05	2.46	2.76
30	0.683	1.31	1.70	2.04	2.46	2.75
35	0.682	1.31	1.69	2.03	2.44	2.73
40	0.681	1.30	1.68	2.02	2.42	2.70
50	0.679	1.30	1.68	2.01	2.40	2.68
70	0.678	1.29	1.67	1.99	2.38	2.65
100	0.677	1.29	1.66	1.98	2.36	2.63
df > 100	0.675	1.28	1.65	1.96	2.33	2.58

Para detalles específicos acerca del uso de esta tabla para hallar: coeficientes de confianza, vea páginas 477-480; valores *p*, páginas 483, 486; valores críticos, páginas 477-478.

A través de los años, *Estadística elemental: Lo esencial* se ha transformado en un libro de texto introductorio muy accesible que promueve el aprendizaje, la comprensión y la motivación al presentar la estadística a los estudiantes en un contexto real, además ha respondido a la aceptación gradual en muchas disciplinas donde la estadística se ha vuelto una herramienta importante. Como resultado de lo anterior, las aplicaciones, los ejemplos, proyectos y ejercicios que se presentan en esta edición contienen datos que abarcan una amplia variedad de áreas de interés, incluyendo la física y las ciencias sociales, la opinión pública y la ciencia política, los negocios, la economía y la medicina.

En resumen los autores continúan esforzándose para darle a la estadística un tono de accesibilidad y sentido común que motive a los estudiantes que están más interesados en las aplicaciones que en la teoría.

Como novedad conviene indicar que varios capítulos de esta edición se revisaron completamente, además al final de cada capítulo se incluyen una serie de proyectos, mismos que incluyen un análisis breve que deberá desarrollarse en forma individual o en pequeños grupos de investigación. También se incluyen exámenes de práctica en los capítulos.

En todo el texto se incluye una cantidad abundante de ejemplos, mismos que presentan el proceso de resolución paso a paso de los conceptos estadísticos clave. También se incluye una amplia cantidad de ejemplos de aplicación que incorporan conceptos estadísticos para demostrar cómo trabaja la estadística en el mundo real.

Por último, la obra incluye instrucciones para el uso de Minitab, Excel y la calculadora TI-83/84.

CENGAGE Learning™

http://latinoamerica.cengage.com

ISBN-13: 978-607481199-5
ISBN-10: 607481199-7

9 786074 811995