Estadística aplicada a los negocios y la economía

LIND

MARCHAL

WATHEN

DECIMOSEXTA EDICIÓN

Mc Graw Hill Education

Estadística aplicada a los
NEGOCIOS
y la
ECONOMÍA

Estadística aplicada a los
NEGOCIOS
Y LA
ECONOMÍA

Decimosexta edición

DOUGLAS A. LIND

Coastal Carolina University y Universidad de Toledo

WILLIAM G. MARCHAL

Universidad de Toledo

SAMUEL A. WATHEN

Coastal Carolina University

Revisión técnica

OFELIA VIZCAÍNO DÍAZ

Escuela de Ingeniería y Arquitectura
Instituto Tecnológico y de Estudios Superiores de Monterrey,
Campus Ciudad de México

PEDRO SILVA VELÁZQUEZ

Universidad de Puerto Rico en Humacao

SONIA COLÓN PARRILLA

Universidad de Puerto Rico en Humacao

AIDA E. CARRASQUILLO SÁNCHEZ

Universidad de Puerto Rico en Humacao

Mc Graw Hill Education

MÉXICO • AUCKLAND • BOGOTÁ • BUENOS AIRES • GUATEMALA • LONDRES
MADRID • MILÁN • MONTREAL • NUEVA DELHI • NUEVA YORK • SAN FRANCISCO
SAN JUAN • SANTIAGO • SAO PAULO • SIDNEY • SINGAPUR • ST. LOUIS • TORONTO

Directora de desarrollo de contenido editorial y digital: Patricia Ledezma Llaca
Coordinador sponsor: Jesús Mares Chacón
Coordinadora editorial: Marcela I. Rocha Martínez
Editora de desarrollo: Karen Estrada Arriaga
Supervisor de producción: Zeferino García García

Traducción: Ricardo Martín Rubio Ruiz, María del Pilar Carril Villarreal,
María del Pilar Obón León y Javier León Cárdenas

ESTADÍSTICA APLICADA A LOS
NEGOCIOS
Y LA
ECONOMÍA

DECIMOSEXTA EDICIÓN

Mc
Graw
Hill
Education

Edificio Punta Santa Fe
Prolongación Paseo de la Reforma 1015, Torre A,
Piso 16, Colonia Desarrollo Santa Fe,
Delegación Álvaro Obregón,
C.P. 01376, México, D.F.
Miembro de la Cámara Nacional de la Industria Editorial Mexicana, Reg. Núm. 736

ISBN: 978-607-15-1303-8
ISBN (décima edición): 978-607-15-0742-6

Traducido de la décima edición de *Statiscal Techniques in Business & Economics* by Douglas A. Lind,
William G. Marchal and Samuel A. Wathen, © 2015 by McGraw-Hill Education.
All rights reserved. ISBN 978-0-07-802052-0.

JUC 05/15

1234567890 2346789015

Impreso en México *Printed in Mexico*

DEDICATORIA

A Jane, mi esposa y mejor amiga, y a nuestros hijos, sus esposas y nuestros nietos: Mike y Sue (Steve y Courtney), Steve y Kathryn (Kennedy, Jane y Brady), y Mark y Sarah (Jared, Drew y Nate).

Douglas A. Lind

A mis nuevos nietos (George Orn Marchal, Liam Brophy Horowitz y Eloise Larae Marchal Murray), a mi nuevo yerno (James Miller Nicholson) y a mi nueva esposa (Andrea).

William G. Marchal

A mi maravillosa familia: Isaac, Hannah y Barb.

Samuel A. Wathen

NOTA DE LOS AUTORES

En el transcurso de los años, hemos recibido muchas felicitaciones por este texto, y hemos comprendido que es un favorito de los estudiantes. Reconocemos que eso es un gran cumplido y seguimos trabajando muy duro para mantener ese estatus.

El objetivo de *Estadística aplicada a los negocios y la economía* consiste en proporcionar a aquellos estudiantes que cursan maestrías en administración, marketing, finanzas, contabilidad, economía y otros campos de la administración de negocios, una visión introductoria de las muchas aplicaciones de las estadísticas descriptivas e inferenciales. Nos enfocamos en sus aplicaciones comerciales, pero también utilizamos muchos ejercicios y ejemplos que se relacionan con el mundo actual del estudiante universitario. No es necesario haber cursado estudios previos en estadística, y los requisitos matemáticos corresponden al álgebra de primer año.

En este texto, mostramos a los estudiantes principiantes los pasos que necesitan para tener éxito en un curso básico de estadística; este enfoque paso a paso aumenta el desempeño, acelera la preparación y mejora significativamente la motivación. Entender los conceptos, ver y realizar muchos ejemplos y ejercicios, así como comprender la aplicación de los métodos estadísticos en los negocios y la economía son el enfoque principal de este libro.

En 1967 se publicó la primera edición de este texto; en aquel entonces era difícil localizar datos relevantes relacionados con los negocios. ¡Todo eso ha cambiado! En la actualidad, encontrar los datos ya no constituyen un problema; el número de artículos que se compran en la tienda de abarrotes se registra de manera automática en la caja en la que se realiza el pago. Las compañías telefónicas rastrean constantemente la fecha y hora de nuestras llamadas, su duración y la identidad de la persona a quien llamamos. Las compañías de tarjetas de crédito conservan la información relacionada con el número, hora, fecha y cantidad de nuestras compras. Los aparatos médicos monitorean nuestro ritmo cardiaco, presión sanguínea y temperatura desde lugares remotos. Una gran cantidad de información de negocios se registra y se reporta casi al instante. CNN, USA Today y MSNBC, por ejemplo, publican en sus sitios web los precios de las acciones con un retraso menor a 20 minutos.

En la actualidad se requieren habilidades para manejar un gran volumen de información numérica. Primero, debemos ser consumidores críticos de la información que nos presentan; segundo, necesitamos ser capaces de reducir grandes cantidades de información en una forma concisa y significativa que nos permita realizar interpretaciones, juicios y decisiones eficaces. Todos los estudiantes tienen calculadoras y la mayoría cuenta con computadoras personales o con acceso a ellas en un laboratorio del campus; el software estadístico, como Microsoft Excel y Minitab, está disponible en esas computadoras, y los comandos necesarios para obtener resultados de dichos programas aparecen en el apéndice C, al final del libro. Utilizamos capturas de pantalla en los capítulos para que el estudiante se familiarice con la naturaleza de la aplicación.

Debido a la disponibilidad de software y computadoras, ya no es necesario perder tiempo haciendo cálculos; así que reemplazamos muchos de los ejemplos de cálculo con ejemplos para ayudar al estudiante a entender e interpretar los resultados estadísticos, además, hacemos mayor hincapié en la naturaleza conceptual de los temas estadísticos. No obstante esos cambios, seguimos presentando, de la mejor forma posible, los conceptos claves junto con ejemplos de apoyo interesantes y relevantes.

¿Qué hay de nuevo en esta decimosexta edición?

Hemos hecho algunos cambios en esta edición, y pensamos que resultarán útiles y oportunos para usted y sus alumnos.

- Reorganizamos los capítulos para que cada sección corresponda a un objetivo de aprendizaje y revisamos cada uno de ellos.
- Extendimos a seis pasos el procedimiento de prueba de hipótesis en el capítulo 10, enfatizando la interpretación de los resultados de la prueba.
- Revisamos los ejemplos de varios capítulos:
    - En el capítulo 5 ahora se incluye un nuevo ejemplo para demostrar las tablas de contingencia y los diagramas en árbol; también revisamos el ejemplo que demuestra la fórmula de combinación.

- En el capítulo 6 se incorporó un ejemplo revisado que demuestra la distribución binomial.
- En el capítulo 15 se agregó un nuevo ejemplo que demuestra el análisis de tabla de contingencia.

- Revisamos el ejemplo de regresión simple en el capítulo 13 y aumentamos el número de observaciones para ilustrar mejor los principios de la regresión lineal simple.
- Reordenamos los capítulos no paramétricos y los ubicamos después de los capítulos de estadísticas tradicionales.
- Movimos las secciones en pruebas de una y dos muestras de proporciones, colocando todos los análisis de datos nominales en el capítulo “Métodos no paramétricos: pruebas de hipótesis del nivel nominal”.
- Combinamos las respuestas de los “Ejercicios de autoevaluación” en un nuevo apéndice.
- Unimos los “Comandos de software” en un nuevo apéndice.
- Conjuntamos los glosarios en los repasos de las secciones en uno solo que se incorpora después de los apéndices al final del texto.
- Mejoramos los gráficos en todo el texto.

¿CÓMO SE ORGANIZAN LOS CAPÍTULOS PARA COMPROMETER A LOS ESTUDIANTES Y PROMOVER EL APRENDIZAJE?

Objetivos de aprendizaje del capítulo

En cada capítulo se inicia con un conjunto de objetivos de aprendizaje, diseñados para enfocarse en los temas tratados y motivar el aprendizaje de los alumnos. Estos se localizan en el margen próximo al tema e indican lo que el estudiante debería ser capaz de hacer después de completar el capítulo.

Recientemente, las tiendas BARNES & NOBLE comenzaron a vender un lector electrónico llamado Nook Color, un dispositivo mediante el cual se pueden descargar de manera electrónica más de dos millones de libros, periódicos y revistas y que, además, despliega los materiales descargados a todo color. Suponga que usted sabe cuántos Nook Color se vendieron por día durante el último mes en la tienda Bar-

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA1-1 Explicar por qué es importante conocer de estadística.

OA1-2 Definir el concepto de estadística y proporcionar un ejemplo de su aplicación.

OA1-3 Diferenciar entre estadística descriptiva y estadística inferencial.

OA1-4 Clasificar las variables como cualitativas o cuantitativas, y discretas o continuas.

OA1-5 Distinguir entre los niveles nominal, ordinal, de intervalo y de razón de la medición de datos.

Ejercicio al inicio del capítulo

En cada capítulo se comienza con un ejercicio representativo que muestra cómo el contenido correspondiente se puede aplicar a una situación de la vida real.

Introducción

En el capítulo 2 se inició el estudio de la estadística descriptiva. Con el fin de transformar datos en bruto o no agrupados en alguna forma significativa, es necesario organizarlos en una distribución de frecuencias, la cual se representa en forma gráfica en un histograma o en un polígono de frecuencias. Este arreglo permite visualizar dónde tienden a acumularse los datos, los valores máximo y mínimo, y la forma general de los datos.

En el capítulo 3, primero se calcularon diversas medidas de ubicación o de localización, tales como la media, la mediana y la moda, que permiten informar un valor típico de un conjunto de observaciones. También se calcularon diversas medidas de localización, como el rango, la varianza y

Introducción al tema

En cada capítulo se incluye una revisión de los conceptos importantes del que le antecedió, que se vinculan con el material del capítulo actual; al proporcionar continuidad al flujo de conceptos, este enfoque paso a paso eleva la comprensión.

EJEMPLO

Hay 42 salidas en la autopista I-75, que atraviesa el estado de Kentucky. A continuación aparece la lista de distancias entre salidas (en millas).

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

¿Por qué esta información representa una población? ¿Cuál es la media aritmética de millas entre salidas?

Ejemplo resuelto

Tras introducir los conceptos importantes, se presenta un ejemplo resuelto que ilustra a los estudiantes sobre “cómo hacerlo” y mostrar una aplicación relevante de negocios o basada en la economía; con este recurso se ayuda a responder la pregunta: “¿Para qué puedo usar esto?”

Autoevaluaciones

A lo largo de cada capítulo se presentan autoevaluaciones muy apegadas a los ejemplos previos; esto ayuda a los estudiantes a monitorear su progreso y les proporciona un refuerzo inmediato en cada técnica.

AUTOEVALUACIÓN 3-1

1. Los ingresos anuales de una muestra de empleados de administración media en Westinghouse son: 62 900, 69 100, 58 300 y 76 800 dólares.
   (a) Proporcione la fórmula de la media muestral.
   (b) Determine la media muestral.
   (c) ¿Es la media que calculó en el inciso anterior un estadístico o un parámetro? ¿Por qué razón?
   (d) ¿Cuál es su mejor aproximación de la media de la población?
2. Todos los estudiantes de la clase 411 del curso de ciencias avanzadas de la computación constituyen una población. Sus calificaciones en el curso son 92, 96, 61, 86, 79 y 84.
   (a) Proporcione la fórmula de la media poblacional.
   (b) Calcule la calificación media del curso.
   (c) ¿Es la media que calculó en el inciso anterior un estadístico o un parámetro? ¿Por qué razón?

Estadística en acción

La sección “Estadística en acción” se incluye a lo largo de todo el libro, por lo general, dos veces por capítulo; en ella se proporcionan aplicaciones únicas e interesantes, así como perspectivas históricas en el campo de la estadística.

ESTADÍSTICA EN ACCIÓN

A Florence Nightingale se le conoce como la fundadora de la profesión de enfermería. Sin embargo, también salvó muchas vidas con la ayuda del aná-

Definiciones

Las definiciones de términos nuevos o exclusivos del ámbito estadístico se sitúan independientemente del texto, y se resaltan para facilitar su referencia y revisión; también aparecen en el glosario que está al final del libro.

TABLA DE FRECUENCIAS Agrupación de datos cualitativos en clases mutuamente excluyentes y colectivamente exhaustivas que muestra el número de observaciones en cada clase.

Fórmulas

Las fórmulas que se utilizan por primera vez están encerradas en un recuadro y numeradas para simplificar su referencia; al final se incluye una lista con todas las fórmulas claves.

VARIANZA MUESTRAL $$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \qquad \textbf{[3.9]}$$

Ejercicios

Los ejercicios se ubican después de las secciones dentro del capítulo y al final de este; con estos se cubre el material que se estudió en cada sección.

EJERCICIOS

Las respuestas a los ejercicios impares se encuentran al final del libro, en el apéndice D.

1. Una gráfica de pastel muestra la porción relativa de mercado de los productos de cola. La “rebanada” de Pepsi-Cola tiene un ángulo central de 90 grados. ¿Cuál es su participación de mercado?
2. En un estudio de mercado se pidió a 100 consumidores que seleccionaran el mejor reproductor musical digital entre iPod, iRiver y Magic Star MP3. Con la finalidad de resumir las respuestas de los consumidores en una tabla de frecuencias, ¿cuántas clases debería tener esta?
3. Se preguntó a un total de 1 000 residentes de Minnesota cuál estación del año preferían. Estos fueron los resultados: a 100 les gustaba más el invierno; a 300, la primavera; a 400, el verano y a 200, el otoño. Desarrolle una tabla de frecuencias y una de frecuencias relativas para resumir esta información.
4. Se preguntó a dos mil viajeros frecuentes (de negocios) qué ciudad de la región central de Estados Unidos prefieren: Indianápolis, San Luis, Chicago o Milwaukee. De ellos, 100 contestaron que Indianápolis; 450, San Luis; 1 300, Chicago y el resto dijo que Milwaukee. Elabore una tabla de frecuencias y una tabla de frecuencias relativas para resumir esta información.
5. Wellstone, Inc., produce y comercializa fundas para teléfonos celulares en cinco diferentes colores: blanco brillante, negro metálico, lima magnético, naranja tangerina y rojo fusión. Para estimar la de-

Capturas de pantalla

El texto incluye muchos ejemplos en software, como Excel, MegaStat® y Minitab.

Independent t test.xlsx

	A	B	C	D	E	F	G
1	Casa	Schadek	Bowyer		Prueba t para dos muestras pareadas		
2	1	235	228				
3	2	210	205			*Schadek*	*Bowyer*
4	3	231	219		Media	226.800	222.200
5	4	242	240		Varianza	208.844	204.178
6	5	205	198		Observaciones	10.000	10.000
7	6	230	223		Varianza conjunta	206.511	
8	7	231	227		Diferencia media hipotética	0.000	
9	8	210	215		gl	18.000	
10	9	225	222		Estadístico t	0.716	
11	10	249	245		P(T<=t) de una cola	0.242	
12					t crítica de una cola	1.734	
13	Media =	226.80	222.20		P(T<=t) de dos colas	0.483	
14	S =	14.45	14.29		t crítica de dos colas	2.101	

¿CÓMO SE REFUERZA EL APRENDIZAJE MEDIANTE ESTE TEXTO?

Por capítulo

Resumen del capítulo

Cada capítulo contiene un breve resumen del material que se estudia en él, incluyendo el vocabulario y las fórmulas más importantes.

RESUMEN DEL CAPÍTULO

I. La distribución uniforme es de probabilidad continua, y tiene las siguientes características:
- **A.** Su forma es rectangular.
- **B.** La media y la mediana son iguales.
- **C.** Su valor mínimo *a* y su valor máximo *b* la describen por completo.
- **D.** La siguiente ecuación de la región de *a* a *b* la describe:

$$P(x) = \frac{1}{b - a} \quad \textbf{[7.3]}$$

- **E.** La media y la desviación estándar de una distribución uniforme se calculan de la siguiente manera:

Clave de pronunciación

Esta herramienta enlista el símbolo matemático, su significado y cómo pronunciarlo; pensamos que esto ayudará al estudiante a retener el significado del símbolo y que, en general, mejorará la comunicación en el curso.

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
H_0	Hipótesis nula	*H, subíndice cero*
H_1	Hipótesis alternativa	*H, subíndice uno*
$\alpha/2$	Nivel de significancia de dos colas	*Alfa sobre dos*
$\bar{x}_c$	Límite de la media muestral	*x barra, subíndice c*
μ_0	Media supuesta de la población	*Mu, subíndice cero*

Ejercicios del capítulo

En términos generales, en los ejercicios de final de capítulo se encuentran los mayores desafíos y se integran los conceptos estudiados. Las respuestas y las soluciones ya trabajadas de todos los ejercicios impares aparecen en el apéndice D al final del texto. Muchos ejercicios se señalan con un ícono de archivo de datos al margen; para ellos se crearon documentos de datos en formato Excel que se localizan en el sitio web del texto, **www.mhhe.com/uni/lind_ae16e**. Estos archivos ayudan a los estudiantes a utilizar el software estadístico para resolver los ejercicios.

EJERCICIOS DEL CAPÍTULO

41. La cantidad de bebida de cola en una lata de 12 onzas tiene una distribución uniforme entre 11.96 onzas y 12.05 onzas.
- **a.** ¿Cuál es la cantidad media de bebida por lata?
- **b.** ¿Cuál es la desviación estándar de la cantidad de bebida por lata?
- **c.** ¿Cuál es la probabilidad de elegir una lata de bebida que contenga menos de 12 onzas?
- **d.** ¿Cuál es la probabilidad de elegir una lata de bebida que contenga más de 11.98 onzas?
- **e.** ¿Cuál es la probabilidad de elegir una lata de bebida que contenga más de 11 onzas?

42. Un tubo de pasta dental Listerine Control Tartar contiene 4.2 onzas. Conforme la gente utiliza la pasta, la cantidad que queda en cualquier tubo es aleatoria. Suponga que la cantidad de pasta restante en el tubo tiene una distribución uniforme. De acuerdo con estos datos, es posible determinar información relativa a la cantidad restante de un tubo de pasta dental sin invadir la privacidad de nadie.
- **a.** ¿Cuánta pasta esperaría que quedara en el tubo?
- **b.** ¿Cuál es la desviación estándar de la pasta que queda en el tubo?
- **c.** ¿Cuál es la probabilidad de que en el tubo queden menos de 3.0 onzas?
- **d.** ¿Cuál es la probabilidad de que en el tubo queden más de 1.5 onzas?

43. Muchas tiendas de menudeo ofrecen sus propias tarjetas de crédito. En el momento de hacer la solicitud de crédito, el cliente recibe 10% de descuento en su compra. El tiempo que se requiere para el proceso de la solicitud de crédito se rige por una distribución uniforme con tiempos que varían entre 4 y 10 minutos.
- **a.** ¿Cuál es el tiempo medio que dura el proceso de la solicitud?
- **b.** ¿Cuál es la desviación estándar del tiempo de proceso?
- **c.** ¿Cuál es la probabilidad de que una solicitud tarde menos de seis minutos?
- **d.** ¿Cuál es la probabilidad de que una solicitud tarde más de cinco minutos?

Ejercicios de base de datos

Los ejercicios que están al final de cada capítulo se basan en tres grandes conjuntos de datos, que aparecen en el apéndice A del texto; estos conjuntos confrontan a los estudiantes con aplicaciones del mundo real mucho más complejas.

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

50. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.
- **a.** Un artículo reciente en el *Arizona Republic* indicó que el precio medio de venta de las casas en esta área es superior a 220 000 dólares. Con el nivel de significancia 0.01, ¿puede concluir que el precio medio de venta en el área de Goodyear es superior a 220 000 dólares? Determine el valor *p*.
- **b.** El mismo artículo informó que el tamaño medio es superior a 2 100 pies cuadrados. Con el nivel de significancia 0.01, ¿puede concluir que el tamaño medio de las casas que se vendieron en Goodyear es superior a 2 100 pies cuadrados? Determine el valor *p*.

51. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012.
- **a.** Lleve a cabo una prueba de hipótesis para determinar si el salario medio de los equipos fue distinto de 80.0 millones de dólares. Aplique el nivel de significancia 0.05.

Comandos de software

A todo lo largo del texto se incluyen ejemplos de software que utilizan Excel, MegaStat® y Minitab, pero las explicaciones de los comandos de cada programa para ingresar los datos están al final del texto, en el apéndice C; esto permite que el estudiante se enfoque en las técnicas estadísticas y no en cómo ingresar los datos.

16-7 a.

		Rango			
x	*y*	*x*	*y*	*d*	d^2
805	23	5.5	1	4.5	20.25
777	62	3.0	9	−6.0	36.00
820	60	8.5	8	0.5	0.25
682	40	1.0	4	−3.0	9.00
777	70	3.0	10	−7.0	49.00
810	28	7.0	2	5.0	25.00
805	30	5.5	3	2.5	6.25
840	42	10.0	5	5.0	25.00
777	55	3.0	7	−4.0	16.00
820	51	8.5	6	2.5	6.25
				0	193.00

CAPÍTULO 5

5.1 En seguida se muestran los comandos de Excel para determinar el número de permutaciones de la página 164:

a. Haga clic en la pestaña en la barra de herramientas y seleccione **Insert Function fx**.

b. En el cuadro **Insert Function,** seleccione **Statistical** como categoría**;** vaya al recuadro de abajo y busque **PERMUT** en la lista **Select a function** y haga clic en **OK.**

c. En el cuadro **PERMUT,** introduzca *8* en **Number** y en el cuadro de **Number_chosen,** inserte *3*. La respuesta correcta, 336, aparece dos veces en el cuadro.

Respuestas a las autoevaluaciones

En el apéndice E se proporcionan las soluciones a los ejercicios de autoevaluación.

Por sección

Repasos de las secciones

Se incluye un repaso de sección en varios grupos selectos de capítulos (1-4, 5-7, 8 y 9, 10-12, 13 y 14, 15 y 16, y 17 y 18) a modo de repaso antes del examen. Se incluye una breve **perspectiva general** de los capítulos, un **glosario** de los principales términos y **problemas para repasar**.

REPASO DE LOS CAPÍTULOS 10 a 12

Esta sección es un repaso de los conceptos y términos importantes que se presentaron en los capítulos 10, 11 y 12. En el capítulo 10 se inició el estudio de la prueba de hipótesis (una afirmación acerca del valor del parámetro de una población). Una prueba de hipótesis estadística comienza con una afirmación respecto del valor del parámetro de la población en la hipótesis nula; esta se establece para realizar la prueba. Al completarla se debe rechazar o no la hipótesis nula; si se rechaza, se concluye que la hipótesis alternativa es verdadera. La hipótesis alternativa (también llamada hipótesis de investigación) se "acepta" solo si se demuestra que la hipótesis nula es falsa. La mayoría de las veces se desea probar la hipótesis alternativa.

En el capítulo 10 se seleccionaron muestras aleatorias de una sola población y se probó si era razonable que el parámetro de la población en estudio igualara un valor en particular; por ejemplo, para investigar si el tiempo medio de duración en el métodos para conducir la prueba cuando la desviación estándar de la población estaba disponible y cuando no lo estaba.

En el capítulo 11 se amplió la idea de prueba de hipótesis para verificar si dos muestras aleatorias independientes provenían de poblaciones con las mismas medias poblacionales (o iguales); por ejemplo, el St. Mathews Hospital opera una sala de urgencias en las zonas norte y sur de Knoxville, Tennessee; la pregunta de investigación es: ¿el tiempo de espera medio de los pacientes es igual en ambas salas? Para responder esta pregunta, se selecciona una muestra aleatoria de cada sala y se calculan las medias muestrales; se prueba la hipótesis nula (el tiempo de espera medio es el mismo en las dos salas); la hipótesis alternativa es que el tiempo medio de espera no es el mismo en las dos salas. Si se conocen las desviaciones estándar de cada población, se utiliza la distribución *z* como la del estadístico de prueba; en caso contrario, este sigue la distribución *t*.

Casos

En el repaso también se incluyen casos continuados y otros más pequeños que permiten que los estudiantes tomen decisiones mediante técnicas y herramientas aprendidas en diversos capítulos.

CASOS

A. Century National Bank

Consulte los datos relativos a Century National Bank. ¿Es razonable que la distribución para verificar los saldos de las cuentas se aproxime a una distribución de probabilidad normal? Determine la media y la desviación estándar de una muestra de 60 clientes. Compare la distribución real con la teórica. Mencione algunos ejemplos específicos y haga comentarios acerca de sus conclusiones.

Divida los saldos de las cuentas en tres grupos de 20 cada uno, y coloque la tercera parte más pequeña en el primer grupo;

B. Auditor de elecciones

Algunos temas, como el incremento de los impuestos, la revocación de funcionarios electos o la expansión de los servicios públicos, pueden someterse a un referéndum si se recaban suficientes firmas válidas para apoyar la petición. Desafortunadamente, muchas personas firman la petición aunque no estén registradas en el distrito correspondiente, o lo hacen más de una vez.

Sara Ferguson, auditora de elecciones del condado de Venango, tiene que certificar la validez de las firmas antes de pre-

Cuestionario de práctica

El cuestionario de práctica se diseñó para dar a los estudiantes una idea del contenido que puede aparecer en un examen y cómo este puede estar estructurado; además, se incluyen preguntas objetivas y problemas que cubren el material que se estudió en la sección.

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. ¿Bajo qué condiciones una probabilidad sería mayor a 1 o 100%? **1.** ______________
2. Un _______ es la observación de alguna actividad o el acto de tomar algún tipo de medida. **2.** ______________
3. Un _______ es la recolección de uno o más resultados de un experimento. **3.** ______________
4. Una probabilidad _______ implica que dos o más eventos ocurrirán al mismo tiempo. **4.** ______________
5. En una (5a) _______, el orden en que se cuentan los eventos es importante, pero en una (5b) _______ no lo es. **5. a.** ______________ **5. b.** ______________
6. En una distribución de probabilidad discreta, la suma de los posibles resultados es igual a _______. **6.** ______________
7. ¿Cuál de los siguientes *NO* es un requisito para la distribución binomial? Probabilidad constante de éxito, tres o más resultados, el resultado de los conteos. **7.** ______________

¿DE QUÉ MANERA SE CONECTA LA TECNOLOGÍA CON LOS ESTUDIANTES DE ESTADÍSTICA PARA LOS NEGOCIOS?

McGraw-Hill Connect®

Menos control, más enseñanza y mejor aprendizaje

Connect® es una solución en línea de evaluación y aprendizaje, que brinda a los estudiantes las herramientas y recursos que necesitan para alcanzar el éxito, pues les permite un aprendizaje más rápido y eficaz con mayor retención del conocimiento. Para mayor información acerca de Connect®, contacte a su representante local.

LearnSmart

LearnSmart es un método de autoestudio adaptativo en el que se combinan la práctica, la evaluación y el repaso de cada concepto que aborda el libro de texto:

- Utiliza un motor de búsqueda inteligente para relacionar conceptos e incluir información nueva cuando el usuario está listo para abordarla.
- Se adapta, de manera automática, a la respuesta de cada estudiante y presenta conceptos que amplían la comprensión de cada tema.
- El estudiante emplea menos tiempo en el estudio de los temas que ya domina y practica más los tópicos que aún no comprende en su totalidad.
- Proporciona un repaso continuo en el que solo se le brinda a cada estudiante la guía que necesita.
- Integra el diagnóstico como parte del proceso de aprendizaje.
- Permite evaluar los conceptos que cada estudiante maneja, lo cual deja más tiempo libre para la discusión y las aplicaciones en clase.
- Promueve un dominio mucho más rápido de los conceptos que se abordan en el capítulo.

Para mayor información acerca de LearnSmart, contacte a su representante local.

Centro de recursos en línea

Centro de aprendizaje en línea

www.mhhe.com/uni/lind_ae16e

El centro de aprendizaje en línea (OLC) cuenta con diversos materiales que apoyan el aprendizaje de la estadística.

Apoyos para el estudiante

1. Conjuntos de datos en Excel
2. Documentos en Excel
3. Capítulo 20
4. Apéndices A y B
5. Liga al sitio de MegaStat®

Apoyos para el profesor

Este libro cuenta con diversos materiales de apoyo para el profesor, los cuales están disponibles para quienes adopten el texto. Para más información acerca de este complemento, contacte a su representante local.

AGRADECIMIENTOS

Esta edición de *Estadística aplicada a los negocios y la economía* es producto del esfuerzo de muchas personas: estudiantes, colegas, revisores y el equipo de McGraw-Hill/Irwin. Nuestro agradecimiento para todos ellos. Deseamos expresar nuestra más sincera gratitud a los participantes del grupo de investigación y enfoque, y a los revisores:

Sung K. Ahn
Washington State University–Pullman

Vaughn S. Armstrong
Utah Valley University

Scott Bailey
Troy University

Douglas Barrett
University of North Alabama

Arnab Bisi
Purdue University

Pamela A. Boger
Ohio University–Athens

Emma Bojinova
Canisius College

Ann Brandwein
Baruch College

Giorgio Canarella
California State University–Los Angeles

Lee Cannell
El Paso Community College

James Carden
University of Mississippi

Mary Coe
St. Mary College of California

Anne Davey
Northeastern State University

Neil Desnoyers
Drexel University

Nirmal Devi
Embry Riddle Aeronautical University

David Doorn
University of Minnesota–Duluth

Ronald Elkins
Central Washington University

Vickie Fry
Westmoreland County Community College

Xiaoning Gilliam
Texas Tech University

Mark Gius
Quinnipiac University

Clifford B. Hawley
West Virginia University

Peter M. Hutchinson
Saint Vincent College

Lloyd R. Jaisingh
Morehead State University

Ken Kelley
University of Notre Dame

Mark Kesh
University of Texas

Melody Kiang
California State University–Long Beach

Morris Knapp
Miami Dade College

David G. Leupp
University of Colorado–Colorado State

Teresa Ling
Seattle University

Cecilia Maldonado
Georgia Southwestern State University

John D. McGinnis
Pennsylvania State–Altoona

Mary Ruth J. McRae
Appalachian State University

Jackie Miller
The Ohio State University

Carolyn Monroe
Baylor University

Valerie Muehsam
Sam Houston State University

Tariq Mughal
University of Utah

Elizabeth J. T. Murff
Eastern Washington University

Quinton Nottingham
Virginia Polytechnic Institute and State University

René Ordonez
Southern Oregon University

Ed Pappanastos
Troy University

Michelle Ray Parsons
Aims Community College

Robert Patterson
Penn State University

Joseph Petry
University of Illinois at Urbana-Champaign

Germain N. Pichop
Oklahoma City Community College

Tammy Prater
Alabama State University

Michael Racer
University of Memphis

Darrell Radson
Drexel University

Steven Ramsier
Florida State University

Emily N. Roberts
University of Colorado–Denver

Christopher W. Rogers
Miami Dade College

Stephen Hays Russell
Weber State University

Martin Sabo
Community College of Denver

Farhad Saboori
Albright College

Amar Sahay
Salt Lake Community College and University of Utah

Abdus Samad
Utah Valley University

Nina Sarkar
Queensborough Community College

Roberta Schini
West Chester University of Pennsylvania

Robert Smidt
California Polytechnic State University

Gary Smith
Florida State University

Stanley D. Stephenson
Texas State University–San Marcos

Debra Stiver
University of Nevada–Reno

Bedassa Tadesse
University of Minnesota–Duluth

Stephen Trouard
Mississippi College

Elzbieta Trybus
California State University–Northridge

Daniel Tschopp
Daemen College

Sue Umashankar
University of Arizona

Bulent Uyar
University of Northern Iowa

Jesus M. Valencia
Slippery Rock University

Joseph Van Matre
University of Alabama at Birmingham

Raja Vatti
St. John's University

Holly Verhasselt
University of Houston–Victoria

Angie Waits
Gadsden State Community College

Bin Wang
St. Edwards University

Kathleen Whitcomb
University of South Carolina

Blake Whitten
University of Iowa

Oliver Yu
San Jose State University

Zhiwei Zhu
University of Louisiana

Participantes del grupo de reconocimiento y enfoque

Nawar Al-Shara
American University

Charles H. Apigian
Middle Tennessee State University

Nagraj Balakrishnan
Clemson University

Philip Boudreaux
University of Louisiana at Lafayette

Nancy Brooks
University of Vermont

Qidong Cao
Winthrop University

Margaret M. Capen
East Carolina University

Robert Carver
Stonehill College

Jan E. Christopher
Delaware State University

James Cochran
Louisiana Tech University

Farideh Dehkordi-Vakil
Western Illinois University

Brant Deppa
Winona State University

Bernard Dickman
*Hofstra University*Casey DiRienzo
Elon University

Erick M. Elder
University of Arkansas at Little Rock

Nicholas R. Farnum
California State University–Fullerton

K. Renee Fister
Murray State University

Gary Franko
Siena College

Maurice Gilbert
Troy State University

Deborah J. Gougeon
University of Scranton

Christine Guenther
Pacific University

Charles F. Harrington
University of Southern Indiana

Craig Heinicke
Baldwin-Wallace College

George Hilton
Pacific Union College

Cindy L. Hinz
St. Bonaventure University

Johnny C. Ho
Columbus State University

Shaomin Huang
Lewis-Clark State College

J. Morgan Jones
University of North Carolina at Chapel Hill

Michael Kazlow
Pace University

John Lawrence
California State University–Fullerton

Sheila M. Lawrence
Rutgers, The State University of New Jersey

Jae Lee
State University of New York at New Paltz

Rosa Lemel
Kean University

Robert Lemke
Lake Forest College

Francis P. Mathur
California State Polytechnic University, Pomona

Ralph D. May
Southwestern Oklahoma State University

Richard N. McGrath
Bowling Green State University

Larry T. McRae
Appalachian State University

Dragan Miljkovic
Southwest Missouri State University

John M. Miller
Sam Houston State University

Cameron Montgomery
Delta State University

Broderick Oluyede
Georgia Southern University

Andrew Paizis
Queens College

Andrew L. H. Parkes
University of Northern Iowa

Paul Paschke
Oregon State University

Srikant Raghavan
Lawrence Technological University

Surekha K. B. Rao
Indiana University Northwest

Timothy J. Schibik
University of Southern Indiana

Carlton Scott
University of California, Irvine

Samuel L. Seaman
Baylor University

Scott J. Seipel
Middle Tennessee State University

Sankara N. Sethuraman
Augusta State University

Daniel G. Shimshak
University of Massachusetts, Boston

Robert K. Smidt
California Polytechnic State University

William Stein
Texas A&M University

Robert E. Stevens
University of Louisiana at Monroe

Debra Stiver
University of Nevada–Reno

Ron Stunda
Birmingham-Southern College

Edward Sullivan
Lebanon Valley College

Dharma Thiruvaiyaru
Augusta State University

Daniel Tschopp
Daemen College

Bulent Uyar
University of Northern Iowa

Lee J. Van Scyoc
University of Wisconsin–Oshkosh

Stuart H. Warnock
Tarleton State University

Mark H. Witkowski
University of Texas at San Antonio

William F. Younkin
University of Miami

Shuo Zhang
State University of New York, Fredonia

Zhiwei Zhu
University of Louisiana at Lafayette

Sus sugerencias y un repaso cuidadoso de la edición anterior y del original de esta edición contribuyen a mejorar el contenido.

En especial estamos agradecidos con las siguientes personas: el profesor Malcolm Gold, de Avila University, quien revisó el original y las pruebas, así como el manual de soluciones, para verificar la precisión de los ejercicios; el profesor José López-Calleja, de Miami Dade College-Kendall, quien elaboró el banco de pruebas; la profesora Vickie Fry, de Westmoreland County Community College, quien comprobó la exactitud de los ejercicios *Connect.*

También deseamos agradecer al personal de McGraw-Hill/Irwin, entre ellos, a Thomas Hayward, editor ejecutivo; a Kaylee Putbrese, editora de desarrollo; Diane Nowaczyk, gerente de proyecto y a quienes no conocemos personalmente y que hicieron valiosas contribuciones.

MEJORAS EN LA DECIMOSEXTA EDICIÓN DE ESTADÍSTICA APLICADA A LOS NEGOCIOS Y LA ECONOMÍA

Principales cambios a los capítulos individuales:

Capítulo 1 ¿Qué es la estadística?

- Se incluyó una fotografía y un ejercicio al inicio del capítulo sobre el Nook Color que vende Barnes & Noble.
- Se agregó una introducción con nuevas gráficas que muestran la creciente cantidad de información recabada y procesada con nuevas tecnologías.
- Se incluyó un ejemplo de la escala ordinal basada en clasificaciones de los estados según el clima de negocios.
- En el capítulo se incluyen varios ejemplos nuevos.
- El capítulo se enfoca más en los objetivos de aprendizaje revisados y en mejorar el flujo del texto.
- El ejercicio 17, revisado, se basa en datos económicos.

Capítulo 2 Descripción de datos: tablas de frecuencias, distribuciones de frecuencias y su representación gráfica

- Se revisó la autoevaluación 2.3 para incluir datos.
- Se actualizó la lista de compañías del ejercicio 38, revisado.
- Se incorporaron ejercicios nuevos o revisados (45, 47 y 48).

Capítulo 3 Descripción de datos: medidas numéricas

- Se reorganizó el capítulo con base en los objetivos de aprendizaje revisados.
- Se reemplazó la desviación media para enfatizar la varianza y la desviación estándar.
- Se actualizaron los recuadros "Estadística en acción".

Capítulo 4 Descripción de datos: presentación y análisis

- Se actualizó el ejercicio 22 con los salarios de los jugadores de los Yankees de Nueva York en 2012.

Capítulo 5 Estudio de los conceptos de la probabilidad

- Se incluyó una nueva explicación de la posibilidad comparada con la probabilidad.
- Se incluyó un nuevo ejercicio (21).
- Se agregó un nuevo ejemplo para demostrar las tablas de contingencia y los diagramas en árbol.
- Se incorporó un nuevo ejercicio (31) para las tablas de contingencia.
- Se revisó el ejemplo que demuestra la fórmula de combinación

Capítulo 6 Distribuciones discretas de probabilidad

- Se revisó la sección de la distribución binomial.
- Se revisó el ejemplo que demuestra la distribución binomial.
- Se revisó la autoevaluación 6.4 aplicando la distribución binomial.
- Se incluyó un nuevo ejercicio (10) utilizando el número de préstamos "por debajo del agua".
- Se incorporó un nuevo ejercicio utilizando un sorteo en un club de golf local para demostrar la probabilidad y los beneficios esperados.

Capítulo 7 Distribuciones de probabilidad continua

- Se actualizaron los recuadros "Estadística en acción".
- Se revisó la autoevaluación 7.2 basada en el consumo personal diario de agua.
- Se revisó la explicación de la regla empírica según se relaciona con la distribución normal.

Capítulo 8 Métodos de muestreo y teorema del límite central

- Se incluyó un nuevo ejemplo del muestreo aleatorio simple y la aplicación de la tabla de números aleatorios.
- Se revisaron las exposiciones de muestreo aleatorio sistemático, muestreo aleatorio estratificado y el muestreo por conglomerados.
- Se revisó el ejercicio 44 basado en el teorema del límite central.

Capítulo 9 Estimación e intervalos de confianza

- Se integró un nuevo recuadro "Estadística en acción", que describe la economía de combustible del EPA.
- Se incorporó una nueva sección sobre estimación de puntos.
- Integración y aplicación del teorema del límite central.
- Se incorporó una nueva presentación sobre el uso de la tabla *t* para encontrar valores *z*.
- Se revisó la exposición acerca de la determinación del intervalo de confianza para la media poblacional.
- Se extendió la sección sobre el cálculo del tamaño de la muestra.
- Se agregó un nuevo ejercicio (12) acerca del consumo de leche.

Capítulo 10 Pruebas de hipótesis de una muestra

- Se incluyó un nuevo ejemplo sobre el estacionamiento del aeropuerto.
- Se revisaron la solución de software y la explicación de los valores *p*.
- Se incorporaron nuevos ejercicios acerca del consumo diario de agua (17) y del número de mensajes de texto entre los adolescentes (19).
- La prueba de hipótesis sobre la proporción de la población se movió al capítulo 15.
- Se incluyó un nuevo ejemplo que introduce el concepto de prueba de hipótesis.
- Se añadió un sexto paso al procedimiento de prueba de hipótesis que enfatiza la interpretación de los resultados.

Capítulo 11 Pruebas de hipótesis de dos muestras

- Se sustituyó la introducción del capítulo.
- La sección de las pruebas de proporción de dos muestras se movió al capítulo 15.
- Se cambiaron los subíndices en el ejemplo para su mejor comprensión.
- Se actualizó el ejercicio con los salarios de los Yankees de Nueva York para 2012.

Capítulo 12 Análisis de la varianza

- Se incorporó una nueva introducción al capítulo.
- Se incluyó un nuevo ejercicio utilizando la velocidad de los buscadores para navegar en internet (24).
- Se revisó el ejercicio 33, comparando el aprendizaje tradicional contra los cursos en línea.
- Se integró una nueva sección sobre la comparación de dos varianzas de población.
- Se incluyó un nuevo ejemplo que ilustra la comparación de las varianzas.
- Se revisó la sección de la ANOVA de dos vías, con interacción con nuevos ejemplos y un ejemplo también revisado.
- Se revisaron los nombres de las aerolíneas en el ejemplo de la ANOVA de una vía.
- Se cambiaron los subíndices en el ejemplo para su mejor comprensión.
- Se incorporó un nuevo ejercicio acerca de los tiempos de vuelo entre Los Ángeles y San Francisco (30).

Capítulo 13 Regresión lineal y correlación

- Se reescribió la introducción del capítulo.
- Se cambiaron los datos utilizados como base para el ejemplo de Copier Sales de Norteamérica que se utiliza a lo largo del capítulo y se extendió a 15 observaciones, para demostrar más claramente los objetivos de aprendizaje del capítulo.
- Se revisó la sección sobre la transformación de datos, utilizando la relación económica entre precio y ventas.
- Se incluyeron nuevos ejercicios acerca de la transformación de datos (35), los precios y puntuaciones del torneo The Masters (36), los puntos de la NFL en 2012 contra los puntos permitidos (43), el tamaño de un almacén y sus ventas (44) y las distancias y tarifas de una aerolínea (61).

Capítulo 14 Análisis de regresión múltiple

- Se reescribió la sección sobre cómo evaluar la ecuación de la regresión múltiple.
- Se hizo mayor hincapié en la tabla de regresión ANOVA.
- Se resaltó la exposición sobre el valor *p* en la toma de decisiones.
- Se enfatizó el cálculo del factor de varianza de la inflación para evaluar la multicolinealidad.

Capítulo 15 Métodos no paramétricos: pruebas de nivel nominal

- Se movió y renombró el capítulo.
- Se movieron a este capítulo las pruebas de proporciones de una y dos muestras de los capítulos 10 y 11.
- Se incluyó un nuevo ejemplo que introduce las pruebas de bondad de ajuste.
- Se retiraron los métodos gráficos para evaluar la normalidad.
- Se revisó la sección de la tabla de análisis de contingencia con un nuevo ejemplo.
- Se revisaron los ejercicios de conjunto de datos.

Capítulo 16 Métodos no paramétricos: análisis de datos ordinales

- Se movió y renombró el capítulo.
- Se incorporaron un ejemplo y una autoevaluación nuevos que demuestran una prueba de hipótesis de la mediana.
- Se integró un nuevo ejemplo que demuestra la correlación entre el rango y el orden.

Capítulo 17 Números índices

- Se movió el capítulo para que quedara después de las estadísticas no paramétricas.
- Se actualizaron los datos, las ilustraciones y los ejemplos.
- Se revisó el ejemplo que demuestra el uso del Índice de Precios al Productor para desinflar los dólares de las ventas.
- Se revisó el ejemplo que demuestra la comparación del Promedio Industrial Dow Jones y el Nasdaq utilizando la indexación.
- Se incluyó una nueva autoevaluación acerca del uso de los índices para comparar dos medidas distintas en el transcurso del tiempo.
- Se revisó el ejercicio de establecimiento de datos.

Capítulo 18 Series de tiempo y proyección

- Se movió el capítulo para que quedara después de las estadísticas no paramétricas y los números índices.
- Se actualizaron los datos, las ilustraciones y los ejemplos.
- Se revisó la sección de los componentes de una serie de tiempo.
- Se revisaron las gráficas para proporcionar una mejor ilustración.

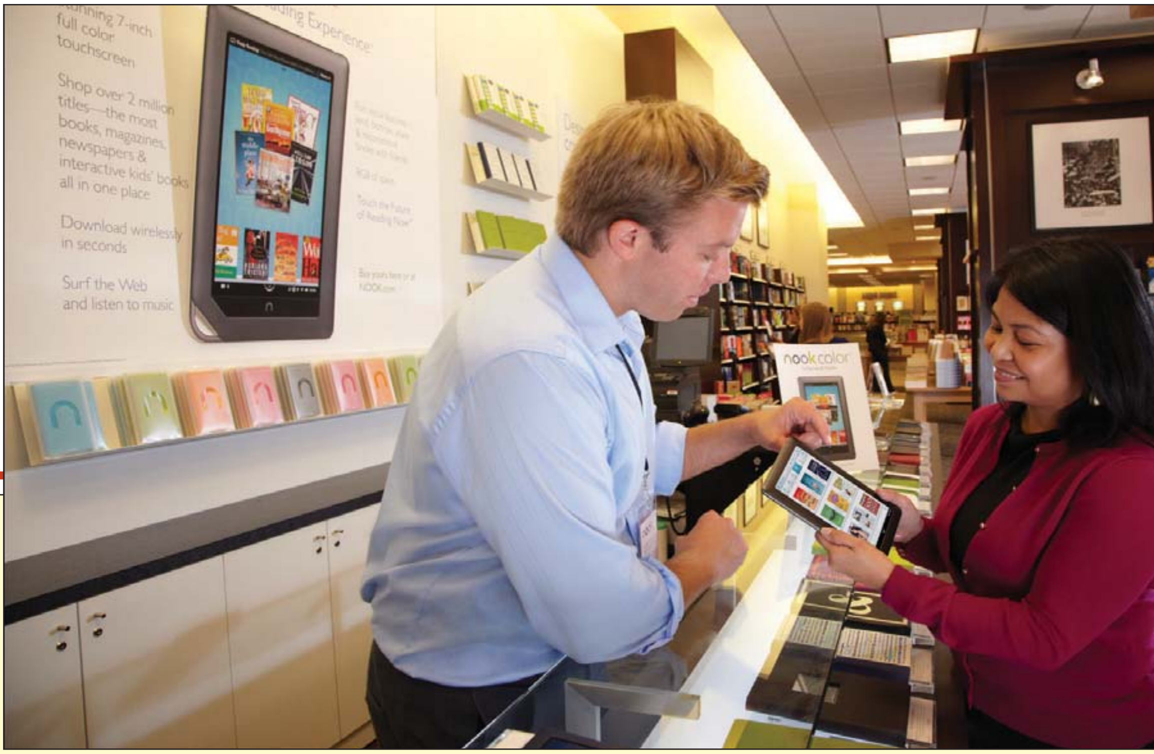
Capítulo 19 Control estadístico del proceso y administración de calidad

- Se actualizó la sección de los ganadores de la Malcolm Baldrige National Quality Award, 2012.

CONTENIDO

¿Qué es la estadística?

1



Recientemente, las tiendas BARNES & NOBLE comenzaron a vender un lector electrónico llamado Nook Color, un dispositivo mediante el cual se pueden descargar de manera electrónica más de dos millones de libros, periódicos y revistas y que, además, despliega los materiales descargados a todo color. Suponga que usted sabe cuántos Nook Color se vendieron por día durante el último mes en la tienda Barnes & Noble del centro comercial Market Commons en Riverside, California. Describa una condición en la que esta información podría considerarse una muestra. Ejemplifique una segunda situación en la que los mismos datos podrían representar una población (vea el ejercicio 11 y el objetivo de aprendizaje OA1-3).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA1-1 Explicar por qué es importante conocer de estadística.

OA1-2 Definir el concepto de estadística y proporcionar un ejemplo de su aplicación.

OA1-3 Diferenciar entre estadística descriptiva y estadística inferencial.

OA1-4 Clasificar las variables como cualitativas o cuantitativas, y discretas o continuas.

OA1-5 Distinguir entre los niveles nominal, ordinal, de intervalo y de razón de la medición de datos.

OA1-6 Enlistar los valores asociados con la práctica de la estadística.

Introducción

Suponga que trabaja para una gran empresa, y su supervisor le pide decidir entre producir y vender una nueva versión de un smartphone o no hacerlo. Usted comienza pensando en las innovaciones y nuevas características del producto. Después, se detiene y se da cuenta del peso de la decisión. El producto deberá ser rentable, por lo que el precio y los costos de producción y distribución son muy importantes. La determinación de introducir el producto se basa en muchas alternativas. Así que, ¿cómo puede usted decidir? ¿Por dónde comenzar?

Al no tener una vasta experiencia en la industria, es esencial empezar a desarrollar una inteligencia que le convierta en experto. Usted elige a tres personas más para trabajar y se reúne con ellas. La conversación se centra en lo que usted necesita saber y en los datos e información que precisa. En esa reunión se plantean muchas preguntas. ¿Cuántos competidores hay en el mercado? ¿Cómo se establece el precio de los smartphones? ¿Qué características de diseño tienen los productos de la competencia? ¿Qué características requiere el mercado? ¿Qué esperan los clientes de un smartphone? ¿Qué características de los productos existentes les gustan a los consumidores? Las respuestas estarán basadas en la inteligencia comercial; es decir, en los datos e información recabados a través de encuestas al consumidor, análisis de ingeniería e investigación de mercado. Al final, la presentación para sustentar su decisión (introducir o no un nuevo smartphone) se basará en la estadística que utilice para resumir y organizar sus datos, comparar el nuevo producto con los ya existentes y estimar las futuras ventas, costos y rendimientos. La estadística será el foco de la futura conversación con su supervisor acerca de esta importante decisión.

Como persona responsable de ciertas decisiones, usted deberá adquirir y analizar datos para sustentar sus determinaciones. El propósito de este libro es desarrollar su conocimiento de técnicas y métodos estadísticos básicos y mostrarle cómo aplicarlos para desarrollar la inteligencia personal y de negocios que le ayuden a tomar decisiones.

OA1-1
Explicar por qué es importante conocer de estadística.

¿Por qué estudiar estadística?

Si revisa el plan de estudios de su universidad, notará que varios programas universitarios incluyen estadística. A medida que investigue distintas carreras, como contabilidad, economía, recursos humanos, finanzas u otros campos de negocios, descubrirá que también incluyen esa materia. ¿Por qué el estudio de la estadística es un requisito en tantas disciplinas?

Una razón de peso para saber de estadística son las tecnologías disponibles para captura de datos. Los ejemplos incluyen la tecnología que utiliza Google para rastrear la forma en que los usuarios de internet acceden a diversos sitios. A medida que la gente utiliza el buscador, Google registra cada consulta y luego emplea estos datos para desplegar y priorizar los resultados de futuras solicitudes de información. Un estimado reciente indica que Google procesa 20 000 terabytes de información por día. Los grandes minoristas como Target, Walmart, Kroger y otros escanean cada compra y utilizan los datos para manejar la distribución de productos, tomar decisiones relacionadas con ventas y marketing y rastrear las ventas por día e incluso por hora. Los departamentos de policía recaban y utilizan datos para proporcionar a los ciudadanos mapas que comunican información acerca de crímenes cometidos y su ubicación. Todas las organizaciones recolectan y utilizan datos para desarrollar el conocimiento y la inteligencia que ayudarán a la gente a tomar decisiones informadas y para rastrear la implementación de estas decisiones. En la ilustración que se presenta en esta página se muestra la cantidad de datos que se generan cada minuto (www.domo.com). Conocer de manera profunda la estadística lo ayudará a resumir y organizar los datos; así proporcionará información útil y sustentable para la toma de decisiones. La estadística se utiliza para realizar comparaciones válidas y predecir los resultados de las decisiones.

En resumen, existen cuando menos tres razones para estudiar estadística: 1) los datos se colectan en todas partes y se requiere de conocimiento estadístico para que la información sea útil; 2) las técnicas estadísticas se utilizan para tomar decisiones personales y profesionales; y 3) sin importar cuál sea su carrera, usted necesitará saber estadística para entender el mundo y desarrollarse en esa carrera. Comprender la estadística y su método le permitirá tomar decisiones personales y profesionales más efectivas.

¿Qué se entiende por estadística?

OA1-2
Definir el concepto de estadística y proporcionar un ejemplo de su aplicación.

Esta pregunta puede replantearse en dos formas sutiles y diferentes: ¿qué son los estadísticos? y, ¿qué es la estadística? Para responder a la primera cuestión, las estadísticas son un número utilizado para comunicar información. Ejemplos de **estadísticos** son:

- El índice de inflación es 2%.
- Su puntuación promedio al graduarse es 3.5.
- El precio del nuevo sedán Tesla Premium eléctrico es 85 400 dólares.

Cada una de estos estadísticos es un hecho numérico y comunica información muy limitada que en sí misma no es muy útil. Sin embargo, al reconocer que cada uno de estos estadísticos es parte de un asunto más grande, entonces aplica la pregunta "¿qué es la estadística?". La estadística es un conjunto de conocimientos y habilidades utilizadas para organizar, resumir y analizar datos. Los resultados del análisis estadístico originan conversaciones interesantes en busca del conocimiento y la inteligencia que sustentan decisiones. Por ejemplo:

- El índice de inflación para el año calendario fue 2%. Al aplicar la estadística podríamos comparar el índice de inflación de este año con observaciones pasadas de la inflación. ¿Es más alto, más bajo o casi el mismo? ¿La tendencia es hacia el aumento o hacia la disminución de la inflación? ¿Existe una relación entre las tasas de interés y los bonos del gobierno?
- Su puntuación promedio al graduarse (PPG) es 3.5. Al recolectar datos y aplicar la estadística es posible determinar el PPG requerido para ser admitido en el programa de la maestría en administración de empresas de la Universidad de Chicago, Harvard, o la Universidad de Michigan. Es posible determinar la probabilidad de ingresar a un programa de estudios en particular. Usted puede estar interesado en entrevistarse para obtener un puesto gerencial en Procter & Gamble. ¿Qué PPG requiere esa empresa para los graduados universitarios con una licenciatura? ¿Existe un rango de PPG aceptable?
- Usted presupuesta un auto nuevo. Le gustaría tener uno eléctrico, con poco impacto ecológico. El precio del sedán Tesla Premium eléctrico es 85 400 dólares. Al recabar datos adicionales y al aplicar la estadística, usted podrá analizar sus opciones. Por ejemplo, otra alternativa es un auto híbrido que funciona con gasolina o electricidad, como el Toyota Prius. Puede comprarlo por casi 27 000 dólares. Otro híbrido, el Chevrolet Volt, cuesta unos 32 000 dólares. ¿Cuáles son las diferencias en las especificaciones de los autos? ¿Qué información puede recabar y resumir para tomar una buena decisión de compra?

Otro ejemplo del uso de la estadística para proporcionar información para evaluar decisiones es la distribución y participación en el mercado de los productos Frito-Lay. Se recaban datos sobre cada una de las líneas de productos de esa marca entre los que se incluyen la participación en el mercado y la cantidad de producto vendido. La estadística se utiliza para presentar la información en una gráfica de barras en la gráfica 1.1, en la cual se muestra claramente el dominio de Frito-Lay en los mercados de frituras de papa, maíz y tortilla. También se muestra la cantidad absoluta de cada línea de producto que se consume en Estados Unidos.

Estos ejemplos muestran que la estadística es más que la presentación de información numérica. La estadística implica reunir y procesar información para crear una conversación, estimular preguntas adicionales y proporcionar la base para la toma de decisiones. La definición específica de la estadística es:

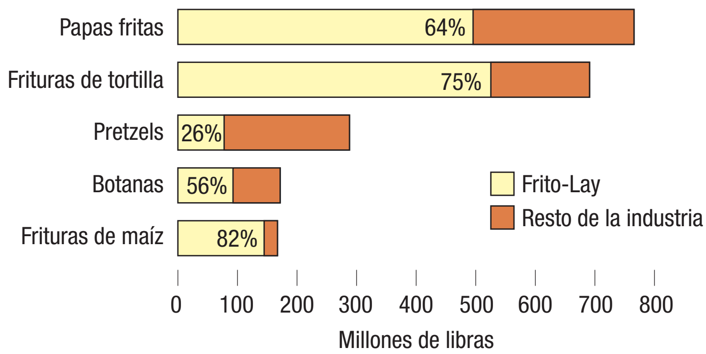
ESTADÍSTICA Ciencia por medio de la cual se recogen, organizan, presentan, analizan e interpretan datos con el fin de propiciar una toma de decisiones más eficaz.

En este libro usted aprenderá a utilizar las técnicas básicas y aplicaciones de la estadística que pueden ayudarle a sustentar sus decisiones, tanto personales como profesionales. Para comenzar, diferenciaremos entre estadística descriptiva e inferencial.

ESTADÍSTICA EN ACCIÓN

Centre su atención en el título de esta sección: "Estadística en acción". Al leer con cuidado obtendrá una idea de la amplia gama de aplicaciones de la estadística en la administración, economía, enfermería, cumplimiento de la ley, deportes y otras disciplinas.

- En 2013, *Forbes* publicó una lista de los estadounidenses más ricos. William Gates, fundador de Microsoft Corporation, aparecía como el número uno. Su fortuna se calculaba en 66 mil millones de dólares (**www.forbes.com**).
- En 2013, las cuatro compañías estadounidenses con mayores ingresos fueron Cargill, Koch Industries, Mars y Bechtel (**www.forbes.com**).
- En Estados Unidos, un típico estudiante graduado de la escuela secundaria gana 652 dólares por semana; el egresado universitario promedio gana 1 066 dólares por semana; y un posgraduado gana 1 300 dólares por semana (**www.bis.gov/emp/ep_chart_001.htm**).



GRÁFICA 1.1 Volumen y acciones de Frito-Lay en las principales categorías de botanas en los supermercados de Estados Unidos

OA1-3 Diferenciar entre estadística descriptiva y estadística inferencial.

Tipos de estadística

Cuando utilizamos la estadística para generar información y tomar decisiones a partir de dichos datos, usamos ya sea la estadística descriptiva o la inferencial. Su aplicación depende de las preguntas planteadas y del tipo de datos disponibles.

Estadística descriptiva

Una masa de datos desorganizados —como un censo de población, los salarios semanales de miles de programadores de computadoras y las respuestas de dos mil votantes registrados para elegir al presidente de Estados Unidos— resulta de poca utilidad. No obstante, las técnicas de la estadística descriptiva permiten organizar esta clase de datos y darles significado. Definimos a la **estadística descriptiva** como:

ESTADÍSTICA DESCRIPTIVA Métodos para organizar, resumir y presentar datos de manera informativa.

A continuación se presentan algunos ejemplos de estadística descriptiva para resumir una gran cantidad de datos y proporcionar información que sea fácil de entender.

- Hay un total de casi 68 859 kilómetros de carreteras interestatales en Estados Unidos. El sistema interestatal representa apenas 1% del total de carreteras de esa nación, aunque alberga a más de 20% del tránsito. La más larga es la autopista I-90, que va de Boston a Seattle, una distancia de 4 957.32 kilómetros. La más corta es la I-878, localizada en Nueva York, cuya longitud es de 1.12 kilómetros. Alaska no cuenta con carreteras interestatales; Texas posee la mayor cantidad de kilómetros interestatales, poco más de 5 201, y Nueva York tiene la mayoría de las rutas interestatales, 28 en total.
- Una persona promedio gastó 103 dólares en mercancía alusiva a San Valentín el 14 de febrero de 2013. Esto representa un aumento de 0.50 dólares con respecto a 2012. Como en años anteriores, los hombres gastaron el doble que las mujeres en esa fecha. El hombre promedio gastó 135.35 dólares para impresionar a sus seres queridos, mientras que las mujeres solo gastaron 72.28. Las mascotas también sienten amor: una persona promedio gastó 3.27 dólares en su amigo peludo, en comparación con los 2.17 del año anterior.

Los métodos y técnicas estadísticos para generar estadística descriptiva se presentan en los capítulos 2 y 4, e incluyen organizar y resumir los datos mediante distribuciones de frecuencia y presentarlas en tablas y gráficas. Además, las medidas estadísticas para resumir las características de una distribución se analizan en el capítulo 3.

Estadística inferencial

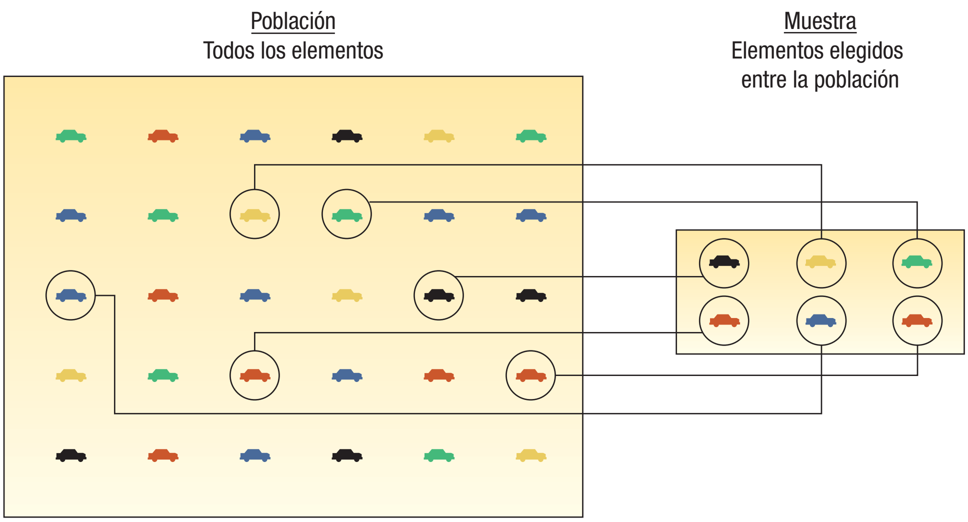
A veces debemos tomar decisiones a partir de un grupo limitado de datos. Por ejemplo, quisiéramos conocer las características de operación tales como la eficiencia del uso de combustible medido en kilómetros por litro de los vehículos deportivos utilitarios (SUV) que se usan actualmente. Si gastamos mucho tiempo, dinero y esfuerzo, podríamos encuestar a todos los dueños de estos vehículos. En este caso, nuestro objetivo sería encuestar a la **población** de dueños de SUV.

POBLACIÓN Conjunto de individuos u objetos de interés o medidas que se obtienen a partir de todos esos individuos u objetos.

Sin embargo, basándonos en la estadística inferencial podríamos encuestar a un número limitado de propietarios de SUV y recabar una **muestra** de la población.

MUESTRA Porción o parte de la población de interés.

A menudo, las muestras se utilizan para obtener estimados confiables de parámetros de población (las muestras se analizan en el capítulo 8). En el proceso se realizan compensaciones entre el tiempo, el dinero y el esfuerzo para recabar los datos y el error de estimar un parámetro de población. El proceso de muestreo de las SUV se ilustra en la siguiente gráfica. En este ejemplo se investiga la media (o promedio) de la eficiencia de combustible del vehículo. Para estimar la media de la población, se muestrean seis SUV y se calcula la media de su rendimiento.



Así, el ejemplo de las seis SUV representa la evidencia de la población que se utiliza para llegar a una inferencia o conclusión acerca del rendimiento de todas las SUV. El proceso de muestreo de una población con el objeto de estimar sus propiedades se llama **estadística inferencial.**

ESTADÍSTICA INFERENCIAL Métodos que se emplean para determinar una propiedad de una población con base en la información de una muestra de esta.

La estadística inferencial se utiliza ampliamente para saber algo acerca de una población en los negocios, la agricultura y el gobierno, como se muestra en los siguientes ejemplos:

- Las cadenas de televisión hacen un monitoreo continuo de la popularidad de sus programas y contratan a Nielsen y otras organizaciones con el fin de que estas tomen muestras sobre las preferencias del auditorio. Por ejemplo, 10.5% de una muestra de hogares con televisión vio *The Big Bang Theory* durante la semana del 25 de febrero de 2012 (**www.nielsen.com**). Estos índices de audiencia se emplean para tomar decisiones acerca de las tarifas de publicidad o para continuar o suspender un programa.
- En 2012, se seleccionó una muestra de 40 sitios del programa de voluntarios de la Administración Federal de Ingresos de Estados Unidos y se preparó a los asesores fiscales voluntarios con tres declaraciones de impuestos estándar. La muestra indicó que las declaraciones se

completaban con un rango de exactitud de 49%. En este ejemplo se utilizó la estadística para tomar decisiones acerca de cómo mejorar el rango de exactitud, corrigiendo los errores más comunes y mejorando la capacitación de los voluntarios (**www.treasury.gov/tigta/auditreports/2012reports/20124008fr.pdf**).

Una característica de este libro son los ejercicios de autoevaluación, los cuales se encuentran intercalados en cada capítulo. A continuación se presenta el primero. Cada uno pone a prueba su comprensión del material precedente. La respuesta y método de solución aparecen en el apéndice E. Le recomendamos resolver primero cada uno y después comparar su respuesta.

AUTOEVALUACIÓN 1-1

Las respuestas están en el apéndice E.

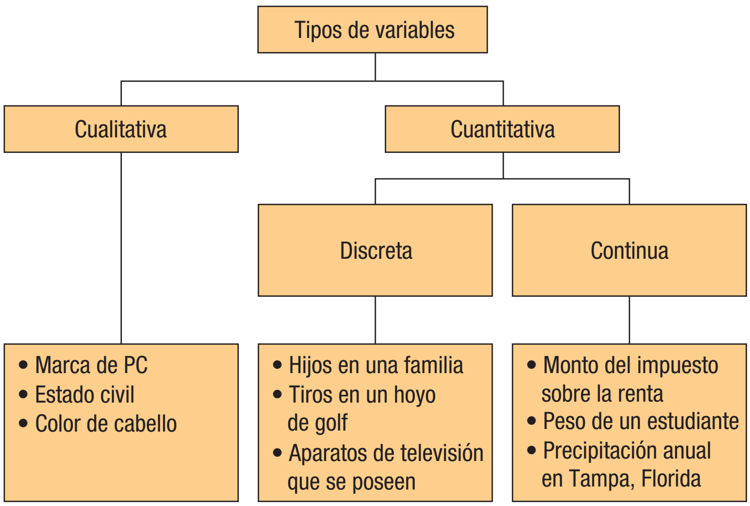
La empresa de publicidad Brandon and Associates, con sede en Atlanta, solicitó a una muestra de 1 960 consumidores que probaran un nuevo platillo con pollo elaborado por Boston Market. De las 1 960 personas de la muestra, 1 176 dijeron que comprarían el alimento si se comercializaba.

(a) ¿Qué podría informar Brandon and Associates a Boston Market respecto de la aceptación en la población del platillo?
(b) ¿Es un ejemplo de estadística descriptiva o estadística inferencial? Explique su respuesta.

OA1-4 Clasificar las variables como cualitativas o cuantitativas, y discretas o continuas.

Tipos de variables

Existen dos tipos básicos de variables: 1) cualitativas y 2) cuantitativas (vea la gráfica 1.2). Cuando el objeto se observa y registra como una característica no numérica, recibe el nombre de *variable cualitativa* o *atributo*. Algunos ejemplos de variables cualitativas son: género, preferencia en bebidas, tipo de automóvil que se posee, estado de nacimiento y color de ojos. Cuando la variable es cualitativa, por lo general se cuenta el número de observaciones para cada categoría y se determina el porcentaje de cada una. Por ejemplo, en la variable color de ojos, ¿qué porcentaje de la población tiene ojos cafés? Si la variable es el tipo de vehículos, ¿qué porcentaje del total de automóviles vendidos el mes pasado eran SUV? Con frecuencia, las variables cualitativas se resumen en tablas y gráficas de barras (capítulo 2).



GRÁFICA 1.2 Resumen de los tipos de variables

Cuando la variable puede presentarse en forma numérica, se le denomina *variable cuantitativa*; por ejemplo, el saldo en su cuenta de cheques, las edades de los presidentes de la compañía, la duración de la batería de un automóvil (aproximadamente 42 meses) y el número de personas empleadas en una empresa.

Las variables cuantitativas pueden ser discretas o continuas. Las variables discretas solo adoptan ciertos valores y existen “brechas” entre ellos. Algunas muestras de variables discretas son: la cantidad de dormitorios en una casa (1, 2, 3, 4, etc.); el número de automóviles que en una hora usan la salida 25 de la carretera I-4 en Florida, cerca del Walt Disney World (326, 421, etc.) y el número de estudiantes en cada sección de un curso de estadística (25 en la sección A, 42 en la sección B y 18 en la sección C). Aquí se cuenta, por ejemplo, el número de automóviles que utilizan la salida 25 de la carretera I-4, o el número de estudiantes de estadística en cada sección. Observe que en una casa hay 3 o 4 dormitorios, pero no 3.56. Por consiguiente, existe una "brecha" entre los valores posibles. En general, las variables discretas son resultado del conteo

Las observaciones de una variable continua toman cualquier valor dentro de un rango específico; por ejemplo, la presión del aire en una llanta y el peso de un cargamento de tomates. Otros ejemplos son las onzas de cereal con pasas que contiene una caja y la duración de los vuelos de Orlando a San Diego. El promedio de puntos al graduarse (PPG) constituye una variable continua. El PPG de determinado estudiante se podría expresar como 3.2576952. Se acostumbra redondear a tres decimales (3.258). Por lo general, las variables continuas son el resultado de mediciones.

Niveles de medición

OA1-5
Distinguir entre los niveles de medición de datos nominal, ordinal, de intervalo y de razón.

Los datos pueden clasificarse por niveles de medición, los cuales determinan cómo se resumirán y presentarán los datos. También establecen cuáles pruebas estadísticas pueden realizarse. A continuación hay dos ejemplos de la relación entre medición y la forma de aplicar la estadística. En una bolsa de M&M hay lunetas de seis diferentes colores. Suponga que asigna los siguientes valores: 1 al café, 2 al amarillo, 3 al azul, 4 al naranja, 5 al verde y 6 al rojo. ¿Qué tipo de variable es el color de un M&M? Suponga que alguien resume los colores de M&M añadiendo los valores asignados a cada color, divide la suma entre el número de lunetas e informa que el color promedio es 3.56. ¿Cómo se interpreta esta estadística? Tiene razón al concluir que no tiene significado como medición del color de M&M. Como variable cualitativa solo es posible reportar el conteo y el porcentaje de cada color en una bolsa de M&M. Como segundo ejemplo, hay ocho competidores en la pista de una escuela secundaria para la carrera de 400 metros. La media del orden en que llegan a la meta es de 4.5. ¿Qué revela este promedio? ¡Nada! En ambos casos, no se empleó la estadística adecuada para cada nivel de medición.

Existen cuatro niveles de medición: nominal, ordinal, de intervalo y de razón. La medición más baja, o primaria, corresponde al nivel nominal. La más alta es la medición de razón.

Datos de nivel nominal

En el caso del **nivel nominal** de medición, las observaciones acerca de una variable cualitativa se miden y se registran como etiquetas o nombres, las cuales solo pueden clasificarse y contarse.

NIVEL NOMINAL DE MEDICIÓN Los datos registrados en el nivel nominal de medición se representan como etiquetas o nombres. No tienen un orden. Solo pueden clasificarse y contarse.

La clasificación de los seis colores de las lunetas de M&M constituye un ejemplo del nivel nominal de medición; estas se clasificaron simplemente por color. No existe un orden natural; es decir, es posible reportar primero las lunetas cafés, las anaranjadas o las de cualquier color. Registrar la variable de género representa otro ejemplo del nivel nominal de medición. Suponga que hace un conteo de los estudiantes que entran con su credencial a un partido de futbol e informa cuántos hombres y cuántas mujeres asistieron. Podría presentar primero a los hombres o a las mujeres. Para obtener datos a nivel nominal, solo basta contar el número en cada categoría de la variable. A menudo, estos conteos se convierten en porcentajes. Por ejemplo, un estudio de las lunetas M&M arroja los resultados que se muestran en el cuadro de la derecha (www.sensationalcolor.com/color-trends/most-popular-color-177/mam-colors.html).

Color	Porcentaje en una bolsa
Azul	24%
Verde	20
Anaranjado	16
Amarillo	14
Rojo	13
Café	13

Es común codificar numéricamente los nombres o etiquetas para procesar los datos de una variable medida a nivel nominal. Por ejemplo, si le interesa investigar el estado de ori-

ESTADÍSTICA EN ACCIÓN

¿Dónde se originó la estadística? En 1662, John Graunt publicó el artículo "Natural and Political Observations Made upon Bills of Mortality". Las "observaciones" del autor eran el resultado del estudio y análisis de una publicación religiosa semanal llamada *Bills of Mortality*, la cual incluía nacimientos, bautizos y muertes junto con sus causas. Graunt observó que *Bills of Mortality* representaba apenas una fracción de los nacimientos y muertes en Londres. Sin embargo, utilizó los datos para llegar a conclusiones relativas al efecto de las enfermedades, como la peste, en la población. Su lógica constituye un ejemplo de inferencia estadística. Su análisis e interpretación de los datos marcaron el inicio de la estadística.

gen de los estudiantes de la Universidad de Carolina del Este, puede asignar el código 1 a los estudiantes de Alabama, el código 2 a los de Alaska, el 3 a los de Arizona, etcétera. Mediante este procedimiento, Wisconsin recibe el código 49 y Wyoming, el 50. Observe que el número asignado a cada estado sigue siendo un nombre o etiqueta. La razón de asignar códigos numéricos es facilitar el conteo del número de estudiantes de cada estado con un software estadístico. Observe que asignar números a los estados no permite manipularlos como información numérica. En este ejemplo específico, 1 + 2 = 3 correspondería a Alabama + Alaska = Arizona. Claramente, el nivel nominal de medición no permite realizar una operación matemática que tenga una interpretación válida.

Datos de nivel ordinal

El nivel inmediato superior de datos es el **nivel ordinal.** Para este nivel de medición, una variable cualitativa o atributo, se clasifica o califica en una escala relativa.

NIVEL ORDINAL DE MEDICIÓN Los datos registrados en el nivel ordinal de medición se basan en una clasificación o calificación relativa de elementos basados en un atributo definido o variable cualitativa. Las variables que se basan en este nivel de medición solo se clasifican o cuentan.

Mejor ambiente de negocios
1. Texas
2. Utah
3. Virginia
4. Florida
5. Luisiana
6. Indiana
7. Carolina del Sur
8. Tennessee
9. Georgia
10. Nebraska

Por ejemplo, muchas empresas toman decisiones acerca de dónde ubicar sus instalaciones; en otras palabras, ¿cuál es el mejor lugar para su negocio? Business Facilities (**www.businessfacilities.com**) publica una lista de los diez mejores estados con "el mejor ambiente de negocios". A la izquierda se muestra la clasificación de 2012. Se basa en la evaluación de 20 factores diferentes, incluyendo el costo de mano de obra, el clima tributario empresarial, la calidad de vida, la infraestructura de transporte, la fuerza de trabajo capacitada y el potencial de crecimiento económico para clasificar a los estados con base en el atributo "mejor ambiente de negocios".

Este es un ejemplo de una escala ordinal porque los estados se clasifican en el orden de mejor a peor ambiente de negocios. Esto es, se conoce el orden relativo de los estados con base en el atributo. Por ejemplo, en 2012, Texas tenía el mejor ambiente de negocios. Luisiana estaba en quinto lugar, y eso era mejor que Carolina del Sur, pero no tan bueno como Virginia. Observe que no se puede decir que el ambiente de negocios de Texas es cinco veces mejor que el de Luisiana, porque la magnitud de las diferencias entre ambos estados es desconocida.

Otro ejemplo del nivel ordinal de medición se basa en una escala que mide un atributo. Este tipo de escala se utiliza cuando los estudiantes califican a sus maestros en una variedad de características; por ejemplo: "En general, ¿cómo califica la calidad de instrucción en esta clase?". La respuesta del estudiante se registra en una escala relativa: inferior, pobre, buena, excelente y superior. Una característica importante de utilizar una escala relativa de medición es que no es posible distinguir la magnitud de las diferencias entre los grupos. No se sabe si la diferencia entre "superior" y "bueno" es la misma que entre "pobre" e "inferior".

TABLA 1.1 Calificaciones asignadas a un profesor de finanzas

Calificación	Frecuencia
Superior	6
Bueno	28
Promedio	25
Malo	12
Inferior	3

En la tabla 1.1 se presentan las calificaciones que los alumnos del profesor James Bruner le otorgaron después de un curso de introducción a las finanzas. Los datos se resumen en el orden de la escala utilizada para calificar al maestro. Esto es, se resumen según el número de estudiantes que indicaron una calificación superior (6), buena (28), etcétera. Las frecuencias también pueden convertirse a porcentajes. Cerca de 37.8% de los estudiantes calificaron al instructor como bueno.

Datos de nivel de intervalo

El **nivel de intervalo** de medición es el nivel inmediato superior. Incluye todas las características del nivel ordinal; además, la diferencia o intervalo entre valores es significativa.

NIVEL DE INTERVALO DE MEDICIÓN El intervalo o distancia entre los valores de los datos registrados en el nivel de intervalo de medición es significativo. El nivel de intervalo de medición se basa en una escala con una unidad conocida de medición.

La escala de temperatura Fahrenheit es un ejemplo del nivel de intervalo de medición. Suponga que las mayores temperaturas durante tres días consecutivos de invierno en Boston son de 28, 31 y 20 grados Fahrenheit. Es fácil clasificar estas temperaturas, pero también es posible determinar la diferencia entre ellas. Es decir, la diferencia entre 10 y 15 grados Fahrenheit es cinco; la diferencia entre 50 y 55 grados también es cinco. Es importante destacar que cero es un punto más en la escala. No representa la ausencia de estado. Cero grados Fahrenheit no representa la ausencia de frío o calor. Pero según nuestra propia escala de medición, ¡hace frío! Una limitante importante de una variable medida en el nivel de intervalo es que no puede afirmarse que 20 grados Fahrenheit es una temperatura dos veces más cálida que 10 grados Fahrenheit.

Otro ejemplo de escala de intervalo de medición consiste en las tallas de ropa para dama. A la derecha se muestran datos referentes a diversas medidas de una prenda de una mujer caucásica típica.

Talla	Busto (pulgadas)	Cintura (pulgadas)	Cadera (pulgadas)
8	32	24	35
10	34	26	37
12	36	28	39
14	38	30	41
16	40	32	43
18	42	34	45
20	44	36	47
22	46	38	49
24	48	40	51
26	50	42	53
28	52	44	55

¿Por qué razón la escala "talla" es una medición de intervalo? Observe que conforme la talla cambia dos unidades (de la talla 10 a la 12, o de la talla 24 a la 26), cada medida aumenta dos pulgadas. En otras palabras, los intervalos son los mismos.

No existe un punto cero natural que represente una talla. Una prenda "talla cero" no está hecha de "cero" material. Más bien, se trata de una prenda con 24 pulgadas de busto, 16 pulgadas de cintura y 27 de cadera. Además, las razones no son proporcionales. Si divide una talla 28 entre una talla 14, no obtiene la misma respuesta que si divide una talla 20 entre una 10. Ninguna razón es igual a dos, como sugeriría el número de "talla". En resumen, si las distancias entre los números tienen sentido, aunque las razones no, entonces se trata de una escala de intervalo de medición.

Datos del nivel de razón

Todos los datos cuantitativos se registran en el nivel de razón de la medición. El **nivel de razón** es el "más alto". Posee todas las características del nivel de intervalo, pero, además, el punto cero tiene sentido y la razón entre dos números es significativa.

NIVEL DE RAZÓN DE LA MEDICIÓN Los datos registrados en el nivel de razón de la medición se basan en una escala que tenga una unidad conocida de medición y una interpretación significativa del cero.

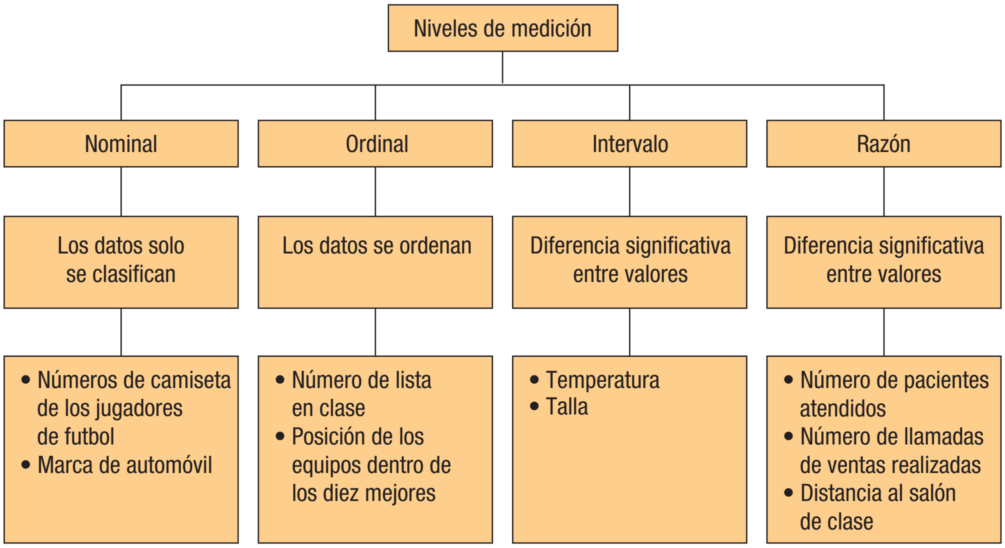
Los salarios, las unidades de producción, el peso, los cambios en los precios de las acciones, la distancia entre sucursales y la altura son algunos ejemplos de la escala de razón de medición. El dinero ilustra bien el caso. Si tiene cero dólares, entonces no tiene dinero, y un salario de 50 dólares por hora es el doble de uno de 25 dólares. El peso también se mide en el nivel de razón de medición. Si el cuadrante de la escala de un dispositivo correctamente calibrado se ubica en cero, entonces hay ausencia total de peso. Más aún, algo que pese un kilo es la mitad de pesado que algo que pese dos kilos.

En la tabla 1.2 se ilustra el uso de la escala de razón de medición para la variable de ingresos anuales de cuatro parejas de padre e hijo. Observe que el señor Lahey gana el doble que su hijo. En la familia Rho, el hijo percibe el doble que el padre.

La gráfica 1.3 resume las principales características de los diversos niveles de medición. El nivel de medición determinará el tipo de métodos estadísticos que pueden utilizarse para analizar una variable. Los métodos estadísticos para analizar variables medidas a nivel nominal se exponen en el capítulo 15; el capítulo 16 se ocupa de los métodos para las variables a nivel ordinal, y los que analizan variables a nivel intervalo o de razón se presentan en los capítulos 9 a 14.

TABLA 1.2 Combinaciones de ingresos entre padre e hijo

Nombre	Padre	Hijo
Lahey	$80 000	$ 40 000
Nale	90 000	30 000
Rho	60 000	120 000
Steele	75 000	130 000



GRÁFICA 1.3 Resumen y ejemplos de las características de los niveles de medición

AUTOEVALUACIÓN 1-2

¿Cuál es el nivel de medición que reflejan los siguientes datos?

(a) La edad de cada persona en una muestra de 50 adultos que escuchan una de las 1 230 estaciones de radio que transmiten entrevistas en Estados Unidos es:

35	29	41	34	44	46	42	42	37	47
30	36	41	39	44	39	43	43	44	40
47	37	41	27	33	33	39	38	43	22
44	39	35	35	41	42	37	42	38	43
35	37	38	43	40	48	42	31	51	34

(b) En una encuesta de 200 propietarios de automóviles de lujo, 100 eran de California, 50 de Nueva York, 30 de Illinois y 20 de Ohio.

EJERCICIOS

Las respuestas a los ejercicios impares se encuentran en el apéndice D.

1. ¿Cuál es el nivel de medición de cada una de las siguientes variables?
   a. Coeficientes intelectuales de los estudiantes.
   b. La distancia que viajan los estudiantes para llegar a clases.
   c. Los números en las camisetas de un equipo universitario femenino de futbol.
   d. Una clasificación de estudiantes por lugar de nacimiento.
   e. Una clasificación de estudiantes que cursan primero, segundo, tercero o último grado.
   f. Cantidad de horas que los alumnos estudian a la semana.
2. El *San Francisco Chronicle* es un gran periódico que se publica diariamente. ¿Cuál es el nivel de medición para cada una de las siguientes variables?
   a. El número de periódicos vendidos todos los domingos durante 2014.
   b. Los diferentes departamentos, como edición, publicidad, deportes, etcétera.
   c. Un resumen del número de periódicos vendidos por condado.
   d. Cantidad de años que cada empleado ha laborado en el periódico.
3. Localice en la última edición de *USA Today* o en el periódico de su localidad ejemplos de cada nivel de medición. Redacte un breve resumen de lo que descubra.
4. En los siguientes casos determine si el grupo representa una muestra o una población.
   a. Los participantes en el estudio de un nuevo fármaco para el colesterol.
   b. Los conductores que recibieron una multa por exceso de velocidad en la ciudad de Kansas durante el último mes.
   c. Beneficiarios del programa de asistencia social en Cook County (Chicago), Illinois.
   d. Las 30 acciones que forman parte del promedio industrial Dow Jones.

Ética y estadística

OA1-6
Enlistar los valores asociados con la práctica de la estadística.

Después de eventos tales como el esquema Ponzi, del administrador de dinero de Wall Street, Bernie Madoff, que estafó miles de millones de dólares a los inversionistas, y las distorsiones financieras de Enron y Tyco, los estudiantes de administración necesitan comprender que estos acontecimientos se debieron a la interpretación equivocada de los datos administrativos y financieros. En cada caso, el personal comunicó a los inversionistas información financiera que indicaba que las compañías se estaban desempeñando mucho mejor de lo que en realidad lo hacían. Cuando se presentó la información verdadera, las compañías tenían un valor muy inferior al que se anunciaba. El resultado fue que muchos inversionistas perdieron todo o casi todo el dinero que invirtieron en estas compañías.

El artículo “Statistics and Ethics: Some Advice for Young Statisticians”, que apareció en *The American Statistician* 57, núm. 1 (2003) proporciona orientación al respecto. Los autores aconsejan la práctica de la estadística con integridad y honestidad, e instan a “hacer lo correcto” cuando se recoja, organice, resuma, analice e interprete información numérica. La contribución real de la estadística a la sociedad es de naturaleza moral. Los analistas financieros necesitan proporcionar información que refleje el verdadero desempeño de una compañía, de tal manera que no desorienten a los inversionistas. La información relativa a defectos de un producto que puede ser dañino debe analizarse y darse a conocer con integridad y honestidad. Los autores del artículo de *The American Statistician* también indicaron que cuando se practique la estadística es necesario mantener “un punto de vista independiente y con principios” al analizar y reportar hallazgos y resultados.

Conforme avance en este libro notará cómo se resaltan cuestiones éticas relacionadas con la recopilación, análisis, presentación e interpretación de información estadística. Es de esperarse, asimismo, que conforme aprenda más estadística se convierta en un consumidor más informado. Por ejemplo, pondrá en tela de juicio un informe basado en datos que no representan fielmente a la población, otro que no contenga estadísticas relevantes, uno que incluya una elección incorrecta de medidas estadísticas o una presentación de datos tendenciosa en un intento deliberado por desorientar o tergiversar los hechos.

Aplicaciones de software

Utilizar la estadística para procesar y analizar los datos requiere un software. A lo largo del texto se muestra la aplicación de Microsoft Excel y, ocasionalmente, Minitab. Las computadoras en las universidades y escuelas de formación profesional suelen tener Microsoft Excel. Su computadora quizá ya lo tenga; en caso contrario, el paquete de Microsoft Office con Excel a menudo se vende a un menor precio en algunas instituciones. En este texto se emplea Excel para la mayoría de las aplicaciones. También se utiliza un complemento de Excel llamado MegaStat. Si su maestro solicita este paquete, está disponible en **www.mhhe.com/megastat**. Este complemento otorga a Excel la capa-

APPLEWOOD AUTO GROUP.xlsx

	A	B	C	D	E	F	G	H
1	Age	Profit	Location	Vehicle-Type	Previous		*Profit*	
2	33	$1,889	Olean	SUV	1			
3	47	$1,461	Kane	Sedan	0		Mean	1843.17
4	44	$1,532	Tionesta	SUV	3		Standard Error	47.97
5	53	$1,220	Olean	Sedan	0		Median	1882.50
6	51	$1,674	Sheffield	Sedan	1		Mode	1915.00
7	41	$2,389	Kane	Truck	1		Standard Deviation	643.63
8	58	$2,058	Kane	SUV	1		Sample Variance	414256.61
9	35	$1,919	Tionesta	SUV	1		Kurtosis	-0.22
10	45	$1,266	Olean	Sedan	0		Skewness	-0.24
11	54	$2,991	Tionesta	Sedan	0		Range	2998
12	56	$2,695	Kane	Sedan	2		Minimum	294
13	41	$2,165	Tionesta	SUV	0		Maximum	3292
14	38	$1,766	Sheffield	SUV	0		Sum	331770
15	48	$1,952	Tionesta	Compact	1		Count	180

cidad de producir reportes estadísticos adicionales. Ocasionalmente se usará Minitab para ilustrar una aplicación. Visite **www.minitab.com** para mayor información. Minitab también ofrece descuentos a los estudiantes. Si utiliza una computadora Mac con Excel, deberá descargar la versión de prueba gratuita de Stat Plus en **www.analystsoft.com**. Es un paquete de software estadístico que se integra con Excel en dichas computadoras.

El ejemplo de la página anterior muestra la aplicación de Excel para realizar un análisis estadístico. En los capítulos 2, 3 y 4 se muestran los métodos de software para resumir y describir datos. Un ejemplo que se utiliza en dichos capítulos se refiere a la base de datos de Applewood Auto Group. Esta y otras bases de datos y archivos están disponibles en el sitio del libro para el estudiante, **www.mhhe.com/uni/lind_ae16e**. El grupo de datos de Applewood tiene diversas variables que fueron medidas para 180 vehículos vendidos por el Applewood Auto Group. A continuación se muestra un cálculo de Excel de diversas estadísticas para la variable ganancia. La siguiente presentación de Excel revela, entre otros aspectos, que se vendieron 180 vehículos; la ganancia media (promedio) por vehículo fue de 1 843.17 dólares; y las ganancias iban desde un mínimo de 294 dólares hasta un máximo de 3 292 dólares. En todo el texto, las ilustraciones de Excel se apoyan en instrucciones para que usted aprenda cómo aplicar Excel para hacer análisis estadísticos. Las instrucciones se presentan en el apéndice C de este texto.

Según el criterio de su instructor y dependiendo del sistema de software disponible, es aconsejable utilizar un paquete de computadora para resolver los ejercicios en los "Ejercicios de la base de datos". Ello le evitará cálculos tediosos y le permitirá concentrarse en el análisis de datos.

RESUMEN DEL CAPÍTULO

I. La estadística es la ciencia que recoge, organiza, presenta, analiza e interpreta datos con el fin de facilitar la toma de decisiones más eficaces.

II. Existen dos clases de estadística.
- **A.** La estadística descriptiva consiste en un conjunto de procedimientos para organizar y resumir datos.
- **B.** La estadística inferencial implica tomar una muestra de una población y llevar a cabo cálculos relativos a esta con base en los resultados de la muestra.
  - **1.** Una población es un conjunto de individuos u objetos de interés o las medidas que se obtienen de todos estos individuos u objetos.
  - **2.** Una muestra es una parte de la población.

III. Existen dos tipos de variables.
- **A.** Una variable cualitativa es de naturaleza no numérica.
  - **1.** Por lo común, lo que interesa es el número o porcentaje de observaciones en cada categoría.
  - **2.** Los datos cualitativos se reúnen en gráficas y diagramas de barras.
- **B.** Existen dos tipos de variables cuantitativas y, en general, se presentan de forma numérica.
  - **1.** Las variables discretas toman ciertos valores, y existen vacíos entre estos.
  - **2.** Una variable continua adopta cualquier valor dentro de un intervalo específico.

IV. Existen cuatro niveles de medición.
- **A.** En el caso del nivel nominal, los datos se distribuyen en categorías sin un orden particular.
- **B.** El nivel ordinal de medición supone que una clasificación se encuentra en un nivel superior a otra.
- **C.** El nivel de medición de intervalo posee la característica de clasificación correspondiente al nivel ordinal de medición; además, la distancia entre valores es constante.
- **D.** El nivel de medición de razón cuenta con todas las características del nivel de intervalo; además, existe un punto cero y la razón entre dos valores resulta significativa.

EJERCICIOS DEL CAPÍTULO

5. Explique la diferencia entre variables cualitativas y cuantitativas. Proporcione un ejemplo de variable cuantitativa y otro de variable cualitativa.

6. Explique la diferencia entre muestra y población.

7. Explique la diferencia entre variable discreta y continua. Proporcione un ejemplo (que no aparezca en el texto) de cada una.

8. En los siguientes problemas indique si reuniría información utilizando una muestra o una población y por qué lo haría.

a. Estadística 201 es un curso que se imparte en una universidad. El profesor Rauch ha enseñado a casi 1 500 estudiantes los pasados cinco años. Usted quiere conocer el grado promedio de los estudiantes que toman el curso.
b. Como parte de un proyecto de investigación, usted necesita dar a conocer la rentabilidad de la compañía líder en *Fortune 500* durante los pasados diez años.
c. Usted espera graduarse y conseguir su primer empleo como vendedor en una de las cinco principales compañías farmacéuticas. Al hacer planes para sus entrevistas, necesitará conocer la misión de la empresa, rentabilidad, productos y mercados.
d. Usted quiere comprar un nuevo reproductor de música MP3, como el iPod de Apple. El fabricante anuncia la cantidad de pistas que almacena la memoria. Considere que los anunciantes toman en cuenta piezas de música popular cortas para calcular la cantidad de pistas que pueden almacenarse. Sin embargo, usted prefiere las melodías de Broadway, que son más largas, y desea calcular cuántas melodías de estas podrá guardar en su reproductor MP3.

9. Antes, las salidas en las carreteras interestatales se numeraban sucesivamente a partir del borde oeste o sur de un estado. Sin embargo, recientemente el Departamento de Transporte de Estados Unidos cambió muchos de estos números para que concordaran con los señalados en los marcadores de millas a lo largo de la carretera.
a. ¿De qué nivel de medición eran los datos sobre los números consecutivos de las salidas?
b. ¿De qué nivel de medición son los datos sobre los números asentados en los marcadores?
c. Exponga las ventajas del nuevo sistema.

10. Un sondeo solicita que un gran número de estudiantes universitarios den información acerca de las siguientes variables: el nombre de su proveedor de servicios de telefonía celular (AT&T, Verizon u otro), los números de minutos que utilizaron durante el último mes (200, 400, por ejemplo) y su nivel de satisfacción con el servicio (terrible, adecuado, excelente, etc.). ¿Cuál es la escala de datos para cada una de estas tres variables?

11. Recientemente, las tiendas Barnes & Noble comenzaron a vender el lector de libros electrónicos llamado Nook Color, un dispositivo mediante el cual se pueden descargar electrónicamente más de dos millones de libros electrónicos, periódicos y revistas; además, despliega los materiales descargados a todo color. Asuma que usted conoce el número de Nook Color vendidos cada día durante el último mes en la tienda de Barnes & Noble del centro comercial Market Commons en Riverside, California. Describa una condición en la que esta información podría considerarse como una muestra. Ejemplifique una segunda situación en la que los mismos datos podrían considerarse como una población.

12. Utilice los conceptos de muestra y población para describir por qué una elección presidencial no es igual a una encuesta "de salida" del electorado.

13. Ubique las variables en las siguientes tablas de clasificación. Resuma en cada tabla sus observaciones y evalúe si los resultados son verdaderos. Por ejemplo, el salario se presenta como una variable cuantitativa continua, pero también es una variable de escala de razón.
a. Salario
b. Género
c. Volumen de ventas de reproductores MP3
d. Preferencia por los refrescos
e. Temperatura
f. Resultados del Salvation Attitude Test (SAT)*
g. Lugar que ocupa un estudiante en clase
h. Calificaciones de un profesor de finanzas
i. Cantidad de computadoras domésticas

	Variables discretas	**Variables continuas**
Variables cualitativas		
Variables cuantitativas		a. Salario

	Discreta	**Continua**
Nominal		
Ordinal		
Intervalo		
Razón		a. Salario

* **Nota de editor:** El SAT es un examen propuesto por E.D. Hirsch, quien argumentaba que de nada servían las técnicas pedagógicas en boga si los estudiantes no contaban con un bagaje de conocimientos que fundamentaran su aprendizaje.

14. A partir de los datos de publicaciones como *Statistical Abstract of the United States*, *The World Almanac*, *Forbes* o del periódico local, proporcione ejemplos de los niveles de medición nominal, ordinal, de intervalo y de razón.

15. Struthers Wells Corporation tiene a más de 10 000 empleados administrativos en sus oficinas de ventas y fabricación en Estados Unidos, Europa y Asia. Una muestra de 300 de esos empleados reveló que 120 aceptarían ser transferidos fuera de Estados Unidos. Con base en estos hallazgos, redacte un breve memorando dirigido a la señora Wanda Carter, vicepresidenta de Recursos Humanos, relacionado con los empleados administrativos de la firma y su disposición para que se les reubique.

16. AVX Stereo Equipment, Inc., recién comenzó a aplicar una política de devolución de artículos "sin complicaciones". Una muestra de 500 clientes que recién habían devuelto artículos mostró que 400 pensaban que la política era justa, 32 opinaban que requería mucho tiempo llevar a cabo la transacción y el resto no opinó. De acuerdo con dicha información, haga una inferencia sobre la reacción del consumidor ante la nueva política.

17. El sitio web de *The Wall Street Journal* (**www.wsj.com**) reportó el número de autos y camionetas vendidos por los ocho principales fabricantes de automóviles en los primeros dos meses de 2013. Compare los datos de 2013 con los correspondientes a los primeros dos meses de 2009, según se reportan a continuación.

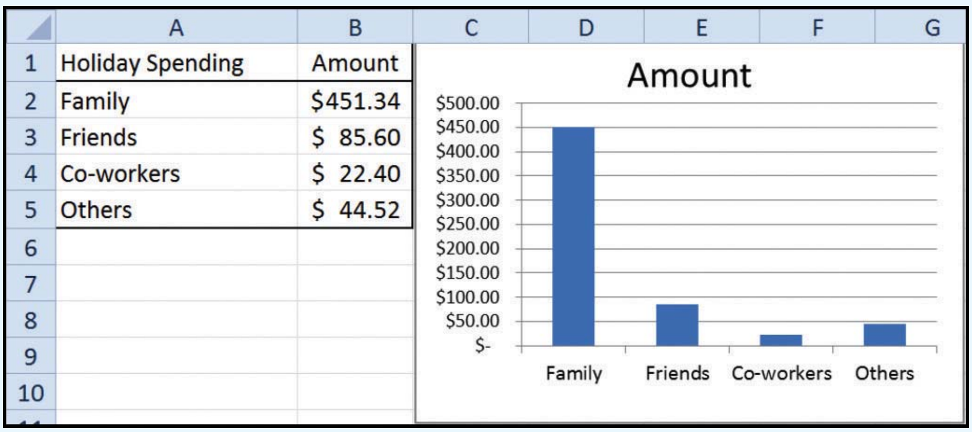
	Ventas en lo que va del año	
Fabricante	**Febrero 2013**	**Febrero 2009**
General Motors Corp.	419 013	252 701
Ford Motor Company	361 713	185 825
Toyota Motor Sales USA Inc.	324 102	226 870
Chrysler LLC	256 746	146 207
American Honda Motor Co. Inc.	201 613	142 606
Nissan North America Inc.	180 555	108 133
Hyundai Motor America	96 024	55 133
Mazda Motor of America Inc.	46 255	31 821

a. Compare el total de ventas de los ocho fabricantes. ¿Ha habido un decremento o un aumento en las ventas de 2013 con respecto al mismo periodo de 2009?

b. El número total de autos y camionetas vendido en 2013 fue 2 234 000, mientras que en 2009 se vendieron 1 346 000. Calcule el porcentaje de mercado que posee cada compañía. ¿Hubo cambios en la participación de mercado de alguna de las compañías?

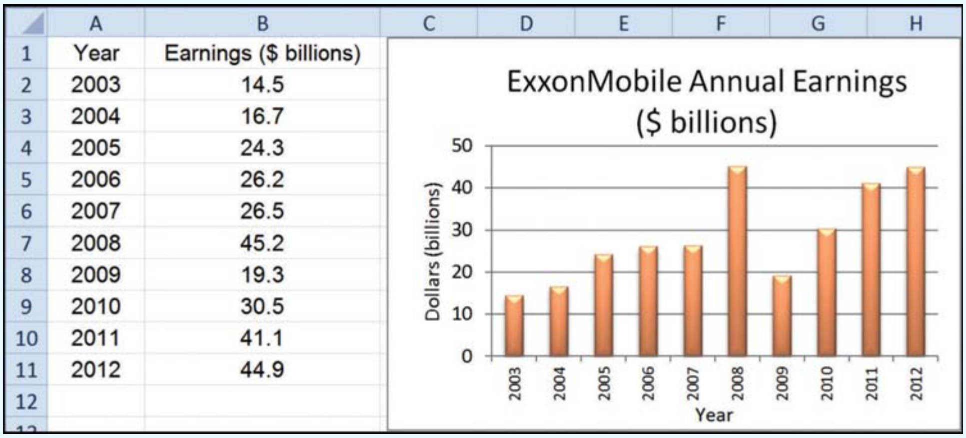
c. Compare el incremento del porcentaje de cada una de las ocho compañías. ¿Qué compañías registraron cambios significativos en sus ventas?

18. La siguiente gráfica describe las cantidades promedio gastadas por los consumidores en regalos de días festivos.



Redacte un breve informe que resuma las cantidades gastadas durante los días festivos. Asegúrese de incluir el total de gastos, así como el porcentaje que corresponde a cada grupo.

19. La siguiente gráfica representa las utilidades en millones de dólares de ExxonMobil en el periodo entre 2003 y 2012. ¿Fueron más altas en un año que en los otros? ¿Las ganancias aumentaron, se redujeron o permanecieron sin cambios durante el periodo?



Year	Earnings ($ billions)
2003	14.5
2004	16.7
2005	24.3
2006	26.2
2007	26.5
2008	45.2
2009	19.3
2010	30.5
2011	41.1
2012	44.9

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro **http://www.mhhe.com/uni/lind_ae16e**)

20. Remítase a los datos sobre el sector inmobiliario que aparecen en el texto e incluyen información acerca de casas vendidas en la zona de Goodyear, Arizona, el año pasado. Considere las siguientes variables: precio de venta, número de recámaras, ubicación y distancia al centro de la ciudad.
  a. De las variables, ¿cuáles son cualitativas y cuáles cuantitativas?
  b. Determine el nivel de medición de cada variable.

21. Consulte los datos sobre Baseball 2012 que contienen información de los treinta equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Considere las siguientes variables: número de victorias, salario del equipo, asistencia durante la temporada, si el equipo jugó los partidos como anfitrión sobre césped, pasto sintético o superficie artificial, así como el número de carreras anotadas.
  a. ¿Cuáles de estas variables son cuantitativas y cuáles cualitativas?
  b. Determine el nivel de medición de cada variable.

22. Remítase a los datos de Buena School District que reportan información sobre la flota de autobuses en el distrito escolar.
  a. ¿Cuáles de las variables son cuantitativas y cuáles cualitativas?
  b. Determine el nivel de medición de cada variable.

2 Descripción de datos:

TABLAS DE FRECUENCIAS, DISTRIBUCIONES DE FRECUENCIAS Y SU REPRESENTACIÓN GRÁFICA

MERRILL LYNCH recién concluyó el estudio de una cartera de inversiones en línea para una muestra de clientes. Elabore una distribución de frecuencias con los datos de los 70 participantes en el estudio (vea el ejercicio 43 y el objetivo de aprendizaje OA2-3).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA2-1 Resumir variables cualitativas con tablas de frecuencias y de frecuencias relativas.

OA2-2 Desplegar una tabla de frecuencias utilizando una gráfica de barras o de pastel.

OA2-3 Resumir variables cuantitativas con distribuciones de frecuencias y de frecuencias relativas.

OA2-4 Desplegar una frecuencia de distribución utilizando un histograma o un polígono de frecuencia.

Introducción

El altamente competitivo negocio de la venta de automóviles al menudeo en Estados Unidos ha sufrido cambios significativos durante los últimos años, los cuales desataron eventos como:

- Las quiebras de General Motors y Chrysler en 2009.
- La eliminación de marcas bien conocidas, como Pontiac y Saturn.
- El cierre de más de 1 500 distribuidoras locales.
- El colapso de la disponibilidad de créditos al consumidor.
- La consolidación de grupos de concesionarias.

Por tradición, una familia local poseía y manejaba la concesionaria de la comunidad, que podía incluir a uno o dos fabricantes, como Pontiac y GMC Trucks o Chrysler y la popular línea Jeep. Sin embargo, algunas compañías hábilmente administradas y bien financiadas han adquirido recientemente las concesionarias locales en extensas regiones de ese país. Al adquirirlas, estos grupos traen consigo sus prácticas de venta, plataformas tecnológicas comunes de software y hardware y técnicas de presentación de informes administrativos. Su objetivo consiste en proporcionar al consumidor una mejor experiencia de compra e incrementar la rentabilidad. Con frecuencia, estas megaconcesionarias emplean alrededor de diez mil personas, que generan varios miles de millones de dólares en ventas anuales, poseen más de cien franquicias y se cotizan en la bolsa de valores de Nueva York o en NASDAQ. Hoy día, la mayor concesionaria es AutoNation (su símbolo bursátil es AN). Otras son Penske Auto Group (PAG, la segunda más grande). Asbury Automotive Group (ABG), y Hendrick Auto Group (empresa privada).

El Applewood Auto Group comprende cuatro concesionarias. El grupo vende una amplia gama de vehículos, entre ellas las marcas económicas de importación Kia y Hyundai, la línea de alta calidad de sedanes BMW y Mercedes Benz y una línea completa de automóviles y camiones Ford y Chevrolet.

La señora Kathryn Ball es miembro del equipo de alta gerencia de Applewood Auto Group, cuyas oficinas corporativas son adyacentes a Kane Motors. Es la responsable de rastrear y analizar los precios de venta y la rentabilidad de los vehículos. A ella le gustaría resumir las ganancias obtenidas de la venta de los vehículos en tablas y gráficas que pudiese revisar cada mes. A partir de estas tablas y gráficas desea conocer la ganancia por vehículo vendido, así como las ganancias mayores y menores. Además, está interesada en describir el perfil demográfico de los compradores. ¿Qué edades tienen? ¿Cuántos vehículos han adquirido previamente de una de las distribuidoras de Applewood? ¿Qué tipo de vehículo compraron?

El Applewood Auto Group opera cuatro distribuidoras:

- ***Tionesta Ford Lincoln Mercury*** vende automóviles y camiones Ford, Lincoln y Mercury.
- ***Olean Automotive, Inc.,*** tiene la franquicia de Nissan y las marcas Chevrolet, Cadillac y camiones GMC.
- ***Sheffield Motors, Inc.,*** vende Buick, camiones GMC, Hyundai y Kia.
- ***Kane Motors*** ofrece Chrysler, Dodge y la línea Jeep, así como BMW y Volvo.

Cada mes, la señora Ball recaba datos de cada una de las cuatro concesionarias y los ingresa en una hoja de cálculo de Excel. El último mes, Applewood Auto Group vendió 180 vehículos entre sus cuatro distribuidoras. Una copia de sus primeras observaciones aparece a la derecha. Las variables que recopiló son:

- ***Edad*:** la edad del comprador en el momento de la compra.
- ***Ganancia*:** la cantidad que la distribuidora obtuvo por la venta de cada vehículo.
- ***Locación*:** la distribuidora donde se adquirió el vehículo.
- ***Tipo de vehículo*:** SUV, sedán, compacto, híbrido o camión.
- ***Previo*:** número de vehículos previamente comprados por el consumidor en cualquiera de las cuatro distribuidoras Applewood.

El conjunto completo de datos se encuentra disponible en el sitio web de McGraw-Hill (**www.mhhe.com/uni/lind_ae16e**) y en el apéndice A.4, que se ubica al final del libro.

APPLEWOOD AUTO GROUP

	A	B	C	D	E
1	Age	Profit	Location	Vehicle-Type	Previous
2	21	$1,387	Tionesta	Sedan	0
3	23	$1,754	Sheffield	SUV	1
4	24	$1,817	Sheffield	Hybrid	1
5	25	$1,040	Sheffield	Compact	0
6	26	$1,273	Kane	Sedan	1
7	27	$1,529	Sheffield	Sedan	1
8	27	$3,082	Kane	Truck	0
9	28	$1,951	Kane	SUV	1
10	28	$2,692	Tionesta	Compact	0
11	29	$1,206	Sheffield	Sedan	0
12	29	$1,342	Kane	Sedan	2
13	30	$443	Kane	Sedan	3
14	30	$754	Olean	Sedan	2
15	30	$1,621	Sheffield	Truck	1

OA2-1
Resumir variables cualitativas con tablas de frecuencias y de frecuencias relativas.

Construcción de una tabla de frecuencias

Recuerde que, en el capítulo 1, al grupo de técnicas que se utilizan para describir un conjunto de datos se les denominó estadística descriptiva. En otras palabras, la estadística descriptiva se encarga de organizar datos con el fin de mostrar su distribución general y el punto donde tienden a concentrarse, además de señalar valores poco usuales o extremos. El primer procedimiento que se emplea para organizar y resumir un conjunto de datos es una **tabla de frecuencias.**

TABLA DE FRECUENCIAS Agrupación de datos cualitativos en clases mutuamente excluyentes y colectivamente exhaustivas que muestra el número de observaciones en cada clase.

En el capítulo 1 se distinguió entre variables cualitativas y cuantitativas. Para recordar, una variable cualitativa es de naturaleza no numérica; es decir, la información puede clasificarse en distintas categorías. Algunos ejemplos de datos cualitativos son la afiliación política (demócrata, conservador, independiente), el lugar de nacimiento (Alabama, Wyoming, etc.) y el método de pago utilizado al comprar en Barnes and Noble (efectivo, cheque o tarjeta de débito o de crédito). Por otra parte, las variables cuantitativas son de índole numérica. Algunos ejemplos de datos cuantitativos relacionados con estudiantes universitarios son el precio de los libros de texto, edad y horas que pasan estudiando cada semana del semestre.

En los datos de Applewood Auto Group existen cinco variables para cada venta de vehículo: 1) edad del comprador, 2) monto de la ganancia, 3) distribuidora que hizo la venta, 4) tipo de vehículo vendido y 5) número de compras previas del consumidor. La distribuidora y el tipo de vehículo son *variables cualitativas*, mientras que el monto de la ganancia, la edad del comprador y el número de compras previas son *variables cuantitativas*.

Suponga que la señora Ball desea resumir las ventas del mes anterior, por locación. Para resumir estos datos cualitativos, clasifique los vehículos que se vendieron el mes previo de acuerdo con la concesionaria: Tionesta, Olean, Sheffield o Kane. Utilice la concesionaria para elaborar una tabla de frecuencias con cuatro clases mutuamente excluyentes (distintivas), lo cual significa que un vehículo no puede pertenecer a dos de ellas. Cada vehículo únicamente se clasifica en una de las cuatro concesionarias mutuamente excluyentes. Además, la tabla de frecuencias debe ser colectivamente exhaustiva, lo cual quiere decir que cada vehículo está representado allí. Esta tabla de frecuencias se muestra en la tabla 2.1. El número de observaciones, que representa las ventas en cada local, se llama *frecuencia de clase*. En este caso, la frecuencia de clase de los vehículos que se vendieron en la locación Kane es 52.

TABLA 2.1 Tabla de frecuencias de los vehículos que vendió Applewood Auto Group por locación

Locación	Número de autos
Kane	52
Olean	40
Sheffield	45
Tionesta	43
Total	180

Frecuencias relativas de clase

Es posible convertir las frecuencias de clase en frecuencias relativas de clase para mostrar la fracción del número total de observaciones en cada una de ellas. Así, una frecuencia relativa captura la relación entre el conjunto de elementos de una clase y el número total de observaciones. En el ejemplo de la venta de vehículos, tal vez desee conocer el porcentaje de automóviles vendidos en cada uno de las cuatro locaciones. Para convertir una distribución de frecuencias en una distribución de frecuencias relativa, cada una de las frecuencias de clase se divide entre el total de observaciones. Por ejemplo, la fracción de vehículos que se vendieron el mes anterior en Kane es 0.289 (que se obtiene al dividir 52 entre 180). La distribución de frecuencias relativas de cada locación se presenta en la tabla 2.2.

TABLA 2.2 Frecuencias relativas de vehículos vendidos por tipo de vehículo en Applewood Auto Group el mes anterior

Locación	Número de autos	Frecuencia relativa	Calculado por
Kane	52	.289	52/180
Olean	40	.222	40/180
Sheffield	45	.250	45/180
Tionesta	43	.239	43/180
Total	180	1.000	

OA2-2
Desplegar una tabla de frecuencias utilizando una gráfica de barras o de pastel.

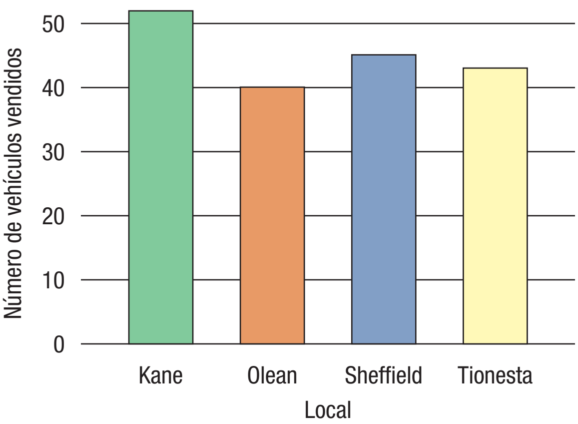
Representación gráfica de datos cualitativos

El instrumento más común para representar una variable cualitativa en forma esquemática es la **gráfica de barras.** En la mayoría de los casos, el eje horizontal muestra la variable de interés y el eje vertical, la frecuencia o fracción de cada uno de los posibles resultados. Una característica distintiva de esta herramienta es que existe una distancia o espacio entre las barras. Es decir, dado que la variable de interés es de naturaleza cualitativa, las barras no son adyacentes. Por consiguiente, una gráfica de barras es una representación gráfica de una tabla de frecuencias mediante una serie de rectángulos de ancho uniforme, cuya altura corresponde a la frecuencia de clase.

GRÁFICA DE BARRAS En ella, las clases cualitativas se representan en el eje horizontal y la frecuencia de clase, en el eje vertical. Las frecuencias de clase son proporcionales a las alturas de las barras.

Se utilizan como ejemplo los datos de Applewood Auto Group (gráfica 2.1). La variable de interés es la locación donde fue vendido el vehículo y la frecuencia de clase, el número de vehículos que se vendieron en cada uno de ellos. Represente los cuatro locales sobre el eje horizontal y el número de vehículos sobre el eje vertical. La altura de las barras, o rectángulos, corresponde a la cantidad de vehículos que se vendieron en cada establecimiento. El mes anterior se vendieron 52 vehículos en Kane, así que la altura de su barra es 52; la altura de la barra de Olean es 40. La variable "local" es de escala nominal, así que no importa el orden de los locales sobre el eje horizontal. También puede ser apropiado enlistar esta variable alfabéticamente, tal como se muestra en la tabla 2.1, o en orden de frecuencias descendentes.

Otro tipo de gráfica útil para describir información cualitativa es la **gráfica de pastel**.



GRÁFICA 2.1 Vehículos vendidos en cada local

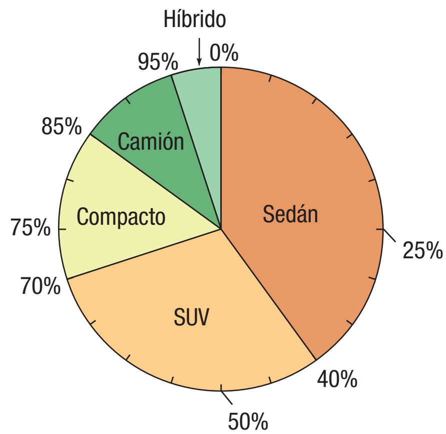
GRÁFICA DE PASTEL Muestra la parte o porcentaje que representa cada clase del total de números de frecuencia.

Los detalles de construcción de una gráfica de pastel se explican empleando la información de la tabla 2.3, en la cual se muestra la frecuencia y porcentaje de autos vendidos en el Applewood Auto Group para cada tipo de vehículo.

El primer paso para elaborar una gráfica de pastel consiste en registrar los porcentajes 0, 5, 10, 15, etcétera, de manera uniforme alrededor de la circunferencia de un círculo (vea la gráfica 2.2). Para indicar la parte de 40% destinada a las ventas totales representadas por los sedanes, trace una línea del centro del círculo a 0, y otra línea del centro del círculo a la marca de 40%. El área de esta "rebanada" representa el número de sedanes vendidos como porcentaje de las ventas totales. Enseguida sume 30% del porcentaje de ventas totales de SUV. El resultado es 70%. Trace una línea del centro del círculo a la marca de 70%; de esta manera el área entre 40% y 70% señala las ventas de SUV como porcentaje de las ventas totales. A continuación, sume 15% de ventas totales de vehículos compactos, lo cual da un total de 85%. Trace una línea del centro del círculo a la marca de 85%; así, la "rebanada" entre 70% y 85% representa el número de vehículos compactos vendidos como porcentaje de las ventas totales. Añada los datos restantes (10% corresponde a las ventas de camiones y 5% a las ventas de híbridos) utilizando el mismo método.

TABLA 2.3 Ventas por tipo de vehículo en Applewood Auto Group

Tipo de vehículo	Unidades vendidas	Porcentaje de ventas
Sedán	72	40
SUV	54	30
Compacto	27	15
Camión	18	10
Híbrido	9	5
Total	180	100

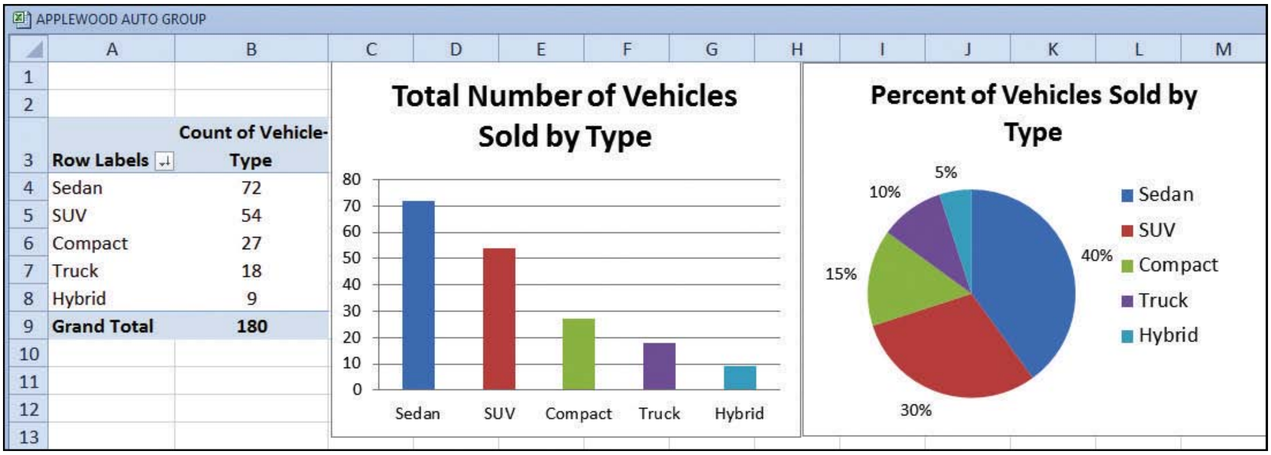


GRÁFICA 2.2 Gráfica de pastel por tipo de vehículos

Dado que cada rebanada de pastel representa la frecuencia relativa de cada tipo de vehículo como porcentaje de las ventas totales, es posible compararlas con facilidad:

- El mayor porcentaje de ventas corresponde a los sedanes.
- Juntos, los sedanes y las SUV representan 70% de las ventas de vehículos.
- Los híbridos representan 5% de las ventas de vehículos, a pesar de haber estado en el mercado solo algunos años.

Es posible utilizar Excel para contar con rapidez el número de autos por tipo de vehículo y crear la tabla de frecuencias, la gráfica de barras y la gráfica de pastel que se muestran a continuación. La herramienta de Excel se llama tabla dinámica. Las instrucciones para producir esta estadística descriptiva y las tablas se incluyen en el apéndice C.



Las gráficas de pastel y las de barras sirven para ilustrar tablas de frecuencias y de frecuencias relativas. ¿Cuándo es preferible usar una gráfica de pastel en vez de una gráfica de barras? En la mayoría de los casos, las gráficas de pastel son más convenientes cuando se trata de mostrar y comparar las diferencias relativas en el porcentaje de observaciones de cada clase o valor de una variable cualitativa. Es preferible usar una gráfica de barras cuando el objetivo es comparar el número o frecuencia de observaciones para cada clase o valor de una variable cualitativa. El siguiente ejemplo ilustra otra aplicación de las gráficas de barras y de pastel.

EJEMPLO

SkiLodges.com realiza una prueba de mercado de su nuevo sitio web y le interesa saber cuánta facilidad de navegación proporciona su diseño. Selecciona al azar 200 usuarios frecuentes de internet y les pide que lleven a cabo una búsqueda en la página web. A cada uno de ellos se le solicita calificar la relativa facilidad para navegar (mala, buena, excelente o sobresaliente). Los resultados aparecen en la siguiente tabla:

Sobresaliente	102
Excelente	58
Buena	30
Mala	10

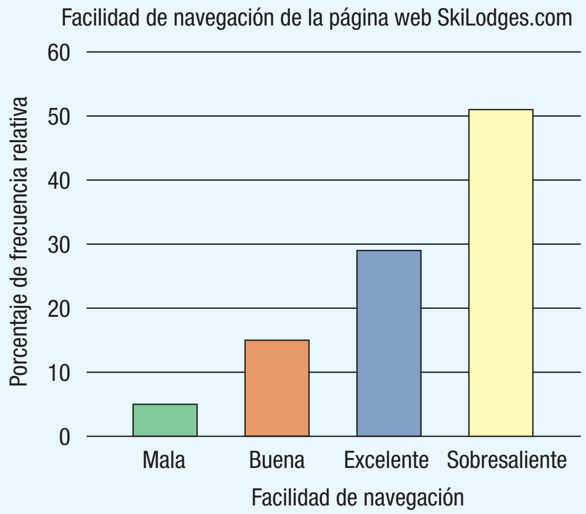
1. ¿Qué tipo de escala de medición se emplea para facilitar la navegación?
2. Elabore una gráfica de barras con los resultados de la encuesta.
3. Dibuje una gráfica de pastel con los resultados de la encuesta.

SOLUCIÓN

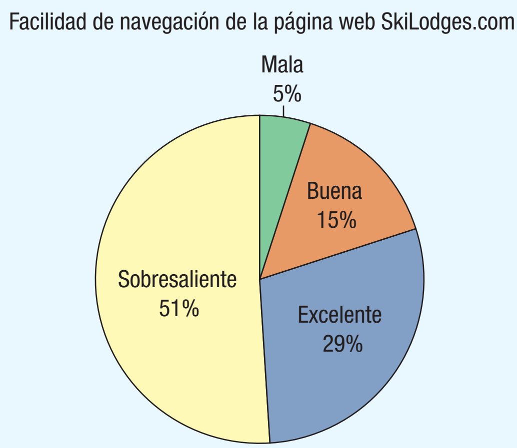
Los datos se miden de acuerdo con una escala ordinal. Es decir, la escala se gradúa en conformidad con la facilidad relativa y abarca de “mala” a “sobresaliente”. Además, se desconoce el intervalo entre

entre cada calificación, así que resulta imposible, por ejemplo, concluir que una buena calificación representa el doble de valor de una mala calificación.

Es posible usar una gráfica de barras para representar los datos. La escala vertical muestra la frecuencia relativa y la horizontal, los valores relativos a la escala de facilidad de navegación.



También es posible emplear una gráfica de pastel para representar estos datos; esta hace hincapié en que más de la mitad de los encuestados consideraron que la relativa facilidad para utilizar el sitio web era sobresaliente.



AUTOEVALUACIÓN 2-1

Las respuestas se encuentran en el apéndice E.

La compañía DeCenzo Specialty Food and Beverage sirve una bebida de cola con un sabor adicional, Cola-Plus, muy popular entre sus clientes. La compañía está interesada en la preferencia de los consumidores por Cola-Plus en comparación con Coca-Cola, Pepsi y una bebida de lima-limón. Se pidió a 100 consumidores seleccionados de forma aleatoria que degustaran una prueba y eligieran la bebida que más les gustaba. Los resultados aparecen en la siguiente tabla:

Bebida	Número
Cola-Plus	40
Coca-Cola	25
Pepsi	20
Lima-limón	15
Total	100

(a) ¿Los datos son de naturaleza cuantitativa o cualitativa? ¿Por qué razón?
(b) ¿Qué nombre recibe la tabla? ¿Qué muestra la tabla?
(c) Diseñe una gráfica de barras para describir la información.
(d) Dibuje una gráfica de pastel utilizando las frecuencias relativas.

EJERCICIOS

Las respuestas a los ejercicios impares se encuentran al final del libro, en el apéndice D.

1. Una gráfica de pastel muestra la porción relativa de mercado de los productos de cola. La “rebanada” de Pepsi-Cola tiene un ángulo central de 90 grados. ¿Cuál es su participación de mercado?

2. En un estudio de mercado se pidió a 100 consumidores que seleccionaran el mejor reproductor musical digital entre iPod, iRiver y Magic Star MP3. Con la finalidad de resumir las respuestas de los consumidores en una tabla de frecuencias, ¿cuántas clases debería tener esta?

3. Se preguntó a un total de 1 000 residentes de Minnesota cuál estación del año preferían. Estos fueron los resultados: a 100 les gustaba más el invierno; a 300, la primavera; a 400, el verano y a 200, el otoño. Desarrolle una tabla de frecuencias y una de frecuencias relativas para resumir esta información.

4. Se preguntó a dos mil viajeros frecuentes (de negocios) qué ciudad de la región central de Estados Unidos preferían: Indianápolis, San Luis, Chicago o Milwaukee. De ellos, 100 contestaron que Indianápolis; 450, San Luis; 1 300, Chicago y el resto dijo que Milwakee. Elabore una tabla de frecuencias y una tabla de frecuencias relativas para resumir esta información.

5. Wellstone, Inc., produce y comercializa fundas para teléfonos celulares en cinco diferentes colores: blanco brillante, negro metálico, lima magnético, naranja tangerina y rojo fusión. Para estimar la demanda de cada color, la compañía montó una isla en el Mall of America durante varias horas y preguntó, a personas elegidas de forma aleatoria, cuál era su color de funda favorito. Los resultados fueron los siguientes:

Blanco brillante	130
Negro metálico	104
Lima magnético	325
Naranja tangerina	455
Rojo fusión	286

a. ¿Qué nombre recibe la tabla?
b. Elabore una gráfica de barras para la tabla.
c. Dibuje una gráfica de pastel.
d. Si Wellstone, Inc., tiene planes de producir un millón de fundas para teléfonos celulares, ¿cuántas de cada color debería producir?

6. Un pequeño negocio de consultoría investiga el desempeño de diversas compañías. Las ventas del cuarto trimestre del año anterior (en miles de dólares) de las compañías seleccionadas fueron las siguientes:

Compañía	Ventas del cuarto trimestre (miles de dólares)
Hoden Building Products	$ 1 645.2
J & R Printing Inc.	4 757.0
Long Bay Concrete Construction	8 913.0
Mancell Electric and Plumbing	627.1
Maxwell Heating and Air Conditioning	24 612.0
Mizelle Roofing & Sheet Metals	191.9

La consultora desea incluir una gráfica en su informe para comparar las ventas de las seis compañías. Utilice una gráfica de barras para comparar las ventas del cuarto trimestre de estas empresas y redacte un breve informe que resuma la gráfica de barras.

OA2-3
Resumir variables cuantitativas con distribuciones de frecuencias y de frecuencias relativas.

Construcción de distribuciones de frecuencias: datos cuantitativos

En el capítulo 1 y al principio de este se ha distinguido entre datos cualitativos y cuantitativos. En la sección anterior, utilizando datos de Applewood Auto Group, aparece un resumen de dos variables cualitativas: 1) local de venta y 2) tipo de vehículo vendido. Se crearon tablas de frecuencias y de frecuencias relativas, y los resultados se reflejaron en gráficas de barras y de pastel.

Los datos de Applewood Auto Group también incluyen variables cuantitativas: la edad del comprador, la ganancia que se obtuvo por la venta del vehículo y el número de compras previas. Suponga que la señora Ball desea resumir las ventas del último mes utilizando las ganancias por venta; en este caso, describirá tal información mediante una **distribución de frecuencias.**

DISTRIBUCIÓN DE FRECUENCIAS Agrupación de datos en clases mutuamente excluyentes y colectivamente exhaustivas, que muestra el número de observaciones que hay en cada clase.

¿Cómo se crea una distribución de frecuencias? El siguiente ejemplo muestra los pasos necesarios. Recuerde que el objetivo es elaborar tablas, diagramas y gráficas que revelen rápidamente la concentración, los valores extremos y la distribución de los datos.

EJEMPLO

Retomemos el ejemplo en que la señora Kathryn Ball, de Applewood Auto Group, desea resumir la variable cuantitativa "ganancias" en una distribución de frecuencias y desplegarla en tablas y gráficas. Con esta información, la señora Ball podrá responder fácilmente a las siguientes preguntas: ¿cuál es la ganancia típica de cada venta?, ¿cuál es la ganancia más alta o máxima?, ¿cuál la ganancia más baja o mínima?, ¿alrededor de qué valor tienden a acumularse las ganancias?

SOLUCIÓN

Para empezar, es preciso conocer las ganancias para cada uno de los 180 vehículos vendidos, que se enlistan en la tabla 2.4. A esta información se le llama *datos en bruto* o *datos no agrupados* porque se trata de un simple listado de las ganancias individuales observadas. Es posible buscar en la lista y encontrar la ganancia más baja o mínima (294 dólares) y la más alta o máxima (3 292 dólares), pero eso es todo. Resulta difícil determinar una ganancia típica o visualizar el punto donde las ganancias tienden a acumularse. Los datos en bruto se interpretan con mayor facilidad si se organizan como una distribución de frecuencias.

TABLA 2.4 Precios de vehículos vendidos el mes anterior en Applewood Auto Group

Máximo

$1 387	$2 148	$2 201	$ 963	$ 820	$2 230	$3 043	$2 584	$2 370
1 754	2 207	996	1 298	1 266	2 341	1 059	2 666	2 637
1 817	2 252	2 813	1 410	1 741	3 292	1 674	2 991	1 426
1 040	1 428	323	1 553	1 772	1 108	1 807	934	2 944
1 273	1 889	352	1 648	1 932	1 295	2 056	2 063	2 147
1 529	1 166	482	2 071	2 350	1 344	2 236	2 083	1 973
3 082	1 320	1 144	2 116	2 422	1 906	2 928	2 856	2 502
1 951	2 265	1 485	1 500	2 446	1 952	1 269	2 989	783
2 692	1 323	1 509	1 549	369	2 070	1 717	910	1 538
1 206	1 760	1 638	2 348	978	2 454	1 797	1 536	2 339
1 342	1 919	1 961	2 498	1 238	1 606	1 955	1 957	2 700
443	2 357	2 127	294	1 818	1 680	2 199	2 240	2 222
754	2 866	2 430	1 115	1 824	1 827	2 482	2 695	2 597
1 621	732	1 704	1 124	1 907	1 915	2 701	1 325	2 742
870	1 464	1 876	1 532	1 938	2 084	3 210	2 250	1 837
1 174	1 626	2 010	1 688	1 940	2 639	377	2 279	2 842
1 412	1 762	2 165	1 822	2 197	842	1 220	2 626	2 434
1 809	1 915	2 231	1 897	2 646	1 963	1 401	1 501	1 640
2 415	2 119	2 389	2 445	1 461	2 059	2 175	1 752	1 821
1 546	1 766	335	2 886	1 731	2 338	1 118	2 058	2 487

Mínimo

ESTADÍSTICA EN ACCIÓN

En 1788, James Madison, John Jay y Alexander Hamilton publicaron de manera anónima una serie de ensayos titulados *The Federalist*. Estos documentos intentaban convencer a la gente de Nueva York de la necesidad de ratificar la Constitución. En el transcurso de la historia se conoció a la mayoría de los autores de estos documentos, aunque otros doce permanecieron en el anonimato. Por medio del análisis estadístico y, en particular, del estudio de la frecuencia con la que se utilizan varias palabras, ahora es posible concluir que James Madison es el autor de los doce documentos. En realidad, la evidencia estadística de la autoría de Madison es abrumadora.

Paso 1: **Defina el número de clases.** Un método útil para determinar la cantidad de clases (k) es la regla de *2 a la k.* Esta guía sugiere que se elija el menor número (k) para el número de clases, de manera que 2^k (en palabras, 2 elevado a la k-ésima potencia) sea mayor que el número de observaciones (n). En el ejemplo de Applewood Auto Group se habían vendido 180 vehículos. Así que $n = 180$. Si supone que $k = 7$, lo cual significa que utilizará 7 clases, entonces $2^7 = 128$, algo menos que 180. De ahí que 7 no represente suficientes clases. Si $k = 8$, entonces $2^8 = 256$, que es mayor que 180. Por lo tanto, el número de clases que se recomienda es 8.

Paso 2: **Determine el intervalo o ancho de clase.** El **intervalo** o **ancho de clase** generalmente es el mismo para todas las clases. Todas las clases juntas deben cubrir por lo menos la distancia del valor más bajo hasta el más alto de los datos. La fórmula para expresar esto es:

$$i \geq \frac{\textit{Valor máximo} - \textit{Valor mínimo}}{k}$$

donde i es el intervalo de clase y k, el número de clases.

En el caso de Applewood Auto Group, el valor más bajo es 294 dólares y el más alto, 3 292 dólares. Si necesitamos 8 clases, el intervalo debería ser:

$$i \geq \frac{\textit{Valor máximo} - \textit{Valor mínimo}}{k} = \frac{\$3\,292 - 294}{8} = \$374.75$$

En la práctica, por lo general este tamaño de intervalo se redondea a una cifra conveniente, tal como un múltiplo de 10 o 100. En este caso, el valor de 400 dólares podría emplearse sin inconvenientes.

Paso 3: **Establezca los límites de cada clase.** Este paso es importante para que sea posible incluir cada observación en una sola categoría. Esto significa que debe evitar la superposición de límites de clase confusos. Por ejemplo, clases como “1 300-1 400 dólares” y “1 400-1 500 dólares” no deberían emplearse porque no resulta claro si el valor de 1 400 dólares pertenece a la primera o a la segunda clase. En general, en este libro se emplea el formato “1 300 **hasta** 1 400 dólares”, “1 400 **hasta** 1 500 dólares” y así sucesivamente. Con este formato resulta claro que 1 399 dólares pertenece a la primera clase y 1 400 dólares, a la segunda.

Dado que siempre se redondea el intervalo de clase hacia arriba para obtener un tamaño conveniente de clase, se cubre un rango más amplio que el necesario. Por ejemplo, 8 clases de 400 dólares de amplitud en el caso de Applewood Auto Group dan como resultado un rango de 8($400) = $3 200. El rango real es de 2 998 dólares, calculado mediante la operación ($3 292 − $294). Al comparar este valor con 3 200 dólares, hay un excedente de 202 dólares. Puesto que solo se necesita abarcar la distancia (*máximo* − *mínimo*), resulta natural poner cantidades aproximadamente iguales al excedente en cada una de las dos colas. Por supuesto, también se deberían elegir límites convenientes de clase. Una directriz consiste en convertir el límite inferior de la primera clase en un múltiplo del intervalo de clase. A veces esto no es posible, pero el límite inferior por lo menos debe redondearse. Ahora bien, estas son las clases que podría utilizar para estos datos:

Clases
$ 200 hasta $ 600
600 hasta 1 000
1 000 hasta 1 400
1 400 hasta 1 800
1 800 hasta 2 200
2 200 hasta 2 600
2 600 hasta 3 000
3 000 hasta 3 400

Paso 4: **Anote las ganancias de venta en las clases.** Para comenzar, la ganancia de venta del primer vehículo en la tabla 2.4 es de 1 387 dólares, cifra que anota en la clase de 1 000 hasta 1 400 dólares. La segunda ganancia de la primera columna de la tabla 2.4 es de 2 148 dólares, la cual se anota en la clase de 1 800 hasta 2 200 dólares. El resto de las ga-

nancias se cuadran de forma similar. Cuando todas las ganancias se hayan registrado, la tabla tendrá la siguiente apariencia:

Ganancia	Frecuencia
$ 200 hasta $ 600	卌 |||
600 hasta 1 000	卌 卌 |
1 000 hasta 1 400	卌 卌 卌 卌 |||
1 400 hasta 1 800	卌 卌 卌 卌 卌 卌 卌 |||
1 800 hasta 2 200	卌 卌 卌 卌 卌 卌 卌 卌 ||||
2 200 hasta 2 600	卌 卌 卌 卌 卌 ||
2 600 hasta 3 000	卌 卌 卌 ||||
3 000 hasta 3 400	||||

Paso 5: **Cuente el número de elementos de cada clase.** El número de elementos que hay en cada clase se llama **frecuencia de clase.** En la clase de 200 hasta 600 dólares hay 8 observaciones, y en la clase de 600 hasta 1 000 dólares hay 11 observaciones. Por lo tanto, la frecuencia de clase de la primera clase es de 8, mientras que en la segunda es de 11. Hay un total de 180 observaciones o frecuencias en todo el conjunto de datos, por lo que la suma de todas las frecuencias debe ser igual a 180.

TABLA 2.5 Distribución de frecuencias de ganancias en Applewood Auto Group sobre los vehículos que se vendieron el mes anterior

Ganancia	Frecuencia
$ 200 hasta $ 600	8
600 hasta 1 000	11
1 000 hasta 1 400	23
1 400 hasta 1 800	38
1 800 hasta 2 200	45
2 200 hasta 2 600	32
2 600 hasta 3 000	19
3 000 hasta 3 400	4
Total	180

Ahora que los datos están organizados en una distribución de frecuencias (vea la tabla 2.5), es posible resumir el patrón de las ganancias por ventas de vehículos de Applewood Auto Group. Observe lo siguiente:

1. Las ganancias por vehículo oscilan entre 200 y 3 400 dólares.
2. Las ganancias por vehículo se clasifican utilizando un intervalo de clase de 400 dólares. El intervalo de clase se determina sustrayendo las clases limítrofes superior o inferior. Por ejemplo, el límite inferior de la primera clase es 200 dólares, y el inferior de la segunda clase es 600 dólares. La diferencia es el intervalo de clase de 400 dólares.
3. Las ganancias se concentran entre 1 000 y 3 000 dólares. Las ganancias de 157 vehículos, u 87%, entran en este rango.
4. Es posible determinar, para cada clase, la ganancia típica o **punto medio de la clase**, el cual está a medio camino entre los límites inferior y superior de dos clases consecutivas. Se determina sumando el límite inferior o superior de clases consecutivas y dividiendo entre 2. En referencia a la tabla 2.5, el límite inferior de la primera clase es 200 dólares, y el de la siguiente es 600 dólares. El punto medio de la clase es 400 dólares, que se determina por ($600 + $200)/2. El punto medio representa de mejor manera, o es típico de, las ganancias provenientes de los vehículos en esa clase. Applewood vendió 8 vehículos con una ganancia típica de 400 dólares.
5. La máxima concentración, o frecuencia más alta, se encuentra en la clase que va de 1 800 hasta 2 200 dólares. Hay 45 vehículos en esta clase. El punto medio de esta clase se ubica en 2 000 dólares; de manera que esa cantidad representa la ganancia típica en la clase con la frecuencia más alta.

Cuando la señora Ball analice esta información tendrá un claro panorama de la distribución de las ganancias de ventas del mes anterior.

Es preciso admitir que la disposición de la información acerca de la venta de precios en una distribución de frecuencias resulta en una pérdida de información detallada. Es decir, al organizar los datos en una distribución de frecuencias no es posible ubicar la ganancia exacta de ningún vehículo, como 1 387, 2 148 o 2 201 dólares. Tampoco es posible afirmar que el monto más bajo de ganancia de cualquier vehículo vendido es de 294 dólares, o que la ganancia máxima fue de 3 292 dólares. Sin embargo, el límite inferior de la primera clase y el límite superior de la clase más grande comunican esencialmente el mismo significado. Lo más probable es que la señora Ball llegue a la misma conclusión si conoce que la ganancia más baja es de casi 200 dólares que si sabe que el monto exacto es de 292. Las ventajas de condensar los datos de forma más entendible y organizada compensa por mucho esta desventaja.

TABLA 2.6 Ingreso bruto ajustado de personas que presentan declaraciones del impuesto sobre la renta

Ingreso bruto ajustado	Número de declaraciones (en miles)
Ingreso bruto no ajustado	178.2
$ 1 hasta 5 000	1 204.6
5 000 hasta 10 000	2 595.5
10 000 hasta 15 000	3 142.0
15 000 hasta 20 000	3 191.7
20 000 hasta 25 000	2 501.4
25 000 hasta 30 000	1 901.6
30 000 hasta 40 000	2 502.3
40 000 hasta 50 000	1 426.8
50 000 hasta 75 000	1 476.3
75 000 hasta 100 000	338.8
100 000 hasta 200 000	223.3
200 000 hasta 500 000	55.2
500 000 hasta 1 000 000	12.0
1 000 000 hasta 2 000 000	5.1
2 000 000 hasta 10 000 000	3.4
10 000 000 o más	0.6

Es preferible utilizar intervalos de clase iguales al resumir datos en bruto con distribuciones de frecuencia. Sin embargo, en ciertos casos se necesita que los intervalos de clase no sean iguales para evitar una gran cantidad de clases vacías, o casi vacías, como en el caso de la tabla 2.6. El Internal Revenue Service de Estados Unidos utilizó intervalos de clase de diferente tamaño para informar el ingreso bruto ajustado sobre declaraciones individuales de impuestos. De haber utilizado intervalos del mismo tamaño, de 1 000 dólares, se habrían requerido más de 1 000 clases para representar todos los impuestos. Al utilizar el método para encontrar intervalos de clase iguales, la regla 2^k resulta en 25 clases, y un intervalo de clase de 400 dólares, asumiendo que cero y 10 000 000 dólares son los valores mínimo y máximo para el ingreso bruto ajustado. Al emplear intervalos de clase iguales, las primeras 13 clases de la tabla 2.6 se combinarían en una clase de aproximadamente 99.9% de todas las declaraciones de impuestos y 24 clases para 0.1% de las declaraciones con un ingreso bruto ajustado por encima de 400 000 dólares. El método de los intervalos de clase iguales no proporciona un buen entendimiento de los datos en bruto. En este caso, se necesita un buen juicio en el uso de los intervalos de clase diferentes, como se expone en la tabla 2.6, para mostrar la distribución del número de declaraciones de impuestos presentadas, especialmente para ingresos por debajo de 500 000 dólares.

AUTOEVALUACIÓN 2-2

Las comisiones en dólares que obtuvieron los once miembros del personal de ventas de Master Chemical Company durante el primer trimestre del año anterior son las siguientes:

1 650, 1 475, 1 670, 1 595, 1 760, 1 540, 1 495, 1 590, 1 625 y 1 510

(a) ¿Cómo se denomina a los valores de 1 650 y 1 475 dólares?
(b) Considere las cantidades que van de 1 400 hasta 1 500 dólares como la primera clase; las que oscilan entre 1 500 hasta 1 600 dólares, como la segunda, y así sucesivamente. Organice las comisiones trimestrales como distribución de frecuencias.
(c) ¿Cómo se denominan los números de la columna derecha de la distribución de frecuencias que elaboró?
(d) Describa la distribución de las comisiones trimestrales con base en la distribución de frecuencias. ¿Cuál es la concentración más grande de comisiones ganadas? ¿Cuál es la menor y cuál es la mayor? ¿Cuál es la típica cantidad ganada?

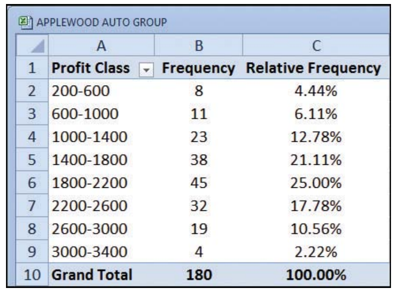
Distribución de frecuencias relativas

Al igual que con los datos cualitativos, quizá resulte conveniente convertir las frecuencias de clase en frecuencias relativas de clase para mostrar la fracción del total de observaciones que hay en cada clase. En el ejemplo de la ganancia por venta de vehículos, podría ser interesante saber qué porcentaje de los precios de vehículos se encuentra en la clase que va de 1 000 hasta 1 400 dólares. En otro estudio, tal vez importe saber qué porcentaje de empleados tomó de 5 hasta 10 días libres el año anterior. Para convertir una distribución de frecuencia en una distribución de frecuencia *relativa,* cada una se divide entre el número total de observaciones. En el caso de la distribución de ganancias por ventas de vehículos (tabla 2.5), la frecuencia relativa de la clase de 1 000 hasta 1 400 dólares es de 0.128, y se determina dividiendo 23 entre 180. Es decir, las ganancias de 12.8% de los vehículos que vendió Applewood Auto Group se encuentran entre 1 000 y 1 400 dólares. Las frecuencias relativas del resto de las clases aparecen en la tabla 2.7.

TABLA 2.7 Distribución de frecuencias relativas de las ganancias por los vehículos vendidos el mes anterior en Applewood Auto Group

Ganancia	Frecuencia	Frecuencia relativa	Determinada por
\$ 200 hasta \$ 600	8	.044	8/180
600 hasta 1 000	11	.061	11/180
1 000 hasta 1 400	23	.128	23/180
1 400 hasta 1 800	38	.211	38/180
1 800 hasta 2 200	45	.250	45/180
2 200 hasta 2 600	32	.178	32/180
2 600 hasta 3 000	19	.106	19/180
3 000 hasta 3 400	4	.022	4/180
Total	180	1.000	

Como se indicó en el capítulo 1, existen diversos paquetes de software que permiten llevar a cabo cálculos estadísticos. A lo largo del libro mostramos los resultados de Microsoft Excel, de MegaStat (un complemento de Excel) y de Minitab. Los pasos necesarios para generar la tabla de esta página aparecen en la sección "Comandos de software" en el apéndice C.


APPLEWOOD AUTO GROUP

	A	B	C
1	Profit Class	Frequency	Relative Frequency
2	200-600	8	4.44%
3	600-1000	11	6.11%
4	1000-1400	23	12.78%
5	1400-1800	38	21.11%
6	1800-2200	45	25.00%
7	2200-2600	32	17.78%
8	2600-3000	19	10.56%
9	3000-3400	4	2.22%
10	Grand Total	180	100.00%

AUTOEVALUACIÓN 2-3

Barry Bonds, jugador de los Gigantes de San Francisco, estableció una nueva marca de cuadrangulares en una sola temporada al conectar 73 durante la temporada 2001. A continuación se enlista la distancia (en pies) de cada cuadrangular.

320	320	347	350	360	360	360	361	365	370
370	375	375	375	375	380	380	380	380	380
380	390	390	391	394	396	400	400	400	400
405	410	410	410	410	410	410	410	410	410
410	410	411	415	415	416	417	417	420	420
420	420	420	420	420	420	429	430	430	430
430	430	435	435	436	440	440	440	440	440
450	480	488							

(a) Para estos datos, muestre las siete clases que podrían usarse para crear la distribución de frecuencias utilizando la regla 2^k.
(b) Demuestre que un intervalo de clase de 30 resumiría los datos en siete clases.
(c) Con siete clases y un intervalo de clase de 30, construya distribuciones de frecuencias y de frecuencias relativas a partir de los datos. Comience la primera clase con un límite inferior a 300.
(d) ¿Cuántos cuadrangulares recorrieron una distancia de 360 hasta 390 pies?
(e) ¿Qué porcentaje de cuadrangulares atravesó una distancia de 360 hasta 390 pies?
(f) ¿Qué porcentaje de cuadrangulares recorrió una distancia de 390 pies o más?

EJERCICIOS

7. Un conjunto de datos consta de 38 observaciones. ¿Cuántas clases recomendaría para la distribución de frecuencias?
8. Un conjunto de datos consta de 45 observaciones entre 0 y 29 dólares. ¿Qué tamaño recomendaría usted para el intervalo de clase?
9. Un conjunto de datos consta de 230 observaciones entre 235 y 567 dólares. ¿Qué intervalo de clase recomendaría?

Este ícono indica que los datos están disponibles en el sitio web del libro, **www.mhhe.com/uni/lind_ae16e**. Usted podrá descargarlos directamente en Excel o en Minitab desde este sitio.

10. Un conjunto de datos contiene 53 observaciones. El valor más bajo es 42 y el más alto, 129. Los datos se van a organizar en una distribución de frecuencias.
a. ¿Cuántas clases sugeriría?
b. ¿Qué cantidad sugeriría como límite inferior de la primera clase?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

11. Wachesaw Manufacturing, Inc. produjo la siguiente cantidad de unidades durante los pasados 16 días.

27	27	27	28	27	25	25	28
26	28	26	28	31	30	26	26

La información se organizará en una distribución de frecuencias.
a. ¿Cuántas clases recomendaría?
b. ¿Qué intervalo de clase sugeriría?
c. ¿Qué límite inferior recomendaría para la primera clase?
d. Organice la información en una distribución de frecuencias y determine la distribución de frecuencias relativas.
e. Comente la forma de la distribución.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

12. La compañía Quick Change Oil cuenta con varios talleres en el área metropolitana de Seattle. Las cantidades diarias de cambios de aceite que se realizaron en el taller de Oak Street durante los pasados veinte días son las siguientes:

65	98	55	62	79	59	51	90	72	56
70	62	66	80	94	79	63	73	71	85

Los datos se organizarán en una distribución de frecuencias.
a. ¿Cuántas clases recomendaría?
b. ¿Qué intervalo de clase sugeriría?
c. ¿Qué límite inferior recomendaría para la primera clase?
d. Organice el número de cambios de aceite como distribución de frecuencias.
e. Comente la forma de la distribución de frecuencias. Determine, asimismo, la distribución de frecuencias relativas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

13. El gerente de Bilo Supermarket, en Mt. Pleasant, Rhode Island, reunió la siguiente información acerca de la cantidad de veces que un cliente visita la tienda durante un mes. Las respuestas de 51 clientes fueron las siguientes:

5	3	3	1	4	4	5	6	4	2	6	6	6	7	1
1	14	1	2	4	4	4	5	6	3	5	3	4	5	6
8	4	7	6	5	9	11	3	12	4	7	6	5	15	1
1	10	8	9	2	12									

a. Comience a partir de 0 como límite inferior de la primera clase, utilice un intervalo de clase de 3 y organice los datos en una distribución de frecuencias.
b. Describa la distribución. ¿Dónde tienden a acumularse los datos?
c. Convierta la distribución en una distribución de frecuencias relativas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14. La división de servicios alimentarios de Cedar River Amusement Park, Inc., estudia la cantidad que gastan al día en alimento y bebida las familias que visitan el parque de diversiones. Una muestra de 40 familias que visitó el parque el día anterior revela que estas gastaron las siguientes cantidades:

$77	$18	$63	$84	$38	$54	$50	$59	$54	$56	$36	$26	$50	$34	$44
41	58	58	53	51	62	43	52	53	63	62	62	65	61	52
60	60	45	66	83	71	63	58	61	71					

a. Organice los datos como distribución de frecuencias utilizando siete clases y el 15 como límite inferior de la primera clase. ¿Qué intervalo de clase eligió?
b. ¿Dónde tienden a acumularse los datos?
c. Describa la distribución.
d. Determine la distribución de frecuencias relativas.

Representación gráfica de una distribución de frecuencias

OA2-4 Desplegar una frecuencia de distribución utilizando un histograma o un polígono de frecuencia.

Es frecuente que los gerentes de ventas, analistas de bolsa, administradores de hospitales y otros ejecutivos ocupados necesiten una vista rápida de las tendencias de las ventas, los precios de las acciones o los costos de hospitalización. A menudo, estas tendencias pueden describirse por medio de tablas y gráficas. Tres herramientas que serán de utilidad para representar gráficamente una distribución de frecuencias son el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas.

Histograma

Un **histograma** de una distribución de frecuencias basada en datos cuantitativos se asemeja mucho a la gráfica de barras, que muestra la distribución de datos cualitativos. Las clases se señalan en el eje horizontal y las frecuencias de clase, en el vertical. Las frecuencias de clase se representan por medio de las alturas de las barras. Ahora bien, a raíz de la naturaleza de los datos, surge una importante diferencia: por lo general, los datos cuantitativos se miden con escalas continuas, no discretas. Por consiguiente, el eje horizontal representa todos los valores posibles, y las barras se colocan de forma adyacente para que muestren la naturaleza continua de los datos.

HISTOGRAMA Gráfica en la que las clases se señalan en el eje horizontal y las frecuencias de clase, en el vertical. Las frecuencias de clase se representan por medio de las alturas de las barras, que se dibujan de manera adyacente.

EJEMPLO

Enseguida aparece la distribución de frecuencias de las ganancias por ventas de vehículos el mes anterior en Applewood Auto Group.

Ganancia	Frecuencia
\$ 200 hasta \$ 600	8
600 hasta 1 000	11
1 000 hasta 1 400	23
1 400 hasta 1 800	38
1 800 hasta 2 200	45
2 200 hasta 2 600	32
2 600 hasta 3 000	19
3 000 hasta 3 400	4
Total	180

Construya un histograma. ¿Qué conclusiones obtiene de la información que se presenta en el histograma?

SOLUCIÓN

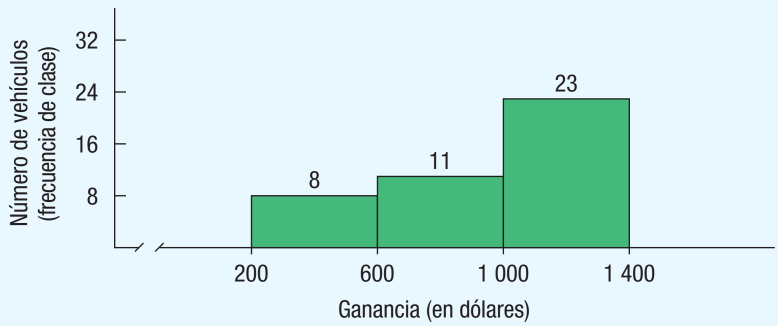
Las frecuencias de clase se colocan en una escala ubicada en el eje vertical (eje *Y*), mientras que a lo largo del eje horizontal se colocan los límites de clase o los puntos medios de clase. Para ilustrar la construcción del histograma, se presentan las primeras tres clases en la gráfica 2.3, página 30.

Observe que, en la gráfica 2.3, la ganancia que produjeron 8 vehículos fue de 200 hasta 600 dólares. Por consiguiente, la altura de la columna de dicha clase es 8. Hay 11 vehículos en los que la ganancia fue de 600 hasta 1 000 dólares. Por consiguiente, es lógico que la altura de dicha columna sea 11. La altura de la barra representa el número de observaciones en la clase.

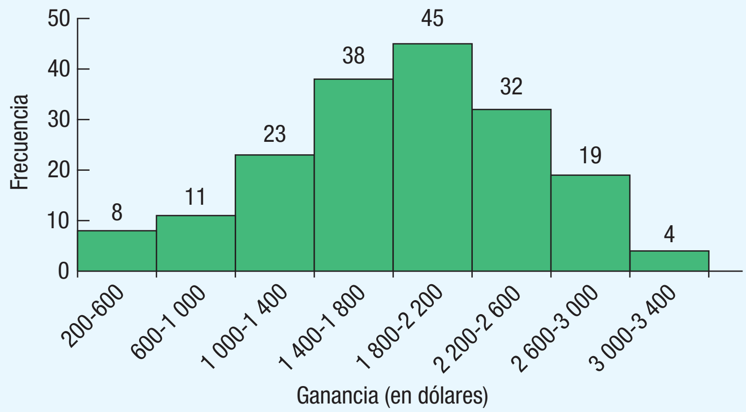
Este procedimiento se aplica en todas las clases. El histograma completo se muestra en la gráfica 2.4, página 30. Advierta que no hay espacio entre las barras, lo cual es una característica del histograma. ¿Por qué? Porque que la variable marcada en el eje horizontal es continua. En una gráfica de barras, la escala de medición es nominal y las barras verticales están separadas. Estas son diferencias importantes entre el histograma y la gráfica de barras.

ESTADÍSTICA EN ACCIÓN

A Florence Nightingale se le conoce como la fundadora de la profesión de enfermería. Sin embargo, también salvó muchas vidas con la ayuda del análisis estadístico. Cuando se encontraba en condiciones poco higiénicas o en un hospital sin suficientes provisiones, mejoraba las condiciones y, enseguida, empleaba los datos estadísticos para documentar las mejoras. De esta manera convenció a otros de la necesidad de una reforma médica, en particular en el área de salubridad. Diseñó gráficas originales para demostrar que, durante la guerra de Crimea, murieron más soldados a causa de las condiciones insalubres que en combate.



GRÁFICA 2.3 Construcción de un histograma



GRÁFICA 2.4 Histograma de ganancias sobre 180 vehículos que vendió Applewood Auto Group

A partir del histograma de la gráfica 2.4, es posible concluir lo siguiente. Son las mismas observaciones basadas en la tabla 2.5.

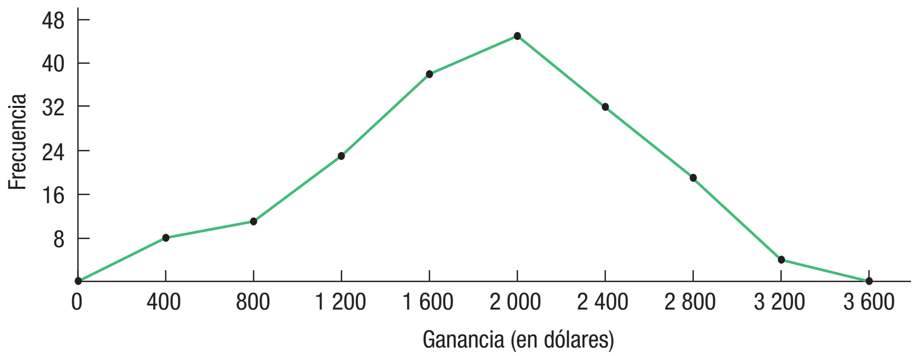
1. La ganancia que se obtuvo por la venta de un vehículo está en un rango de 200 hasta 3 400 dólares.
2. Las ganancias por vehículo se clasifican utilizando un intervalo de clase de 400 dólares. El intervalo de clase se determina sustrayendo los límites de las clases inferior o superior consecutivos. Por ejemplo, el límite inferior de la primera clase es 200 dólares, y el inferior de la segunda clase es 600 dólares. La diferencia es el intervalo de clase, o 400 dólares.
3. Las ganancias se concentran entre 1 000 y 3 000 dólares. Las ganancias sobre 157 vehículos, u 87%, entran en este rango.
4. Para cada clase es posible determinar la ganancia típica o punto medio de la clase. Está a medio camino entre los límites inferior y superior de dos clases consecutivas. Se determina al sumar el límite inferior o superior de clases consecutivas y dividir entre 2. En referencia a la gráfica 2.4, el límite inferior de la primera clase es 200 dólares y el de la siguiente es 600 dólares. El punto medio de la clase es 400 dólares, que se determina por ($600 + $200)/2. El punto medio representa mejor (o es típico de) las ganancias provenientes de los vehículos en esa clase. Applewood vendió 8 vehículos con una ganancia típica de 400 dólares.
5. La máxima concentración, o frecuencia más alta, se encuentra en la clase que va de 1 800 hasta 2 200 dólares. Hay 45 vehículos en esta clase y su punto medio se ubica en 2 000 dólares, cantidad que representa la ganancia típica en la clase con la frecuencia más alta.

Por consiguiente, el histograma proporciona una representación visual de una distribución de frecuencias de fácil interpretación. Cabe señalar que las conclusiones y la forma del histograma hubieran sido las mismas en caso de haber empleado una distribución de frecuencias relativas en lugar de las frecuencias reales. Es decir, el histograma tendría la misma forma que la gráfica 2.4 si se hubieran empleado las frecuencias relativas de la tabla 2.7. La única diferencia consiste en que el eje vertical representaría el porcentaje en lugar de la cantidad de vehículos. Los comandos de Excel para obtener este resultado se incluyen en el apéndice C.

Polígono de frecuencias

Un **polígono de frecuencias** también muestra la forma que tiene una distribución y es similar a un histograma. Consiste en segmentos de recta que conectan los puntos que forman las intersecciones de los puntos medios de clase y las frecuencias de clase. En la gráfica 2.5 se ilustra la construcción de un polígono de frecuencias. Se emplearon las ganancias sobre los vehículos vendidos el mes previo en Applewood Auto Group. El punto medio de cada clase se indica en una escala en el eje *X* y las frecuencias de clase, en el eje *Y.* Recuerde que el punto medio de clase es el valor localizado en el centro de una clase y representa los valores típicos de ella. La frecuencia de clase es el número de observaciones que hay en una clase particular. Las ganancias que se obtuvieron por la venta de vehículos en Applewood Auto Group el mes anterior se repiten a la derecha.

Ganancia	Punto medio	Frecuencia
\$ 200 hasta \$ 600	\$ 400	8
600 hasta 1 000	800	11
1 000 hasta 1 400	1 200	23
1 400 hasta 1 800	1 600	38
1 800 hasta 2 200	2 000	45
2 200 hasta 2 600	2 400	32
2 600 hasta 3 000	2 800	19
3 000 hasta 3 400	3 200	4
Total		180

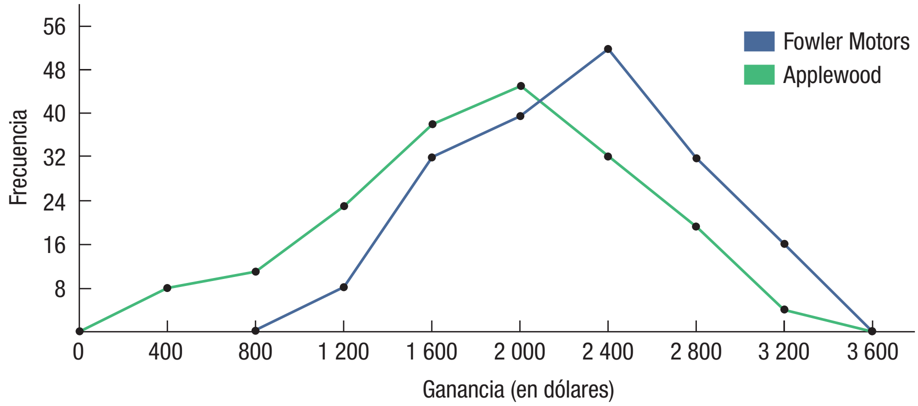


GRÁFICA 2.5 Polígono de frecuencias de las ganancias sobre 180 vehículos que vendió Applewood Auto Group

Como se señaló antes, la clase que va de 200 hasta 600 dólares se representa por el punto medio (400 dólares). Para construir un polígono de frecuencias, hay que desplazarse horizontalmente sobre la gráfica al punto medio y enseguida de manera vertical al 8 (la frecuencia de clase), donde se coloca un punto. Los valores de *x* y de *y* de este punto reciben el nombre de *coordenadas.* Las coordenadas del siguiente punto son $x = \$800$ y $y = 11$. El proceso continúa con todas las clases. Posteriormente, los puntos se conectan de manera ordenada. Es decir, el punto que representa la clase más baja se une al que representa la segunda clase y así en lo sucesivo. Observe que, para completar el polígono de frecuencias en la gráfica 2.5, se añaden los puntos medios de 0 y 3 600 dólares para "anclar" el polígono en la frecuencia cero. Ambos valores (\$0 y \$3 600) se obtuvieron al restar el intervalo de clase al punto medio más bajo (400 dólares) y al sumarlo al punto medio más alto (3 200 dólares) en la distribución de frecuencias.

Tanto el histograma como el polígono de frecuencias permiten tener una vista rápida de las principales características de los datos (máximos, mínimos, puntos de concentración, etc.). Aunque ambas representaciones tienen un propósito similar, el histograma posee la ventaja de describir cada clase como un rectángulo cuya barra de altura representa el número de elementos que hay en cada clase. El polígono de frecuencias, en cambio, tiene una ventaja con respecto al histograma: permite comparar directamente dos o más distribuciones de frecuencias. Suponga que la señora Ball desea comparar las ganancias por vehículo vendido en Applewood Auto Group con las que obtuvo un grupo similar, Fowler Motors, ubicado en Grayling, Michigan. Para hacerlo, debe construir dos polígonos de frecuencias, uno sobre el otro, como lo muestra la gráfica 2.6, página 32. A partir de la gráfica, dos aspectos resultan evidentes:

- Que la ganancia típica que obtiene Fowler es más alta: cercana a 2 000 dólares para Applewood Auto Group y de casi 2 400 dólares para Fowler.
- Existe menos dispersión en las ganancias en Fowler Motors que en Applewood. El límite inferior de la primera clase de Applewood es 0 y el superior, 3 600 dólares. En el caso de Fowler Motors, el límite inferior es 800 dólares y el superior es el mismo: 3 600 dólares.



GRÁFICA 2.6 Distribución de ganancias de vehículos en Applewood Auto Group y en Fowler Motors

El número total de autos vendidos en ambas concesionarias es aproximadamente el mismo, así que es posible llevar a cabo una comparación directa. Si la diferencia entre el número total de autos vendidos es mayor, convertir las frecuencias en frecuencias relativas y representar enseguida las dos distribuciones permitiría obtener una comparación más clara.

AUTOEVALUACIÓN 2-4

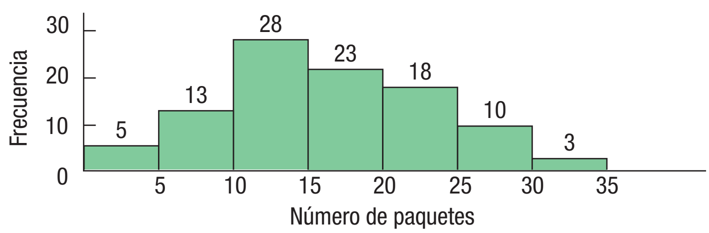
Las importaciones anuales de un grupo de proveedores del sector electrónico aparecen en la siguiente distribución de frecuencias.

Importaciones (millones de dólares)	Número de proveedores
2 hasta 5	6
5 hasta 8	13
8 hasta 11	20
11 hasta 14	10
14 hasta 17	1

(a) Represente las importaciones por medio de un histograma.
(b) Muestre las importaciones por medio de un polígono de frecuencias relativas.
(c) Resuma las facetas importantes de la distribución (como clases, incluyendo las frecuencias más alta y más baja).

EJERCICIOS

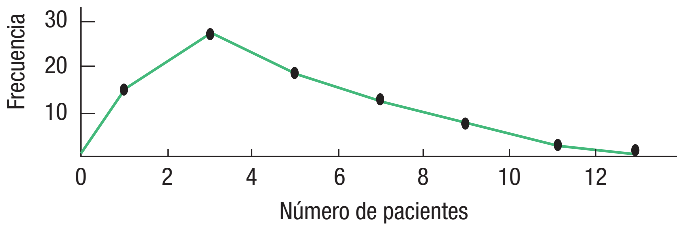
15. Molly's Candle Shop tiene diversas tiendas de venta de menudeo en las áreas costeras de Carolina del Norte y Carolina del Sur. Muchos de los clientes de Molly's han solicitado que se les envíen sus compras. La siguiente gráfica muestra el número de paquetes enviados por día durante los pasados 100 días.



a. ¿Qué nombre recibe la gráfica?
b. ¿Cuál es el número total de frecuencias?
c. ¿Cuál es el intervalo de clase?
d. ¿Cuál es la frecuencia de clase en las clases 10 hasta 15?
e. ¿Cuál es la frecuencia relativa en las clases 10 hasta 15?

f. ¿Cuál es el punto medio de las clases 10 hasta 15?
g. ¿En cuántos días se enviaron 25 o más paquetes?

16. La siguiente gráfica muestra el número de pacientes que admite diariamente el Memorial Hospital en la sala de urgencias.



a. ¿Cuál es el punto medio de la clase que va de 2 hasta 4?
b. ¿Cuántos días se admitieron entre 2 hasta 4 pacientes?
c. ¿Aproximadamente cuántos días fueron estudiados?
d. ¿Cuál es el intervalo de clase?
e. ¿Qué nombre recibe esta gráfica?

17. La siguiente distribución de frecuencias muestra el número de millas de viajero frecuente, expresado en miles de millas, de empleados de Brumley Statistical Consulting, Inc., durante el trimestre más reciente.

Millas de viajero frecuente (en miles)	Número de empleados
0 hasta 3	5
3 hasta 6	12
6 hasta 9	23
9 hasta 12	8
12 hasta 15	2
Total	50

a. ¿Cuántos empleados se estudiaron?
b. ¿Cuál es el punto medio de la primera clase?
c. Construya un histograma.
d. Dibuje un polígono de frecuencias. ¿Cuáles son las coordenadas de la marca correspondientes a la primera clase?
e. Construya un polígono de frecuencias.
f. Interprete las millas de viajero frecuente acumuladas utilizando ambas gráficas.

18. Ecommerce.com, un gran minorista de internet, estudia el tiempo de entrega (el periodo que transcurre desde que se hace un pedido hasta que se entrega) en una muestra de pedidos recientes. Los tiempos de espera se expresan en días.

Tiempo de espera (días)	Frecuencia
0 hasta 5	6
5 hasta 10	7
10 hasta 15	12
15 hasta 20	8
20 hasta 25	7
Total	40

a. ¿Cuántos pedidos se estudiaron?
b. ¿Cuál es el punto medio de la primera clase?
c. ¿Cuáles son las coordenadas de la primera clase en un polígono de frecuencias?
d. Trace un histograma.
e. Dibuje un polígono de frecuencias.
f. Interprete los tiempos de espera mediante ambas gráficas.

Distribuciones de frecuencia acumulativas

Considere nuevamente la distribución de las ganancias sobre vehículos que vendió Applewood Auto Group. Suponga que el interés radica en la cantidad de vehículos que se vendieron con una ganancia menor a 1 400 dólares o la ganancia que se obtuvo en el valor debajo del cual se vendió 40% de los vehículos. Estas cantidades se aproximan mediante una **distribución de frecuencias acumulativas** con representación gráfica de un **polígono de frecuencias acumulativas.**

EJEMPLO

La distribución de frecuencias de las ganancias que obtuvo Applewood Auto Group se toma de la tabla 2.5.

Ganancia	Frecuencia
$ 200 hasta $ 600	8
600 hasta 1 000	11
1 000 hasta 1 400	23
1 400 hasta 1 800	38
1 800 hasta 2 200	45
2 200 hasta 2 600	32
2 600 hasta 3 000	19
3 000 hasta 3 400	4
Total	180

Construya un polígono de frecuencias acumulativas para responder las siguientes dos preguntas: ¿en menos de qué cantidad se sitúa la ganancia que se obtuvo por 75% de los vehículos? y ¿en menos de qué cantidad se sitúa la ganancia que se obtuvo por 60 vehículos?

SOLUCIÓN

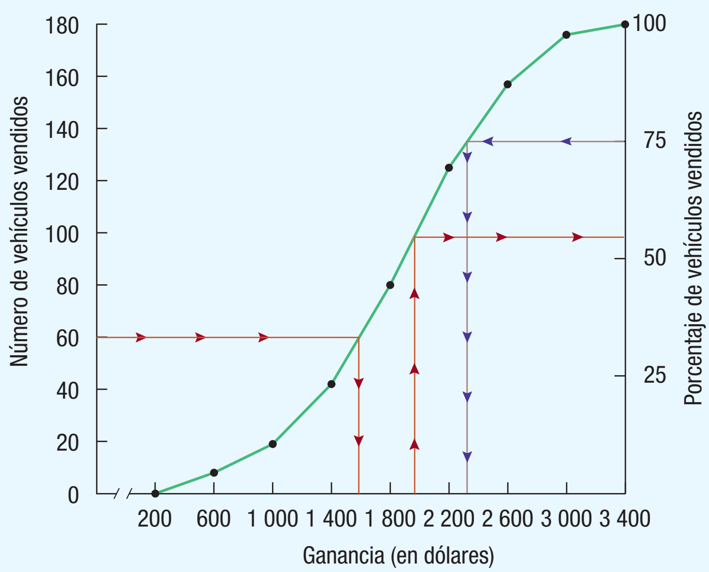
Como su nombre lo indica, una distribución de frecuencias acumulativas y un polígono de frecuencias acumulativas implican *frecuencias acumulativas.* Para construir una distribución de frecuencias acumulativas, consulte la tabla anterior y observe que 8 vehículos se vendieron con una ganancia menor a 600 dólares; dichos vehículos, más 11 de la clase inmediatamente superior, que suman 19, rindieron una ganancia menor a 1 000 dólares. La frecuencia acumulativa de la siguiente clase superior consecutiva es de 42, calculada mediante la operación 8 + 11 + 23. Este proceso se repite en todas las clases. Todos los vehículos produjeron una ganancia menor a 3 400 dólares (vea la tabla 2.8).

TABLA 2.8 Distribución de frecuencias acumulativas de las ganancias obtenidas por vehículos vendidos el mes anterior en Applewood Auto Group

Ganancia	Frecuencia acumulativa	Calculada así
Menos de $ 600	8	8
Menos de 1 000	19	8 + 11
Menos de 1 400	42	8 + 11 + 23
Menos de 1 800	80	8 + 11 + 23 + 38
Menos de 2 200	125	8 + 11 + 23 + 38 + 45
Menos de 2 600	157	8 + 11 + 23 + 38 + 45 + 32
Menos de 3 000	176	8 + 11 + 23 + 38 + 45 + 32 + 19
Menos de 3 400	180	8 + 11 + 23 + 38 + 45 + 32 + 19 + 4

Para trazar una distribución de frecuencias acumulativas, ubique el límite superior de cada clase en una escala a lo largo del eje *X*, y las correspondientes frecuencias acumulativas a lo largo del eje *Y.* Para incluir información adicional, gradúe el eje vertical a la izquierda en unidades y el eje vertical a la derecha, en porcentajes. En el ejemplo de Applewood Auto Group, el eje vertical que se localiza a la izquierda se gradúa desde 0 hasta 180 y, a la derecha, de 0% a 100%. El valor de 50% corresponde a 90 vehículos.

Para comenzar, la primera marca se coloca en $x = 200$ y $y = 0$. Ninguno de los vehículos se vendió con una ganancia menor a 200 dólares. La ganancia de 8 vehículos fue menor de 600 dólares, así que la siguiente marca es $x = 600$ y $y = 8$. A continuación, la próxima marca es $x = 1\,000$ y



GRÁFICA 2.7 Distribución de frecuencias acumulativas por ganancia en vehículos que el mes anterior vendió Applewood Auto Group

$y = 19$. Se registraron 19 vehículos vendidos con una ganancia menor a 1 000 dólares. Se dibuja el resto de los puntos y enseguida se conectan para formar la gráfica 2.7.

Utilice la gráfica 2.7 para determinar el monto de la ganancia que se obtuvo en 75% de los autos vendidos, trace una línea horizontal en la marca de 75%, ubicada en el eje vertical de la derecha, hasta el polígono; enseguida baje al eje *X* y lea el monto de ganancias. El valor sobre el eje *X* es de aproximadamente 2 300 dólares, así que se estima que 75% de los vehículos rindieron una ganancia de 2 300 dólares o menos para Applewood Group.

Para determinar la ganancia que obtuvo en 60 de los 180 vehículos, vehículos, utilice la misma gráfica para localizar el valor de 60 en el eje vertical de la derecha. Luego, trace una línea horizontal a partir de dicho valor hasta el polígono y después baje al eje *X* y lea el monto. Este es de aproximadamente 1 600 dólares, así que se estima que 60 vehículos se vendieron con una ganancia menor a esa cantidad. También es posible hacer aproximaciones del porcentaje de vehículos vendidos en menos de cierta cantidad. Por ejemplo, suponga que desea calcular el porcentaje de vehículos que se vendieron con una ganancia menor a 2 000 dólares. Para comenzar, localice el valor de ese monto en el eje *X,* desplácese por la vertical hasta el polígono y enseguida por la horizontal hasta el eje vertical de la derecha. El valor es cercano a 56%, así que se concluye que 56% de los vehículos se vendieron con una ganancia menor a 2 000 dólares.

AUTOEVALUACIÓN 2-5

En la siguiente tabla se organizó una muestra de salarios por hora de 15 empleados de Home Depot, ubicada en Brunswick, Georgia:

Salario por hora	Número de empleados
\$ 8 hasta \$10	3
10 hasta 12	7
12 hasta 14	4
14 hasta 16	1

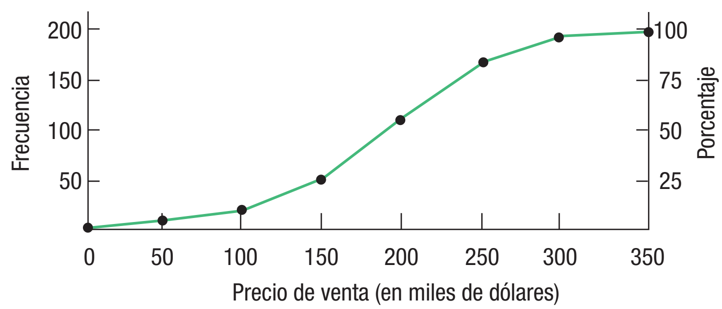
(a) ¿Qué nombre recibe la tabla?
(b) Elabore una distribución de frecuencias acumulativas y represente la distribución en un polígono de frecuencias acumulativas.
(c) De acuerdo con el polígono de frecuencias acumulativas, ¿cuántos empleados ganan 11 dólares o menos por hora?

EJERCICIOS

19. El siguiente polígono de frecuencias muestra los salarios por hora que percibe una muestra de soldadores en el área de Atlanta, Georgia.

a. ¿A cuántos soldadores se estudió?
b. ¿Cuál es el intervalo de clase?
c. ¿Aproximadamente cuántos soldadores ganan menos de 10 dólares por hora?
d. Alrededor de 75% de los soldadores ganan menos de cierta cantidad. ¿Qué cantidad es esta?
e. Diez de los soldadores estudiados ganan menos de cierta cantidad. ¿Qué cantidad es esta?
f. ¿Qué porcentaje de soldadores gana menos de 20 dólares por hora?

20. El siguiente polígono de frecuencias acumulativas muestra los precios de venta (en miles de dólares) de casas que se vendieron en la zona de Billings, Montana.



a. ¿Cuántas casas se estudiaron?
b. ¿Cuál es el intervalo de clase?
c. ¿En menos de qué cantidad se vendieron 100 casas?
d. ¿En menos de qué cantidad se vendió alrededor de 75% de las casas?
e. Calcule el número de casas que se vendieron en la clase que va de 150 000 hasta 200 000 dólares.
f. ¿Cuántas casas se vendieron en menos de 225 000 dólares?

21. Considere la distribución de frecuencias del ejercicio 17, que representa el número de millas de viajero frecuente acumuladas por empleados de Brumley Statistical Consulting Company.

Millas de viajero frecuente (miles)	Frecuencia
0 hasta 3	5
3 hasta 6	12
6 hasta 9	23
9 hasta 12	8
12 hasta 15	2
Total	50

a. ¿Cuántos empleados acumularon menos de 3 000 millas?
b. Convierta la distribución en una distribución de frecuencias acumulativas.
c. Represente la distribución de frecuencias acumulativa en forma de polígono de frecuencias acumulativas.
d. De acuerdo con el polígono de frecuencias, ¿cuántas millas acumuló 75% de los empleados?

22. La distribución de frecuencias de los tiempos de espera en Ecommerce.com, en el ejercicio 18, se repite a continuación.

Tiempo de espera (días)	Frecuencia
0 hasta 5	6
5 hasta 10	7
10 hasta 15	12
15 hasta 20	8
20 hasta 25	7
Total	40

a. ¿Cuántos pedidos se despacharon en menos de 10 días? ¿En menos de 15 días?
b. Convierta la distribución de frecuencias en una distribución de frecuencias acumulativas.
c. Diseñe un polígono de frecuencias acumulativas.
d. ¿En menos de cuántos días se despachó alrededor de 60% de los pedidos?

RESUMEN DEL CAPÍTULO

I. Una tabla de frecuencias es una agrupación de datos cualitativos en clases mutuamente excluyentes y colectivamente exhaustivas que muestra el número de observaciones que hay en cada clase.
II. Una tabla de frecuencias relativas muestra la fracción del número de frecuencias en cada clase.
III. Una gráfica de barras es una representación de una tabla de frecuencias.
IV. Una gráfica de pastel muestra la parte que cada clase representa del número total de observaciones.
V. Una distribución de frecuencias es una agrupación de datos en clases mutuamente excluyentes y colectivamente exhaustivas que muestra el número de observaciones que hay en cada clase.
- **A.** Los pasos para construir una distribución de frecuencias son los siguientes:
  - **1.** Decidir el número de clases.
  - **2.** Determinar el intervalo de clase.
  - **3.** Establecer los límites de cada clase.
  - **4.** Anotar los datos en bruto de las clases.
  - **5.** Enumerar los elementos en cada clase.
- **B.** La frecuencia de clase es el número de observaciones que hay en cada clase.
- **C.** El intervalo de clase es la diferencia entre los límites de dos clases consecutivas.
- **D.** El punto medio de clase se encuentra a la mitad de los límites de clases consecutivas.

VI. Una distribución de frecuencias relativas muestra el porcentaje de observaciones de cada clase.
VII. Existen tres métodos para hacer una representación gráfica de una distribución de frecuencias.
- **A.** Un histograma representa, en forma de rectángulo, el número de frecuencias en cada clase.
- **B.** Un polígono de frecuencias consiste en segmentos de recta que unen los puntos formados por la intersección entre el punto medio de clase y la frecuencia de clase.
- **C.** Una distribución de frecuencias acumulativas muestra el número o porcentaje de observaciones por debajo de valores dados.

EJERCICIOS DEL CAPÍTULO

23. Describa las similitudes y diferencias entre las variables cualitativas y cuantitativas. Asegúrese de considerar lo siguiente:
a. El nivel de medición que se requiere para cada tipo de variable.
b. Si ambos tipos sirven para describir muestras y poblaciones.

24. Describa las similitudes y diferencias entre una tabla de frecuencias y una distribución de frecuencias. Asegúrese de aclarar cuál requiere datos cualitativos y cuál, datos cuantitativos.

25. Alexandra Damonte construirá un nuevo centro vacacional en Myrtle Beach, Carolina del Sur. Debe decidir la manera de diseñar la obra con base en el tipo de actividades que el centro vacacional ofrecerá a sus clientes. Una encuesta reciente de 300 posibles clientes mostró los siguientes resultados relacionados con las preferencias de los consumidores en lo que se refiere a actividades recreativas:

Les gustan las actividades planeadas	63
No les gustan las actividades planeadas	135
No están seguros	78
No respondieron	24

a. ¿Qué nombre recibe la tabla?
b. Diseñe una gráfica de barras para representar los resultados de la encuesta.
c. Trace una gráfica de pastel que muestre los resultados de la encuesta.
d. Si usted se está preparando para presentarle los resultados a la señora Damonte como parte de un informe, ¿qué gráfica preferiría mostrar? ¿Por qué?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

26. Speedy Swift es un servicio de reparto de mercancía que atiende la zona metropolitana más grande de Atlanta, Georgia. Uno de sus objetivos de desempeño para conservar la lealtad del consumidor es la entrega a tiempo. Con el fin de supervisar su desempeño, cada entrega se mide de acuerdo con la siguiente escala: anticipada (mercancía entregada antes del tiempo prescrito); a tiempo (mercancía entregada con una diferencia de cinco minutos en relación con el tiempo prescrito); tarde (mercancía entregada más de cinco minutos después del tiempo prescrito); extraviada (mercancía no entregada). El objetivo de Speedy Swift consiste en entregar 99% de la mercancía en forma anticipada o a tiempo. Otro objetivo es jamás perder un paquete.

Speedy recogió los siguientes datos de desempeño del mes anterior:

A tiempo	A tiempo	Anticipada	Tarde	A tiempo	A tiempo	A tiempo	A tiempo	Tarde	A tiempo
Anticipada	A tiempo	A tiempo	Anticipada	A tiempo	A tiempo	A tiempo	A tiempo	A tiempo	A tiempo
Anticipada	A tiempo	Anticipada	A tiempo	A tiempo	A tiempo	Anticipada	A tiempo	A tiempo	A tiempo
Anticipada	A tiempo	A tiempo	Tarde	Anticipada	Anticipada	A tiempo	A tiempo	A tiempo	Anticipada
A tiempo	Tarde	Tarde	A tiempo	A tiempo	A tiempo	A tiempo	A tiempo	A tiempo	A tiempo
A tiempo	Tarde	Anticipada	A tiempo	Anticipada	A tiempo	Extraviada	A tiempo	A tiempo	A tiempo
Anticipada	Anticipada	A tiempo	A tiempo	Tarde	Anticipada	Extraviada	A tiempo	A tiempo	A tiempo
A tiempo	A tiempo	Anticipada	A tiempo	Anticipada	A tiempo	Anticipada	A tiempo	Tarde	A tiempo
A tiempo	Anticipada	A tiempo	A tiempo	A tiempo	Tarde	A tiempo	Anticipada	A tiempo	A tiempo
A tiempo	A tiempo	A tiempo	A tiempo	A tiempo	Anticipada	Anticipada	A tiempo	A tiempo	A tiempo

a. ¿Qué escala se empleó para medir el desempeño del reparto? ¿Qué clase de variable es dicho desempeño?
b. Construya una tabla de frecuencias que muestre su desempeño durante el mes anterior.
c. Construya una tabla de frecuencias relativas del desempeño de la empresa durante el mes anterior.
d. Dibuje una gráfica de barras de la tabla de frecuencias del desempeño durante el mes anterior.
e. Construya una gráfica de pastel de las entregas a tiempo durante el mes anterior.
f. Analice los resúmenes de datos y redacte una evaluación del desempeño de la empresa durante el mes anterior en relación con sus objetivos. Elabore una recomendación general para realizar un análisis posterior.

27. Un conjunto de datos incluye 83 observaciones. ¿Cuántas clases recomendaría para elaborar una distribución de frecuencias?

28. Un conjunto de datos consta de 145 observaciones que van de 56 a 490. ¿Qué tamaño de intervalo de clase recomendaría?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

29. A continuación se muestra el número de minutos que emplea un grupo de ejecutivos para viajar en automóvil de su casa al trabajo.

28	25	48	37	41	19	32	26	16	23	23	29	36
31	26	21	32	25	31	43	35	42	38	33	28	

a. ¿Cuántas clases recomendaría?
b. ¿Cuántos intervalos de clase sugeriría?
c. ¿Qué intervalo de clase sugeriría como límite inferior de la primera clase?
d. Organice los datos en una distribución de frecuencias.
e. Justifique la forma de la distribución de frecuencias.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

30. Los siguientes datos proporcionan las cantidades semanales que gasta en abarrotes una muestra de hogares.

$271	$363	$159	$ 76	$227	$337	$295	$319	$250
279	205	279	266	199	177	162	232	303
192	181	321	309	246	278	50	41	335
116	100	151	240	474	297	170	188	320
429	294	570	342	279	235	434	123	325

a. ¿Cuántas clases recomendaría?
b. ¿Qué intervalo de clase sugeriría?

c. ¿Cuál sería el límite inferior de la primera clase que usted recomendaría?
d. Organice los datos en una distribución de frecuencias.

31. Un científico social investiga el uso de iPods entre estudiantes universitarios. Una muestra de 45 estudiantes reveló que el día anterior escucharon el siguiente número de canciones.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

4	6	8	7	9	6	3	7	7	6	7	1	4	7	7
4	6	4	10	2	4	6	3	4	6	8	4	3	3	6
8	8	4	6	4	6	5	5	9	6	8	8	6	5	10

Organice esa información en una distribución de frecuencias.
a. ¿Cuántas clases sugiere?
b. ¿Cuál es el intervalo de clase más apropiado?
c. ¿Cuál es el límite inferior de la clase inicial?
d. Elabore la distribución de frecuencias.
e. Describa el perfil de la distribución.

32. David Wise ha manejado su propio portafolio de inversiones durante muchos años. Abajo se enlista el periodo de tenencia (registrado al último año completo) entre la compra y la venta de su colección de acciones.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

8	8	6	11	11	9	8	5	11	4	8	5	14	7	12	8	6	11	9	7
9	15	8	8	12	5	9	8	5	9	10	11	3	9	8	6				

a. ¿Cuántas clases propone?
b. ¿Qué intervalo de clase sugiere?
c. ¿Qué cantidad utilizaría para el límite inferior de la clase inicial?
d. Con base en sus respuestas a los puntos a, b y c, construya una distribución de frecuencias.
e. Identifique la apariencia de la distribución de frecuencias.

33. Está usted escuchando la música en su librería de iTunes. El número total de reproducciones de las 27 canciones que están en su lista de "favoritas" durante el último año se muestra a continuación. Elabore una distribución de frecuencias de las reproducciones y describa su forma. A menudo se dice que una pequeña fracción de las canciones de una persona representa la mayoría de sus reproducciones totales. ¿Este es el caso aquí?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

128	56	54	91	190	23	160	298	445	50
578	494	37	677	18	74	70	868	108	71
466	23	84	38	26	814	17			

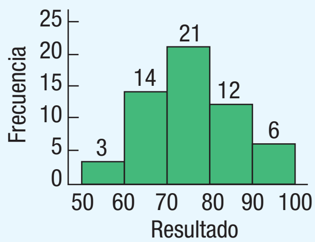
34. A partir de julio de 2005, el *Journal of Finance* puso su contenido a disposición de los lectores en internet. La siguiente tabla muestra el número de veces que se descargó una versión mensual, y el número de artículos que se leyeron cada mes. Suponga que desea hacer una distribución de frecuencias del número de descargas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

312	2 753	2 595	6 057	7 624	6 624	6 362	6 575	7 760	7 085	7 272
5 967	5 256	6 160	6 238	6 709	7 193	5 631	6 490	6 682	7 829	7 091
6 871	6 230	7 253	5 507	5 676	6 974	6 915	4 999	5 689	6 143	7 086

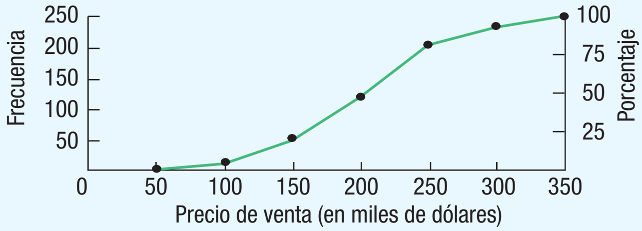
a. ¿Cuántas clases propondría?
b. Sugiera un intervalo de clase.
c. ¿Qué cantidad usaría para el límite inferior de la clase inicial?
d. Con base en sus respuestas a los puntos a, b y c, cree una distribución de frecuencias.
e. Identifique la apariencia de la distribución de frecuencias.

35. El siguiente histograma muestra los resultados del primer examen de una clase de estadística.



a. ¿Cuántos estudiantes presentaron el examen?
b. ¿Cuál es el intervalo de clase?
c. ¿Cuál es el punto medio de la primera clase?
d. ¿Cuántos estudiantes obtuvieron un resultado inferior a 70?

36. La siguiente gráfica resume el precio de venta de casas vendidas el mes anterior en la zona de Sarasota, Florida.



a. ¿Qué nombre recibe la gráfica?
b. ¿Cuántas casas se vendieron el mes anterior?
c. ¿Cuál es el intervalo de clase?
d. ¿En menos de qué cantidad se vendió 75% de las casas?
e. ¿En menos de qué precio se vendieron 175 casas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

37. Una cadena de tiendas deportivas que satisface las necesidades de los esquiadores principiantes, con matriz en Aspen, Colorado, planea llevar a cabo un estudio acerca de la cantidad de dinero que un esquiador novato gasta en su compra inicial de equipo y provisiones. Con base en estas cantidades, desea analizar la posibilidad de ofrecer equipo, como un par de botas y un par de esquíes, para inducir a los clientes a comprar más. Una muestra de 44 comprobantes de la caja registradora reveló las siguientes compras iniciales:

$140	$ 82	$265	$168	$ 90	$114	$172	$230	$142
86	125	235	212	171	149	156	162	118
139	149	132	105	162	126	216	195	127
161	135	172	220	229	129	87	128	126
175	127	149	126	121	118	172	126	

a. Obtenga el intervalo de clase sugerido.
b. Organice los datos en una distribución de frecuencias utilizando un límite inferior de 70 dólares.
c. Interprete sus hallazgos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

38. Las siguientes son las cantidades de acciones emitidas de un grupo selecto de 24 compañías:

Compañía	Cantidad de acciones emitidas (millones)	Compañía	Cantidad de acciones emitidas (millones)
Southwest Airlines	738	Costco	436
FIRSTENERGY	418	Home Depot	1 495
Harley Davidson	226	DTE Energy	172
Entergy	178	Dow Chemical	1 199
Chevron	1 957	Eastman Kodak	272
Pacific Gas and Electric	430	American Electric Power	485
DuPont	932	ITT Corporation	93
Westinghouse Solar	22	Ameren	243
Northeast Utilities	314	Virginia Electric and Power	575
Facebook	1 067	Public Service Electric & Gas	506
Google, Inc.	64	Consumers Energy	265
Apple	941	Starbucks	744

a. Utilizando el número de acciones emitidas, resuma las compañías con una distribución de frecuencias.
b. Despliegue la distribución de frecuencias con un polígono de frecuencias.
c. Cree una distribución de frecuencias acumulativas de las compañías.

d. Despliegue la distribución de frecuencias acumulativas con un polígono de frecuencias acumulativas.

e. Basándose en la distribución de frecuencias acumulativas, ¿75% de las compañías tienen menos de "qué número" de acciones emitidas?

f. Con base en sus resúmenes estadísticos del "número de acciones emitidas", redacte un breve análisis de este grupo de compañías.

39. Una encuesta reciente mostró que el estadounidense típico que posee automóvil invierte 2 950 dólares anuales en gastos operativos. En seguida aparece un desglose detallado de los gastos en artículos. Diseñe una gráfica adecuada que represente los datos y resuma sus hallazgos en un breve informe.

Artículo que genera el gasto	Gasto
Gasolina	$ 603
Intereses de crédito del automóvil	279
Reparaciones	930
Seguro y licencia	646
Depreciación	492
Total	$2 950

40. Midland National Bank seleccionó una muestra de 40 cuentas de cheques de estudiantes. A continuación aparecen sus saldos de fin de mes.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

$404	$ 74	$234	$149	$279	$215	$123	$ 55	$ 43	$321
87	234	68	489	57	185	141	758	72	863
703	125	350	440	37	252	27	521	302	127
968	712	503	489	327	608	358	425	303	203

a. Organice los datos en una distribución de frecuencias utilizando 100 dólares como intervalo de clase y cero como punto de partida.

b. Elabore un polígono de frecuencias acumulativas.

c. El banco considera a cualquier estudiante con un saldo final de 400 dólares o más como "cliente preferente". Calcule el porcentaje de clientes preferentes.

d. El banco hace un cargo por servicio de 10% a los saldos finales más bajos. ¿Qué cantidad recomendaría como punto límite entre los que pagan un cargo por servicio y los que no lo hacen?

41. Los residentes de Carolina del Sur ganaron un total de 69 500 millones de dólares por concepto de ingreso bruto ajustado. Del total, 73% correspondía a sueldos y salarios; 11% a dividendos, intereses y utilidades sobre capital; 8% a fondos para el retiro y pensiones sujetas a impuestos; 3% a pensiones de ingresos por negocio; 2% a seguridad social y 3% a otras fuentes. Genere una gráfica de pastel que describa el desglose del ingreso bruto ajustado. Redacte un párrafo que resuma la información.

42. Un estudio reciente de tecnologías domésticas informó el número de horas de uso semanal de las computadoras personales en una muestra de 60 personas. El estudio excluyó a las personas que laboran fuera del hogar y emplean la computadora como parte de su trabajo.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

9.3	5.3	6.3	8.8	6.5	0.6	5.2	6.6	9.3	4.3
6.3	2.1	2.7	0.4	3.7	3.3	1.1	2.7	6.7	6.5
4.3	9.7	7.7	5.2	1.7	8.5	4.2	5.5	5.1	5.6
5.4	4.8	2.1	10.1	1.3	5.6	2.4	2.4	4.7	1.7
2.0	6.7	1.1	6.7	2.2	2.6	9.8	6.4	4.9	5.2
4.5	9.3	7.9	4.6	4.3	4.5	9.2	8.5	6.0	8.1

a. Organice los datos en una distribución de frecuencias. ¿Cuántas clases sugeriría? ¿Qué valor sugeriría para un intervalo de clase?

b. Elabore un histograma e interprete el resultado que obtenga.

43. Merrill Lynch concluyó un estudio relacionado con el tamaño de las carteras de inversión en línea (acciones, bonos, fondos mutuos y certificados de depósito) en una muestra de clientes del grupo de 40 hasta 50 años de edad. En la página 42 aparece el valor de las inversiones en miles de dólares de los 70 participantes.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

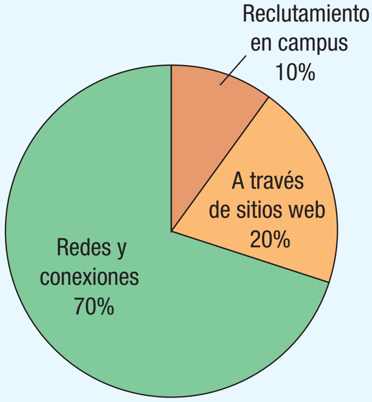
$669.9	$ 7.5	$ 77.2	$ 7.5	$125.7	$516.9	$ 219.9	$645.2
301.9	235.4	716.4	145.3	26.6	187.2	315.5	89.2
136.4	616.9	440.6	408.2	34.4	296.1	185.4	526.3
380.7	3.3	363.2	51.9	52.2	107.5	82.9	63.0
228.6	308.7	126.7	430.3	82.0	227.0	321.1	403.4
39.5	124.3	118.1	23.9	352.8	156.7	276.3	23.5
31.3	301.2	35.7	154.9	174.3	100.6	236.7	171.9
221.1	43.4	212.3	243.3	315.4	5.9	1 002.2	171.7
295.7	437.0	87.8	302.1	268.1	899.5		

a. Organice los datos en una distribución de frecuencias. ¿Cuántas clases sugeriría? ¿Qué valor propondría para un intervalo de clase?
b. Diseñe un histograma e interprete el resultado que obtenga.

44. Un total de 5.9% del público que vio la televisión durante las horas de mayor audiencia se concentró en programas de la ABC; 7.6%, de la CBS; 5.5%, de Fox; 6.0%, de la NBC; 2.0%, de Warner Brothers y 2.2%, de UPN. Un total de 70.8% de la audiencia vio programas de otras cadenas televisivas de cable, como CNN y ESPN. El siguiente sitio web contiene información reciente sobre la audiencia televisiva: **http://tv.zap2it.com/news/ratings**. Diseñe una gráfica de pastel o de barras para describir esta información. Redacte un párrafo que resuma sus hallazgos.

45. Remítase a la siguiente gráfica:

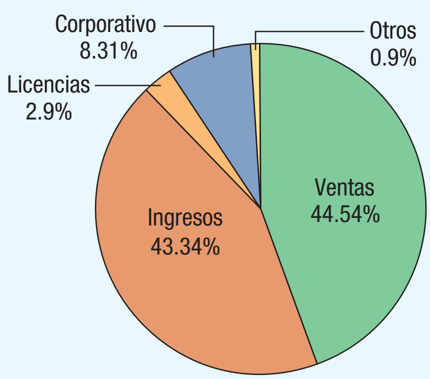
Contacto para obtención de empleo en la Universidad Wake Forest



a. ¿Cuál es el nombre de este tipo de gráfica?
b. Suponga que 1 000 graduados comenzarán en un nuevo empleo poco después de titularse. Estime el número de graduados cuyo primer contacto para empleares ocurrió a través de redes y otras conexiones.
c. ¿Sería razonable concluir que casi 90% de los empleos se realizaron a través de redes, conexiones y sitios web? Proporcione evidencia.

46. La siguiente gráfica representa los ingresos anuales, por tipo de impuesto, del estado de Georgia. La gráfica se desarrolló usando Kids Zone, un proyecto de NCES (Centro Nacional de Estadísticas de la Educación). Su sitio web es: **nces.ed.gov/nceskids/creategraph/**.

Ingresos anuales del estado de Georgia



a. ¿Qué porcentaje de los ingresos estatales representa el impuesto a la venta y el impuesto al ingreso individual?
b. ¿Qué categoría genera más ingresos: los impuestos corporativos o las licencias?
c. El ingreso anual total del estado de Georgia es de 6 300 millones de dólares. Estime el ingreso en miles de millones de dólares que generaron los impuestos a la venta y al ingreso individual.

47. En 2011, Estados Unidos exportó productos a Canadá por un valor de 281 000 millones de dólares. Los cinco productos principales fueron:

Producto	Cantidad
Vehiculos	$46.9
Maquinaria	44.2
Maquinaria eléctrica	27.1
Combustible y aceite mineral	18.4
Plásticos	12.6

a. Utilice un paquete de software para desarrollar una gráfica de barras.
b. ¿Qué porcentaje de las exportaciones totales de Estados Unidos a Canadá representan las categorías "Combustible y aceite mineral" y "Vehículos"?
c. De los cinco principales productos de exportación, ¿qué porcentaje del total representan "Combustible y aceite mineral" y "Vehículos"?

48. La revolución industrial cambió la vida en las granjas de Estados Unidos, haciéndola más eficiente desde principios del siglo xx. Por ejemplo, en 1910 las granjas de Estados Unidos emplearon 24.2 millones de caballos y mulas, y solo alrededor de 1 000 tractores. En 1960 se empleaban 4.6 millones de tractores y solo 3.2 millones de caballos y mulas. En 1920 había más de 6 millones de granjas en Estados Unidos. Hoy hay menos de 2.2 millones. En la lista que sigue aparece el número de granjas, en miles, en cada uno de los 50 estados. Redacte un párrafo en el que resuma sus hallazgos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

49	1	16	49	82	36	5	2	48	47
8	26	76	62	92	66	86	30	8	13
8	55	81	42	108	29	47	3	4	10
21	36	52	32	75	87	39	63	1	27
32	78	248	17	7	47	40	23	78	11

49. Las lunetas de M&M, fabricadas por Mars Company, son uno de los dulces más populares en Estados Unidos. Al principio, todas eran cafés; ahora se fabrican en rojo, verde, azul, naranja, café y amarillo. Si desea leer la historia del producto, obtener ideas para preparar pasteles con lunetas, comprarlas en los diferentes colores de su escuela o equipo favorito y conocer el porcentaje de cada color que contienen las bolsas normales, visite www.m-ms.com. Hace poco, una bolsa de 14 onzas de M&M en su presentación regular contenía 444 dulces distribuidos por colores de la siguiente manera: 130 cafés, 98 amarillos, 96 rojos, 35 anaranjados, 52 azules y 33 verdes. Elabore una gráfica que describa esta información y redacte un párrafo en el que resuma los resultados.

50. Durante un periodo de 30 días se registró el número de familias que usaron el servicio de guardería de la YWCA de Minneápolis. Los resultados son los siguientes:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

31	49	19	62	24	45	23	51	55	60
40	35	54	26	57	37	43	65	18	41
50	56	4	54	39	52	35	51	63	42

a. Construya una distribución de frecuencias acumulativas.
b. Diseñe una gráfica del polígono de frecuencias acumulativas.
c. ¿Cuántos días se registró que menos de 30 familias utilizaron el servicio de guardería?
d. ¿Cuál fue el nivel de ocupación de 80% de los días más concurridos?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**)

51. Consulte los datos de inmobiliarias que contienen información acerca de las casas vendidas en el área de Goodyear, Arizona, el año anterior. Seleccione un intervalo de clase apropiado y organice los precios de venta en una distribución de frecuencias. Escriba un breve reporte que resuma sus resultados. Asegúrese de contestar las siguientes preguntas en dicho reporte.

a. ¿Alrededor de cuáles valores tienden a acumularse los datos?

b. Basándose en la distribución de frecuencias, ¿cuál es el precio de venta típico en la primera clase y en la última clase?

c. Elabore una distribución de frecuencias acumulativas. ¿Cuántas casas se vendieron en menos de 200 000 dólares? Calcule el porcentaje de casas que se vendieron en más de 220 000 dólares. ¿Qué porcentaje de casas se vendió en menos de 125 000 dólares?

d. Remítase a la variable con respecto a los municipios. Elabore una gráfica de barras que muestre el número de casas vendidas en cada municipio. ¿Existen diferencias o el número de casas que se vendieron en cada municipio es similar?

52. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Cree una distribución de frecuencias para la variable "nómina" y responda las siguientes preguntas:

a. ¿Cuál es la nómina típica de un equipo? ¿Cuál es el rango de nóminas?

b. Comente la forma de la distribución. ¿Parece que alguna de las nóminas de los equipos no se encuentra en línea con las demás?

c. Diseñe una distribución de frecuencias acumulativas. ¿Treinta por ciento de los equipos pagan menos que cuál cantidad de la nómina total del equipo? ¿Cuántos equipos aproximadamente tienen nóminas totales inferiores a 100 000 000 dólares?

53. Consulte los datos de los autobuses del Distrito Escolar Buena. Seleccione la variable que se refiere al número de millas que recorrieron el mes anterior y organice estos datos en una distribución de frecuencias.

a. ¿Cuál es la cantidad típica de millas recorridas? ¿Cuál es el rango?

b. Comente la forma de la distribución. ¿Existen valores atípicos en términos de millas conducidas?

c. Diseñe una distribución de frecuencias acumulativas. ¿Cuarenta por ciento de los autobuses fueron conducidos durante menos de cuántas millas? ¿Cuántos autobuses fueron conducidos menos de 850 millas?

d. Consulte las variables con respecto al tipo de autobús y al número de asientos en cada uno. Elabore una gráfica de pastel de cada variable y comente sus hallazgos.

Descripción de datos:

MEDIDAS NUMÉRICAS

3

EL DERBY DE KENTUCKY se celebra el primer sábado de mayo en Churchill Downs, Louisville, Kentucky. La pista mide una milla y cuarto. En la tabla del ejercicio 82 se muestran los ganadores desde 1990, su margen de victoria, el tiempo del ganador y las ganancias sobre una apuesta de dos dólares. Determine la media y la mediana de estas dos últimas variables (vea el ejercicio 82 y el objetivo de aprendizaje OA3-1).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA3-1 Calcular e interpretar la media, la mediana y la moda.

OA3-2 Calcular la media ponderada.

OA3-3 Calcular e interpretar la media geométrica.

OA3-4 Calcular e interpretar el rango, la varianza y la desviación estándar.

OA3-5 Explicar y aplicar el teorema de Chebyshev y la regla empírica.

OA3-6 Calcular la media y la desviación estándar de datos agrupados.

ESTADÍSTICA EN ACCIÓN

¿Alguna vez ha conocido a un estadounidense promedio? Pues bien, se llama Robert (nivel nominal de la medición); tiene 31 años (nivel de razón); mide 1.77 metros (otro nivel de razón de la medición); pesa 78 kilogramos; calza del 9½; su cintura mide 85 cm de diámetro y viste trajes talla 40. Además, cada año, Robert come 1.8 kg de papas fritas; mira 2 567 horas el televisor y se come 11.77 kg de plátanos; también duerme 7.7 horas por noche.

Una estadounidense promedio mide 1.64 metros de estatura y pesa 64 kg, mientras que una modelo estadounidense promedio mide 1.65 metros y pesa 53 kg. Un día cualquiera, casi la mitad de las mujeres en Estados Unidos está a dieta. Marilyn Monroe, idolatrada en la década de 1950, se consideraría con sobrepeso según los estándares actuales; ella usaba vestidos de las tallas 14 a la 18, y era una mujer saludable y atractiva.

Introducción

En el capítulo 2 se inició el estudio de la estadística descriptiva. Para transformar un cúmulo de datos en bruto en algo con significado, los datos cuantitativos se organizaron en una distribución de frecuencias y, después, los resultados se representaron en una gráfica de barras. De manera similar, los datos cuantitativos se organizaron en una distribución de frecuencias y se presentaron gráficamente en un histograma. Se mostraron otras técnicas para graficar, como las gráficas de pastel para representar datos cualitativos, y polígonos de frecuencias para representar datos cuantitativos.

En este capítulo se presentan dos formas numéricas de describir datos cuantitativos: las **medidas de ubicación** o **medidas de localización** y las **medidas de dispersión.** A las medidas de ubicación a menudo se les llama *promedios***.** El propósito de una medida de ubicación consiste en señalar el centro de un conjunto de valores. Los promedios aparecen a diario en televisión, en el periódico y otras publicaciones. He aquí algunos ejemplos:

- La casa promedio en Estados Unidos cambia de dueño cada 11.8 años.
- Un estadounidense recibe un promedio de 568 piezas de correspondencia cada año.
- El hogar estadounidense promedio tiene más televisores que personas. Hay 2.73 televisores y 2.55 personas en el hogar típico.
- La pareja estadounidense promedio gasta 20 398 dólares en su boda, mientras que su presupuesto es 50% menor. Esta cifra no incluye el costo de la luna de miel ni del anillo de compromiso.
- El precio promedio de un boleto de teatro en Estados Unidos es de 7.50 dólares, según la Asociación Nacional de Propietarios de Teatros.

Si únicamente toma en cuenta las medidas de ubicación de un conjunto de datos o si compara varios conjuntos de datos utilizando valores centrales, llegará a una conclusión incorrecta. Además de las medidas de ubicación, debe tomar en cuenta la **dispersión** —denominada con frecuencia *variación* o *propagación*— de los datos. Por ejemplo, suponga que el ingreso anual promedio de los ejecutivos de compañías relacionadas con internet es de 80 000 dólares, igual que el ingreso promedio de ejecutivos de compañías farmacéuticas. Si solo atiende a los ingresos promedio, podría concluir, equivocadamente, que ambas distribuciones de salarios son idénticas. Un vistazo a los rangos salariales indica que esta conclusión no es correcta. Los salarios de los ejecutivos de las empresas de internet oscilan entre 70 000 y 90 000 dólares; en cambio, los salarios de los ejecutivos de marketing de la industria farmacéutica van de 40 000 a 120 000 dólares. Por consiguiente, aunque los salarios promedios son los mismos en ambas industrias, hay más propagación o dispersión en los que perciben los ejecutivos de la industria farmacéutica. Para describir la dispersión, considere el rango, la desviación media, la varianza y la desviación estándar.

OA3-1
Calcular e interpretar la media, la mediana y la moda.

Medidas de ubicación

En principio se explican las medidas de ubicación. No existe una única medida de ubicación; de hecho, existen varias. Se considerarán cinco: 1) la media aritmética, la media ponderada, la mediana, la moda y la media geométrica. La media aritmética es la medida de ubicación que más se utiliza y que se publica con mayor frecuencia, por lo cual se le considerará como parámetro para una población y como estadístico para una muestra.

La media poblacional

Muchos estudios incluyen todos los valores que hay en una población. Por ejemplo, la tienda de menudeo Reynolds Road Carpet tiene 12 empleados. El monto promedio de comisiones que ganaron el mes anterior fue de 1 345 dólares. Este es el valor poblacional, puesto que considera la comisión de *todos* los asociados de ventas. Otros ejemplos de media poblacional serían los siguientes:

- El precio de cierre promedio de las acciones de Johnson & Johnson durante los últimos cinco días fue de 64.75 dólares.
- La semana pasada, los seis soldadores del departamento de soldadura de Butts Welding, Inc. trabajaron, en promedio, 6.45 horas extras.
- Caryn Tisrch inició el mes anterior un sitio web dedicado a la jardinería orgánica. La media aritmética de visitas a su sitio durante los 31 días de julio fue de 84.36.

En el caso de los datos en bruto, que no están agrupados en una distribución de frecuencias, la media poblacional es la suma de todos los valores observados en la población dividida entre el número de valores de la población. Para determinar la media poblacional, aplique la siguiente fórmula:

$$\text{Media poblacional} = \frac{\text{Suma de todos los valores observados en la población}}{\text{Número de valores en la población}}$$

En lugar de escribir las instrucciones completas para calcular la media poblacional (o cualquier otra medida), resulta más conveniente utilizar símbolos matemáticos adecuados. La media de una población con símbolos matemáticos es:

MEDIA POBLACIONAL $$\mu = \frac{\Sigma x}{N} \quad [3.1]$$

en la cual:

μ es la letra minúscula griega *mu*, y representa la media poblacional;
N es el número de valores en la población;
x representa cualquier valor particular;
Σ es la letra mayúscula griega *sigma*, e indica la operación de suma;
Σx es la suma de *C* valores en la población.

Cualquier característica medible de una población recibe el nombre de **parámetro.** Una media de una población es un parámetro.

PARÁMETRO Característica de una población.

EJEMPLO

Hay 42 salidas en la autopista I-75, que atraviesa el estado de Kentucky. A continuación aparece la lista de distancias entre salidas (en millas).

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

¿Por qué esta información representa una población? ¿Cuál es la media aritmética de millas entre salidas?

SOLUCIÓN

Es una población porque se toman en cuenta a todas las salidas en la autopista I-75 de Kentucky. Sume las distancias entre cada una de las 42 salidas. La distancia total es de 192 millas. Para determinar la media aritmética, divida este total entre 42. Así, la media aritmética es 4.57 millas, calculada mediante la operación 192/42. De acuerdo con la fórmula [3.1]:

$$\mu = \frac{\Sigma x}{N} = \frac{11 + 4 + 10 + \cdots + 1}{42} = \frac{192}{42} = 4.57$$

¿Cómo se interpreta el valor 4.57? Es el número típico de millas entre salidas. Como se han tomado en cuenta todas las salidas de la autopista I-75 de Kentucky, este valor es un parámetro poblacional.

Media muestral

Como se explicó en el capítulo 1, con frecuencia se selecciona una muestra de la población para estimar una característica específica de ésta. Por ejemplo, el departamento de control de calidad de Smucker's necesita garantizar que la cantidad de mermelada de fresa en un recipiente cuya etiqueta indica que contiene 12 onzas, realmente contenga dicha cantidad. Sería muy costoso y lento revisar el peso de cada recipiente; por tanto, se selecciona una muestra de 20 recipientes, se determina la media de ella y se utiliza ese valor para estimar la cantidad de mermelada que hay en cada uno.

En el caso de los datos en bruto; es decir, los datos no agrupados, *la media es la suma de los valores de la muestra dividida entre el número total de valores de esta.* La media de una muestra se determina así:

$$\text{Media de la muestra} = \frac{\text{Suma de todos los valores de la muestra}}{\text{Número de valores de la muestra}}$$

La media de una muestra y la media de una población se calculan de la misma forma pero la notación abreviada que se utiliza es diferente. La fórmula de la media de una muestra es:

MEDIA MUESTRAL $$\overline{x} = \frac{\Sigma x}{n} \quad [3.2]$$

donde:

$\overline{x}$ es la media de la muestra; se lee: "*x* barra";
n es el número de valores de la muestra;
X representa cualquier valor particular;
Σ es la letra mayúscula griega *sigma*, e indica la operación de suma;
Σx es la suma de *C* valores de la muestra.

La media de una muestra o cualquier otra medición basada en una muestra de datos recibe el nombre de **estadístico.** Si el peso promedio de una muestra de 10 contenedores de mermelada de naranja Smucker's es de 11.5 onzas, se trata de un ejemplo de estadístico.

ESTADÍSTICO Característica de una muestra.

EJEMPLO

Verizon estudia la cantidad de minutos que consumen sus clientes que cuentan con un plan tarifario para teléfono celular. Una muestra aleatoria de 12 clientes arroja la siguiente cantidad de minutos empleados el mes anterior.

90	77	94	89	119	112
91	110	92	100	113	83

¿Cuál es valor de la media aritmética de los minutos consumidos?

SOLUCIÓN

De acuerdo con la fórmula [3.2], la media muestral es:

$$\text{Media muestral} = \frac{\text{Suma de todos los valores en la muestra}}{\text{Número de valores en la muestra}}$$

$$\overline{x} = \frac{\Sigma x}{n} = \frac{90 + 77 + \cdots + 83}{12} = \frac{1\,170}{12} = 97.5$$

El valor de la media aritmética de los minutos consumidos el mes anterior por los usuarios de teléfonos celulares de la muestra es de 97.5 minutos.

Propiedades de la media aritmética

La media aritmética es una medida de ubicación muy utilizada. Cuenta con algunas propiedades importantes:

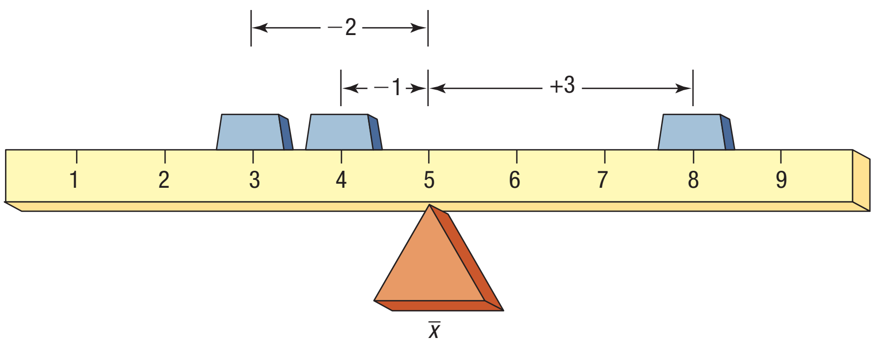
1. **Para calcular una media, los datos deben pertenecer al nivel de intervalo o de razón.** Recuerde, del capítulo 1, que los datos del nivel de razón incluyen información como edades, ingresos y pesos, y que la distancia entre los números es constante.
2. **Todos los valores se incluyen en el cálculo de la media.**
3. **La media es única.** Solo existe una media para un conjunto de datos. Más adelante en este capítulo descubriremos un promedio que podría aparecer dos o más veces en un conjunto de datos.
4. **La suma de las desviaciones de cada valor a la media es cero.** Expresado simbólicamente:

$$\Sigma(x - \bar{x}) = 0$$

Como ejemplo, la media de 3, 8 y 4 es 5. De esta manera:

$$\begin{aligned}\Sigma(x - \bar{x}) &= (3 - 5) + (8 - 5) + (4 - 5)\\ &= -2 + 3 - 1\\ &= 0\end{aligned}$$

De esta manera, la media es un punto de equilibrio de un conjunto de datos. Para ilustrarlo, imagine una regla con los números 1, 2, 3,…, 9 uniformemente espaciados. Suponga que se colocan tres barras del mismo peso sobre la regla en los números 3, 4 y 8 y que el punto de equilibrio se colocara en el 5, la media de los tres números. Descubriría que la regla se equilibra perfectamente. Las desviaciones debajo de la media (−3) son iguales a las desviaciones por encima de la media (+3). El esquema es:



La media tiene un punto débil; recuerde que el valor de cada de una muestra, o población, se utiliza al calcularla. Si uno o dos de estos valores son extremadamente grandes o pequeños en comparación con la mayoría de los datos, la media podría no ser un promedio adecuado para representar los datos. Por ejemplo, suponga que los ingresos anuales de un pequeño grupo de corredores de bolsa en Merrill Lynch son de 62 900, 61 600, 62 500, 60 800 y 1 200 000 dólares. El ingreso medio es de 289 560 dólares; claro, no es representativo del grupo, ya que todos, salvo un corredor, tienen ingresos entre 60 000 y 63 000 dólares. Un ingreso (1 200 millones de dólares) afecta en exceso la media.

AUTOEVALUACIÓN 3-1

1. Los ingresos anuales de una muestra de empleados de administración media en Westinghouse son: 62 900, 69 100, 58 300 y 76 800 dólares.
   (a) Proporcione la fórmula de la media muestral.
   (b) Determine la media muestral.
   (c) ¿Es la media que calculó en el inciso anterior un estadístico o un parámetro? ¿Por qué razón?
   (d) ¿Cuál es su mejor aproximación de la media de la población?
2. Todos los estudiantes de la clase 411 del curso de ciencias avanzadas de la computación constituyen una población. Sus calificaciones en el curso son 92, 96, 61, 86, 79 y 84.
   (a) Proporcione la fórmula de la media poblacional.
   (b) Calcule la calificación media del curso.
   (c) ¿Es la media que calculó en el inciso anterior un estadístico o un parámetro? ¿Por qué razón?

EJERCICIOS

Las respuestas a los ejercicios impares se encuentran en el apéndice D.

1. Calcule la media de la siguiente población de valores: 6, 3, 5, 7, 6.
2. Calcule la media de la siguiente población de valores: 7, 5, 7, 3, 7, 4.
3. **a.** Calcule la media de los siguientes valores muestrales: 5, 9, 4, 10.
   b. Demuestre que $\Sigma(x - \bar{x}) = 0$.
4. **a.** Calcule la media de los siguientes valores muestrales: 1.3, 7.0, 3.6, 4.1, 5.0.
   b. Demuestre que $\Sigma(x - \bar{x}) = 0$.
5. Calcule la media de los siguientes valores muestrales: 16.25, 12.91, 14.58.
6. Suponga que va a la tienda y gasta 61.85 dólares en 14 artículos. ¿Cuál es el precio promedio por artículo?

En los ejercicios 7 a 10, a) calcule la media aritmética y b) señale si se trata de un estadístico o de un parámetro.

7. Midtown Ford emplea a 10 vendedores. El número de automóviles nuevos que cada uno vendió el mes anterior fue: 15, 23, 4, 19, 18, 10, 10, 8, 28, 19.
8. El departamento de contabilidad en una compañía de ventas por catálogo contó las siguientes cantidades de llamadas recibidas por día en el número gratuito de la empresa durante los primeros siete días de mayo de 2006: 14, 24, 19, 31, 36, 26, 17.
9. Cambridge Power and Light Company seleccionó una muestra aleatoria de 20 clientes residenciales. Enseguida aparecen las sumas, redondeadas al dólar más próximo, que se cobraron a los clientes por el servicio de luz el mes anterior:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

54	48	58	50	25	47	75	46	60	70
67	68	39	35	56	66	33	62	65	67

10. El director de relaciones humanas de Ford inició un estudio de las horas de trabajo extra en el departamento de inspección. Una muestra de 15 trabajadores reveló que estos laboraron la siguiente cantidad de horas extras el mes anterior:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

13	13	12	15	7	15	5	12
6	7	12	10	9	13	12	

11. AAA Heating and Air Conditioning concluyó 30 trabajos el mes anterior con un ingreso medio de 5 430 dólares por trabajo. El presidente desea conocer el ingreso total del mes. Con base en la información limitada que se proporciona, ¿puede calcular el ingreso total? ¿A cuánto asciende?
12. Una gran compañía farmacéutica contrata graduados de administración de empresas para vender sus productos. La compañía se expande con rapidez y dedica un día a capacitar a los nuevos vendedores. El objetivo que la compañía fija a cada nuevo vendedor es de 10 000 dólares mensuales, cifra que refleja las ventas promedio actuales por mes de la empresa. Después de revisar las retenciones de impuestos de los nuevos empleados, la compañía encuentra que solo uno de cada 10 permanece más de tres meses en la empresa. Comente la utilización de las ventas promedio actuales mensuales como objetivo de ventas para los nuevos empleados. ¿Por qué abandonan los empleados la compañía?

La mediana

Se ha enfatizado que si los datos contienen uno o dos valores muy grandes o muy pequeños, la media aritmética no resulta representativa. Es posible describir el centro de dichos datos a partir de una medida de ubicación denominada **mediana**.

Para ilustrar la necesidad de una medida de ubicación diferente de la media aritmética, suponga que busca un condominio en Palm Aire. Su agente de bienes raíces le dice que el precio típico de las unidades disponibles en este momento es de 110 000 dólares. ¿Aún insiste en seguir buscando? Si usted tiene un presupuesto máximo de 75 000 dólares, podría pensar que los condominios se encuentran fuera de su presupuesto. Sin embargo, la verificación de los precios de las unidades individuales podría hacerle cambiar de parecer. Los costos son de 60 000, 65 000, 70 000, 80 000 y, en el caso de un lujoso *penthouse*, 275 000 dólares. El importe promedio aritmético es de 110 000 dólares, como le informó el agente de bienes raíces, pero un precio (275 000 dólares) eleva la media aritmética y lo convierte en un promedio no representativo. Parece que un precio de más o menos 70 000 dólares es un promedio más típico o representativo, y así es. En casos como este, la mediana proporciona una medida de ubicación más válida.

MEDIANA Punto medio de los valores una vez que se han ordenado de menor a mayor o de mayor a menor.

La mediana del precio de las unidades disponibles es de 70 000 dólares. Para determinarla, ordene los precios de menor (60 000 dólares) a mayor (275 000 dólares) y seleccione el valor medio (70 000 dólares). En el caso de la mediana, los datos deben ser por lo menos de un nivel ordinal de medición.

Precios ordenados de menor a mayor		Precios ordenados de mayor a menor
$ 60 000		$275 000
65 000		80 000
70 000	← Mediana →	70 000
80 000		65 000
275 000		60 000

Observe que existe el mismo número de precios bajo la mediana de 70 000 dólares que sobre ella. Por consiguiente, a la mediana no le afectan precios bajos o altos. Si el precio más alto fuera de 90 000, 300 000, o incluso de un millón de dólares, la mediana del precio aún sería de 70 000 dólares. Asimismo, si el precio más bajo fuera de 20 000 o 50 000 dólares, su mediana todavía sería de 70 000 dólares.

En el ejemplo anterior hay un número *impar* de observaciones (cinco). ¿Cómo se determina la mediana en el caso de un número *par* de observaciones? Como antes, se ordenan las observaciones. Enseguida, con el fin de obtener un único valor por convención, calcule la media de las dos observaciones medias. Así, en el caso de un número par de observaciones, la mediana quizá no sea uno de los valores dados.

EJEMPLO

Facebook es una popular red social en internet. Los usuarios pueden agregar amigos y enviarles mensajes, así como actualizar sus perfiles personales para informar a sus amigos sobre sí mismos y sus actividades. Una muestra de 10 adultos reveló que pasaron las siguientes cantidades de horas utilizando Facebook el mes anterior.

3	5	7	5	9	1	3	9	17	10

Encuentre la media aritmética de horas.

SOLUCIÓN

Observe que el número de adultos muestreados es par (10). Como antes, el primer paso es ordenar las horas durante las cuales se usó Facebook de menor a mayor. Identifique los dos tiempos medios. La media aritmética de las dos observaciones del medio proporciona la mediana de horas. Si organiza los valores de menor a mayor tendrá que:

1	3	3	5	5	7	9	9	10	17

Para encontrar la media se promedian los dos valores centrales, que en este caso son 5 y 7 horas; la media de estos dos valores es 6. Se concluye que el usuario de Facebook típico pasa seis horas al mes en el sitio. Observe que la mediana no es uno de los valores. Asimismo, la mitad de los tiempos se encuentran por debajo de la mediana y la mitad, sobre ella.

Las principales propiedades de la mediana son las siguientes:

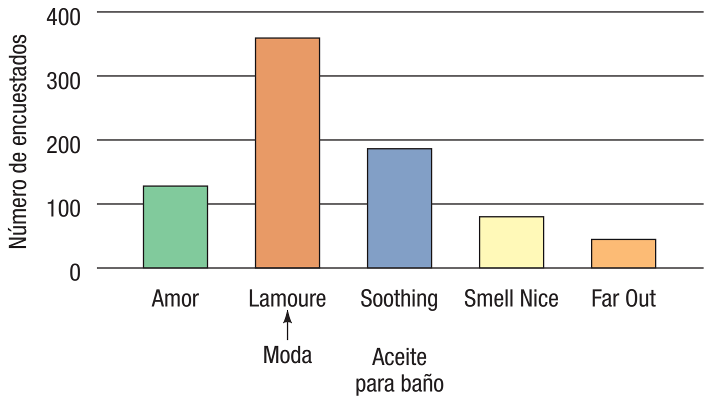
1. **Los valores extremadamente grandes o pequeños no la afectan.** Por consiguiente, la mediana es una valiosa medida de ubicación cuando dichos valores se presentan.
2. **Es calculable en el caso de datos de nivel ordinal o mayor.** Recuerde, del capítulo 1, que los datos de nivel ordinal pueden ordenarse de menor a mayor.

La moda

La **moda** es otra medida de ubicación.

MODA Valor de la observación que aparece con mayor frecuencia.

La moda es de especial utilidad para resumir datos de nivel nominal. Un ejemplo de su aplicación en este tipo de datos es: una compañía creó cinco aceites para baño. En la gráfica 3.1 se muestran los resultados de una encuesta de mercado que se diseñó para determinar qué aceite para baño prefieren los consumidores. La mayoría de los encuestados se inclinó por Lamoure, según lo evidencia la barra más grande. Por consiguiente, Lamoure representa la moda.



GRÁFICA 3.1 Número de encuestados que prefieren ciertos aceites para baño

EJEMPLO

Recuerde los datos con respecto a la distancia en millas entre las salidas en la autopista I-75 que atraviesa Kentucky. Esa información se repite a continuación.

11	4	10	4	9	3	8	10	3	14	1	10	3	5
2	2	5	6	1	2	2	3	7	1	3	7	8	10
1	4	7	5	2	2	5	1	1	3	3	1	2	1

¿Cuál es la distancia modal?

SOLUCIÓN

El primer paso es organizar las distancias en una tabla de frecuencias para determinar la distancia que se presenta más a menudo.

Distancia en millas entre salidas	Frecuencia
1	8
2	7
3	7
4	3
5	4
6	1
7	3
8	2
9	1
10	4
11	1
14	1
Total	42

La distancia que se presenta con mayor frecuencia es una milla. Se repite ocho veces, es decir, hay ocho salidas separadas por una milla. Así que la distancia modal entre salidas es una milla.

¿Cuál de estas tres medidas de ubicación (media, mediana o moda) representa mejor la ubicación central de estos datos? ¿Es la moda la mejor medida de ubicación para representar los datos de

Kentucky? No. La moda solo toma en cuenta la escala nominal de medición, y la variable “millas” se mide utilizando la escala de razón. Se ha calculado que la media es de 4.57 millas. ¿Es la media la mejor medida de ubicación para representar estos datos? Probablemente no. Hay muchos casos en que la distancia entre salidas es larga. Estos valores afectan la media, pues hacen que sea demasiado grande y no es representativa de las distancias entre salidas. ¿Y qué hay de la mediana? La distancia mediana es de tres millas. Esto es, la mitad de las distancias entre salidas es de tres millas o menos. En este caso, la mediana de tres millas entre salidas probablemente es una medida más representativa.

En resumen, es posible determinar la moda para todos los niveles de datos —nominal, ordinal, de intervalo y de razón—. La moda también tiene la ventaja de no verse afectada por valores extremadamente grandes o pequeños.

No obstante, la moda tiene sus desventajas, por las cuales se le utiliza con menor frecuencia que a la media o la mediana. En muchos conjuntos de datos no existe la moda, porque ningún valor se presenta más de una vez. Por ejemplo, no hay moda en el siguiente conjunto de datos de precios porque cada valor solo aparece en una ocasión: 19, 21, 23, 20 y 18 dólares. Sin embargo, como cada valor es diferente, podría argumentar que cada valor es la moda. Por el contrario, en el caso de algunos conjuntos de datos hay más de una moda. Suponga que las edades de los miembros de un club de inversionistas son 22, 26, 27, 27, 31, 35 y 35 años. Las edades 27 y 35 son modas. Así, este agrupamiento de edades se denomina *bimodal* (tiene dos modas). Alguien podría cuestionar la utilización de dos modas para representar la ubicación de este conjunto de datos de edades.

AUTOEVALUACIÓN 3-2

1. Una muestra de personas solteras, residentes en Towson, Texas, que reciben pagos por seguridad social reveló los siguientes subsidios mensuales: 852, 598, 580, 1 374, 960, 878 y 1 130 dólares.
   (a) ¿Cuál es la mediana del subsidio mensual?
   (b) ¿Cuántas observaciones se encuentran debajo de la mediana? ¿Por encima de ella?
2. El número de interrupciones de trabajo en la industria del automóvil en meses muestreados son de 6, 0, 10, 14, 8 y 0.
   (a) ¿Cuál es la mediana del número de interrupciones?
   (b) ¿Cuántas observaciones se encuentran por debajo de la mediana? ¿Por encima de ella?
   (c) ¿Cuál es el número modal de interrupciones de trabajo?

EJERCICIOS

13. ¿Qué informaría usted como valor modal de un conjunto de observaciones si hubiera un total de:
- **a.** Diez observaciones y no hubiera dos valores iguales;
- **b.** Seis observaciones, todas iguales;
- **c.** Seis observaciones con valores de 1, 2, 3, 3, 4 y 4?

En los ejercicios 14 a 16, determine a) la media, b) la mediana y c) la moda.

14. Los siguientes son los números de cambios de aceite de los últimos siete días en Jiffy Lube, que se ubica en la esquina de Elm Street y Pennsylvania Avenue.

41	15	39	54	31	15	33

15. El siguiente es el cambio porcentual en el ingreso neto del año anterior al presente en una muestra de 12 compañías constructoras de Denver.

5	1	−10	−6	5	12	7	8	6	5	−1	11

16. Las siguientes son las edades de 10 personas que se encuentran en la sala de videojuegos del Southwyck Shopping Mall a las 10 de la mañana.

12	8	17	6	11	14	8	17	10	8

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

17. Abajo se enlistan diversos indicadores del crecimiento económico a largo plazo de Estados Unidos.

Indicador económico	Cambio porcentual	Indicador económico	Cambio porcentual
Inflación	4.5%	PIB real	2.9%
Exportaciones	4.7	Inversión (residencial)	3.6
Importacioness	2.3	Inversión (no residencial)	2.1
Ingreso	2.9	Productividad (total)	1.4
Consumo	2.7	Productividad (fabricación)	5.2

a. ¿Cuál es la mediana del cambio porcentual?
b. ¿Cuál es el cambio porcentual modal?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

18. Sally Reynolds vende bienes raíces en el área costera del norte de California. Enseguida se muestra la cantidad total de las comisiones que ha ganado entre 2002 y 2012. Encuentre la media, la mediana y la moda de las comisiones que ha ganado en esos años.

Año	Cantidad (miles de dólares)
2002	$237.51
2003	233.80
2004	206.97
2005	248.14
2006	164.69
2007	292.16
2008	269.11
2009	225.57
2010	255.33
2011	202.67
2012	206.53

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

19. La empresa de contabilidad de Rowatti y Koppel se especializa en la elaboración de declaraciones del impuesto sobre la renta de profesionales independientes, como médicos, dentistas, arquitectos y abogados. La firma emplea a 11 contadores que preparan declaraciones. El año previo, el número de declaraciones que elaboró cada contador fue la siguiente:

58	75	31	58	46	65	60	71	45	58	80

Determine la media, la mediana y la moda de las cantidades de declaraciones que cada contador elaboró. Si usted elaborara una, ¿qué medida de ubicación recomendaría?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

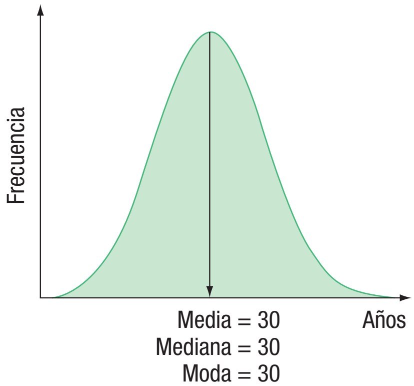
20. La demanda de videojuegos que suministra Mid-Tech Video Games, Inc., se ha disparado en los últimos siete años. De ahí que el propietario requiera técnicos que se mantengan a la par con la demanda. Mid-Tech proporciona a cada solicitante una prueba que el doctor McGraw, diseñador de la prueba, cree que se relaciona estrechamente con la habilidad para crear videojuegos. Para la población en general, la media de esta prueba es de 100. Enseguida aparecen los resultados de la prueba en el caso de los aspirantes.

95	105	120	81	90	115	99	100	130	10

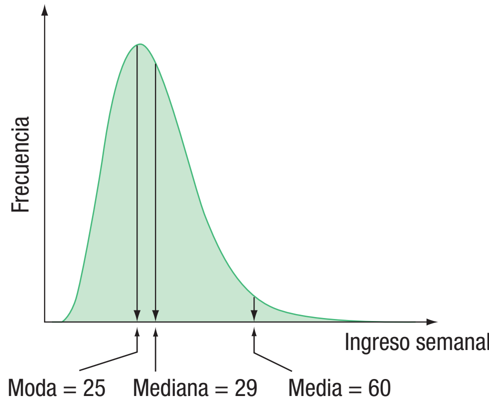
Con base en la prueba, el presidente se encuentra interesado en las cualidades generales de los aspirantes al puesto. Calcule la media y la mediana de los resultados de los 10 aspirantes. ¿Qué informaría usted al presidente? ¿Le parece que los aspirantes son mejores que el resto de la población?

Posiciones relativas de la media, la mediana y la moda

Observe el histograma de la figura 3.2 (página siguiente); se trata de una distribución simétrica que también tiene forma de campana. Esta distribución *posee la misma forma en cualquier lado desde el centro.* Si el histograma estuviera doblado a la mitad, las dos mitades serían idénticas. En cualquier distribución simétrica, la moda, la mediana y la media siempre son iguales. Se consideran equivalentes a 30 años en la gráfica 3.2. Cabe mencionar que hay distribuciones simétricas que no tienen forma de campana.



GRÁFICA 3.2 Distribución simétrica



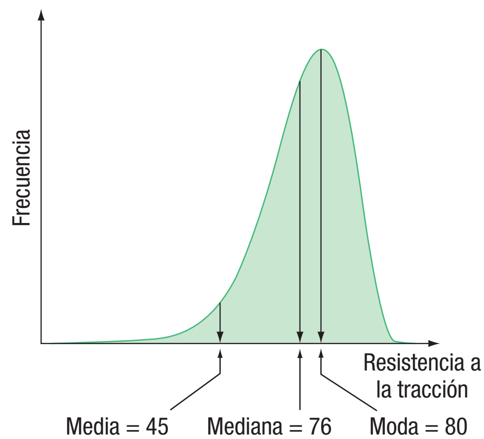
GRÁFICA 3.3 Distribución con sesgo positivo

El número de años correspondiente al punto más alto de la curva es la *moda* (30 años). Como la distribución es simétrica, la *mediana* corresponde al punto en el que la distribución se divide a la mitad (30 años). Asimismo, y dado que la media aritmética es el punto de equilibrio de una distribución, y la distribución es simétrica, la media aritmética es 30. Lógicamente, cualquiera de las tres medidas sería apropiada para representar el centro de la distribución.

Si una distribución es asimétrica o **sesgada,** la relación entre las tres medidas cambia. En una **distribución con sesgo positivo**, como la distribución del ingreso semanal de la gráfica 3.3, la media aritmética es la mayor de las tres medidas. ¿Por qué? Porque en ella influyen, más que sobre la mediana o la moda, unos cuantos valores extremadamente altos. Por lo general, la mediana es la siguiente medida más grande en una distribución de frecuencias con sesgo positivo. La moda es la menor de las tres medidas.

Si la distribución tiene un sesgo muy pronunciado, como en el caso de los ingresos semanales de la gráfica 3.4, la media no sería una medida adecuada. La mediana y la moda serían más representativas.

Por el contrario, si una distribución tiene un **sesgo negativo**, la media es la menor medida de las tres. Por supuesto, la media es sensible a la influencia de una cantidad extremadamente pequeña de observaciones. La mediana es mayor que la media aritmética y la moda es la más grande de las tres medidas. De nuevo, si la distribución tiene un sesgo muy pronunciado, la media no se utilizaría para representar a los datos.



GRÁFICA 3.4 Distribución con sesgo negativo

AUTOEVALUACIÓN 3-3

Las ventas semanales de una muestra de tiendas de suministros electrónicos de alta tecnología se organizaron en una distribución de frecuencias. La media de las ventas semanales que se calculó fue de 105 900 dólares; la mediana, de 105 000 dólares y la moda, de 104 500 dólares.

(a) Trace una gráfica de las ventas por medio de un polígono de frecuencias suavizado. Observe la ubicación de la media, la mediana y la moda sobre el eje *X*.
(b) ¿La distribución es simétrica, tiene un sesgo positivo o negativo? Explique su respuesta.

EJERCICIOS

21. La tasa de desempleo en el estado de Alaska por mes aparece en la siguiente tabla:

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
8.7	8.8	8.7	7.8	7.3	7.8	6.6	6.5	6.5	6.8	7.3	7.6

a. ¿Cuál es la media aritmética de la tasa de desempleo en Alaska?
b. Encuentre la media y la moda de la tasa de desempleo.
c. Calcule la media aritmética y la mediana solo de los meses de invierno (de diciembre a marzo). ¿Es muy diferente?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

22. Big Orange Trucking diseña un sistema de información que se utiliza para comunicaciones en cabina. Debe resumir datos de ocho sitios de cierta zona para describir condiciones típicas. Calcule una medida adecuada de ubicación central de cada una de las tres variables que aparecen en la siguiente tabla:

Ciudad	Dirección del viento	Temperatura	Pavimento
Anniston, AL	Oeste	89	Seco
Atlanta, GA	Noreste	86	Mojado
Augusta, GA	Suroeste	92	Mojado
Birmingham, AL	Sur	91	Seco
Jackson, MS	Suroeste	92	Seco
Meridian, MS	Sur	92	Sendero
Monroe, LA	Suroeste	93	Seco
Tuscaloosa, AL	Suroeste	93	Sendero

Solución con software

Es posible utilizar un paquete de software de estadística para determinar varias medidas de ubicación.

EJEMPLO

En la tabla 2.4 (página 23) se muestra la ganancia que obtuvo Applewood Auto Group el mes anterior por la venta de 180 vehículos. Determine la media y la mediana de los precios de venta.

SOLUCIÓN

La media, la mediana y la moda del monto de las ganancias se presentan en el informe de la siguiente salida de Excel (resaltados en la toma de pantalla; recuerde que las instrucciones para crear la salida aparecen en la sección “Comandos de software” localizada en el apéndice C). En el estudio se incluyen 180 vehículos, así que los cálculos con una calculadora resultarían tediosos y propensos a error.

APPLEWOOD AUTO GROUP

	A	B	C	D	E	F	G	H
1	Age	Profit	Location	Vehicle-Type	Previous		*Profit*	
2	21	$1,387	Tionesta	Sedan	0			
3	23	$1,754	Sheffield	SUV	1		Mean	1843.17
4	24	$1,817	Sheffield	Hybrid	1		Standard Error	47.97
5	25	$1,040	Sheffield	Compact	0		Median	1882.50
6	26	$1,273	Kane	Sedan	1		Mode	1915.00
7	27	$1,529	Sheffield	Sedan	1		Standard Deviation	643.63
8	27	$3,082	Kane	Truck	0		Sample Variance	414256.61
9	28	$1,951	Kane	SUV	1		Kurtosis	-0.22
10	28	$2,692	Tionesta	Compact	0		Skewness	-0.24
11	29	$1,342	Kane	Sedan	2		Range	2998
12	29	$1,206	Sheffield	Sedan	0		Minimum	294
13	30	$443	Kane	Sedan	3		Maximum	3292
14	30	$1,621	Sheffield	Truck	1		Sum	331770
15	30	$754	Olean	Sedan	2		Count	180

La ganancia promedio es de 1 843.17 dólares y la mediana, de 1 882.50 dólares. La diferencia entre estos valores es menor a 40 dólares, así que cualquiera de ellos es razonable. En la salida de Excel también es posible ver que se vendieron 180 vehículos y que la ganancia total fue de 331 700 dólares. Más adelante (en este y en otros capítulos) se explicará el significado de error estándar, desviación estándar y otras medidas reportadas en esta salida.

¿Cuál es la conclusión? La ganancia típica de un vehículo es de aproximadamente 1 850 dólares. La gerencia de Applewood puede usar este valor para realizar la proyección de sus ingresos. Por ejemplo, si la distribuidora incrementa el número de ventas en un mes, de 180 a 200, obtiene una estimación adicional de 37 000 dólares de ganancia, calculada mediante 20($1 850).

La media ponderada

OA3-2 Calcular la media ponderada.

La media ponderada, que constituye un caso especial de la media aritmética, se presenta cuando hay varias observaciones con el mismo valor. Para entender este tema, suponga que el restaurante Wendy´s vende refrescos medianos, grandes y gigantes a 0.90, 1.25 y 1.50 dólares, respectivamente. De las 10 últimas bebidas que se vendieron 3 eran medianas, 4 eran grandes y 3 eran gigantes. Para determinar el precio promedio de las últimas 10 bebidas vendidas recurra a la fórmula [3.2].

$$\bar{x} = \frac{\$.90 + \$.90 + \$.90 + \$1.25 + \$1.25 + \$1.25 + \$1.25 + \$1.50 + \$1.50 + \$1.50}{10}$$

$$\bar{x} = \frac{\$12.20}{10} = \$1.22$$

El precio promedio de venta de las últimas 10 bebidas es de 1.22 dólares.

Una forma más fácil de calcular el precio promedio de venta consiste en determinar la media ponderada: multiplique cada observación por el número de veces que aparece. La media ponderada se representa como $\bar{x}_w$, que se lee: "*x* barra subíndice *w*".

$$\bar{x} = \frac{3(\$0.90) + 4(\$1.25 + 3(\$1.50)}{10} = \frac{\$12.20}{10} = \$1.22$$

En este caso, las ponderaciones son conteos de frecuencias. Sin embargo, cualquier medida de importancia podría utilizarse como una ponderación. En general, la media ponderada del conjunto de números representados como $x_1, x_2, x_3, \ldots, x_n$, con las ponderaciones correspondientes $w_1, w_2, w_3, \ldots, w_n$, se calcula de la siguiente manera:

MEDIA PONDERADA

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n}{w_1 + w_2 + w_3 + \cdots + w_n} \qquad \textbf{[3.3]}$$

que se abrevia de la siguiente manera:

$$\bar{x}_w = \frac{\Sigma(wx)}{\Sigma w}$$

Observe que el denominador de una media ponderada siempre es la suma de las ponderaciones.

EJEMPLO

Carter Construction Company paga a sus empleados que trabajan por hora 16.50, 19.00 o 25.00 dólares por hora. Hay 26 empleados contratados para trabajar por hora, 14 de los cuales reciben la tarifa de 16.50 dólares; 10, la tarifa de 19.00 dólares y 2, la de 25.00 dólares. ¿Cuál es la tarifa promedio por hora que se paga a los 26 empleados?

SOLUCIÓN

Para determinar la tarifa media por hora, multiplique cada una de las tarifas por hora por el número de empleados que ganan dicha tarifa. De acuerdo con la fórmula [3.3], la tarifa media por hora es

$$\bar{x}_w = \frac{14(\$16.50) + 10(\$19.00 + 2(\$25.00)}{10} = \frac{\$471.00}{10} = \$18.1154$$

El salario promedio ponderado por hora se redondea a 18.12 dólares.

AUTOEVALUACIÓN 3-4

Springers vendió 95 trajes para caballero Antonelli a un precio normal de 400 dólares. Durante la venta de primavera rebajaron los trajes a 200 dólares y vendieron 126. Al final de la venta de liquidación, redujeron el precio a 100 dólares y se vendieron los restantes 79 trajes.

(a) ¿Cuál fue el precio promedio ponderado de un traje Antonelli?

(b) Springers pagó 200 dólares por cada uno de los 300 trajes. Haga algún comentario sobre la ganancia de la tienda por traje, si un vendedor recibe 25 dólares de comisión por cada unidad que vende.

EJERCICIOS

23. En junio, una inversionista compró 300 acciones de Oracle (una compañía de tecnología de la información) a 20 dólares cada una. En agosto compró 400 acciones más a 25 dólares. En noviembre compró otras 400 acciones, pero el precio bajó a 23 dólares por título. ¿Cuál es el precio promedio ponderado de cada acción?

24. Bookstall, Inc., es una librería especializada que se dedica a la venta de libros usados por internet. Los libros de pasta blanda cuestan un dólar cada uno y los de pasta dura, 3.50 dólares cada uno. De los 50 libros que se vendieron el pasado martes por la mañana, 40 eran de pasta blanda y el resto, de pasta dura. ¿Cuál fue el precio promedio ponderado por libro?

25. Loris Healthcare System tiene 200 empleados en su personal de enfermería; 50 son auxiliares de enfermería, 50 son enfermeras practicantes y 100 son enfermeras tituladas. Las auxiliares de enfermería ganan 8 dólares por hora; las enfermeras practicantes, 15; y las tituladas, 24. ¿Cuál es el salario promedio ponderado por hora?

26. Andrews and Associates se especializa en leyes empresariales. Cobran 100 dólares por hora de investigación de un caso; 75 dólares por hora de asesoría y 200 dólares por hora de redacción de un expediente. La semana pasada uno de los socios dedicó 10 horas a dar asesoría a una clienta, 10 horas a investigar el caso y 20 horas a redactar el expediente. ¿Cuál fue el monto medio ponderado por hora de honorarios por servicios legales?

OA3-3
Calcular e interpretar la media geométrica.

La media geométrica

La media geométrica es útil para determinar el cambio promedio de porcentajes, razones, índices o tasas de crecimiento. Posee amplias aplicaciones en la administración y la economía, ya que con frecuencia hay interés en determinar los cambios porcentuales de ventas, salarios o cifras económicas, como el producto interno bruto, los cuales se combinan o son la base de otros. La media geométrica de un conjunto de *n* números positivos se define como la raíz enésima de un producto de *n* valores. La fórmula de la media geométrica se escribe de la siguiente manera:

MEDIA GEOMÉTRICA $$GM = \sqrt[n]{(x_1)(x_2)\cdots(x_n)}$$ **[3.4]**

La media geométrica siempre es menor o igual (nunca mayor) que la media aritmética. Todos los datos deben ser positivos.

Como ejemplo de media geométrica, suponga que usted recibe 5% de incremento salarial este año y 15% de incremento el siguiente. El incremento porcentual anual promedio es de 9.886, no de 10.0. ¿Por qué razón? Comience calculando la media geométrica. Recuerde, por ejemplo, que 5% de incremento salarial equivale a 105%, que expresa como 1.05.

$$GM = \sqrt{(1.05)\,(1.15)} = 1.09886$$

Este resultado puede verificarse suponiendo que su ingreso mensual fue de 3 000 dólares al comienzo y que recibió dos incrementos de 5% y 15%.

Incremento 1 =	$3 000(.05) =	$150.00
Incremento 2 =	$3 150(.15) =	472.50
Total		$622.50

El incremento total de su salario es de 622.50 dólares. Esto equivale a:

$3 000.00(.09886) = $296.58
$3 296.58(.09886) = $325.90
$622.48 que es alrededor de $622.50 dólares

El siguiente ejemplo muestra la media geométrica de diversos porcentajes.

EJEMPLO

La recuperación de una inversión que realizó Atkins Construction Company durante cuatro años consecutivos fue de 30%, 20%, −40% y 200%. ¿Cuál es la media geométrica de la recuperación de la inversión?

SOLUCIÓN

El número 1.3 representa 30% de la recuperación de la inversión, que es la inversión *original* de 1.0 más la *recuperación* de 0.3. Por otra parte, el número 0.6 representa la pérdida de 40%, que es la inversión original de 1.0 menos la pérdida de 0.4. Este cálculo supone que el total de la inversión de cada periodo se reinvierte o se convierte en la base de la siguiente. En otras palabras, la base del segundo periodo es 1.3 y la base del tercer periodo es (1.3)(1.2) y así sucesivamente.

En consecuencia, la media geométrica de la tasa de recuperación es de 29.4%, determinada por medio del siguiente cálculo:

$$GM = \sqrt[n]{(x_1)(x_2) \cdots (x_n)} = \sqrt[4]{(1.3)(1.2)(0.6)(3.0)} = \sqrt[4]{2.808} = 1.294$$

De esta manera, la media geométrica es la raíz cuarta de 2.808. Así, la tasa promedio de recuperación (tasa de crecimiento anual compuesta) es de 29.4%.

Observe, asimismo, que si calcula la media aritmética $[(30 + 20 - 40 + 200)/4 = 52.5]$, obtendrá un número mucho más grande, lo que dispararía la tasa de recuperación real.

Otro modelo de aplicación de la media geométrica se relaciona con la determinación de un cambio porcentual promedio durante cierto periodo. Por ejemplo, si usted ganó 30 000 dólares en el año 2000 y 50 000 dólares en 2010, ¿cuál es la tasa anual de incremento durante el periodo? Esta es de 5.24%. La tasa de incremento se determina a partir de la siguiente fórmula.

TASA DE INCREMENTO CON EL TIEMPO

$$GM = \sqrt[n]{\frac{\text{Valor al final del periodo}}{\text{Valor al inicio del periodo}}} - 1 \qquad \textbf{[3.5]}$$

En el recuadro anterior, *n* es el número de periodos. Un ejemplo mostrará los detalles para determinar el incremento porcentual anual.

EJEMPLO

Durante la década de 1991 y hasta los primeros años del siguiente decenio, Las Vegas, Nevada, fue la ciudad de mayor crecimiento en Estados Unidos. La población se incrementó de 258 295 en 1990 a 584 539 en 2011, o 126.3% durante el periodo. La población es más del doble. ¿Cuál es el incremento *anual* promedio?

SOLUCIÓN

Hay 21 años entre 1990 y 2011, así que $n = 21$. De esta manera, la fórmula [3.5] de la media geométrica, aplicada a este problema, se transforma en:

$$GM = \sqrt[21]{\frac{\text{Valor al final del periodo}}{\text{Valor al inicio del periodo}}} - 1.0 = \sqrt[21]{\frac{584\,539}{258\,295}} - 1.0 = 1.0397 - 1.0 = .0397$$

Para resumir, los pasos para calcular una media geométrica son:

1. Dividir el valor al final del periodo entre el valor al comienzo del periodo.
2. Encontrar la enésima del rango, donde *n* es el número de periodos.
3. Restar uno.

El valor de 0.0397 indica que el crecimiento anual promedio durante el periodo fue de 3.97%. Expresado en otros términos, la población de Las Vegas creció a una tasa de 3.97% por año de 1990 a 2011.

AUTOEVALUACIÓN 3-5

1. El incremento porcentual de ventas de los últimos cuatro años en Combs Cosmetics fue de 4.91, 5.75, 8.12 y 21.60.
   (a) Determine la media geométrica del incremento porcentual.
   (b) Determine la media aritmética del incremento porcentual.
   (c) ¿La media aritmética es igual o mayor que la media geométrica?
2. La producción de camiones Cablos se elevó de 23 000 unidades en el año 2000 a 120 520 unidades en 2013. Calcule la media geométrica del incremento porcentual anual.

EJERCICIOS

27. Calcule la media geométrica de los siguientes incrementos porcentuales: 8, 12, 14, 26 y 5.
28. Estime la media geométrica de los siguientes incrementos porcentuales: 2, 8, 6, 4, 10, 6, 8 y 4.
29. A continuación se enlista el incremento porcentual de ventas de MG Corporation durante los últimos cinco años. Determine la media geométrica del incremento porcentual de ventas en ese periodo.

9.4	13.8	11.7	11.9	14.7

30. En 1996, en Estados Unidos, un total de 14 968 000 contribuyentes presentaron en forma electrónica sus declaraciones de impuestos. En el año 2010 el número se había incrementado a 99 000 000. ¿Cuál es la media geométrica del incremento anual del periodo?
31. El U.S. Bureau of Labor Statistics publica mensualmente el índice de precios al consumidor e informa el cambio de precios de una canasta de artículos en el mercado de un periodo a otro. El índice del año 2000 fue de 172.2. Para 2012 se incrementó a 229.6. ¿Cuál es la media geométrica del incremento anual de dicho periodo?
32. JetBlue Airways es una aerolínea estadounidense de bajo costo con sede en la ciudad de Nueva York. Su base principal está en el Aeropuerto Internacional John F. Kennedy. La ganancia de JetBlue en 2002 fue de 635.2 millones de dólares. En 2012 se incrementó a 3 788 millones de dólares. ¿Cuál es la media geométrica del incremento anual en dicho periodo?
33. En 1985 había 340 213 suscriptores a compañías de telefonía celular en Estados Unidos. En 2012, el número de suscriptores aumentó a 327 577 529. ¿Cuál es la media geométrica del incremento anual en dicho periodo?
34. La siguiente información muestra el costo de un año de estudios en universidades públicas y privadas entre 2002-03 y 2012-13. ¿Cuál es la media geométrica del incremento anual en dicho periodo en el caso de ambos tipos de escuelas? Compare las tasas de incremento.

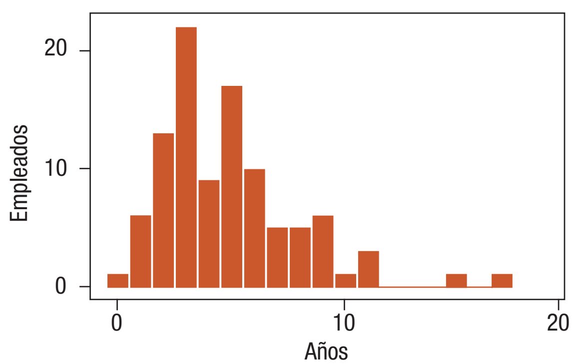
Tipo de universidad	2002-03	2012-13
Pública	$ 4 960	$ 8 655
Privada	18 056	29 056

OA3-4
Calcular e interpretar el rango, la varianza y la desviación estándar.

¿Por qué estudiar la dispersión?

Una medida de ubicación, como la media, la mediana o la moda, solo describe el centro de los datos. Desde este punto de vista resulta valiosa, pero no revela nada acerca de la dispersión de los datos. Por ejemplo, si un guía de turismo ecológico dice que el río que se encuentra a pocos pasos tiene en promedio tres pies de profundidad, ¿querría usted cruzarlo a pie sin más información? Quizá no. Usted desearía saber algo sobre la variación de la profundidad. ¿La máxima profundidad es de 3.25 pies y la mínima, de 2.75 pies? En dicho caso, usted estaría de acuerdo en cruzar. ¿Qué hay si usted supo que la profundidad del río varía de 0.50 a 5.5 pies? Su decisión probablemente sería no cruzar. Antes de decidir, usted desea información tanto de la profundidad típica como de la dispersión de la profundidad del río.

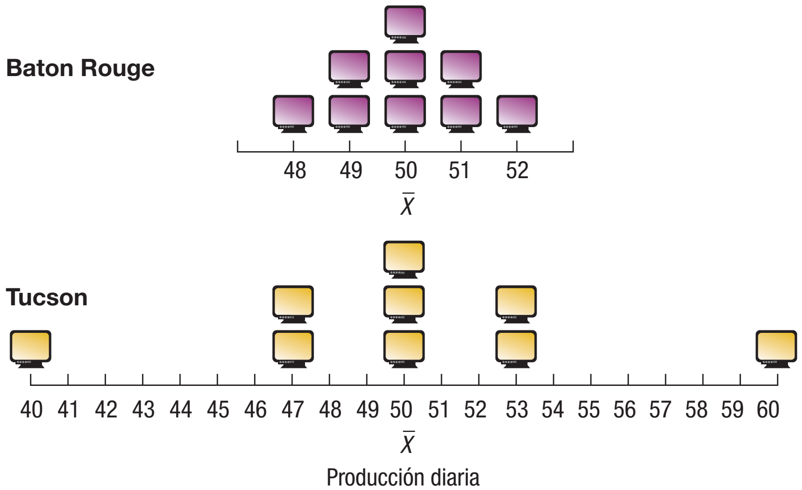
Una medida pequeña de dispersión indica que los datos se acumulan con proximidad alrededor de la media aritmética. Por consiguiente, la media se considera representativa de los datos. Por el contrario, una medida grande de dispersión indica que la media no es confiable (vea la gráfica 3.5). Los 100 empleados de Hammond Iron Works, Inc., una compañía que fabrica acero, se organizan en un histograma basado en el número de años que cada uno ha laborado en la compañía. La media es de 4.9 años, pero la dispersión de los datos es de 6 meses a 16.8 años. La media de 4.9 años no es representativa de todos los empleados.



GRÁFICA 3.5 Histograma de los años laborados para Hammond Iron Works, Inc.

Una segunda razón para estudiar la dispersión en un conjunto de datos consiste en comparar la propagación en dos o más distribuciones. Por ejemplo, suponga que el nuevo monitor de computadora Vision Quest LCD se arma en Baton Rouge y también en Tucson. La producción media aritmética por hora, en ambas plantas, es de 50.

Con base en estas dos medias, podría concluir que las distribuciones de las producciones por hora son idénticas. Sin embargo, los registros de producción de 9 horas en las dos plantas revelan que esta conclusión no es correcta (vea la gráfica 3.6 en la página 61). La producción de Baton Rouge varía de 48 a 52 montajes por hora,



GRÁFICA 3.6 Producción por hora de monitores de computadora en las plantas de Baton Rouge y Tucson

mientras que la producción en la planta de Tucson es más errática, ya que varía de 40 a 60 por hora. Por lo tanto, la producción por hora en Baton Rouge se acumula cerca de la media (50); la producción por hora de Tucson es más dispersa.

Se considerarán diversas medidas de dispersión. El rango se sustenta en los valores máximo y mínimo del conjunto de datos, es decir, solo se consideran dos valores. La desviación media, la varianza y la desviación estándar se basan en desviaciones de la media aritmética.

Rango

La medida más simple de dispersión es el **rango.** Representa la diferencia entre los valores máximo y mínimo de un conjunto de datos. En forma de ecuación:

RANGO	Rango = valor máximo − valor mínimo	**[3.6]**

El rango se emplea mucho en aplicaciones de control de procesos estadísticos (CPE) debido a que es fácil calcularlo y entenderlo.

EJEMPLO

Consulte la gráfica 3.6. Determine el rango del número de monitores de computadora que se producen por hora en las plantas de Baton Rouge y Tucson. Interprete ambos rangos.

SOLUCIÓN

El rango de la producción por hora de monitores de computadora en la planta de Baton Rouge es de 4, el cual se determina por la diferencia entre la producción máxima por hora (52) y la mínima (48). El rango de la producción por hora en la planta de Tucson es de 20 monitores, que se obtiene con el cálculo 60 − 40. Por tanto: 1) existe menos dispersión en la producción por hora en la planta de Baton Rouge que en la de Tucson, porque el rango de 4 monitores es menor que el de 20; 2) la producción se acumula más alrededor de la media de 50 en la planta de Baton Rouge que en la planta de Tucson (ya que un rango de 4 es menor que un rango de 20). Por ello, la producción media en la planta de Baton Rouge (50 monitores) resulta una medida de ubicación más representativa que la media de 50 monitores en la planta de Tucson.

Varianza

Un problema que presenta el rango estriba en que parte de dos valores, el máximo y el mínimo, es decir, no los toma en cuenta a todos. La **varianza** sí lo hace; mide la cantidad media respecto de la cual los valores de una población o muestra varían. Expresado en forma de definición:

ESTADÍSTICA EN ACCIÓN

El servicio postal de Estados Unidos ha intentado ser "más amigable con el usuario" durante los últimos años. Una encuesta reciente mostró que los consumidores estaban interesados en que hubiera más *regularidad* en los tiempos de entrega. Antes, una carta local podría tardar en llegar uno o varios días. "Solo díganme con cuántos días de anticipación tengo que enviar una tarjeta de felicitación a mi mamá para que llegue el día de su cumpleaños, ni antes ni después", era una queja común. El nivel de regularidad se mide a partir de la desviación estándar de los tiempos de entrega.

VARIANZA Media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética.

El siguiente ejemplo ilustra cómo se usa la varianza para medir la dispersión.

EJEMPLO

La siguiente tabla muestra el número de capuchinos que se vendieron en los locales de Starbucks de los aeropuertos de Orange County y Ontario, California, entre las 16:00 y las 17:00 horas, de una muestra de cinco días el mes anterior.

California Airports	
Orange County	Ontario
20	20
40	45
50	50
60	55
80	80

Determine la media, la mediana, el rango y la desviación media de cada local. Compare las similitudes y las diferencias.

SOLUCIÓN

La media, la mediana y el rango de cada aeropuerto se reportan a continuación como parte de una hoja de cálculo de Excel.

	A	B	C
1		California Airports	
2		Orange County	Ontario
3		20	20
4		40	45
5		50	50
6		60	55
7		80	80
8			
9	Mean	50	50
10	Median	50	50
11	Range	60	60

Observe que las tres medidas son exactamente iguales. ¿Indica esto que no hay diferencias entre ambos grupos de datos? Calculando las desviaciones medias se obtiene un panorama más claro. Primero, Orange County:

F	G	H
Calculation of Variance for Orange County		
Number Sold	Each Value - Mean	Squared Deviation
20	20 - 50 = -30	900
40	40 - 50 = -10	100
50	50 - 50 = 0	0
60	60 - 50 = 10	100
80	80 - 50 = 30	900
	Total	2000

$$\text{Varianza} = \frac{\Sigma(x - \mu)^2}{N} = \frac{(-30^2) + (-10^2) + 0^2 + 10^2 + 30^2}{5} = \frac{2\,000}{5} = 400$$

La varianza es 400. Esto es, la desviación cuadrada promedio desde la media es 400.

La siguiente tabla muestra los detalles para determinar la varianza para el número de capuchinos vendidos en el aeropuerto de Ontario.

Calculation of Variance for Ontario		
Number Sold	Each Value - Mean	Squared Deviation
20	20 - 50 = -30	900
45	45 - 50 = -5	25
50	50 - 50 = 0	0
55	55 - 50 = 5	25
80	80 - 50 = 30	900
	Total	1850

$$\text{Varianza} = \frac{\Sigma(x - \mu)^2}{N} = \frac{(-30^2) + (-5^2) + 0^2 + 5^2 + 30^2}{5} = \frac{1\,850}{5} = 370$$

Así que la media, la mediana y el rango de los capuchinos que se vendieron en ambos aeropuertos son los mismos, pero las varianzas son distintas. La varianza de Orange County es 400, la de Ontario es 370.

Interprete y compare los resultados de las medidas en el caso de las tiendas de Starbucks. La media y la mediana de ambas tiendas son exactamente las mismas, 50 capuchinos al día. Por consiguiente, la ubicación de ambas distribuciones es la misma. El rango en ambas tiendas también es igual: 60. Sin embargo, recuerde que el rango proporciona información limitada sobre la dispersión de la distribución, porque se basa solo en dos observaciones.

Las varianzas no son las mismas en ambos aeropuertos porque se basan en las diferencias entre todas las observaciones y la media aritmética, que muestra la relativa proximidad o acumulación de los datos concerniente a la media o centro de la distribución. Compare la varianza de Orange County (400) con la de Ontario (370). Con base en la varianza, es posible decir que la dispersión de la distribución de ventas de Starbucks Ontario se encuentra más concentrada, cerca de la media de 50, que en la tienda de Orange County.

La varianza tiene una importante ventaja sobre el rango: utiliza todos los valores en el cálculo. Recuerde que el rango solo incluye los valores más alto y más bajo.

AUTOEVALUACIÓN 3-6

Los pesos de los contenedores enviados a Irlanda son (en miles de libras):

95	103	105	110	104	105	112	90

(a) ¿Cuál es el rango de los pesos?
(b) Calcule el peso medio aritmético.
(c) Estime la desviación media de los pesos.

EJERCICIOS

En los ejercicios 35 a 38, calcule a) el rango, b) la media aritmética, c) la varianza; y d) analice los valores que obtenga.

35. Hubo cinco representantes de servicio al cliente que trabajaron en Electronic Super Store durante la pasada venta de fin de semana. Las cantidades de HDTV que vendieron estos representantes son: 5, 8, 4, 10 y 3.

36. El Departamento de Estadística de la Western State University ofrece ocho secciones de estadística básica. Enseguida aparecen los números de estudiantes matriculados en estas secciones: 34, 46, 52, 29, 41, 38, 36 y 28.

37. Dave's Automatic Door instala puertas automáticas para cocheras. La siguiente lista indica el número de minutos que se requieren para instalar 10 puertas automáticas: 28, 32, 24, 46, 44, 40, 54, 38, 32 y 42.

38. Las ocho compañías de la industria aeronáutica participaron en una encuesta sobre la recuperación de la inversión que tuvieron el año anterior. Los resultados (en porcentaje) son los siguientes: 10.6, 12.6, 14.8, 18.2, 12.0, 14.8, 12.2 y 15.6.

39. Diez adultos jóvenes que viven en California, elegidos al azar, calificaron el sabor de una nueva pizza de sushi con atún, arroz y algas en una escala de 1 a 50, en la que 1 indica que no les gusta el sabor y 50, que sí les gusta. Las calificaciones fueron las siguientes:

34	39	40	46	33	31	34	14	15	45

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

En un estudio paralelo, 10 adultos jóvenes de Iowa, elegidos al azar, calificaron el sabor de la misma pizza. Las calificaciones fueron las siguientes:

28	25	35	16	25	29	24	26	17	20

Como investigador de mercado, compare los mercados potenciales para la pizza de sushi.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

40. Los archivos de personal de ocho empleados en las instalaciones de Pawnee de Acme Carpet Cleaners, Inc., revelaron que durante el último semestre estos perdieron las siguientes cantidades de días por enfermedad:

2	0	6	3	10	4	1	2

Durante el mismo periodo, los archivos revelaron que los ocho empleados que trabajaron en la planta de Chickpee de Acme Carpets perdieron las siguientes cantidades de días por enfermedad:

2	0	1	0	5	0	1	0

Como director de relaciones humanas, compare las ausencias en las dos plantas. ¿Qué recomendaría?

Varianza de la población

En el ejemplo anterior se desarrolló el concepto de varianza como una medida de dispersión. En forma similar a como se hace con la media, es posible calcular la varianza de una población o de una muestra. La varianza de la población se determina de la siguiente manera:

VARIANZA DE LA POBLACIÓN $$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} \quad [3.7]$$

En esta fórmula:

- σ^2 es la varianza de la población (σ es la letra minúscula griega sigma); se lee *sigma al cuadrado;*
- x es el valor de una observación de la población;
- μ es la media aritmética de la población;
- N es el número de observaciones de la población.

Observe el proceso de cálculo de la varianza, implícito en la fórmula:

1. Comience por determinar la media.
2. Calcule la diferencia entre cada observación y la media, y eleve al cuadrado dicha diferencia.
3. Sume todas las diferencias elevadas al cuadrado.
4. Divida la suma de las diferencias elevadas al cuadrado entre el número de elementos de la población.

Así, la varianza de la población es la media de las diferencias elevadas al cuadrado entre cada valor y la media. En poblaciones cuyos valores son cercanos a la media, la varianza puede ser pequeña. En poblaciones cuyos valores se apartan de la media, la varianza de la población puede ser grande.

La varianza compensa el inconveniente que presenta el rango al utilizar todos los valores de la población, mientras que el rango incluye solo los valores máximo y mínimo. El problema de que $\Sigma(x-\mu) = 0$, se corrige elevando al cuadrado las diferencias, lo que siempre dará como resultado valores no negativos. He aquí otro ejemplo que ilustra el cálculo e interpretación de la varianza.

EJEMPLO

El número de multas de tránsito que se aplicaron el año anterior, por mes, en Beaufort County, Carolina del Sur, se reporta en la siguiente tabla.

Multas por mes											
Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
19	17	22	18	28	34	45	39	38	44	34	10

Determine la varianza de la población.

SOLUCIÓN

Dado que el objetivo es estudiar todas las multas que se aplicaron en un año, los datos integran una población. Para determinar la varianza de la población se utiliza la fórmula [3.7]. La siguiente tabla detalla los cálculos.

Mes	Multas (x)	$x - \mu$	$(x - \mu)^2$
Enero	19	−10	100
Febrero	17	−12	144
Marzo	22	−7	49
Abril	18	−11	121
Mayo	28	−1	1
Junio	34	5	25
Julio	45	16	256
Agosto	39	10	100
Septiembre	38	9	81
Octubre	44	15	225
Noviembre	34	5	25
Diciembre	10	−19	361
Total	348	0	1 488

1. Para comenzar, es necesario determinar la media aritmética de la población. El número total de multas aplicadas en el año es de 348, así que la media aritmética por mes es 29.

$$\mu = \frac{\Sigma x}{N} = \frac{19 + 17 + \cdots + 10}{12} = \frac{348}{12} = 29$$

2. Enseguida se calcula la diferencia entre la media y cada observación. Esta se muestra en la tercera columna de la tabla. Recuerde que previamente en este capítulo se indicó que la suma de las diferencias entre cada valor y la media es 0. En la hoja de cálculo, la suma de las diferencias entre la media y el número de multas de cada mes es 0.
3. El siguiente paso es elevar al cuadrado la diferencia entre cada valor mensual. Todas las diferencias elevadas al cuadrado serán positivas. Observe que al elevar al cuadrado un valor negativo, o multiplicar un valor negativo por sí mismo, siempre resulta en un valor positivo.
4. Se suman las diferencias elevadas al cuadrado. El total de la cuarta columna es 1 488. A esto se refiere la ecuación $\Sigma(x - \mu)^2$.
5. Finalmente, las diferencias elevadas al cuadrado se dividen entre N; es decir, el número de observaciones que se realizaron.

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N} = \frac{1\,488}{12} = 124$$

Así, la varianza de la población con respecto al número de multas es de 124.

Como en el caso del rango, la varianza se emplea para comparar la dispersión entre dos o más conjuntos de observaciones. Por ejemplo, se calculó que la varianza del número de multas levantadas en Beaufort County fue de 124. Si la varianza del número de multas aplicadas en Marlboro County, Carolina del Sur, es de 342.9, se concluye que: 1) hay menos dispersión en la distribución del número de multas levantadas en Beaufort (ya que 124 es menor que 342.9) y 2) el número de infracciones en Beaufort County se encuentra más apiñado en torno a la media (29) que el número de multas levantadas en Marlboro County. Por consiguiente, la media de multas aplicadas en Beaufort County constituye una medida de ubicación más representativa que la media de multas en Marlboro County.

Desviación estándar de la población

Al calcular la varianza es importante entender la unidad de medida y lo que ocurre cuando las diferencias en el numerador se elevan al cuadrado. Esto implica, en el ejemplo anterior, que el número de multas mensuales es la variable. Al calcular la varianza, las multas al cuadrado representan la unidad de medida por la varianza. Pero utilizar "multas al cuadrado" como unidad de medida es algo torpe.

Existe una forma de salir del problema. Si extrae la raíz cuadrada de la varianza de la población puede convertirla a las mismas unidades de medición que emplean los datos originales. La raíz cuadrada de 124 multas elevadas al cuadrado es de 11.4 multas. Las unidades ahora son, sencillamente, multas. La raíz cuadrada de la varianza de la población es la **desviación estándar de la población.**

DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN

$$\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}} \quad [3.8]$$

AUTOEVALUACIÓN 3-7

Este año la oficina en Filadelfia de PricewaterhouseCoopers LLP contrató a cinco contadores que están haciendo prácticas. Sus salarios mensuales iniciales fueron de 3 536, 3 173, 3 448, 3 121 y 3 622 dólares.

(a) Calcule la media de la población.
(b) Estime la varianza de la población.
(c) Aproxime la desviación estándar de la población.
(d) La oficina de Pittsburgh contrató a cinco empleados que están haciendo prácticas. El salario mensual promedio fue de 3 550 dólares y la desviación estándar, de 250 dólares. Compare ambos grupos.

EJERCICIOS

41. Considere en una población los siguientes cinco valores: 8, 3, 7, 3 y 4.
- **a.** Determine la media de la población.
- **b.** Determine la varianza.

42. Considere a los siguientes seis valores como una población: 13, 3, 8, 10, 8 y 6.
- **a.** Determine la media de la población.
- **b.** Determine la varianza.

43. El informe anual de Dennis Industries incluyó las siguientes ganancias primarias por acción común durante los últimos cinco años: 2.68, 1.03, 2.26, 4.30 y 3.58 dólares. Si supone que estos son los valores poblacionales:
- **a.** ¿Cuáles son las medias aritméticas de las ganancias primarias por acción común?
- **b.** ¿Cuál es la varianza?

44. Con respecto al ejercicio anterior, el informe anual de Dennis Industries también arrojó estos rendimientos sobre valores de renta variable durante el mismo periodo de cinco años (en porcentaje): 13.2, 5.0, 10.2, 17.5 y 12.9.
- **a.** ¿Cuál es la media aritmética del rendimiento?
- **b.** ¿Cuál es la varianza?

45. Plywood, Inc. dio a conocer las siguientes utilidades sobre valores de renta variable durante los últimos cinco años: 4.3, 4.9, 7.2, 6.7 y 11.6. Considere estos valores como poblacionales.
- **a.** Calcule el rango, la media aritmética, la varianza y la desviación estándar.
- **b.** Compare las utilidades sobre valores de renta variable de Playwood, Inc., con las de Dennis Industries que se citaron en el ejercicio 44.

46. Los ingresos anuales de cinco vicepresidentes de TMV Industries son: 125 000, 128 000, 122 000, 133 000 y 140 000 dólares. Considere estos valores como una población.
- **a.** ¿Cuál es el rango?
- **b.** ¿Cuál es el ingreso medio aritmético?
- **c.** ¿Cuál es la varianza poblacional y la desviación estándar?
- **d.** También se estudiaron los ingresos anuales del personal de otra empresa similar a TMV. La media fue de 129 000 dólares y la desviación estándar, de 8 612 dólares. Compare las medias y dispersiones de ambas firmas.

Varianza muestral y desviación estándar

La fórmula para determinar la media poblacional es $\mu = \Sigma x/N$. Sencillamente, cambie los símbolos de la media de la muestra; es decir, $\bar{x} = \Sigma x/N$. Por desgracia, la conversión de una varianza poblacional en una varianza muestral no es tan directa. Requiere un cambio en el denominador. En lugar de sustituir n (el número de la muestra) por N (el número de la población), el denominador es $n - 1$. Así, la fórmula de la **varianza muestral** es:

VARIANZA MUESTRAL

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \qquad [3.9]$$

donde:

- s^2 es la varianza muestral;
- x es el valor de cada observación de la muestra;
- $\bar{x}$ es la media de la muestra;
- n es el número de observaciones realizadas.

¿Por qué se hizo este cambio en el denominador? Aunque el empleo de n se entiende en virtud del uso de $\bar{x}$ para calcular μ, esto tiende a subestimar la varianza poblacional, σ^2. La inclusión de $(n - 1)$ en el denominador proporciona la corrección adecuada para esta tendencia. Como la aplicación fundamental de estadísticos muestrales como s^2 es calcular parámetros de población como σ^2, se prefiere $(n - 1)$ en lugar de n para definir la varianza muestral. También se emplea esta convención al calcular la desviación estándar de una muestra.

EJEMPLO

Los salarios por hora de una muestra de empleados de medio tiempo de Home Depot son: 12, 20, 16, 18 y 19 dólares. ¿Cuál es la varianza de la muestra?

SOLUCIÓN

La varianza de la muestra se calcula con la fórmula [3.9].

$$\bar{x} = \frac{\Sigma x}{n} = \frac{\$85}{5} = \$17$$

Salario por hora (x)	$x - \bar{x}$	$(x - \bar{x})^2$
\$12	−\$5	25
20	3	9
16	−1	1
18	1	1
19	2	4
\$85	0	40

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} = \frac{40}{5 - 1}$$

$$= 10 \text{ en dólares al cuadrado.}$$

La desviación estándar de la muestra se utiliza para estimar la desviación estándar de la población. Como se hizo notar, la desviación estándar de la población es la raíz cuadrada de la varianza de la población. Asimismo, la *desviación estándar de la muestra es la raíz cuadrada de la varianza de la muestra.* La desviación estándar de la muestra se calcula con mayor facilidad de la siguiente manera:

DESVIACIÓN ESTÁNDAR DE LA MUESTRA

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \qquad [3.10]$$

EJEMPLO

La varianza de la muestra en el ejemplo anterior, que incluye salarios por hora, se calculó en 10. ¿Cuál es la desviación estándar?

SOLUCIÓN

La desviación estándar de la muestra es 3.16 dólares, que se determina con $\sqrt{10}$. Observe nuevamente que la varianza de la muestra se expresa en términos de dólares al cuadrado, pero al extraer la raíz cuadrada a 10 se obtiene $3.16, que se encuentra en las mismas unidades (dólares) que los datos originales.

Solución con software

En el ejemplo de la sección "Solución con software" se utilizó Excel para determinar la media y la mediana de los datos de Applewood Auto Group y para presentar la desviación estándar de la muestra. Como la mayoría de los paquetes de software de estadística, Excel supone que los datos corresponden a una muestra.

APPLEWOOD AUTO GROUP

	A	B	C	D	E	F	G	H
1	Age	Profit	Location	Vehicle-Type	Previous		Profit	
2	21	$1,387	Tionesta	Sedan	0			
3	23	$1,754	Sheffield	SUV	1		Mean	1843.17
4	24	$1,817	Sheffield	Hybrid	1		Standard Error	47.97
5	25	$1,040	Sheffield	Compact	0		Median	1882.50
6	26	$1,273	Kane	Sedan	1		Mode	1915.00
7	27	$1,529	Sheffield	Sedan	1		Standard Deviation	643.63
8	27	$3,082	Kane	Truck	0		Sample Variance	414256.61
9	28	$1,951	Kane	SUV	1		Kurtosis	-0.22
10	28	$2,692	Tionesta	Compact	0		Skewness	-0.24
11	29	$1,342	Kane	Sedan	2		Range	2998
12	29	$1,206	Sheffield	Sedan	0		Minimum	294
13	30	$443	Kane	Sedan	3		Maximum	3292
14	30	$1,621	Sheffield	Truck	1		Sum	331770
15	30	$754	Olean	Sedan	2		Count	180

AUTOEVALUACIÓN 3-8

Los años de servicio de una muestra de siete empleados en la oficina de quejas de State Farm Insurance, de Cleveland, Ohio, son: 4, 2, 5, 4, 5, 2 y 6. ¿Cuál es la varianza de la muestra? Calcule la desviación estándar de la muestra.

EJERCICIOS

En los ejercicios 47 a 52, efectúe lo siguiente:

a. Calcule la varianza de la muestra.
b. Determine la desviación estándar de la muestra.

47. Considere los siguientes valores como una muestra: 7, 2, 6, 2 y 3.
48. Los siguientes cinco valores son una muestra: 11, 6, 10, 6 y 7.
49. Dave's Automatic Door, que se mencionó en el ejercicio 37, instala puertas automáticas para cocheras. Con base en una muestra, los siguientes son los tiempos, en minutos, que se requieren para instalar 10 puertas automáticas: 28, 32, 24, 46, 44, 40, 54, 38, 32 y 42.
50. A la muestra de ocho compañías en la industria aeronáutica (ejercicio 38), se le aplicó una encuesta referente a su recuperación de inversión del año anterior. Los resultados son los siguientes: 10.6, 12.6, 14.8, 18.2, 12.0, 14.8, 12.2 y 15.6.

51. La Asociación de Propietarios de Moteles de Houston, Texas, llevó a cabo una encuesta relativa a las tarifas de los moteles entre semana en el área. Enseguida aparece la tarifa por cuarto para huéspedes de negocios en una muestra de 10 moteles.

\$101	\$97	\$103	\$110	\$78	\$87	\$101	\$80	\$106	\$88

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

52. Una organización de protección al consumidor se ocupa de las deudas de las tarjetas de crédito. Una encuesta entre 10 adultos jóvenes con una deuda en su la tarjeta de crédito mayor a 2 000 dólares mostró que estos pagan en promedio un poco más de 100 dólares mensuales como abono a sus saldos. En la siguiente lista aparecen las sumas que cada adulto joven pagó el mes anterior.

\$101	\$97	\$103	\$110	\$78	\$87	\$101	\$80	\$106	\$88

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Interpretación y usos de la desviación estándar

La desviación estándar normalmente se utiliza como medida para comparar la dispersión de dos o más conjuntos de observaciones. Por ejemplo, se calcula que la desviación estándar de las sumas quincenales invertidas en el plan de reparto de utilidades de Dupree Saint Company es de 7.51 dólares. Suponga que estos empleados se ubican en Georgia. Si la desviación estándar de un grupo de empleados en Texas es de 10.47 dólares y las medias son casi las mismas, esto indica que las sumas invertidas por los empleados de Georgia no se encuentran tan dispersas como las de los empleados en Texas (ya que $\$7.51 < \$10.47$). Como las sumas invertidas por los empleados de Georgia se acumulan más cerca de la media, su media es una medida más confiable que la del grupo de Texas.

Teorema de Chebyshev

OA3-5 Explicar y aplicar el teorema de Chebyshev y la regla empírica.

Ya se ha insistido en el hecho de que una desviación estándar pequeña de un conjunto de valores indica que estos se localizan cerca de la media. Por lo contrario, una desviación grande revela que las observaciones se encuentran muy dispersas con respecto a la media. El matemático ruso P. L. Chebyshev (1821-1894) estableció un teorema que permite determinar la mínima porción de valores que se encuentran a cierta cantidad de desviaciones estándares de la media. Por ejemplo, de acuerdo con el **teorema de Chebyshev,** por lo menos tres de cuatro valores, o 75%, deben encontrarse entre la media más dos desviaciones estándares y la media menos dos desviaciones estándares. Esta relación se cumple independientemente de la forma de la distribución. Además, por lo menos ocho de los nueve valores, 88.9%, se encontrarán a más de tres desviaciones estándares y menos tres desviaciones estándares de la media. Por lo menos 24 de 25 valores, o 96%, se encontrará entre más y menos cinco desviaciones estándares de la media.

El teorema de Chebyshev establece lo siguiente:

TEOREMA DE CHEBYSHEV En cualquier conjunto de observaciones (muestra o población), la proporción de valores que se encuentran a k desviaciones estándares de la media es, por lo menos, de $1 - 1/k^2$, siendo k cualquier constante mayor que 1.

ESTADÍSTICA EN ACCIÓN

La mayoría de las universidades informan el *tamaño promedio de los grupos*. Esta información puede inducir a error porque el tamaño promedio se determina de diversas formas. Si calcula la cantidad de estudiantes en *cada clase* en cierta universidad, el resultado es la cantidad promedio de estudiantes por clase. Si recaba una lista de tamaños de grupos y calcula el tamaño de grupo promedio, podría hallar que la media es muy diferente. Una escuela descubrió que el promedio de estudiantes en cada una de sus 747 clases era de 40.

(continúa)

EJEMPLO

La media aritmética de la suma quincenal que aportan los empleados de Dupree Saint al plan de reparto de utilidades de la compañía es de 51.54 dólares y la desviación estándar, de 7.51 dólares. ¿Por lo menos qué porcentaje de las aportaciones se encuentra a más de 3.5 desviaciones estándares y a menos 3.5 desviaciones de la media?

SOLUCIÓN

Alrededor de 92%, que se determina de la siguiente manera:

$$1 = \frac{1}{k^2} = 1 - \frac{1}{(3.5)^2} = 1 - \frac{1}{12.25} = 0.92$$

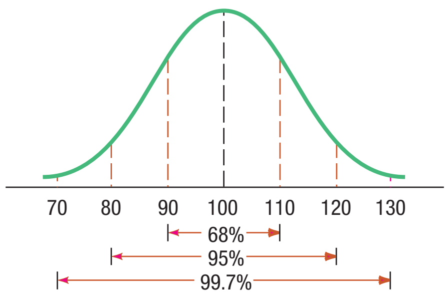
(*continuación*)

Pero cuando calculó la media a partir de una lista de tamaños de grupo, esta resultó ser de 147. ¿Por qué la discrepancia? Hay menos estudiantes en los grupos pequeños y una gran cantidad de estudiantes en los grupos grandes, lo cual tiene el efecto de incrementar el tamaño promedio de los grupos cuando se calcula de esta manera. Una universidad podría reducir su tamaño promedio de grupo si reduce el número de estudiantes en cada grupo. Esto significa eliminar las cátedras en las que hay muchos estudiantes de primer grado.

La regla empírica

El teorema de Chebyshev se relaciona con cualquier conjunto de valores; es decir, la distribución de valores puede tener cierta forma. Sin embargo, en cualquier distribución simétrica con forma de campana, como se muestra en la gráfica 3.7, es posible ser más precisos al explicar la dispersión en torno a la media. Estas relaciones implican la desviación estándar y la media, y se encuentran descritas en la **regla empírica,** a veces denominada **regla normal.**

REGLA EMPÍRICA En cualquier distribución de frecuencias simétrica con forma de campana, aproximadamente 68% de las observaciones se encontrarán entre más y menos una desviación estándar de la media; cerca de 95% de las observaciones se encontrarán entre más y menos dos desviaciones estándares de la media y, de hecho, todas (99.7%) estarán entre más y menos tres desviaciones estándares de la media.



GRÁFICA 3.7 Curva simétrica con forma de campana que muestra las relaciones entre la desviación estándar y las observaciones

Estas relaciones se representan en la gráfica 3.7 en el caso de una distribución con forma de campana con una media de 100 y una desviación estándar de 10.

Aplicando la regla empírica, si una distribución es simétrica y tiene forma de campana, todas las observaciones se encuentran entre la media más y menos tres desviaciones estándares. Por consiguiente, si $\bar{x} = 100$ y $s = 10$, todas las observaciones se encuentran entre $100 + 3(10)$ y $100 - 3(10)$, o 70 y 130. Por tanto, el rango es de 60, que se calcula restando $130 - 70$.

Por el contrario, si sabe que el rango es de 60, puede aproximar la desviación estándar dividiendo el rango entre 6. En este caso: rango $\div 6 = 60 \div 6 = 10$; es decir, la desviación estándar.

EJEMPLO

Una muestra de tarifas de renta de los departamentos de University Park se asemeja a una distribución simétrica con forma de campana. La media de la muestra es de 500 dólares; la desviación estándar, de 20. De acuerdo con la regla empírica conteste las siguientes preguntas:

1. ¿Entre qué dos cantidades se encuentra aproximadamente 68% de los gastos mensuales en alimentos?
2. ¿Entre qué dos cantidades se encuentra alrededor de 95% de los gastos mensuales en alimentos?
3. ¿Entre qué dos cantidades se encuentran casi todos los gastos mensuales en alimentos?

SOLUCIÓN

1. Cerca de 68% se encuentra entre 480 y 520 dólares, calculado de la siguiente manera: $\bar{x} \pm 1s = \$500 \pm 1(\$20)$.
2. Aproximadamente 95% se encuentra entre 460 y 540 dólares, calculado de la siguiente manera: $\bar{x} \pm 2s = \$500 \pm 2(\$20)$.
3. Casi todas (99.7%) se encuentra entre 440 y 560 dólares, calculado de la siguiente manera: $\bar{x} \pm 3s = \$500 \pm 3(\$20)$.

AUTOEVALUACIÓN 3-9

Pitney Pipe Company es un fabricante estadounidense de tubos de PVC. El departamento de control de calidad tomó una muestra de 600 tubos de 10 pies de longitud. A una distancia de un pie del extremo del tubo, se midió el diámetro externo. La media fue de 14.0 pulgadas y la desviación estándar, de 0.1 pulgadas.

(a) Si no conoce la forma de la distribución, ¿por lo menos qué porcentaje de las observaciones se encontrará entre 13.85 y 14.15 pulgadas?

(b) Si supone que la distribución de los diámetros es simétrica y tiene forma de campana, ¿entre qué dos valores se encontrará aproximadamente 95% de las observaciones?

EJERCICIOS

53. De acuerdo con el teorema de Chebyshev, ¿por lo menos qué porcentaje de cualquier conjunto de observaciones se encontrará a 1.8 desviaciones estándares de la media?

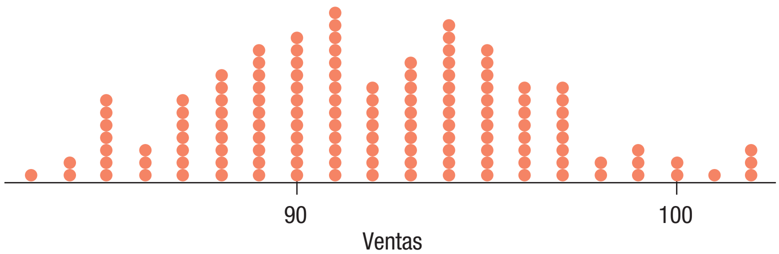
54. El ingreso medio de un grupo de observaciones de una muestra es de 500 dólares; la desviación estándar es de 40 dólares. De acuerdo con el teorema de Chebyshev, ¿por lo menos qué porcentaje de ingresos se encontrará entre 400 y 600 dólares?

55. La distribución de pesos de una muestra de 1 400 contenedores de carga es simétrica y tiene forma de campana. Considere la regla empírica para determinar qué porcentaje de pesos se encontrará entre:

a. $\bar{x} - 2s$ y $\bar{x} + 2s$;

b. $\bar{x}$ y $\bar{x} + 2s$; debajo de $\bar{x} - 2s$.

56. La siguiente gráfica representa la distribución del número de refrescos tamaño gigante que vendió un restaurante Wendy's los últimos 141 días. La cantidad promedio de refrescos vendidos por día es de 91.9 y la desviación estándar, de 4.67.



Si utiliza la regla empírica, ¿entre cuáles dos valores de 68% de los días se encontrarán las ventas? ¿Entre cuáles dos valores de 95% de los días se encontrarán las ventas?

Media y desviación estándar de datos agrupados

OA3-6 Calcular la media y la desviación estándar de datos agrupados.

La mayoría de las medidas de ubicación, como la media, y las medidas de dispersión, como la desviación estándar, se determinan utilizando valores individuales. Los paquetes de software de estadística facilitan el cálculo de estos valores, incluso en el caso de conjuntos grandes de datos. Sin embargo, algunas veces solo se cuenta con la distribución de frecuencias y se desea calcular la media o la desviación estándar. A continuación se le explicará cómo calcular la media y la desviación estándar a partir de datos organizados en una distribución de frecuencias. Hay que insistir en que una media o una desviación estándar de datos agrupados es una *estimación* de los valores reales correspondientes.

Media aritmética de datos agrupados

Para aproximar la media aritmética de datos organizados en una distribución de frecuencia, comience suponiendo que las observaciones en cada clase se representan a través del *punto medio* de la clase. La media de una muestra de datos organizados en una distribución de frecuencias se calcula de la siguiente manera:

MEDIA ARITMÉTICA DE DATOS AGRUPADOS

$$\bar{x} = \frac{\Sigma fM}{n} \qquad \textbf{[3.11]}$$

donde:

- $\bar{x}$ designa la media muestral;
- M es el punto medio de cada clase;
- f es la frecuencia en cada clase;
- fM es la frecuencia en cada clase multiplicada por el punto medio de la clase;
- ΣfM es la suma de estos productos;
- n es el número total de frecuencias.

ESTADÍSTICA EN ACCIÓN

Buster Posey, de los Gigantes de San Francisco, ostentó el máximo promedio de bateo (0.336) durante la temporada 2012 de las Ligas Mayores de Béisbol. Tony Gwynn bateó 0.394 en la temporada 1994, en la que hubo pocos strikes, y Ted Williams bateó 0.406 en 1941. Nadie ha bateado arriba de 0.400 desde 1941. El promedio de bateo se ha mantenido constante alrededor de 0.260 durante más de 100 años, pero la desviación estándar se redujo de 0.049 a 0.031. Esto indica que hay menos dispersión en el

(*continúa*)

(*continuación*)
promedio de bateo de hoy y permite explicar que no haya bateadores con promedio de 0.400 recientemente.

EJEMPLO

Los cálculos de la media aritmética de datos agrupados en una distribución de frecuencias que aparecen enseguida se basan en los datos de las ganancias de Applewood Auto Group. Recuerde que en el capítulo 2, tabla 2.7, página 27, construyó una distribución de frecuencias de precios de venta de vehículos. La información se repite abajo. Determine la ganancia media aritmética por vehículo.

Ganancia	Frecuencia
$ 200 hasta $ 600	8
600 hasta 1 000	11
1 000 hasta 1 400	23
1 400 hasta 1 800	38
1 800 hasta 2 200	45
2 200 hasta 2 600	32
2 600 hasta 3 000	19
3 000 hasta 3 400	4
Total	180

SOLUCIÓN

La ganancia media de los vehículos se calcula a partir de datos agrupados en una distribución de frecuencias. Para calcular la media, suponga que el punto medio de cada clase es representativo de los valores incluidos en dicha clase. Recuerde que el punto medio de una clase se encuentra a la mitad de los límites de dos clases consecutivas. Para determinar el punto medio de una clase en particular, sume los límites de clase superior e inferior y divida entre 2. Por consiguiente, el punto medio de la primera clase es 400 dólares, que se calcula con la operación ($200 + $600)/2. Suponga que el valor de 400 dólares es representativo de los ocho valores incluidos en dicha clase. En otras palabras, se asume que la suma de los ocho valores en esta clase es de 3 200 dólares, que se calcula por medio del producto 8($400). Continúe con el proceso de multiplicación del punto medio de clase por la frecuencia de clase de cada clase y enseguida sume estos productos. Los resultados se resumen en la tabla 3.1.

TABLA 3.1 Ganancia sobre los 180 vehículos que se vendieron el mes anterior en Applewood Auto Group

Ganancia	Frecuencia (f)	Punto medio (M)	fM
$ 200 hasta $ 600	8	$ 400	$ 3 200
600 hasta 1 000	11	800	8 800
1 000 hasta 1 400	23	1 200	27 600
1 400 hasta 1 800	38	1 600	60 800
1 800 hasta 2 200	45	2 000	90 000
2 200 hasta 2 600	32	2 400	76 800
2 600 hasta 3 000	19	2 800	53 200
3 000 hasta 3 400	4	3 200	12 800
Total	180		$333 200

Al despejar la media aritmética de la fórmula [3.11] se obtiene:

$$\bar{x} = \frac{\Sigma fM}{n} = \frac{\$333\ 200}{180} = \$1\ 851.11$$

Así, se concluye que la ganancia media por vehículo es de aproximadamente 1 851 dólares.

Desviación estándar de datos agrupados

Para calcular la desviación estándar de datos agrupados en una distribución de frecuencias es necesario ajustar ligeramente la fórmula [3.10]. Pondere cada una de las diferencias cuadradas por el número de frecuencias en cada clase. La fórmula es:

DESVIACIÓN ESTÁNDAR, DATOS AGRUPADOS $$s = \sqrt{\frac{\Sigma f(M - \bar{x})^2}{n - 1}}$$ **[3.12]**

donde:

- s es el símbolo de la desviación estándar de la muestra;
- M es el punto medio de la clase;
- f es la frecuencia de clase;
- n es el número de observaciones en la muestra;
- $\bar{x}$ designa la media muestral.

EJEMPLO

Consulte la distribución de frecuencias de los datos de la ganancia de Applewood Auto Group que se muestran en la tabla 3.1, página 72. Calcule la desviación estándar de las ganancias que generó cada vehículo.

SOLUCIÓN

De acuerdo con la misma técnica que se empleó para calcular la media de los datos agrupados en una distribución de frecuencias, f es la frecuencia de clase, M es el punto medio de clase y n es el número de observaciones.

Ganancia	Frecuencia (f)	Punto medio (M)	fM	$(M - \bar{x})$	$(M - \bar{x})^2$	$f(M - \bar{x})^2$
\$ 200 hasta \$ 600	8	400	3 200	−1 451	2 105 401	16 843 208
600 hasta 1 000	11	800	8 800	−1 051	1 104 601	12 150 611
1 000 hasta 1 400	23	1 200	27 600	−651	423 801	9 747 423
1 400 hasta 1 800	38	1 600	60 800	−251	63 001	2 394 038
1 800 hasta 2 200	45	2 000	90 000	149	22 201	999 045
2 200 hasta 2 600	32	2 400	76 800	549	301 401	9 644 832
2 600 hasta 3 000	19	2 800	53 200	949	900 601	17 111 419
3 000 hasta 3 400	4	3 200	12 800	1 349	1 819 801	7 279 204
Total	180		333 200			76 169 780

Para determinar la desviación estándar:

Paso 1: reste la media del punto medio de clase. Es decir, encuentre $(M - \bar{x})$. Para la primera clase (\$400 − \$1 851 = −\$1 451); para la segunda (\$800 − \$1 851 = −\$1 051), y así en lo sucesivo.

Paso 2: eleve al cuadrado la diferencia entre el punto medio de clase y la media. En el caso de la primera clase sería $(\$400 - \$1\,851)^2 = 2\,105\,401$; en el de la segunda $(\$800 - \$1\,851)^2 = 1\,104\,601$, y así en lo sucesivo.

Paso 3: multiplique la diferencia al cuadrado entre el punto medio de clase y la media por la frecuencia de clase. En el caso de la primera clase el valor es $8(\$400 - \$1\,851)^2 = 16\,843\,208$; en el de la segunda $11(\$800 - \$1\,851)^2 = 12\,150\,611$, y así sucesivamente.

Paso 4: sume $f(M - \bar{x})^2$. El total es 76 169 920. Para determinar la desviación estándar, inserte estos valores en la fórmula [3.12].

$$s = \sqrt{\frac{\Sigma f(M - \bar{x})^2}{n - 1}} = \sqrt{\frac{76\,169\,780}{180 - 1}} = 652.33$$

Por lo general, la media y la desviación estándar que se calculan a partir de datos agrupados en una distribución de frecuencias se encuentran cerca de los valores calculados a partir de los datos en bruto. Los datos agrupados originan la pérdida de alguna información. En el ejemplo de la ganancia por vehículo, la ganancia media que aparece en la hoja de Excel del ejemplo de la sección “Solución con software” es de 1 843.17 dólares; y la desviación estándar, de 643.63 dólares. Los valores respectivos calculados a partir de datos agrupados en una distribución de frecuencias son 1 851.11 y 652.33 dólares. La diferencia entre las medias es de 7.94 dólares o, aproximadamente, 0.4%. Las desviaciones estándares difieren en 8.70 dólares o 1.4%. Con base en la diferencia porcentual, las aproximaciones se acercan mucho a los valores reales.

AUTOEVALUACIÓN 3-10

Los ingresos netos de una muestra de grandes importadores de antigüedades se organizaron en la siguiente tabla:

Ingreso neto (millones de dólares)	Número de importadores
2 hasta 6	1
6 hasta 10	4
10 hasta 14	10
14 hasta 18	3
18 hasta 22	2

(a) ¿Qué nombre recibe la tabla?
(b) Con base en la distribución, ¿cuál es el cálculo aproximado del ingreso neto medio aritmético?
(c) Con base en la distribución, ¿cuál es el cálculo aproximado de la desviación estándar?

EJERCICIOS

57. Al calcular la media de una distribución de frecuencia, ¿por qué se hace referencia a esta como una media *aproximada?*

58. Determine la media y la desviación estándar de la siguiente distribución de frecuencias.

Clase	Frecuencia
0 hasta 5	2
5 hasta 10	7
10 hasta 15	12
15 hasta 20	6
20 hasta 25	3

59. Determine la media y la desviación estándar de la siguiente distribución de frecuencias.

Clase	Frecuencia
20 hasta 30	7
30 hasta 40	12
40 hasta 50	21
50 hasta 60	18
60 hasta 70	12

60. SCCoast, un proveedor de internet del sureste de Estados Unidos, elaboró una distribución de frecuencias sobre la edad de los usuarios de internet. Determine la media y la desviación estándar.

Edad (años)	Frecuencia
10 hasta 20	3
20 hasta 30	7
30 hasta 40	18
40 hasta 50	20
50 hasta 60	12

61. El IRS (Internal Revenue Service) estaba interesado en el número de formas fiscales individuales que preparan las pequeñas empresas de contabilidad. El IRS tomó una muestra aleatoria de 50 empresas de contabilidad pública con 10 o más empleados que operan en la zona de Dallas-Fort Worth. En la siguiente tabla de frecuencias se muestran los resultados del estudio. Calcule la media y la desviación estándar.

Cantidad de clientes	Frecuencia
20 hasta 30	1
30 hasta 40	15
40 hasta 50	22
50 hasta 60	8
60 hasta 70	4

62. Los gastos en publicidad constituyen un elemento significativo del costo de los artículos vendidos. Enseguida aparece una distribución de frecuencias que muestra los gastos en publicidad de 60 compañías manufactureras ubicadas en el suroeste de Estados Unidos. Calcule la media y la desviación estándar de los gastos en publicidad.

Gastos en publicidad (millones de dólares)	Número de compañías
25 hasta 35	5
35 hasta 45	10
45 hasta 55	21
55 hasta 65	16
65 hasta 75	8
Total	60

Ética e informe de resultados

En el capítulo 1 se analizó la manera de informar resultados estadísticos con ética e imparcialidad. Aunque usted está aprendiendo a organizar, resumir e interpretar datos mediante la estadística, también es importante que comprenda esta disciplina para que se convierta en un consumidor inteligente de información.

En este capítulo se demostró la forma de calcular estadísticas descriptivas de naturaleza numérica. En particular, la manera de calcular e interpretar medidas de ubicación de un conjunto de datos: la media, la mediana y la moda. También se estudiaron las ventajas y desventajas de cada estadístico. Por ejemplo, si un agente de bienes raíces le dice a un cliente que la casa promedio de determinada parcela se vendió en 150 000 dólares, quizá suponga que 150 000 dólares es un precio de venta representativo de todas las casas. Pero si el cliente pregunta, además, cuál es la mediana del precio de venta y resulta ser 60 000 dólares, ¿por qué el agente informó solo el precio promedio? Esta información es de suma importancia cuando una persona toma la decisión de comprar una casa. Conocer las ventajas y desventajas de la media, la mediana y la moda es importante al dar un informe estadístico y cuando se emplea información estadística para tomar decisiones.

También se expuso cómo calcular medidas de dispersión: el rango, la desviación media y la desviación estándar. Cada uno de estos estadísticos tiene ventajas y desventajas. Recuerde que el rango proporciona información sobre la dispersión total de una distribución. Sin embargo, no aporta información acerca de la forma en que se acumulan los datos o se concentran en torno al centro de la distribución. Conforme aprenda más estadística, necesitará recordar que cuando emplee esta disciplina deberá mantener un punto de vista independiente y basado en principios. Cualquier informe estadístico requiere la comunicación honesta y objetiva de los resultados.

RESUMEN DEL CAPÍTULO

I. Una medida de ubicación es un valor que sirve para describir el centro de un conjunto de datos.
   A. La media aritmética es la medida de ubicación que más se informa.
      1. Se calcula mediante la suma de los valores de las observaciones, que luego se divide entre el número total de observaciones.
         a. La fórmula de una media poblacional de datos no agrupados o en bruto es:

$$\mu = \frac{\Sigma x}{N} \qquad \textbf{[3.1]}$$

         b. La fórmula de la media muestral es

$$\bar{x} = \frac{\Sigma x}{n} \qquad \textbf{[3.2]}$$

         c. La fórmula de la media muestral en una distribución de frecuencias es

$$\bar{x} = \frac{\Sigma fM}{n} \qquad \textbf{[3.11]}$$

2. Las principales características de la media aritmética son:
   a. Por lo menos se requiere la escala de medición de intervalo.
   b. Todos los valores de los datos se incluyen en el cálculo.
   c. Un conjunto de datos solo posee una media. Es decir, es única.
   d. La suma de las desviaciones de la media es igual a 0.

B. La mediana es el valor que se encuentra al centro de un conjunto de datos ordenados.

1. Para determinar la mediana se ordenan las observaciones de menor a mayor y se identifica el valor intermedio.
2. Las principales características de la mediana son:
   a. Se requiere al menos la escala ordinal de medición.
   b. No influyen sobre esta valores extremos.
   c. Cincuenta por ciento de las observaciones son más grandes que la mediana.
   d. Es única de un conjunto de datos.

C. La moda es el valor que se presenta con mayor frecuencia en un conjunto de datos.

1. La moda se determina en el caso de datos de nivel nominal.
2. Un conjunto de datos puede tener más de una moda.

D. La media ponderada se encuentra al multiplicar cada observación por su correspondiente ponderación.

1. La fórmula para determinar la media ponderada es:

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_nx_n}{w_1 + w_2 + w_3 + \cdots + w_n} \quad \textbf{[3.3]}$$

E. La media geométrica es la enésima raíz del producto de *n* valores positivos.

1. La fórmula de la media geométrica es:

$$GM = \sqrt[n]{(x_1)(x_2)(x_3)\cdots(x_n)} \quad \textbf{[3.4]}$$

2. La media geométrica también se emplea para determinar la razón de cambio de un periodo a otro.

$$GM = \sqrt[n]{\frac{\text{Valor al final del periodo}}{\text{Valor al principio del periodo}}} - 1 \quad \textbf{[3.5]}$$

3. La media geométrica siempre es igual o menor que la media aritmética.

II. La dispersión es la variación o propagación en un conjunto de datos.

A. El rango es la diferencia entre el valor máximo y el mínimo en un conjunto de datos.

1. La fórmula del rango es la siguiente:

$$\text{Rango} = \text{valor máximo} - \text{valor mínimo} \quad \textbf{[3.6]}$$

2. Las principales características del rango son:
   a. Solo se emplean dos valores en su cálculo.
   b. Recibe la influencia de los valores extremos.
   c. Es fácil de calcular y definir.

B. La varianza es la media de las desviaciones al cuadrado de la media aritmética.

1. La fórmula de la varianza de la población es:

$$\sigma^2 = \frac{\Sigma(x - \mu)^2}{N} \quad \textbf{[3.7]}$$

2. La fórmula de la varianza muestral es:

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1} \quad \textbf{[3.9]}$$

3. Las principales características de la varianza son:
   a. Todas las observaciones se utilizan para realizar el cálculo.
   b. Resulta de alguna manera difícil trabajar con las unidades, pues son las unidades originales elevadas al cuadrado.

C. La desviación estándar es la raíz cuadrada de la varianza.

1. Las principales características de la desviación estándar son:
   a. Se expresa en las mismas unidades de los datos originales.
   b. Es la raíz cuadrada de la distancia promedio al cuadrado de la media.
   c. No puede ser negativa.
   d. Es la medida de dispersión que se informa con más frecuencia.

2. La fórmula de la desviación estándar de la muestra es:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} \qquad \textbf{[3.10]}$$

3. La fórmula de la desviación estándar para datos agrupados es:

$$s = \sqrt{\frac{\Sigma f(M - \bar{x})^2}{n - 1}} \qquad \textbf{[3.12]}$$

III. La desviación estándar se utiliza para describir la distribución de frecuencias aplicando el teorema de Chebyshev o la regla empírica.

A. El teorema de Chebyshev establece que, independientemente de la forma de la distribución, por lo menos $1 - 1/k^2$ de las observaciones se encontrarán a k desviaciones estándares de la media, siendo k mayor que 1.

B. La regla empírica afirma que en el caso de una distribución en forma de campana, alrededor de 68% de los valores se encontrarán a una desviación estándar de la media; 95%, a dos y casi todas, a tres.

CLAVE DE PRONUNCIACIÓN

SÍMBOLO	SIGNIFICADO	PRONUNCIACIÓN
μ	Media de población	*mu*
Σ	Operación de suma	*sigma*
Σx	Suma de un grupo de valores	*sigma x*
$\bar{x}$	Media de la muestra	*x barra*
$\bar{x}_w$	Media ponderada	*x barra subíndice w*
GM	Media geométrica	*GM*
ΣfM	Suma del producto de las frecuencias y los puntos medios de clase	*sigma f M*
σ^2	Varianza de la población	*sigma al cuadrado*
σ	Desviación estándar de la población	*sigma*

EJERCICIOS DEL CAPÍTULO

63. La empresa de contabilidad Crawford and Associates está formada por cinco socios. El día anterior, estos atendieron a 6, 4, 7 y 5 clientes, respectivamente.

a. Calcule la media y la mediana de la cantidad de clientes que cada socio atendió.

b. La media, ¿es muestral o poblacional?

c. Verifique que $\Sigma(x - \mu) = 0$.

64. Owens Orchards vende manzanas por peso en bolsas grandes. Una muestra de siete bolsas contenía las siguientes cantidades de manzanas: 23, 19, 26, 17, 21, 24 y 22.

a. Calcule la media y la mediana de las manzanas que hay en una bolsa.

b. Verifique que $\Sigma(x - \bar{x}) = 0$.

65. Una muestra de familias que ha contratado los servicios de la United Bell Phone Company reveló que cada familia recibió la siguiente cantidad de llamadas la semana pasada. Determine la media y la mediana de las llamadas que recibieron.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

52	43	30	38	30	42	12	46	39	37
34	46	32	18	41	5				

66. La Citizens Banking Company estudia la cantidad de veces que se utiliza al día el cajero automático ubicado en uno de los supermercados de Loblaws, en Market Street; enseguida figuran las cantidades de ocasiones que se utilizó al día durante los últimos 30 días; determine la media de la cantidad de veces que este se utilizó al día.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

83	64	84	76	84	54	75	59	70	61
63	80	84	73	68	52	65	90	52	77
95	36	78	61	59	84	95	47	87	60

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

67. Un estudio reciente sobre los hábitos de lavado de ropa de los estadounidenses incluyó el tiempo en minutos del ciclo de lavado. A continuación hay una muestra de 40 observaciones. Determine la media y la mediana de un ciclo de lavado típico.

35	37	28	37	33	38	37	32	28	29
39	33	32	37	33	35	36	44	36	34
40	38	46	39	37	39	34	39	31	33
37	35	39	38	37	32	43	31	31	35

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

68. Trudy Green trabaja en la True-Green Lawn Company. Su tarea consiste en ofrecer mantenimiento de césped vía telefónica. Enseguida aparece una lista de la cantidad de citas por hora que hizo durante las últimas 25 horas de llamadas. ¿Cuál es la media aritmética de citas que hace por hora? ¿Cuál es la mediana de la cantidad de citas que hace por hora? Redacte un breve informe que resuma sus conclusiones.

9	5	2	6	5	6	4	4	7	2	3	6	3
4	4	7	8	4	4	5	5	4	8	3	3	

69. La Split-A-Rail Fence Company vende tres tipos de cerca a propietarios de los suburbios de Seattle, Washington. El precio por pie e instalación de las cercas grado A es de 5.00 dólares; el de las cercas grado B, de 6.50 dólares, y el de las de grado C (las de alta calidad), de 8.00 dólares. Ayer, Split-A-Rail instaló 270 pies de cerca grado A, 300 pies de cerca grado B y 100 pies de cerca grado C. ¿Cuál es la media del costo por pie de cerca instalada?

70. Rolland Poust es un estudiante de primer grado de la Facultad de Administración del Scandia Tech. El semestre anterior tomó dos cursos de estadística y contabilidad de tres horas cada uno y obtuvo A en ambos, mientras que recibió una B en un curso de historia de cinco horas y B en un curso de historia del jazz de dos horas. Además, tomó un curso de una hora relativo a las reglas de basquetbol con el fin de obtener su licencia para arbitrar partidos de este deporte en escuelas secundarias en el cual obtuvo una A. ¿Cuál fue su promedio semestral? Suponga que le dan 4 puntos por una A; 3 por una B y así sucesivamente. ¿Qué medida de ubicación calculó?

71. La siguiente tabla muestra el porcentaje de fuerza laboral desempleada y el tamaño de la fuerza laboral en tres condados del noroeste de Ohio. Jon Elsas es director regional de desarrollo económico. Debe presentar un informe a varias compañías que piensan ubicarse en el noroeste del estado. ¿Cuál sería el índice de desempleo adecuado que debe reportar para toda la región?

Condado	Porcentaje de desempleo	Tamaño de la fuerza laboral
Wood	4.5	15 300
Ottawa	3.0	10 400
Lucas	10.2	150 600

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

72. La Asociación Americana de Diabetes recomienda una lectura de valores de glucosa sanguínea menor a 130 para quienes tienen diabetes tipo 2. La glucosa sanguínea mide la cantidad de azúcar en la sangre. A continuación se presentan las lecturas de febrero de una persona que fue recientemente diagnosticada con este tipo de diabetes.

112	122	116	103	112	96	115	98	106	111
106	124	116	127	116	108	112	112	121	115
124	116	107	118	123	109	109	106		

a. ¿Cuál es la media aritmética de la lectura de glucosa sanguínea?
b. ¿Cuál es la mediana de la lectura de glucosa sanguínea?
c. ¿Cuál es la moda de la lectura de glucosa sanguínea?

73. Se esperaba que el área metropolitana de Los Angeles-Long Beach, California, mostrara el mayor incremento del número de puestos de trabajo de 1989 a 2010. Se pensaba que el número de trabajos se incrementaría de 5 164 900 a 6 286 800. ¿Cuál es la media geométrica de la tasa de incremento anual esperada?

74. Un artículo reciente sugirió que, si en la actualidad alguien gana 25 000 dólares anuales y la tasa de inflación se mantiene en 3% anual, esa persona necesitaría ganar 33 598 dólares en 10 años para tener el mismo poder adquisitivo, y 44 771 dólares si la tasa de inflación se elevara a 6%. Confirme si estas afirmaciones son exactas determinando la tasa media geométrica de incremento.

75. Las edades de una muestra de turistas canadienses que vuelan de Toronto a Hong Kong fueron las siguientes: 32, 21, 60, 47, 54, 17, 72, 55, 33 y 41 años.
   a. Calcule el rango.
   b. Calcule la desviación estándar.

76. Los pesos (en libras) de una muestra de cinco cajas enviadas por UPS son: 12, 6, 7, 3 y 10.
   a. Calcule el rango.
   b. Calcule la desviación estándar.

77. La siguiente tabla presenta las inscripciones a 13 universidades públicas del estado de Ohio.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Universidad	Inscripciones
University of Akron	26 666
Bowling Green State University	17 298
Central State University	2 152
University of Cincinnati	33 347
Cleveland State University	17 529
Kent State University	27 706
Miami University	16 924
Ohio State University	56 387
Ohio University	25 223
Shawnee State University	4 630
University of Toledo	21 500
Wright State University	16 762
Youngstown State University	13 813

   a. ¿Es una muestra o una población?
   b. ¿Cuál es la media de las inscripciones?
   c. ¿Cuál es la mediana de las inscripciones?
   d. ¿Cuál es el rango de las inscripciones?
   e. Calcule la desviación estándar.

78. Los temas de salud representan una preocupación para los gerentes, en especial cuando deben evaluar el costo del seguro médico. Una encuesta reciente entre 150 ejecutivos de Elvers Industries, una importante empresa financiera y de seguros, ubicada en el suroeste de Estados Unidos, informó la cantidad de libras de sobrepeso de los ejecutivos. Calcule la media y la desviación estándar.

Libras de sobrepeso	Frecuencia
0 hasta 6	14
6 hasta 12	42
12 hasta 18	58
18 hasta 24	28
24 hasta 30	8

79. El programa espacial Apolo duró de 1967 hasta 1972, e incluyó 13 misiones; las cuales tuvieron una duración de entre 7 y 301 horas. Enseguida aparece la duración de cada vuelo.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

9	195	241	301	216	260	7	244	192	147
10	295	142							

   a. Explique por qué los tiempos de vuelo constituyen una población.
   b. Calcule la media y la mediana de los tiempos de vuelo.
   c. Estime el rango y la desviación estándar de los tiempos de vuelo.

80. Creek Ratz es un restaurante muy popular, localizado en la costa del norte de Florida, que sirve una variedad de alimentos con carne de res y mariscos. Durante la temporada de vacaciones de verano no se aceptan reservaciones. La gerencia está interesada en conocer el tiempo que un cliente tiene que esperar antes de pasar a la mesa. A continuación aparece la lista de tiempos de espera, en minutos, de las 25 mesas que se ocuparon la noche del sábado anterior.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

28	39	23	67	37	28	56	40	28	50
51	45	44	65	61	27	24	61	34	44
64	25	24	27	29					

a. Explique por qué los tiempos constituyen una población.
b. Calcule la media y la mediana de los tiempos de espera.
c. Estime el rango y la desviación estándar de los tiempos de espera.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

81. Una muestra de 25 estudiantes universitarios reportó las siguientes cifras en dólares de gastos por concepto de entretenimiento el año anterior.

684	710	688	711	722	698	723	743	738	722	696	721	685
763	681	731	736	771	693	701	737	717	752	710	697	

a. Encuentre la media, la mediana y la moda de esa información.
b. Determine el rango y la desviación estándar.
c. Emplee la regla empírica para establecer un intervalo que incluya aproximadamente 95% de las observaciones.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

82. El Derby de Kentucky se celebra el primer sábado de mayo en Churchill Downs, Louisville, Kentucky. La pista mide una milla y cuarto. En la siguiente tabla se muestran los ganadores desde 1990, su margen de victoria, el tiempo del ganador y las ganancias sobre una apuesta de dos dólares.

Año	Ganador	Margen de ganancia (longitudes)	Tiempo ganador (minutos)	Ganancia sobre una apuesta de dos dólares
1990	Unbridled	3.5	2.03333	10.80
1991	Strike the Gold	1.75	2.05000	4.80
1992	Lil E. Tee	1	2.05000	16.80
1993	Sea Hero	2.5	2.04000	12.90
1994	Go For Gin	2	2.06000	9.10
1995	Thunder Gulch	2.25	2.02000	24.50
1996	Grindstone	nariz	2.01667	5.90
1997	Silver Charm	cabeza	2.04000	4.00
1998	Real Quiet	0.5	2.03667	8.40
1999	Charismatic	cuello	2.05333	31.30
2000	Fusaichi Pegasus	1.5	2.02000	2.30
2001	Monarchos	4.75	1.99950	10.50
2002	War Emblem	4	2.01883	20.50
2003	Funny Cide	1.75	2.01983	12.80
2004	Smarty Jones	2.75	2.06767	4.10
2005	Giacomo	0.5	2.04583	50.30
2006	Barbaro	6.5	2.02267	6.10
2007	Street Sense	2.25	2.03617	4.90
2008	Big Brown	4.75	2.03033	6.80
2009	Mine That Bird	6.75	2.04433	103.20
2010	Super Saver	2.50	2.07417	18.00
2011	Animal Kingdom	2.75	2.034	43.80
2012	I'll Have Another	1.5	2.03050	32.60
2013	Orb	2.5	2.04817	12.80

a. Determine la media y la mediana de las variables tiempo ganador y ganancia sobre apuesta de dos dólares.
b. Determine el rango y la desviación estándar de las variables tiempo ganador y ganancia.
c. Refiérase a la variable "tiempo ganador". ¿Cuál es el nivel de medición? ¿Qué medida de ubicación sería la más adecuada?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

83. El gerente de la tienda Wal-Mart de la localidad estudia la cantidad de artículos que compran los consumidores en el horario de la tarde. A continuación aparece la cantidad de artículos de una muestra de 30 consumidores.

15	8	6	9	9	4	18	10	10	12
12	4	7	8	12	10	10	11	9	13
5	6	11	14	5	6	6	5	13	5

a. Calcule la media y la mediana de la cantidad de artículos.
b. Estime el rango y la desviación estándar de la cantidad de artículos.

c. Organice la cantidad de artículos en una distribución de frecuencias. Quizá desee repasar las instrucciones del capítulo 2 para establecer el intervalo de clase y el número de clases.
d. Calcule la media y la desviación estándar de los datos organizados en una distribución de frecuencias. Compare estos valores con los que calculó en el punto **a.** ¿Por qué son diferentes?

84. La siguiente distribución de frecuencias contiene los costos de electricidad de una muestra de 50 departamentos de dos recámaras en Albuquerque, Nuevo México, durante el mes de mayo del año anterior.

Costos de electricidad	Frecuencia
$80 hasta $100	3
100 hasta 120	8
120 hasta 140	12
140 hasta 160	16
160 hasta 180	7
180 hasta 200	4
Total	50

a. Calcule el costo medio.
b. Determine la desviación estándar.
c. Utilice la regla empírica para calcular la fracción de costos que se encuentra a dos desviaciones estándares de la media. ¿Cuáles son estos límites?

85. Bidwell Elctronics, Inc. tomó una muestra de empleados para determinar a qué distancia viven de las oficinas centrales de la empresa. Los resultados aparecen a continuación. Calcule la media y la desviación estándar.

Distancia (en millas)	Frecuencia	*M*
0 hasta 5	4	2.5
5 hasta 10	15	7.5
10 hasta 15	27	12.5
15 hasta 20	18	17.5
20 hasta 25	6	22.5

EJERCICIOS DE LA BASE DE DATOS

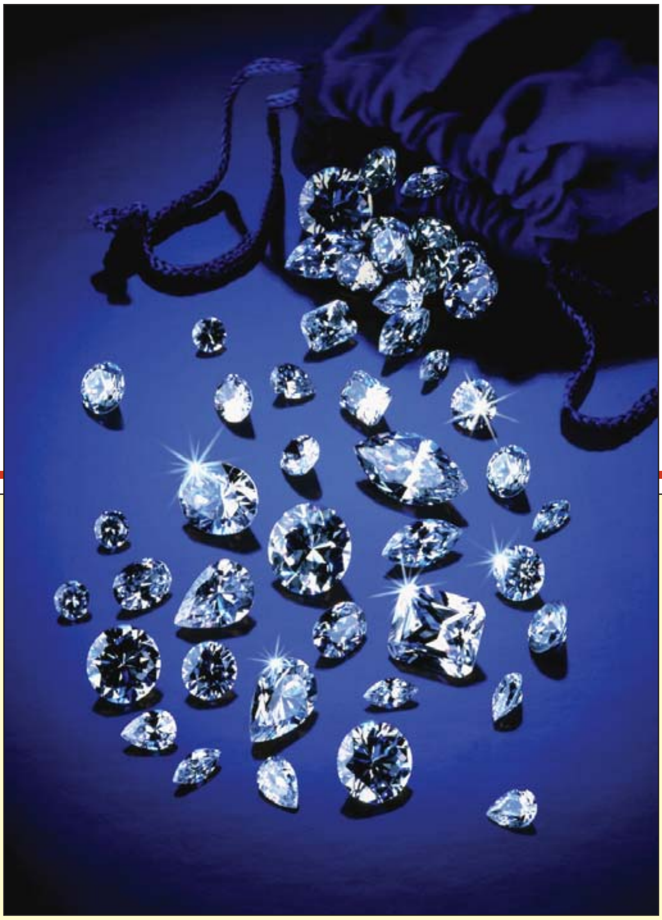
86. Consulte los datos sobre Real Estate, que contienen información acerca de casas que se vendieron en el área de Goodyear, Arizona, el año anterior. Redacte un breve informe sobre la distribución de los precios de venta. Asegúrese de contestar, en dicho reporte, las siguientes preguntas:
a. ¿Alrededor de cuáles variables tienden a concentrarse los datos? ¿Cuál es el precio medio de venta? ¿Cuál es la mediana del precio de venta? ¿Es una medida más representativa que otras de los precios típicos de venta?
b. ¿Cuál es el rango de los precios de venta? ¿Cuál es la desviación estándar? ¿Entre cuáles valores se ubica cerca de 95% de los precios de venta?

87. Consulte los datos sobre béisbol 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Refiérase a la variable “salario del equipo”.
a. Prepare un reporte sobre los salarios de los equipos que responda las siguientes preguntas:
1. ¿Alrededor de cuáles valores tienden a acumularse los datos? En específico, ¿cuál es el salario medio? ¿Cuál es la mediana del salario? ¿Es una medida más representativa que otras de los salarios típicos de los equipos?
2. ¿Cuál es el rango de los salarios? ¿Cuál es la desviación estándar? ¿Entre cuáles valores se ubica cerca de 95% de los salarios?
b. Refiérase a la información sobre el salario promedio de cada año. En 1989, el salario promedio de un jugador fue de 512 930 dólares. En 2012, el salario promedio de un jugador se incrementó a 3 440 000 dólares. ¿Cuál fue el rango de incremento en el periodo?

88. Consulte los datos de los autobuses del Distrito Escolar Buena. Prepare un reporte sobre el costo de mantenimiento del mes anterior. Responda las siguientes preguntas en dicho informe:
a. ¿Alrededor de cuáles valores tienden a acumularse los datos? En específico, ¿cuál fue el costo medio de mantenimiento el mes previo? ¿Cuál es la mediana del costo? ¿Es una medida más representativa que otras del costo típico?
b. ¿Cuál es el rango de los costos de mantenimiento? ¿Cuál es la desviación estándar? ¿Entre cuáles valores se ubica cerca de 95% de estos costos?

4 Descripción de datos:

PRESENTACIÓN Y ANÁLISIS



Recientemente, McGIVERN JEWELERS publicó un anuncio en el periódico local en el que indicaba la forma, tamaño, precio y grado de corte de 33 de sus diamantes en existencia. Elabore el diagrama de caja de la variable "precio" y comente el resultado (vea el ejercicio 37 y el objetivo de aprendizaje OA4-4).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA4-1 Elaborar e interpretar un diagrama de puntos.

OA4-2 Crear e interpretar una gráfica de tallo y hojas.

OA4-3 Identificar y calcular medidas de posición.

OA4-4 Construir e interpretar diagramas de caja.

OA4-5 Calcular y entender el coeficiente de sesgo.

OA4-6 Trazar e interpretar un diagrama de dispersión.

OA4-7 Construir e interpretar una tabla de contingencia.

Introducción

En el capítulo 2 se inició el estudio de la estadística descriptiva. Con el fin de transformar datos en bruto o no agrupados en alguna forma significativa, es necesario organizarlos en una distribución de frecuencias, la cual se representa en forma gráfica en un histograma o en un polígono de frecuencias. Este arreglo permite visualizar dónde tienden a acumularse los datos, los valores máximo y mínimo, y la forma general de los datos.

En el capítulo 3, primero se calcularon diversas medidas de ubicación o de localización, tales como la media, la mediana y la moda, que permiten informar un valor típico de un conjunto de observaciones. También se calcularon diversas medidas de localización, como el rango, la varianza y la desviación estándar, que permiten describir la variación o la dispersión en un conjunto de observaciones.

En este capítulo se continúa el estudio de la estadística descriptiva y se presentan los siguientes temas: 1) diagramas de puntos, 2) gráficas de tallo y hojas, 3) percentiles y 4) diagramas de caja. Estos diagramas y estadísticas proporcionan una idea adicional de dónde se concentran los valores, así como de la forma general de los datos. Enseguida se consideran datos bivariados de cada una de las observaciones individuales o seleccionadas. Algunos ejemplos de esto son la cantidad de horas que estudia un alumno y la calificación que obtiene en un examen; si un producto que se toma de una muestra es aceptable o no y el horario en el que se le fabrica; y la cantidad de electricidad que consume una casa en un mes, así como la temperatura alta media diaria de la región durante ese mes.

Diagramas de puntos

OA4-1
Elaborar e interpretar un diagrama de puntos.

Recuerde que en los datos de Applewood Auto Group, la ganancia obtenida por la venta de 180 vehículos se resumió en ocho clases. Al organizar así los datos se perdió el valor exacto de las observaciones. Por su parte, un **diagrama de puntos** agrupa los datos de la menor manera posible y evita la pérdida de identidad de cada observación. Para crear un diagrama de puntos se coloca un punto que representa cada observación a lo largo de una recta numérica horizontal, la cual indica los valores posibles de los datos. Si hay observaciones idénticas o si las observaciones se encuentran muy próximas, los puntos se "apilan" uno sobre otro para que se puedan ver de manera individual. Esto permite distinguir la forma de la distribución, el valor en torno al cual tienden a acumularse los datos y las observaciones máxima y mínima. Los diagramas de puntos son más útiles en el caso de conjuntos de datos pequeños, mientras que los histogramas lo son para conjuntos grandes. En el siguiente ejemplo se muestra cómo construir e interpretar los diagramas de puntos.

EJEMPLO

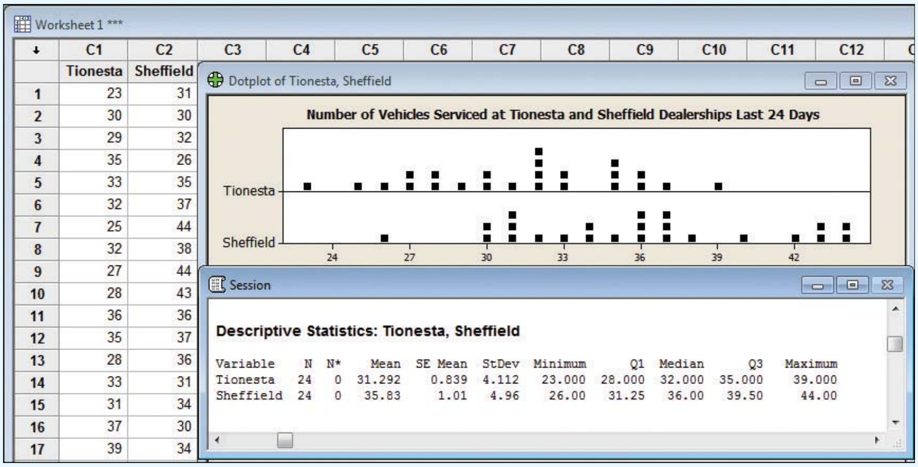
Los departamentos de servicio de Tionesta Ford Lincoln Mercury y Sheffield Motors, Inc., dos de las cuatro distribuidoras de Applewood Auto Group, abrieron 24 días hábiles el mes anterior. A continuación aparece el número de vehículos que recibieron servicio durante ese mes en ambas distribuidoras. Elabore un diagrama de puntos y presente un resumen estadístico para comparar ambas distribuidoras.

Tionesta Ford Lincoln Mercury					
Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
23	33	27	28	39	26
30	32	28	33	35	32
29	25	36	31	32	27
35	32	35	37	36	30

Sheffield Motors Inc.					
Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
31	35	44	36	34	37
30	37	43	31	40	31
32	44	36	34	43	36
26	38	37	30	42	33

SOLUCIÓN

El sistema Minitab proporciona un diagrama de puntos y permite calcular la media, la mediana, los valores máximo y mínimo y la desviación estándar de la cantidad de automóviles que recibieron servicio en cada concesionaria durante los últimos 24 días hábiles.



Los esquemas de puntos, que se muestran al centro de la pantalla, ilustran gráficamente las distribuciones de ambas concesionarias. Los puntos muestran las diferencias en la ubicación y la dispersión de las observaciones. Al observar los esquemas de puntos se puede ver que el número de vehículos que recibieron servicio en la distribuidora Sheffield están más dispersos y tienen una media mayor que los de Tionesta. Otras características del número de vehículos que recibieron servicio son:

- Tionesta fue el que menos servicios dio en un día: 23.
- Sheffield dio servicio a 26 autos en su día más bajo, cuatro autos menos que en su siguiente día más bajo.
- Tionesta dio servicio exactamente a 32 vehículos en cuatro días diferentes.
- Los números de autos que recibieron servicio se acumulan alrededor del 36 en el caso de Sheffield y 32 en el de Tionesta.

A partir de la estadística descriptiva es posible visualizar que Sheffield dio servicio a un promedio de 35.83 vehículos diarios y Tionesta, un promedio de 31.292 autos al día en el mismo periodo. También existe mayor dispersión, o variación, en el número diario de vehículos que recibieron servicio en Sheffield que en Tionesta. ¿Cómo se llega a esta conclusión? La desviación estándar de Sheffield es mayor (4.96 automóviles por día) que la de Tionesta (4.112 por día).

OA4-2
Crear e interpretar una gráfica de tallo y hojas.

Gráficas de tallo y hojas

En el capítulo 2 se ilustró la manera de organizar datos en una distribución de frecuencias de tal manera que los datos brutos se puedan resumir de forma significativa. La ventaja principal de organizar los datos en una distribución de frecuencias estriba en que permite visualizar de manera rápida la forma de la distribución sin necesidad de llevar a cabo ningún cálculo. En otras palabras, es posible ver dónde se concentran los datos y, asimismo, determinar si hay valores extremadamente grandes o pequeños. Sin embargo, hay dos desventajas al organizar los datos en la distribución de frecuencias: 1) se pierde la identidad exacta de cada valor y 2) no es clara la forma en que se distribuyen los valores de cada clase. Para mayor precisión, en la distribución de frecuencias de la izquierda se muestra la cantidad de espacios publicitarios que compraron los 45 miembros de la Greater Buffalo Automobile Dealers Association durante el año 2010. Observe que 7 de las 45 concesionarias compraron de 90 a 100 espacios. Sin embargo, ¿los espacios comprados en esta clase se acumulan en torno a 90, se distribuyen uniformemente a lo largo de la clase o se acumulan cerca de 99? No es posible afirmar nada.

Cantidad de espacios comprados	Frecuencia
80 hasta 90	2
90 hasta 100	7
100 hasta 110	6
110 hasta 120	9
120 hasta 130	8
130 hasta 140	7
140 hasta 150	3
150 hasta 160	3
Total	45

Otra técnica que se utiliza para representar información cuantitativa en forma condensada es el **diagrama de tallo y hojas.** Una de sus ventajas sobre la distribución de frecuencias es que la identidad de cada observación no se pierde. En el ejemplo anterior no se conoce la identidad de los valores en la clase de 90 hasta 100. Para ilustrar la forma de construir un diagrama de tallo y hojas a partir de la cantidad de espacios publicitarios comprados, suponga que las siete observaciones en la clase de 90 hasta 100 son: 96, 94, 93, 94, 95, 96 y 97. El valor de **tallo** es el dígito o dígitos principales, en este caso 9. Las **hojas** son los dígitos secundarios. El tallo se coloca a la izquierda de una línea vertical y los valores de las hojas, a la derecha.

Los valores en la clase de 90 hasta 100 se verían de la siguiente manera:

9 |	6	4	3	4	5	6	7

También es costumbre ordenar los valores en cada tallo de menor a mayor. Por consiguiente, la segunda fila del diagrama de tallo y hojas quedaría de la siguiente manera:

9 |	3	4	4	5	6	6	7

Con un diagrama de tallo y hojas es más fácil observar que dos concesionarias compraron 94 espacios y que el número de espacios comprados varía de 93 hasta 97. Este tipo de diagrama se parece a una distribución de frecuencias pero con mayor información, es decir, la identidad de las observaciones se conserva.

DIAGRAMA DE TALLO Y HOJAS Técnica estadística para presentar un conjunto de datos. Cada valor numérico se divide en dos partes. El dígito principal se convierte en el tallo y los dígitos secundarios, en las hojas. El tallo se localiza a lo largo del eje vertical y los valores de las hojas se apilan a lo largo del eje horizontal.

En el siguiente ejemplo se explican los detalles para elaborar un diagrama de tallo y hojas.

EJEMPLO

La tabla 4.1 contiene la lista de la cantidad de espacios publicitarios de 30 segundos en radio que compró cada uno de los 45 miembros de la Greater Buffalo Automobile Dealers Association el año anterior. Organice los datos en un diagrama de tallo y hojas. ¿Alrededor de cuáles valores tiende a acumularse el número de espacios publicitarios? ¿Cuál es el número menor de espacios publicitarios comprados? ¿Cuál es el número máximo de espacios comprados?

TABLA 4.1 Número de espacios publicitarios que compraron los miembros de la Greater Buffalo Automobile Dealers Association

96	93	88	117	127	95	113	96	108	94	148	156
139	142	94	107	125	155	155	103	112	127	117	120
112	135	132	111	125	104	106	139	134	119	97	89
118	136	125	143	120	103	113	124	138			

SOLUCIÓN

De acuerdo con los datos de la tabla 4.1, el número mínimo de espacios publicitarios comprados es 88. Por ello, el primer valor de tallo es 8. El número máximo es 156, así que los valores de tallo comienzan en 8 y continúan hasta 15. El primer número de la tabla 4.1 es 96, que tendrá 9 como valor de tallo y 6 como valor de hoja. Al recorrer por el renglón superior, el segundo valor es 93 y el tercero, 88. Después de considerar los primeros tres valores, el diagrama queda de la siguiente manera:

Tallo	Hoja
8	8
9	6 3
10	
11	
12	
13	
14	
15	

Al organizar los datos, el diagrama de tallo y hojas queda de la siguiente manera:

Tallo	Hoja
8	8 9
9	6 3 5 6 4 4 7
10	8 7 3 4 6 3
11	7 3 2 7 2 1 9 8 3
12	7 5 7 0 5 5 0 4
13	9 5 2 9 4 6 8
14	8 2 3
15	6 5 5

El procedimiento acostumbrado consiste en ordenar los valores de las hojas de menor a mayor. La última línea, la fila que se refiere a los valores próximos a 150, se vería de la siguiente manera:

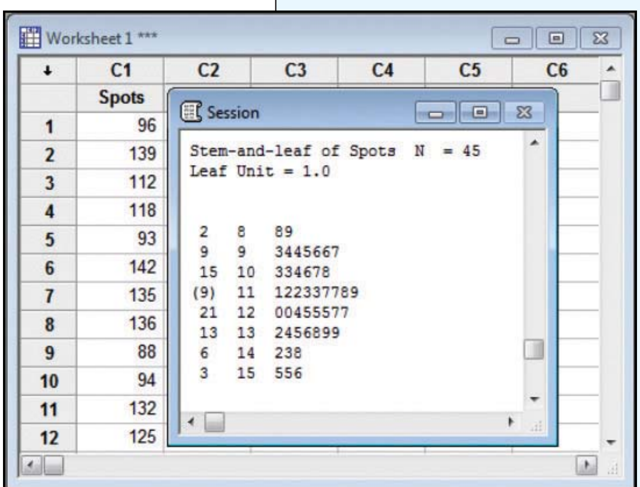
15 |	5	5	6

La tabla final sería la siguiente, en la cual se ordenan todos los valores de las hojas:

Tallo	Hoja
8	8 9
9	3 4 4 5 6 6 7
10	3 3 4 6 7 8
11	1 2 2 3 3 7 7 8 9
12	0 0 4 5 5 5 7 7
13	2 4 5 6 8 9 9
14	2 3 8
15	5 5 6

Es posible deducir algunas conclusiones del diagrama de tallo y hojas. Primero, la cantidad mínima de espacios publicitarios comprados es 88 y la máxima, 156. Dos concesionarias compraron menos de 90, y tres compraron 150 o más. Observe, por ejemplo, que las tres concesionarias que compraron más de 150, en realidad compraron 155, 155 y 156. La concentración de la cantidad de espacios se encuentra entre 110 y 130. Hubo nueve concesionarias que compraron entre 110 y 119 y ocho compraron entre 120 y 129. También note que en el grupo ubicado entre 120 y 129 el número real de espacios comprados se distribuyó uniformemente; es decir, dos concesionarias compraron 120, una compró 124, tres compraron 125 y dos compraron 127.

Además, es posible generar esta información en el sistema de software Minitab. La variable se llama “spots”. Enseguida aparece el resultado de Minitab (encontrará los comandos de Minitab que lo generan en el apéndice C).

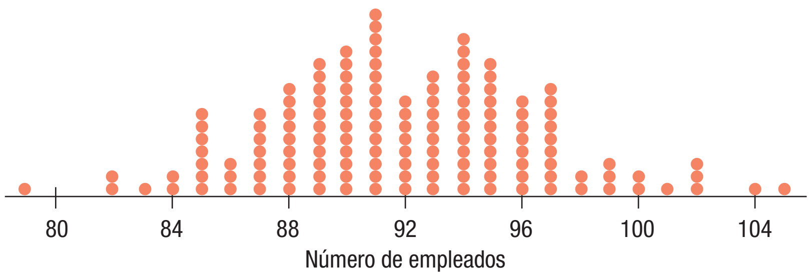


La solución de Minitab proporciona información adicional relacionada con los totales acumulados. En la columna a la izquierda de los valores de tallo se encuentran números como 2, 9, 15, y así sucesivamente. El número 9 indica que se presentaron nueve observaciones antes del valor 100. El 15 muestra que se presentó ese número de observaciones antes de 110. Más o menos a la mitad de la columna aparece el número 9 entre paréntesis, que indica que el valor de medio o mediana aparece en dicha fila y que hay nueve valores en este grupo. En este caso, el valor medio es el número debajo del cual se presenta la mitad de las observaciones. Hay un total de 45 observaciones, así que el valor medio, en caso de ordenar los datos en orden ascendente, sería la observación vigésima tercera; este valor es 118. Después de la mediana, los valores comienzan a decrecer. Estos valores representan los totales acumulados *más que.* Hay 21 observaciones de 120 o más, 13 de 130 o más, y así sucesivamente.

¿Cuál es la mejor opción, el esquema de puntos o el diagrama de tallo y hojas? En realidad, este dilema es cuestión de elección y conveniencia personal. Para presentar datos, en especial con una gran cantidad de observaciones, usted observará que los diagramas de puntos se utilizan con mayor frecuencia. Encontrará diagramas de puntos en la literatura analítica, informes de marketing y, en ocasiones, en informes anuales. Si realiza un análisis rápido para usted mismo, los diagramas de tallo y hojas son accesibles y fáciles, en particular en relación con un conjunto pequeño de datos.

AUTOEVALUACIÓN 4-1

1. En el siguiente diagrama se muestra el número de empleados en cada una de las 142 tiendas de Home Depot ubicadas sureste de Estados Unidos.



   (a) ¿Cuáles son los números máximo y mínimo de empleados por tienda?
   (b) ¿Cuántas tiendas emplean a 91 personas?
   (c) ¿Alrededor de cuáles valores tiende a acumularse el número de empleados por tienda?

2. La tasa de recuperación de 21 acciones es la siguiente:

8.3	9.6	9.5	9.1	8.8	11.2	7.7	10.1	9.9	10.8	
10.2	8.0	8.4	8.1	11.6	9.6	8.8	8.0	10.4	9.8	9.2

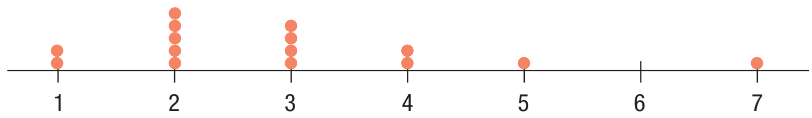
   Organice esta información en un diagrama de tallo y hojas.
   (a) ¿Cuántas tasas son menores que 9.0?
   (b) Haga una lista de las tasas en la categoría que va de 10.0 hasta 11.0.
   (c) ¿Cuál es la mediana?
   (d) ¿Cuáles son las tasas máxima y mínima de recuperación?

EJERCICIOS

1. Describa las diferencias entre un histograma y un diagrama de puntos. ¿Cuándo podría resultar mejor un diagrama de puntos que un histograma?

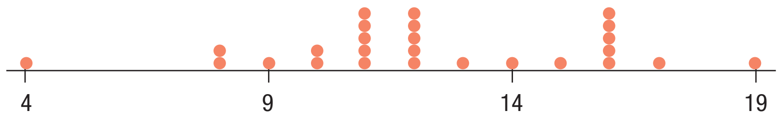
2. Explique las diferencias entre un histograma y un diagrama de tallo y hojas.

3. Considere el siguiente diagrama.



a. ¿Qué nombre recibe este diagrama?
b. ¿Cuántas observaciones hay en el estudio?
c. ¿Cuáles son los valores máximo y mínimo?
d. ¿En torno a cuáles valores tienden a acumularse las observaciones?

4. En el siguiente diagrama se registra el número de teléfonos celulares que Radio Shack vendió durante los últimos 26 días.



a. ¿Cuáles son los números máximo y mínimo de teléfonos celulares vendidos en un día?
b. ¿Cuál es el número más frecuente de teléfonos celulares vendidos?

5. La primera fila del diagrama de tallo y hojas es la siguiente: 62|1 3 3 7 9. Suponga que se trata de números enteros.
a. ¿Cuál es el "posible rango" de los valores de esta fila?
b. ¿Cuántos valores hay en esta fila?
c. Haga una lista de los valores reales de esta fila de datos.

6. La tercera fila de un diagrama de tallo y hojas aparece de la siguiente manera: 21|0 1 3 5 7 9. Suponga que los valores son números enteros.
a. ¿Cuál es el "posible rango" de los valores de esta fila?
b. ¿Cuántos datos hay en esta fila?
c. Elabore una lista de los datos reales de esta fila.

7. En el siguiente diagrama de tallo y hojas de Minitab se muestra el número de unidades producidas por día en una fábrica.

1	3	8
1	4	
2	5	6
9	6	0133559
(7)	7	0236778
9	8	59
7	9	00156
2	10	36

a. ¿Cuántos días se registraron?
b. ¿Cuántas observaciones hay en la primera clase?
c. ¿Cuál es el valor mínimo y el valor máximo?
d. Elabore una lista de los valores reales de la cuarta fila.
e. Elabore una lista de los valores reales de la segunda fila.
f. ¿Cuántos valores son menores que 70?
g. ¿Cuántos valores son iguales o mayores que 80?
h. ¿Cuál es la mediana?
i. ¿Cuántos valores se encuentran entre 60 y 89 (incluidos estos)?

8. En el siguiente diagrama de tallo y hojas se presenta la cantidad de películas rentadas por día en Video Connection, ubicado en la esquina de las calles Forth y Main.

3	12	689
6	13	123
10	14	6889
13	15	589
15	16	35
20	17	24568
23	18	268
(5)	19	13456
22	20	034679
16	21	2239
12	22	789
9	23	00179
4	24	8
3	25	13
1	26	
1	27	0

a. ¿Cuántos días se registraron?
b. ¿Cuántas observaciones hay en la última clase?
c. ¿Cuáles son los valores máximo y mínimo de todo el conjunto de datos?
d. Elabore una lista de valores reales de la cuarta fila.
e. Elabore una lista de valores reales que aparecen en la penúltima fila.
f. ¿En cuántos días se rentaron menos de 160 películas?
g. ¿En cuántos días se rentaron 220 películas o más?
h. ¿Cuál es el valor medio?
i. ¿En cuántos días se rentaron entre 170 y 210 películas?

9. Una encuesta sobre el número de llamadas telefónicas por celular realizada con una muestra de suscriptores de Verizon la semana pasada reveló la siguiente información. Elabore un diagrama de tallo y hojas. ¿Cuántas llamadas hizo un suscriptor promedio? ¿Cuáles fueron los números máximo y mínimo de llamadas que realizaron?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

52	43	30	38	30	42	12	46	39
37	34	46	32	18	41	5		

10. Aloha Banking, Co., estudia el uso de cajeros automáticos en los suburbios de Honolulu. Una muestra de 30 cajeros mostró que estos se utilizaron la siguiente cantidad de veces el día anterior. Elabore un diagrama de tallo y hojas. Resuma la cantidad de veces que se utilizó cada cajero automático. ¿Cuáles son los números mínimo y máximo de veces que se utilizó cada uno?

83	64	84	76	84	54	75	59	70	61
63	80	84	73	68	52	65	90	52	77
95	36	78	61	59	84	95	47	87	60

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Otras medidas de posición

OA4-3 Identificar y calcular medidas de posición.

La desviación estándar es la medida de dispersión que más se utiliza. No obstante, existen otras formas de describir la variación o dispersión de un conjunto de datos. Un método consiste en determinar la *ubicación* de los valores que dividen un conjunto de observaciones en partes iguales. Estas medidas incluyen los **cuartiles**, **deciles** y **percentiles**.

Los cuartiles dividen a un conjunto de observaciones en cuatro partes iguales. Para explicarlo mejor, piense en un conjunto de valores ordenados de menor a mayor. En el capítulo 3 se denominó *mediana* al valor intermedio de un conjunto de datos ordenados de menor a mayor. Es decir, 50% de las observaciones son mayores que la mediana y 50% son menores. La mediana constituye una medida de ubicación porque señala el centro de los datos. De igual manera, los **cuartiles** dividen a un conjunto de observaciones en cuatro partes iguales. El primer cuartil, que se representa mediante Q_1, es el valor bajo el cual se presenta 25% de las observaciones, y el tercer cuartil, simbolizado por Q_3, es el valor bajo el cual se presenta 75% de las observaciones.

Asimismo, los **deciles** dividen un conjunto de observaciones en 10 partes iguales y los **percentiles**, en 100. Por lo tanto, si su promedio general en la universidad se encuentra en el octavo decil, usted podría concluir que 80% de los estudiantes tuvieron un promedio general inferior al suyo y 20%, uno superior. Si su promedio estuvo en el 92° percentil, entonces 92% de los estudiantes tuvo un promedio general menor que el suyo, y solo 8% de ellos tuvo uno mayor. Con frecuencia, en Estados Unidos, las calificaciones que se expresan en percentiles se utilizan para dar a conocer resultados relacionados con pruebas estandarizadas como SAT, ACT, GMAT (que se emplean para determinar el ingreso en algunas maestrías de administración de empresas) y LSAT (que sirve para determinar el ingreso a la escuela de leyes).

Cuartiles, deciles y percentiles

Para formalizar el proceso de cálculo, suponga que L_p representa la ubicación de cierto percentil que se busca. De esta manera, si quiere encontrar el 92° percentil, utilizaría L_{92}; y si buscara la mediana, el percentil 50°, entonces emplearía L_{50}. Para un número de observaciones n, la ubicación del percentil P° puede encontrarse con la siguiente fórmula:

LOCALIZACIÓN DE UN PERCENTIL $$L_p = (n + 1)\frac{P}{100} \quad \textbf{[4.1]}$$

Un ejemplo ayudará explicar la fórmula anterior.

EJEMPLO

Enseguida aparecen las comisiones que ganó el último mes una muestra de 15 corredores de bolsa de la oficina de Morgan Stanley Smith Barney's Oakland, California. Esta compañía de inversiones tiene oficinas en todo Estados Unidos.

\$2 038	\$1 758	\$1 721	\$1 637	\$2 097	\$2 047	\$2 205	\$1 787	\$2 287
1 940	2 311	2 054	2 406	1 471	1 460			

Localice la mediana, el primer y tercer cuartiles de las comisiones ganadas.

SOLUCIÓN

El primer paso consiste en ordenar las comisiones ganadas de menor a mayor.

\$1 460	\$1 471	\$1 637	\$1 721	\$1 758	\$1 787	\$1 940	\$2 038
2 047	2 054	2 097	2 205	2 287	2 311	2 406	

El valor mediano es la observación que está al centro, y es el mismo que 50° percentil, así que P es igual a 50. La mediana, o L_{50}, se localiza en $(n + 1)(50/100)$, en donde n representa el número de observaciones. En este caso es la octava posición, determinada por $(15 + 1)(50/100)$. La octava comisión más grande es de 2 038 dólares. Así que se concluye que esta es la mediana y que la mitad de los corredores obtiene comisiones mayores que 2 038 dólares, y la mitad gana menos de esa cantidad. El resultado, usando la fórmula [4.1] para determinar la mediana, es el mismo que con el método presentado en el capítulo 3.

Recuerde la definición de cuartil. Los cuartiles dividen a un conjunto de observaciones en cuatro partes iguales. Por consiguiente, 25% de las observaciones serán menores que el primer cuartil y 75% de ellas serán menores que el tercer cuartil. Para localizar el primer cuartil utilice la fórmula [4.1], en la cual $n = 15$ y $P = 25$:

$$L_{25} = (n + 1)\frac{P}{100} = (15 + 1)\frac{25}{100} = 4$$

para localizar el tercer cuartil, $n = 15$ y $P = 75$:

$$L_{75} = (n + 1)\frac{P}{100} = (15 + 1)\frac{75}{100} = 12$$

Por lo tanto, los valores del primer y tercer cuartiles se localizan en las posiciones 4 y 12. El cuarto valor en la serie ordenada es 1 721 dólares y el decimosegundo es 2 205 dólares. Estos constituyen el primer y tercer cuartiles.

ESTADÍSTICA EN ACCIÓN

En 1939, John W. Tukey (1915-2000) recibió un doctorado en matemáticas de Princeton. Sin embargo, cuando se unió a la Fire Control Research Office, durante la Segunda Guerra Mundial, su interés en las matemáticas abstractas se desvió hacia la estadística aplicada. Desarrolló métodos numéricos y gráficos eficaces para estudiar los patrones que subyacían a los datos. Entre las gráficas que creó se encuentran el diagrama de tallo y hojas y el diagrama de caja y bigotes (o diagrama de caja). De 1960 a 1980, Tukey encabezó la división de estadística electoral del equipo de proyección nocturno de la NBC. En 1960 se hizo famoso cuando evitó el anuncio de la victoria anticipada de Richard Nixon en las elecciones presidenciales que ganó John F. Kennedy.

En el ejemplo anterior, la fórmula de localización arrojó un número entero. Es decir, al buscar el primer cuartil había 15 observaciones, así que la fórmula de localización indica que debería encontrar el cuarto valor ordenado. ¿Qué haría si hubiera 20 observaciones en la muestra, es decir $n = 20$, y quisiera localizar el primer cuartil? De acuerdo con la fórmula de localización [4.1]:

$$L_{25} = (n + 1)\frac{P}{100} = (20 + 1)\frac{25}{100} = 5.25$$

Localizar el quinto valor en la serie ordenada y en seguida desplazarse una distancia de 0.25 entre los valores quinto y sexto, y señalar a este como el primer cuartil. Como en el caso de la mediana, el cuartil no necesita ser uno de los valores exactos del conjunto de datos.

Para explicarlo más a fondo, suponga que un conjunto de datos contiene los seis valores: 91, 75, 61, 101, 43 y 104. Trate de localizar el primer cuartil. Ordene los valores de menor a mayor: 43, 61, 75, 91, 101 y 104. El primer cuartil se localiza en

$$L_{25} = (n + 1)\frac{P}{100} = (6 + 1)\frac{25}{100} = 1.75$$

La fórmula de localización indica que el primer cuartil se ubica entre el primer valor y el segundo, y es 0.75 de la distancia entre ellos.

El primer valor es 43 y el segundo, 61. De esta manera, la distancia entre estos valores es 18. Al localizar el primer cuartil necesita desplazarse una distancia de 0.75 entre el primer y segundo valores; así, 0.75 (18) = 13.5. Para completar el procedimiento, sume 13.5 al primer valor e indique que el primer cuartil es 56.5.

Es posible ampliar la idea para incluir tanto deciles como percentiles. Para localizar el 23° percentil en una muestra de 88 observaciones, busque la posición 18.63.

$$L_{23} = (n + 1)\frac{P}{100} = (80 + 1)\frac{23}{100} = 18.63$$

Para determinar el valor correspondiente al 23° percentil, localice el 18° valor y el 19°, y determine la distancia entre ambos. Luego, multiplique esta diferencia por 0.63 y sume el resultado al valor más pequeño. El resultado es el 23° percentil.

El software estadístico es muy útil para describir y resumir los datos. Excel, Minitab y Megastat, un complemento de análisis estadístico de Excel, proporcionan resúmenes estadísticos que incluyen a los cuartiles. Por ejemplo, a continuación se presenta una salida de Minitab que resume los datos de las comisiones de Smith Barney. Incluye el primer y el tercer cuartiles y otras estadísticas. Con base en los cuartiles reportados, se concluye que 25% de las comisiones fueron menores a 1 721 dólares y 75% fueron menores a 2 205 dólares. Son los mismos valores obtenidos mediante la fórmula [4.1], del ejemplo previo.

Sheet1 ***

↓	C1
	Commissions
1	1460
2	1471
3	1637
4	1721

Session

Descriptive Statistics: Commissions

```
Variable     N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3  Maximum
Commissions  15  0  1947.9     77.1  298.8   1460.0  1721.0  2038.0  2205.0   2406.0
```

Existen otras formas además de la fórmula [4.1] para localizar los valores de cuartiles. Por ejemplo, otro método utiliza $0.25n + 0.75$ para descubrir la posición del primer cuartil y $0.75n + 0.25$ para hallar la posición del tercer cuartil. A esto se le llama *Método Excel.* En el caso de los datos de Smith Barney, este método colocaría el primer cuartil en la posición 4.5 ($0.25 \times 15 + 0.75$); y el tercer cuartil, en la posición 11.5 ($0.75 \times 15 + 0.25$). El primer cuartil se interpolaría en 0.5, o la mitad de la diferencia entre los valores ubicados en el cuarto y quinto lugar. Utilizando este método, el primer cuartil es 1 739.5 dólares, determinado por ($1 721 + 0.5[$1 758 − $1 721]). El tercer cuartil, en la posición 11.5, es 2 151 dólares, o la mitad de la distancia entre los valores ubicados en el undécimo y duodécimo lugar, determinado por ($2 097 + 0.5[$2 205 − $2 097]). Como se muestra en los ejemplos de Smith Barney y Applewood, es posible calcular los cuartiles utilizando cualquiera de ambos métodos. **Observe que el texto utiliza la fórmula [4.1] para calcular los cuartiles.**

SmithBarney.xlsx

	A	B	C	D
1	Commissions			
2	1460			
3	1471		Equation 4-1	
4	1637		Quartile 1	1721
5	1721			
6	1758		Quartile 3	2205
7	1787			
8	1940		Excel Method	
9	2038		Quartile 1	1739.5
10	2047			
11	2054		Quartile 3	2151
12	2097			
13	2205			
14	2287			
15	2311			
16	2406			

¿Es importante la diferencia entre ambos métodos? No, en realidad suele ser solo una molestia. Por lo general, los estadísticos prefieren el primer método aquí expuesto. Cuando la muestra es grande, la diferencia entre los resultados de ambos métodos es pequeña. Por ejemplo, en los datos de Applewood Auto Group hay 180 vehículos. Los cuartiles calculados con ambos métodos se muestran a la derecha. Utilizando la variable "ganancia", 45 de los 180 valores (25%) son menores que ambos valores del primer cuartil, y 135 de los 180 valores (75%) son menores que los valores del tercer cuartil.

Al utilizar Excel procure entender el método utilizado para calcular cuartiles. En Excel 2007, los cuartiles se calculan con el *Método Excel*. Los comandos de Excel 2010 para calcular los cuartiles se muestran en el apéndice C.

APPLEWOOD AUTO GROUP.xlsx

	A	B	C	D	E
1	**Age**	**Profit**			
2	44	$294			
3	40	$323		Equation 4-1	
4	42	$335		1st Quartile	1415.5
5	40	$352			
6	46	$369		3rd Quartile	2275.5
7	53	$377			
8	**30**	**$443**			
9	40	$482		Excel Method	
10	37	$732		1st Quartile	1422.5
11	**30**	**$754**			
12	62	$783		3rd Quartile	2268.5

AUTOEVALUACIÓN 4-2

El departamento de control de calidad de Plainsville Peanut Company verifica el peso de un frasco de crema de cacahuate de ocho onzas. Los pesos de la muestra de nueve frascos fabricados durante la última hora son los siguientes:

7.69	7.72	7.8	7.86	7.90	7.94	7.97	8.06	8.09

(a) ¿Cuál es el peso mediano?
(b) Determine los pesos correspondientes del primer y tercer cuartiles.

EJERCICIOS

11. Determine la mediana y los valores correspondientes al primer y tercer cuartiles en los siguientes datos.

46	47	49	49	51	53	54	54	55	55	59

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

12. Determine la mediana y los valores correspondientes al primer y tercer cuartiles en los siguientes datos.

5.24	6.02	6.67	7.30	7.59	7.99	8.03	8.35	8.81	9.45
9.61	10.37	10.39	11.86	12.22	12.71	13.07	13.59	13.89	15.42

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

13. Thomas Supply Company, Inc., es un distribuidor de generadores de gas. Como en cualquier negocio, el tiempo que emplean los clientes para pagar sus recibos es importante. La siguiente lista está en orden ascendente y registra el tiempo (en días) de una muestra de facturas de Thomas Supply Company, Inc.

13	13	13	20	26	27	31	34	34	34	35	35	36	37	38
41	41	41	45	47	47	47	50	51	53	54	56	62	67	82

a. Determine el primer y tercer cuartiles.
b. Determine el segundo y octavo deciles.
c. Determine el 67° percentil.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14. Kevin Horn es gerente nacional de ventas en National Textbooks, Inc. Cuenta con un personal de ventas conformado por 40 personas, las cuales hacen visitas a profesores universitarios en todo Estados Unidos. Cada sábado por la mañana solicita a su personal que le envíe un informe, el cual debe incluir, entre otros detalles, la cantidad de profesores que visitaron la semana anterior. En la lista siguiente, en orden ascendente, se muestra la cantidad de visitas de la semana pasada.

38	40	41	45	48	48	50	50	51	51	52	52	53	54	55	55	55	56	56	57
59	59	59	62	62	62	63	64	65	66	66	67	67	69	69	71	77	78	79	79

a. Determine la cantidad mediana de visitas.
b. Determine el primer y tercer cuartiles.
c. Determine el primero y noveno deciles.
d. Determine el 33° percentil.

OA4-4
Construir e interpretar diagramas de caja.

Diagramas de caja

Un **diagrama de caja** es una representación gráfica basada en cuartiles que ayuda a presentar un conjunto de datos. Para construir un diagrama de caja solo se necesitan cinco estadísticos: valor mínimo, Q_1 (primer cuartil), mediana, Q_3 (tercer cuartil) y valor máximo. Un ejemplo ayudará a explicarlo.

EJEMPLO

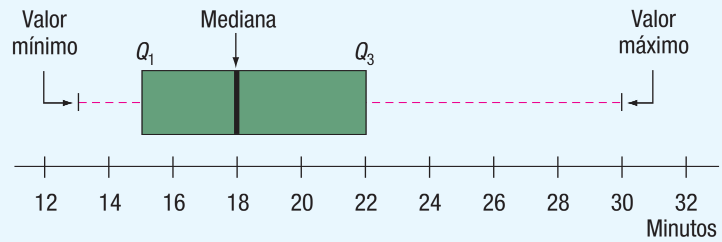
Alexander's Pizza ofrece entregas gratuitas de pizza a 15 millas a la redonda. Alex, el propietario, desea información relacionada con el tiempo de entrega. ¿Cuánto tiempo tarda una entrega típica? ¿En qué margen de tiempo deben completarse la mayoría de las entregas? Alex recopiló la siguiente información de una muestra de 20 entregas:

$$\begin{aligned} \text{Valor mínimo} &= 13 \text{ minutos} \\ Q_1 &= 15 \text{ minutos} \\ \text{Mediana} &= 18 \text{ minutos} \\ Q_3 &= 22 \text{ minutos} \\ \text{Valor máximo} &= 30 \text{ minutos} \end{aligned}$$

Elabore un diagrama de caja de los tiempos de entrega, ¿qué conclusiones tiene acerca de estos?

SOLUCIÓN

El primer paso para elaborar un diagrama de caja consiste en crear una escala adecuada a lo largo del eje horizontal. Luego, se debe dibujar una caja que inicie en Q_1 (15 minutos) y termine en Q_3 (22 minutos). Dentro de la caja trace una línea vertical para representar a la mediana (18 minutos). Por último, prolongue líneas horizontales a partir de la caja dirigidas al valor mínimo (13 minutos) y al valor máximo (30 minutos). Estas líneas horizontales que salen de la caja, a veces reciben el nombre de *bigotes,* en virtud de su parecido a los bigotes de un gato.



El diagrama de caja también muestra el **rango intercuartil** de los tiempos de entrega entre Q1 y Q3; este es de siete minutos e indica que 50% de las entregas se realizan entre 15 y 22 minutos.

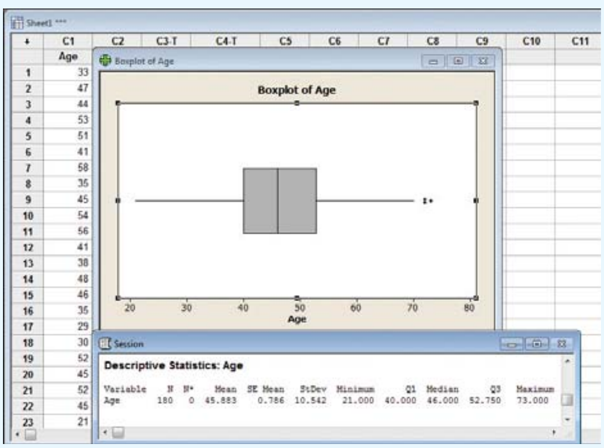
El diagrama de caja también revela que la distribución de los tiempos de entrega tiene un sesgo positivo. Recuerde que en el capítulo 3 se definió el sesgo como la falta de simetría en un conjunto de datos. ¿Cómo saber que esta distribución tiene un sesgo positivo? En este caso hay dos tipos de información que lo sugieren. Primero, la línea punteada a la derecha de la caja, que va de 22 minutos (Q_3) al tiempo máximo de 30 minutos, es más larga que la línea punteada a la izquierda que va de 15 minutos (Q_1) al valor mínimo de 13 minutos. En otras palabras, 25% de los datos mayores que el tercer cuartil se encuentra más disperso que el 25% que es menor al primer cuartil. Una segunda indicación del sesgo positivo es que la mediana no se encuentra al centro de la caja. La distancia del primer cuartil a la mediana es menor que la distancia de la mediana al tercer cuartil. El número de tiempos de entrega entre 15 y 18 minutos es el mismo que el número de tiempos de entrega entre 18 y 22 minutos.

EJEMPLO

Consulte los datos de Applewood Auto Group. Elabore un diagrama de caja con base en la variable “edad del comprador”. ¿Cuál es la conclusión respecto de la distribución de las edades de los compradores?

SOLUCIÓN

Se utilizó Minitab para desarrollar la siguiente gráfica y resumir las estadísticas.



La edad mediana de los compradores es 46 años; 25% de ellos tienen menos de 40 años de edad y 25%, más de 52.75. Con base en la información resumida y en el diagrama de caja es posible concluir que:

- Cincuenta por ciento de los compradores tiene entre 40 y 52.75 años.
- La distribución de edades es simétrica. Existen dos razones para esta conclusión. La longitud del bigote por encima de 52.75 años (Q_3) tiene aproximadamente el mismo largo que el margen que está por debajo de 40 años (Q_1). Asimismo, el área de la caja entre 40 años y la mediana (46 años) es más o menos la misma que el área entre la mediana y 52.75 años.

Hay tres asteriscos (*) por encima de la marca de 70 años. ¿Qué es lo que indican? En un diagrama de caja, un asterisco identifica un **dato atípico**; es decir, se trata de un valor que no concuerda con el resto de los datos. Se define como un valor más de 1.5 veces la amplitud del rango intercuartil más pequeño que Q_1, o mayor que Q_3. En este ejemplo, un dato atípico sería un valor mayor que 71.875 años, el cual se determina con el siguiente cálculo:

$$\textit{Dato atípico} > Q_3 + 1.5(Q_3 + Q_1) = 52.75 + 1.5(52.75 - 40) = 71.875$$

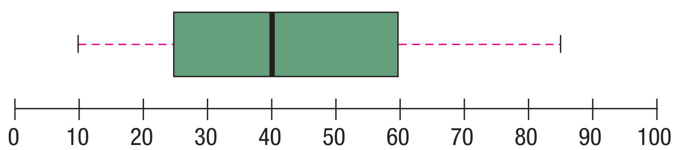
Un valor menor que 20.875 años también es un dato atípico.

$$\textit{Dato atípico} < Q_1 - 1.5(Q_3 - Q_1) = 40 - 1.5(52.75 - 40) = 20.875$$

Con base en el diagrama de caja se concluye que hubo tres compradores de 72 años o mayores, y ninguno menor de 21 años. Nota técnica: en algunos casos, un solo asterisco puede representar más de una observación debido a las limitaciones del software y del espacio disponible. Es buena idea verificar los datos reales. En este caso, hubo tres compradores de 72 años o mayores: dos tienen 72 y uno tiene 73.

AUTOEVALUACIÓN 4-3

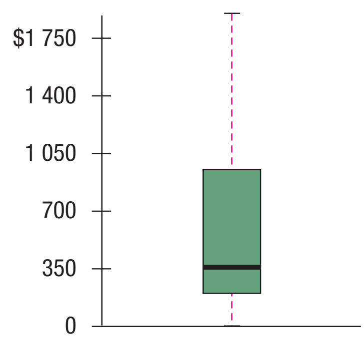
En el siguiente diagrama de caja se muestran los activos en millones de dólares de cooperativas de crédito en Seattle, Washington.



¿Cuáles son los valores mínimo y máximo, los cuartiles primero y tercero, y la mediana? ¿Estaría usted de acuerdo en que la distribución es simétrica? ¿Hay datos atípicos?

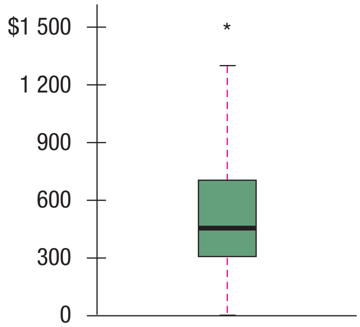
EJERCICIOS

15. El diagrama de caja muestra la suma que los estudiantes de cuarto año de universidades públicas gastaron en libros y suministros durante un año.



a. Calcule la mediana de la suma que se gastó.
b. Calcule el primero y tercer cuartiles de la cantidad que se gastó.
c. Calcule el rango intercuartil de la cantidad que se gastó.
d. ¿Más allá de qué punto un valor se considera dato atípico?
e. Identifique cualesquiera datos atípicos y calcule su valor.
f. ¿Es simétrica la distribución, o tiene sesgo positivo o negativo?

16. El diagrama de caja muestra el cargo interestatal de crédito por hora para carreras de cuatro años de estudiantes graduados en universidades públicas.



a. Calcule la mediana.
b. Calcule el primer y tercer cuartiles.
c. Determine el rango intercuartil.
d. ¿Más allá de qué punto un valor se considera dato atípico?
e. Identifique cualesquiera datos atípicos y calcule su valor.
f. ¿Es simétrica la distribución, o tiene sesgo positivo o negativo?

17. La media de un estudio sobre el rendimiento en millas por galón de gasolina de automóviles modelo 2013 fue 27.5 y la mediana, 26.8. El menor valor fue 12.70 millas por galón y el mayor fue 50.20. El primer y tercer cuartiles fueron 17.95 y 35.45 millas por galón, respectivamente. Elabore un diagrama de caja y haga algún comentario sobre la distribución. ¿Es una distribución simétrica?

18. Una muestra de 28 departamentos de tiempo compartido en el área de Orlando, Florida, reveló las siguientes tarifas diarias de una suite con una recámara. Por comodidad, los datos se encuentran ordenados de manera ascendente. Construya un diagrama de caja para representar los datos. Haga algún comentario sobre la distribución. Identifique el primer y tercer cuartiles, y la mediana.

$116	$121	$157	$192	$207	$209	$209
229	232	236	236	239	243	246
260	264	276	281	283	289	296
307	309	312	317	324	341	353

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

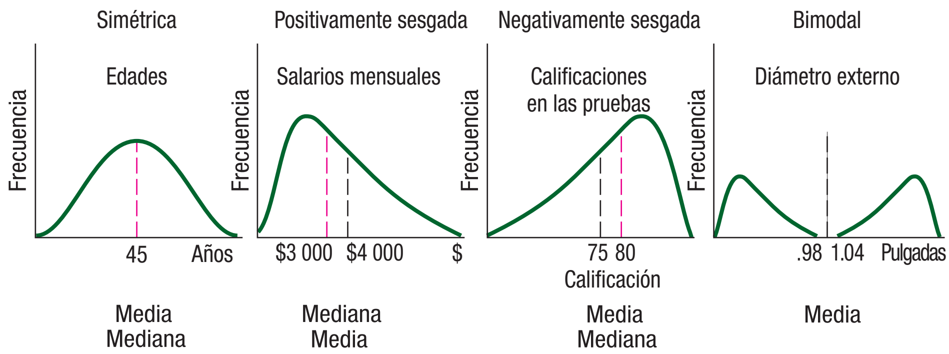
Sesgo

OA4-5 Calcular y entender el coeficiente de sesgo.

En el capítulo 3 se trataron las medidas de ubicación central de un conjunto de observaciones por medio de la presentación de un informe sobre la media, la mediana y la moda. También se describieron medidas que muestran el grado de propagación o variación de un conjunto de datos, como el rango y la desviación estándar.

Otra característica de un conjunto de datos es la forma. Hay cuatro formas: 1) simétrica, 2) con sesgo positivo, 3) con sesgo negativo y 4) bimodal. En un conjunto **simétrico** de observaciones la media y la mediana son iguales, y los datos se dispersan uniformemente en torno a estas. Los valores debajo de la media y la mediana constituyen una imagen especular de los datos que están sobre estas medidas. Un conjunto de valores se encuentra **sesgado a la derecha** o es **positivamente sesgado** si existe un solo pico y los valores se extienden mucho más allá a la derecha del pico que a la izquierda de este. En este caso, la media es más grande que la mediana. En una distribución **negativamente sesgada** existe un solo pico, pero las observaciones se extienden más a la izquierda, en dirección negativa. En una distribución negativamente sesgada, la media es menor que la mediana. Las distribuciones positivamente sesgadas son más comunes. Con frecuencia, los salarios obedecen este patrón. Piense en los salarios del personal de una pequeña compañía con casi 100 empleados. El presidente y unos cuantos altos ejecutivos recibirían mucho más que los demás trabajadores, por lo que la distribución de salarios mostraría un sesgo positivo. Una **distribución bimodal** tendrá dos o más picos. Este es un caso frecuente cuando los valores provienen de dos o más poblaciones. Esta información se resume en la gráfica 4.1 (página siguiente).

En la literatura estadística se utilizan diversas fórmulas para calcular el sesgo. La más sencilla, ideada por el profesor Karl Pearson (1857-1936), se basa en la diferencia entre la media y la mediana.



GRÁFICA 4.1 Formas de los polígonos de frecuencias

COEFICIENTE DE SESGO DE PEARSON

$$sk = \frac{3(\bar{x} - \text{Mediana})}{s} \qquad \textbf{[4.2]}$$

De acuerdo con esta expresión, el sesgo puede variar de -3 hasta 3. Un valor próximo a -3, como -2.57, indica un sesgo negativo considerable; otro como 1.63 indica un sesgo positivo moderado; y un valor de 0, que ocurre cuando la media y la mediana son iguales, indica que la distribución es simétrica y que no hay ningún sesgo.

En esta obra se presentan resultados que se obtuvieron con Minitab y Excel. Con ambos se calcula un valor del coeficiente de sesgo basado en las desviaciones de la media elevadas al cubo. La fórmula es la siguiente:

COEFICIENTE DE SESGO CALCULADO CON SOFTWARE

$$sk = \frac{n}{(n-1)(n-2)}\left[\sum\left(\frac{x-\bar{x}}{s}\right)^3\right] \qquad \textbf{[4.3]}$$

La fórmula [4.3] permite comprender la idea de sesgo. El miembro derecho de la fórmula es la diferencia entre cada valor y la media, dividida entre la desviación estándar. Esto corresponde a la porción $(x - \bar{x})/s$ de la fórmula. A esto se le llama **estandarización.** El concepto de estandarización de un valor se analiza con más detalle en el capítulo 7, cuando se describe la distribución de probabilidad normal. En este punto, observe que el resultado consiste en la diferencia entre cada valor y la media en unidades de desviación estándar. Si la diferencia es positiva, el valor particular es más grande que la media; si el valor es negativo, la cantidad estandarizada es menor que la media. Al elevar al cubo estos valores se conserva la información relativa a la diferencia. Recuerde que en la fórmula de la desviación estándar (vea la fórmula [3.10]) la diferencia entre cada valor y la media se elevó al cuadrado, de manera que, como resultado, ningún valor era negativo.

Si el conjunto de valores que se estudia es simétrico, al elevar al cubo y sumar todos los valores estandarizados, el resultado se aproximaría a cero. Si hay varios valores grandes, claramente separados unos de otros, la suma de las diferencias al cubo sería positiva y grande. Si hay valores pequeños claramente separados de otros, la suma de las diferencias al cubo será negativa.

Un ejemplo ilustrará la idea de sesgo.

ESTADÍSTICA EN ACCIÓN

El difunto Stephen Jay Gould (1941-2002) fue profesor de zoología y geología en la Universidad de Harvard. En 1982 se le diagnosticó cáncer y le dieron ocho meses de vida. No obstante, y sin darse por vencido, mostró en su investigación que la distribución de tiempos de supervivencia se encuentra drásticamente sesgada a la derecha y que no solo 50% de pacientes de tipos similares de cáncer sobreviven más de ocho meses, sino que el tiempo de supervivencia podía ser de años, no de meses. De hecho, el doctor Gould vivió otros 20 años. Con base en su experiencia escribió un ensayo varias veces publicado que se tituló "The Median Is not the Message" (la mediana no es el mensaje).

EJEMPLO

En seguida aparecen las utilidades por acción que obtuvo una muestra de 15 compañías de software durante el año 2013. Las utilidades por acción se encuentran ordenadas de manera ascendente.

\$0.09	\$0.13	\$0.41	\$0.51	\$ 1.12	\$ 1.20	\$ 1.49	\$3.18
3.50	6.36	7.83	8.92	10.13	12.99	16.40	

Calcule la media, la mediana y la desviación estándar. Determine el coeficiente de sesgo utilizando los métodos de Pearson y de software. ¿Qué concluye respecto de la forma de la distribución?

SOLUCIÓN

Estos son los datos de la muestra, así que aplique la fórmula [3.2] para determinar la media:

$$\bar{x} = \frac{\Sigma x}{n} = \frac{\$74.26}{15} = \$4.95$$

La mediana es el valor intermedio de un conjunto de datos ordenados de manera ascendente. En este caso, el valor medio es 3.18 dólares; así, la mediana de las utilidades por acción es $3.18.

Emplee la fórmula [3.10] para calcular la desviación estándar de la muestra:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{(\$0.09 - \$4.95)^2 + \cdots + (\$16.40 - \$4.95)^2}{15 - 1}} = \$5.22$$

El coeficiente de sesgo de Pearson es 1.017, calculado de la siguiente manera:

$$sk = \frac{3(\bar{x} - \text{Mediana})}{s} = \frac{3(\$4.95 - \$3.18)}{\$5.22} = 1.017$$

Esto indica que existe un sesgo positivo moderado en los datos de las utilidades por acción.

Cuando se utiliza el método del software resulta un valor similar, aunque no exactamente el mismo. Los detalles de los cálculos aparecen en la tabla 4.2. Para comenzar, determine la diferencia entre las utilidades por acción y la media; divida el resultado entre la desviación estándar. Recuerde que esto se llama estandarización. Luego, eleve al cubo, es decir, eleve a la tercera potencia el resultado del primer paso. Por último, sume los valores elevados al cubo. Los detalles en el caso de la primera compañía (la que tiene utilidades de 0.09 dólares por acción) son:

$$\left(\frac{x - \bar{x}}{s}\right)^3 = \left(\frac{0.09 - 4.95}{5.22}\right)^3 = (-0.9310)^3 = -0.8070$$

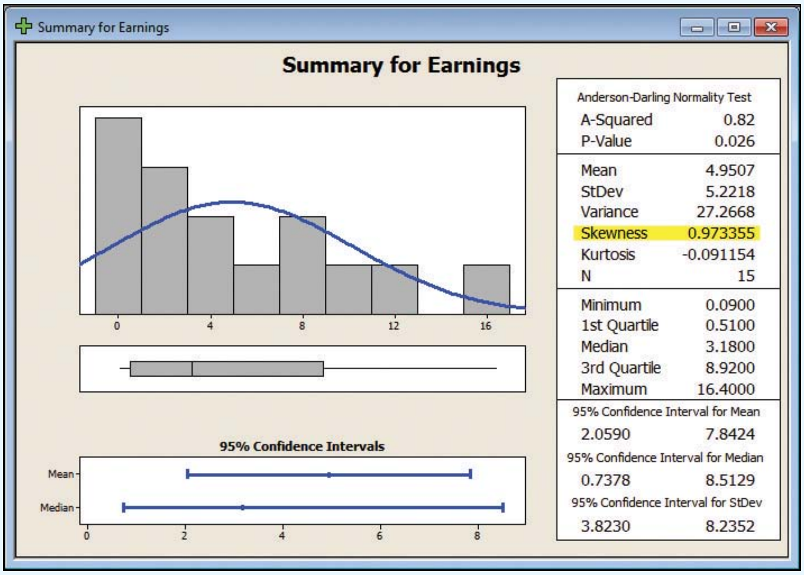
TABLA 4.2 Cálculo del coeficiente de sesgo

Utilidades por acción	$\frac{x - \bar{x}}{s}$	$\left(\frac{x - \bar{x}}{s}\right)^3$
0.09	−0.9310	−0.8070
0.13	−0.9234	−0.7873
0.41	−0.8697	−0.6579
0.51	−0.8506	−0.6154
1.12	−0.7337	−0.3950
1.20	−0.7184	−0.3708
1.49	−0.6628	−0.2912
3.18	−0.3391	−0.0390
3.50	−0.2778	−0.0214
6.36	0.2701	0.0197
7.83	0.5517	0.1679
8.92	0.7605	0.4399
10.13	0.9923	0.9772
12.99	1.5402	3.6539
16.40	2.1935	10.5537
		11.8274

Al sumar los 15 valores cúbicos el resultado es 11.8274. Es decir, el término $\Sigma[(x - \bar{x})/s]^3 = 11.8274$. Para determinar el coeficiente de sesgo utilice la fórmula [4.3], con $n = 15$.

$$sk = \frac{n}{(n - 1)(n - 2)} \Sigma\left(\frac{x - \bar{x}}{s}\right)^3 = \frac{15}{(15 - 1)(15 - 2)} = (11.8274) = 0.975$$

La conclusión es que los valores de la página 99, las utilidades por acción se encuentran un tanto sesgados positivamente. En el siguiente diagrama, de Minitab, se muestran las medidas descriptivas (media, mediana y desviación estándar) de los datos por utilidades por acción. También incluye el coeficiente de sesgo y un histograma con una curva con forma de campana superpuesta.



AUTOEVALUACIÓN 4-4

Una muestra de cinco capturistas de datos que laboran en la oficina de impuestos de Horry County revisó las siguientes cantidades de expedientes fiscales durante la última hora: 73, 98, 60, 92 y 84.

(a) Calcule la media, la mediana y la desviación estándar.
(b) Calcule el coeficiente de sesgo con el método de Pearson.
(c) Calcule el coeficiente de sesgo usando un paquete de software.
(d) ¿Qué conclusión obtiene respecto del sesgo de los datos?

EJERCICIOS

En el caso de los ejercicios 19 a 22:

a. Calcule la media, la mediana y la desviación estándar.
b. Calcule el coeficiente de sesgo con el método de Pearson.
c. Estime el coeficiente de sesgo con un paquete de software.

19. Los siguientes valores son los sueldos iniciales (en miles de dólares) de una muestra de cinco graduados de contabilidad, quienes aceptaron puestos de contaduría pública el año anterior.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

36.0	26.0	33.0	28.0	31.0

20. En la siguiente lista se registran los salarios (en miles de dólares) de una muestra de 15 directores de finanzas de la industria electrónica

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

$516.0	$548.0	$566.0	$534.0	$586.0	$529.0
546.0	523.0	538.0	523.0	551.0	552.0
486.0	558.0	574.0			

21. A continuación aparece una lista de las comisiones (en miles de dólares) que percibieron el año anterior los representantes de ventas de Furniture Patch, Inc.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

$ 3.9	$ 5.7	$ 7.3	$10.6	$13.0	$13.6	$15.1	$15.8	$17.1
17.4	17.6	22.3	38.6	43.2	87.7			

22. La lista que sigue indica los salarios de los Yankees de Nueva York en 2012.

Jugador	Salario	Jugador	Salario
Alex Rodriguez	$30 000 000	Boone Logan	$1 875 000
Mark Teixeira	23 125 000	Joba Chamberlain	1 675 000
CC Sabathia	23 000 000	David Robertson	1 600 000
Derek Jeter	15 729 364	Raul Ibanez	1 100 000
Mariano Rivera	14 940 025	Eric Chavez	900 000
Robinson Cano	14 000 000	Michael Pineda	528 475
Rafael Soriano	11 000 000	Ivan Nova	527 200
Nick Swisher	10 250 000	Clay Rapada	525 000
Curtis Granderson	10 000 000	Eduardo Nunez	523 800
Hiroki Kuroda	10 000 000	Cory Wade	508 925
Russell Martin	7 500 000	David Aardsma	500 000
Freddy Garcia	4 000 000	Chris Stewart	482 500
Pedro Feliciano	3 750 000	Austin Romine	482 000
Phil Hughes	3 200 000	Cesar Cabral	480 000
Brett Gardner	2 800 000	Brad Meyers	480 000
Andruw Jones	2 000 000		

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Descripción de la relación entre dos variables

OA4-6
Trazar e interpretar un diagrama de dispersión.

En el capítulo 2 y en la primera sección de este se han expuesto técnicas gráficas para resumir la distribución de una sola variable. En el capítulo 2 se empleó un histograma para resumir las ganancias por vehículos vendidos en Applewood Auto Group. Las herramientas que se usaron en este capítulo fueron los diagramas de puntos y las gráficas de tallo y hojas para representar visualmente un conjunto de datos. En tanto que aparece una sola variable, se habla de datos **univariados.**

Hay situaciones en las que la relación entre dos variables se estudia y representa visualmente. Al estudiar la relación entre ellas, se hace referencia a los datos como **bivariados.** Con frecuencia, los analistas de datos tratan de entender este tipo de relación. He aquí algunos ejemplos:

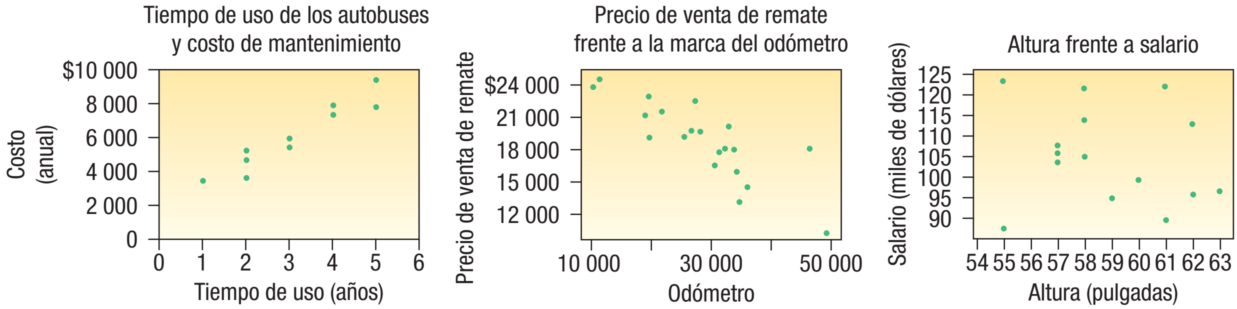
- Tybo and Associates es una firma de abogados que se anuncia mucho en televisión. Los socios están considerando la forma de incrementar su presupuesto publicitario. Antes de hacerlo, les gustaría conocer la relación entre la cantidad que se gasta al mes en publicidad y la cantidad total de cuentas por cobrar en dicho mes. En otras palabras, ¿un incremento de la suma que se gasta en publicidad dará como resultado un incremento de las cuentas por cobrar?
- Coastal Realty estudia sus precios de venta de casas. ¿Qué variables parecen estar relacionadas con ellos? Por ejemplo, ¿las casas más grandes se venden a un precio superior que las más pequeñas? Es probable. Por ello, Coastal tendría que estudiar la relación entre el área en pies cuadrados y el precio de venta.
- El doctor Stephen Givens es experto en desarrollo humano. Estudia la relación entre la altura de los padres y la de sus hijos. Es decir, ¿los padres altos tienden a tener hijos altos? ¿Esperaría usted que Dwight Howard, el basquetbolista profesional de seis pies y once pulgadas de altura y 250 libras de peso tuviera hijos relativamente altos?

Una técnica gráfica útil para mostrar la relación entre variables es el **diagrama de dispersión.**

Para trazar un diagrama de dispersión se necesitan dos variables. Una de las variables se escala sobre el eje horizontal (eje *X*) de una gráfica y la otra variable, a lo largo del eje vertical (eje *Y*). Por lo general, una de las variables depende hasta cierto grado de la otra. En el tercer ejemplo citado, la altura del hijo *depende* de la altura del padre. Así que la altura del padre se representa en el eje horizontal y la del hijo, sobre el eje vertical.

Un software de estadística, como Excel, sirve para ejecutar la función de trazo. *Precaución:* siempre se debe tener cuidado en la escala. Al cambiarla, ya sea del eje vertical o del eje horizontal, se afecta la fuerza de la relación visual.

En la siguiente página aparecen tres diagramas de dispersión (gráfica 4.2). El de la izquierda muestra una mayor relación entre el tiempo de uso y el costo de mantenimiento durante el año anterior de una muestra de 10 autobuses propiedad de la ciudad de Cleveland, Ohio. Observe que a medida que se incrementa el tiempo de uso del autobús, también aumenta el costo anual de mantenimiento. El ejemplo del centro, relativo a una muestra de 20 vehículos, muestra una fuerte relación indirecta entre la lectura del odómetro y el precio de venta de remate. Es decir, conforme aumente el número de millas recorridas, el precio de venta de remate se reduce. El ejemplo de la derecha describe la relación entre la altura y el salario anual de una muestra de 15 supervisores de turno. Esta gráfica indica que existe poca relación entre la altura y el salario anual.



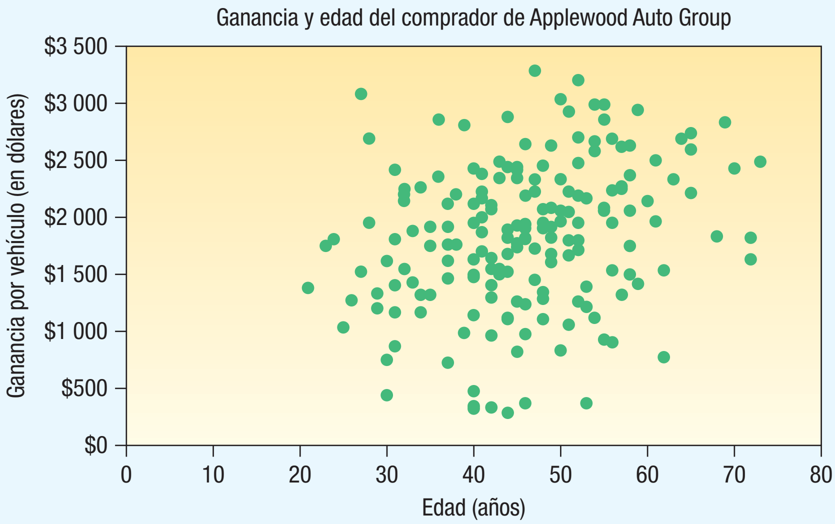
GRÁFICA 4.2 Tres ejemplos de diagramas de dispersión

EJEMPLO

En la introducción del capítulo 2 aparecen datos de Applewood Auto Group. Se reunió información sobre diversas variables, entre ellas la ganancia que se obtuvo por la venta de 180 vehículos el mes anterior. Además del monto de la ganancia en cada venta, otra de las variables es la edad del comprador. ¿Existe alguna relación entre la ganancia que se obtuvo por la venta de un vehículo y la edad del comprador? ¿Sería razonable concluir que se gana más en los vehículos que adquieren los compradores de más edad?

SOLUCIÓN

Es posible investigar la relación entre la ganancia por vehículo vendido y la edad del comprador con un diagrama de dispersión. Represente la escala de edad sobre el eje horizontal (*X*) y la ganancia sobre el eje vertical (*Y*). Asuma que la ganancia depende de la edad del comprador. A medida que la gente se hace mayor, sus ingresos se incrementan y compra autos más caros, lo que a su vez produce mayores ganancias. Utilice Excel para crear un diagrama de dispersión (los comandos están en el apéndice C).



El diagrama de dispersión muestra una relación positiva entre ambas variables. No parece haber mucha relación entre la ganancia por vehículo y la edad del comprador. En el capítulo 13 se estudiará de manera más amplia la relación entre variables, incluso se calcularán varias medidas numéricas para expresar la relación entre variables.

En el ejemplo anterior hay una débil relación positiva, o directa, entre las variables. Sin embargo, hay muchos casos en los que existe una relación entre las variables, pero dicha relación es inversa o negativa. Por ejemplo:

- El valor de un vehículo y la cantidad de millas recorridas. Conforme esta se incrementa, el valor del vehículo desciende.
- La prima de un seguro de automóvil y la edad del conductor. Las cuotas de automóvil tienden ser más altas para los adultos jóvenes y menores para personas de más edad.
- En el caso de muchos oficiales encargados de hacer que se cumpla la ley, conforme aumentan los años de trabajo, la cantidad de multas de tránsito disminuye. Esto puede deberse a que el personal se torna más liberal en sus interpretaciones o a que quizá tengan puestos de supervisión y no un cargo en el que puedan levantar tantas multas. Pero en cualquier caso, conforme la edad aumenta, la cantidad de multas se reduce.

Tablas de contingencia

OA4-7
Construir e interpretar una tabla de contingencia.

Un diagrama de dispersión requiere que las dos variables sean por lo menos de escala de intervalo. En el ejemplo de Applewood Auto Group, tanto la edad como la ganancia de la venta son variables de escala de razón. La altura también es una escala de razón, según la manera en la que se utilizó en el estudio de la relación entre la estatura de los padres y la de los hijos. ¿Y si desea estudiar la relación entre dos variables cuando una o ambas son de escala nominal u ordinal? En este caso, debe registrar los resultados en una **tabla de contingencia.**

TABLA DE CONTINGENCIA Se utiliza para clasificar observaciones de acuerdo con dos características identificables.

Una tabla de contingencia es una tabulación cruzada que resume simultáneamente dos variables de interés. Por ejemplo:

- Los estudiantes en una universidad se clasifican por género y clase (de primero, segundo, penúltimo o último año).
- Un producto se clasifica como aceptable o inaceptable y de acuerdo con el turno (matutino, vespertino, nocturno) en el que se le fabrica.
- Un votante de una escuela que lleva a cabo un referendo para otorgar becas se clasifica de acuerdo con su afiliación partidista (demócrata, republicano u otro), y el número de hijos que asisten a la escuela del distrito (0, 1, 2, etc.).

EJEMPLO

Hay cuatro distribuidoras en Applewood Auto Group. Suponga que desea comparar la ganancia que se obtuvo por cada vehículo vendido en una concesionaria en particular. Dicho de otra forma, ¿existe una relación entre el monto de ganancia y la distribuidora?

SOLUCIÓN

En una tabla de contingencia, ambas variables solo necesitan ser nominales u ordinales. En este ejemplo, la variable “concesionaria” es nominal y la variable “ganancia” es de razón. Para convertir “ganancia” en una variable ordinal hay que clasificarla en dos categorías: aquellos casos en los que la ganancia que se obtuvo es mayor a la mediana, y aquellos en que es menor. En el ejemplo del subtema

Tabla de contingencia sobre la relación entre ganancia y concesionaria

Abajo/Arriba Ganancia mediana	Kane	Olean	Sheffield	Tionesta	Total
Por abajo	25	20	19	26	90
Por arriba	27	20	26	17	90
Total	52	40	45	43	180

"Solución con software" del capítulo 3 se calcula que la ganancia mediana por todas las ventas del mes anterior en Applewood Auto Group es de 1 882.50 dólares.

Si se organiza la información en una tabla de contingencia, es posible comparar la ganancia de las cuatro distribuidoras. Se observa lo siguiente:

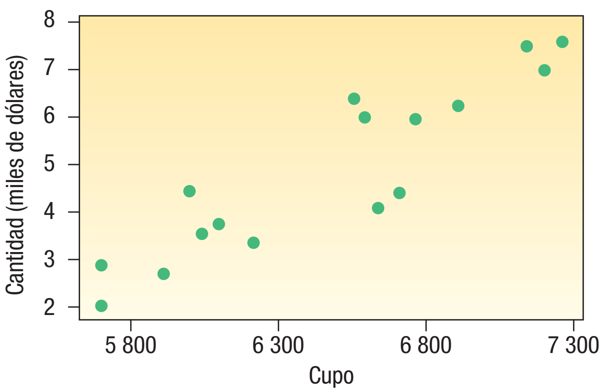
- De la columna "Total" (derecha), 90 de los 180 autos vendidos dieron una ganancia por encima de la mediana, y la otra mitad, por debajo. Esto era lo esperado, dada la definición de mediana.
- En el caso de la distribuidora Kane, 25 de los 52 vehículos, o 48%, se vendieron con una ganancia mayor a la mediana.
- El porcentaje de ganancias por encima de la mediana de las otras concesionarias es 50% en el caso de Olean, 42% en el de Sheffield y 60% en el de Tionesta.

En el capítulo 5 se amplía el tema de las tablas de contingencia al estudiar la probabilidad y en el capítulo, 15 al abordar los métodos no paramétricos de análisis.

AUTOEVALUACIÓN 4-5

El grupo de rock Blue String Beans está de gira por Estados Unidos. En el siguiente diagrama se muestra la relación entre el cupo para el concierto y el ingreso en miles de dólares en una muestra de conciertos.

(a) ¿Qué nombre recibe el diagrama?
(b) ¿Cuántos conciertos se estudiaron?
(c) Calcule los ingresos de un concierto con lleno total.
(d) ¿Cómo caracterizaría la relación entre ingresos y cupo? ¿Es fuerte o débil, directa o inversa?



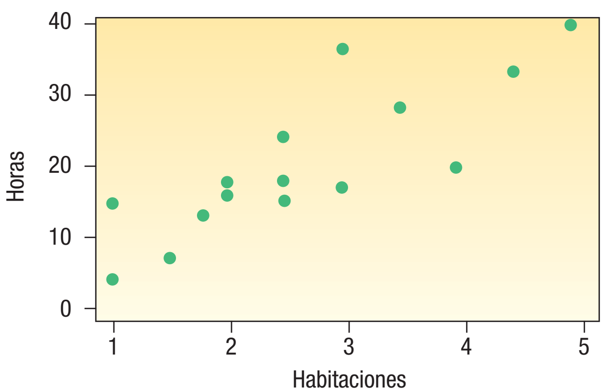
EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

23. Elabore el diagrama de dispersión de los siguientes datos tomados de una muestra. ¿Cómo describiría la relación entre los valores?

Valor *x*	Valor *y*	Valor *x*	Valor *y*
10	6	11	6
8	2	10	5
9	6	7	2
11	5	7	3
13	7	11	7

24. Silver Springs Moving and Storage, Inc., estudia la relación que existe entre el número de habitaciones en una mudanza y el número de horas que se requieren de trabajo para completarla. Como parte del análisis, el director de finanzas de Silver Springs creó el siguiente diagrama de dispersión.



a. ¿Cuántas mudanzas se incluyen en la muestra?
b. ¿Parecería que se requieren más horas de trabajo si la cantidad de habitaciones aumenta, o las horas de trabajo disminuyen si la cantidad de habitaciones se reduce?

25. El director de planeación de Devine Dining, Inc., desea estudiar la relación entre el género de un huésped y si este ordena postre. Para investigar esta relación, recopiló la siguiente información de 200 consumidores.

	Género		
Orden de postre	**Hombre**	**Mujer**	**Total**
Sí	32	15	47
No	68	85	153
Total	100	100	200

a. ¿Cuál es el nivel de medición de las dos variables?
b. ¿Qué nombre recibe esta tabla?
c. A partir de la evidencia que ofrece la tabla, ¿los hombres piden más postre que las mujeres? Explique su respuesta.

26. Sky Resorts, Inc., de Vermont, considera su fusión con Gulf Shores, Inc., de Alabama. El consejo directivo encuestó a 50 accionistas acerca de su posición sobre la fusión. Los resultados aparecen enseguida.

	Opinión			
Cantidad de acciones que posee	**A favor**	**En contra**	**Indeciso**	**Total**
Hasta 200	8	6	2	16
200 hasta 1 000	6	8	1	15
Arriba de 1 000	6	12	1	19
Total	20	26	4	50

a. ¿Cuál es el nivel de medición que se empleó en la tabla?
b. ¿Qué nombre recibe esta tabla?
c. ¿Qué grupo parece oponerse con más fuerza a la fusión?

RESUMEN DEL CAPÍTULO

I. Un diagrama de puntos muestra el rango de valores sobre el eje horizontal y la cantidad de observaciones de cada valor sobre el eje vertical.
  A. Un diagrama de puntos muestra los detalles de cada observación.
  B. Es útil para comparar dos o más conjuntos de datos.

II. Un diagrama de tallo y hojas constituye una alternativa al histograma.
  A. El dígito principal es el tallo y el dígito secundario, la hoja.
  B. Algunas de las ventajas de un diagrama de tallo y hojas sobre un histograma son las siguientes:
    1. La identidad de cada observación no se pierde.
    2. Los dígitos proporcionan una representación de la distribución.
    3. Las frecuencias acumulativas también se exhiben.

III. Las medidas de localización describen la forma de un conjunto de observaciones.
  A. Los cuartiles dividen un conjunto de observaciones en cuatro partes iguales.
    1. Veinticinco por ciento de las observaciones son menores que el primer cuartil, 50% son menores que el segundo cuartil y 75% son menores que el tercer cuartil.
    2. El rango intercuartil es la diferencia entre el tercer y primer cuartiles.
  B. Los deciles dividen a un conjunto de observaciones en diez partes iguales y los percentiles, en 100 partes iguales.

IV. Un diagrama de caja es una representación gráfica de un conjunto de datos.
  A. La caja se traza encerrando las regiones entre el primer y tercer cuartiles.
    1. Se dibuja una línea en el interior de la caja en el valor intermedio.
    2. Los segmentos punteados se prolongan a partir del tercer cuartil hasta el valor más alto con el fin de mostrar el 25% más alto, y a partir del primer cuartil hasta el valor más bajo con el fin de mostrar el 25% más bajo de los valores.

B. Un diagrama de caja se basa en cinco estadísticos: los valores máximo y mínimo, el primer y tercer cuartiles, y la mediana.

V. El coeficiente de sesgo es una medida de la simetría de una distribución.

A. Existen dos fórmulas para determinar el coeficiente de sesgo.

1. La fórmula que elaboró Pearson es:

$$sk = \frac{3(\bar{x} - \text{Mediana})}{s} \qquad \textbf{[4.2]}$$

2. El coeficiente de sesgo calculado con un software estadístico es:

$$sk = \frac{n}{(n-1)(n-2)}\left[\Sigma\left(\frac{x-\bar{x}}{s}\right)^3\right] \qquad \textbf{[4.3]}$$

VI. Un diagrama de dispersión es una herramienta gráfica para representar la relación entre dos variables.

A. Ambas variables se miden con escalas de intervalo o de razón.

B. Si la propagación de los puntos se dirige de la parte inferior izquierda hacia la parte superior derecha, las variables que se estudian se encuentran directa o positivamente relacionadas.

C. Si la dispersión de los puntos se orienta de la parte superior izquierda hacia la parte inferior derecha, las variables se encuentran relacionadas inversa o negativamente.

VII. Una tabla de contingencia se utiliza para clasificar observaciones de escala nominal de acuerdo con dos características.

CLAVE DE PRONUNCIACIÓN

Símbolo	**Significado**	**Pronunciación**
L_p	Ubicación del percentil	*L sub p*
Q_1	Primer cuartil	*Q sub 1*
Q_2	Tercer cuartil	*Q sub 2*

EJERCICIOS DEL CAPÍTULO

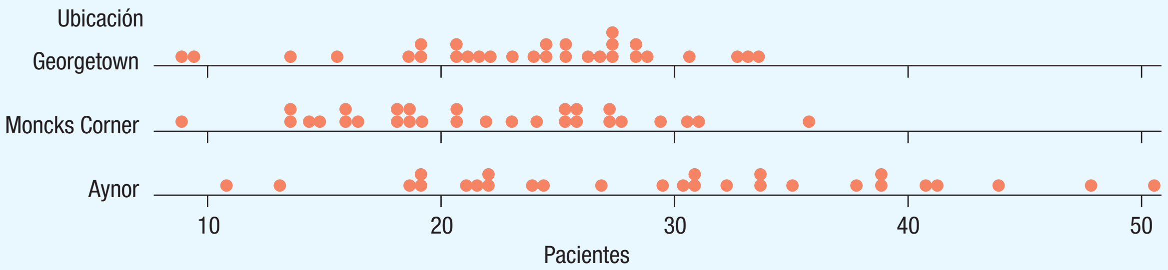
27. Se le preguntó a una muestra de estudiantes que asiste a la Southern Florida University por la cantidad de actividades sociales en las que participaron la semana anterior. El diagrama que aparece enseguida se construyó a partir de datos tomados de la muestra.

a. ¿Cuál es el nombre que se da a este diagrama?

b. ¿Cuántos estudiantes se incluyeron en el estudio?

c. ¿Cuántos estudiantes informaron que no asistían a ninguna actividad social?

28. Doctor's Care es una clínica ambulatoria que tiene sucursales en Georgetown, Monks Corners y Aynor; en esta los pacientes reciben tratamiento por lesiones menores, resfriados y gripes, y se les practican exámenes físicos. En los siguientes diagramas se muestra la cantidad de pacientes que se trataron en las tres sucursales el mes anterior.



Describa el número de pacientes atendidos en las tres sucursales cada día. ¿Cuáles son los números máximo y mínimo de pacientes que se atendieron en cada una de las sucursales?

29. A continuación se proporciona el tamaño de pantalla de 23 televisores LCD. Elabore un diagrama de tallos y hojas de esta variable.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

46	52	46	40	42	46	40	37	46	40	52	32	37	32	52
40	32	52	40	52	46	46	52							

30. En la siguiente tabla se muestran las 25 compañías (ordenadas por capitalización del mercado) que operan en el área de Washington, DC, junto al año en que se fundaron y la cantidad de empleados. Elabore un diagrama de tallo y hojas de estas variables y escriba una breve descripción de sus hallazgos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Compañía	Año de fundación	Empleados
AES Corp.	1981	30 000
American Capital Strategies Ltd.	1986	484
AvalonBay Communities Inc.	1978	1 767
Capital One Financial Corp.	1995	31 800
Constellation Energy Group Inc.	1816	9 736
Coventry Health Care Inc.	1986	10 250
Danaher Corp.	1984	45 000
Dominion Resources Inc.	1909	17 500
Fannie Mae	1938	6 450
Freddie Mac	1970	5 533
Gannett Co.	1906	49 675
General Dynamics Corp.	1952	81 000
Genworth Financial Inc.	2004	7 200
Harman International Industries Inc.	1980	11 246
Host Hotels & Resorts Inc.	1927	229
Legg Mason Inc.	1899	3 800
Lockheed Martin Corp.	1995	140 000
Marriott International Inc.	1927	151 000
MedImmune Inc.	1988	2 516
NII Holdings Inc.	1996	7 748
Norfolk Southern Corp.	1982	30 594
Pepco Holdings Inc.	1896	5 057
Sallie Mae	1972	11 456
Sprint Nextel Corp.	1899	64 000
T. Rowe Price Group Inc.	1937	4 605
The Washington Post Co.	1877	17 100

31. En años recientes, como consecuencia de las bajas tasas de interés, muchos propietarios de casas refinanciaron sus créditos. Linda Lahey es agente hipotecaria de Down River Federal Savings and Loan. A continuación aparecen las sumas refinanciadas de 20 préstamos a los que les dio curso la semana pasada. Los datos se expresan en miles de dólares y se encuentran ordenados de menor a mayor.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

59.2	59.5	61.6	65.5	66.6	72.9	74.8	77.3	79.2
83.7	85.6	85.8	86.6	87.0	87.1	90.2	93.3	98.6
100.2	100.7							

a. Calcule la mediana, el primer y tercer cuartiles.
b. Determine los percentiles 26° y 83°.
c. Trace un diagrama de caja de los datos.

32. La industria disquera de Estados Unidos lleva a cabo un estudio sobre la cantidad de discos compactos de música que poseen las personas de la tercera edad y los adultos jóvenes. La información se muestra enseguida.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Adultos de la tercera edad									
28	35	41	48	52	81	97	98	98	99
118	132	133	140	145	147	153	158	162	174
177	180	180	187	188					

Adultos jóvenes									
81	107	113	147	147	175	183	192	202	209
233	251	254	266	283	284	284	316	372	401
417	423	490	500	507	518	550	557	590	594

a. Calcule la mediana, el primer y tercer cuartiles del número de discos que poseen los ciudadanos de la tercera edad. Diseñe un diagrama de caja de la información.

b. Calcule la mediana, el primer y tercer cuartiles del número de discos que poseen los adultos jóvenes. Diseñe un diagrama de caja de la información.

c. Compare las cantidades de discos que poseen ambos grupos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

33. Las oficinas centrales de la empresa Bank.com, una empresa nueva de internet que realiza todas las transacciones bancarias a través de la red, se localizan en el centro de Filadelfia. El director de recursos humanos lleva a cabo un estudio relacionado con el tiempo que los empleados invierten en llegar al trabajo. La ciudad hace planes para ofrecer incentivos a las empresas que se ubiquen en el centro si estimulan a sus empleados a utilizar el transporte público. A continuación aparece una lista donde se muestra el tiempo requerido en la mañana para llegar al trabajo según el empleado haya utilizado el transporte público o su automóvil.

Transporte público									
23	25	25	30	31	31	32	33	35	36
37	42								

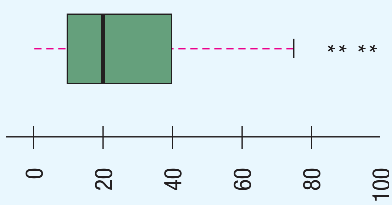
Particular									
32	32	33	34	37	37	38	38	38	39
40	44								

a. Calcule la mediana, el primer y tercer cuartiles del tiempo de desplazamiento de los empleados que utilizaron el transporte público. Elabore un diagrama de caja para la información.

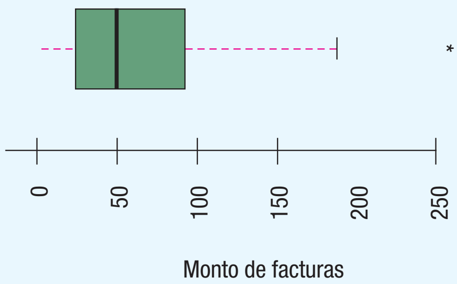
b. Calcule la mediana, el primer y tercer cuartiles del tiempo de desplazamiento de los empleados en su propio vehículo. Elabore un diagrama de caja para la información.

c. Compare los tiempos de ambos grupos.

34. En el siguiente diagrama de caja se muestra la cantidad de diarios que se publican en cada estado y en el distrito de Columbia. Redacte un breve informe que resuma la cantidad que se publicó. Cerciórese de incluir información sobre los valores del primer y tercer cuartiles, la mediana y si existe algún sesgo. Si hay datos atípicos, calcule su valor.



35. Walter Gogel Company es un proveedor industrial de cinturones de seguridad, herramientas y resortes. Las sumas de sus ingresos varían mucho, desde menos de 20 dólares hasta más de 400 dólares. Durante enero enviaron 80 facturas. El siguiente es un diagrama de caja de estas facturas. Redacte un breve informe que resuma los montos de las facturas. Cerciórese de incluir información sobre los valores del primer y tercer cuartiles, la mediana y si existe algún sesgo. Si hay datos atípicos, aproxime el valor de estas facturas.



36. La American Society of PeriAnesthesia Nurses (ASPAN: **www.aspan.org**) es una organización estadounidense que agrupa a enfermeras que se desempeñan en el cuidado preanestesia y posanestesia en cirugías ambulatorias. La organización comprende 40 componentes que se enlistan a continuación.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Estado/Región	Membresía	Estado/Región	Membresía
Alabama	95	Illinois	562
Arizona	399	Indiana	270
Maryland, Delaware, DC	531	Iowa	117
Connecticut	239	Kentucky	197
Florida	631	Louisiana	258
Georgia	384	Michigan	411
Hawaii	73	Massachusetts	480
Maine	97	California	1 165
Minnesota, Dakotas	289	New Mexico	79
Missouri, Kansas	282	Pennsylvania	575
Mississippi	90	Rhode Island	53
Nebraska	115	Colorado	409
North Carolina	542	South Carolina	237
Nevada	106	Texas	1 026
New Jersey, Bermuda	517	Tennessee	167
Alaska, Idaho, Montana, Oregon, Washington	708	Utah	67
New York	891	Virginia	414
Ohio	708	Vermont, New Hampshire	144
Oklahoma	171	Wisconsin	311
Arkansas	68	West Virginia	62

Utilice un software estadístico para realizar las siguientes instrucciones.

a. Encuentre la media, la mediana y la desviación estándar del número de miembros por componente.

b. Ubique el coeficiente de sesgo mediante el software. ¿Cuál es su conclusión con respecto a la forma de la distribución del tamaño del componente?

c. Calcule el primer y tercer cuartiles utilizando la fórmula [4.1].

d. Desarrolle un diagrama de caja. ¿Hay datos atípicos? ¿Cuáles componentes son atípicos? ¿Cuáles son los límites de los componentes atípicos?

37. McGivern Jewelers se ubica en Levis Square Mall, al sur de Toledo, Ohio. Recientemente publicó un anuncio en el periódico local en el que indicaba la forma, tamaño, precio y grado de corte de 33 de sus diamantes en existencia. La información se muestra en el cuadro de la página siguiente.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

a. Diseñe un diagrama de caja con la variable “precio” y haga algún comentario sobre el resultado. ¿Hay valores atípicos? ¿Cuál es la mediana del precio? ¿Cuál es el valor del primer y tercer cuartiles?

b. Diseñe un diagrama de caja de la variable “tamaño” y haga comentarios sobre el resultado. ¿Hay valores atípicos? ¿Cuál es la mediana del precio? ¿Cuál es el valor del primer y tercer cuartiles?

c. Diseñe un diagrama de dispersión entre las variables “precio” y “tamaño”. Coloque el precio en el eje vertical y el tamaño en el eje horizontal. ¿Le parece que hay alguna relación entre las dos variables? ¿La relación es directa o indirecta? ¿Parece que alguno de los puntos es diferente de los demás?

Forma	Tamaño (quilates)	Precio	Grado de corte	Forma	Tamaño (quilates)	Precio	Grado de corte
Princesa	5.03	$44 312	Corte ideal	Redonda	0.77	$2 828	Corte ultraideal
Redonda	2.35	20 413	Corte perfeccionado	Oval	0.76	3 808	Corte perfeccionado
Redonda	2.03	13 080	Corte ideal	Princesa	0.71	2 327	Corte perfeccionado
Redonda	1.56	13 925	Corte ideal	Marquesa	0.71	2 732	Buen corte
Redonda	1.21	7 382	Corte ultraideal	Redonda	0.70	1 915	Corte perfeccionado
Redonda	1.21	5 154	Corte promedio	Redonda	0.66	1 885	Corte perfeccionado
Redonda	1.19	5 339	Corte perfeccionado	Redonda	0.62	1 397	Buen corte
Emerald	1.16	5 161	Corte ideal	Redonda	0.52	2 555	Corte perfeccionado
Redonda	1.08	8 775	Corte ultraideal	Princesa	0.51	1 337	Corte ideal
Redonda	1.02	4 282	Corte perfeccionado	Redonda	0.51	1 558	Corte perfeccionado
Redonda	1.02	6 943	Corte ideal	Redonda	0.45	1 191	Corte perfeccionado
Marquesa	1.01	7 038	Buen corte	Princesa	0.44	1 319	Corte promedio
Princesa	1.00	4 868	Corte perfeccionado	Marquesa	0.44	1 319	Corte perfeccionado
Redonda	0.91	5 106	Corte perfeccionado	Redonda	0.40	1 133	Corte perfeccionado
Redonda	0.90	3 921	Buen corte	Redonda	0.35	1 354	Buen corte
Redonda	0.90	3 733	Corte perfeccionado	Redonda	0.32	896	Corte perfeccionado
Redonda	0.84	2 621	Corte perfeccionado				

d. Diseñe una tabla de contingencia con las variables "forma" y "grado de corte". ¿Cuál es el grado de corte más común? ¿Cuál es la forma más común? ¿Cuál es la combinación más común de grado de corte y forma?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

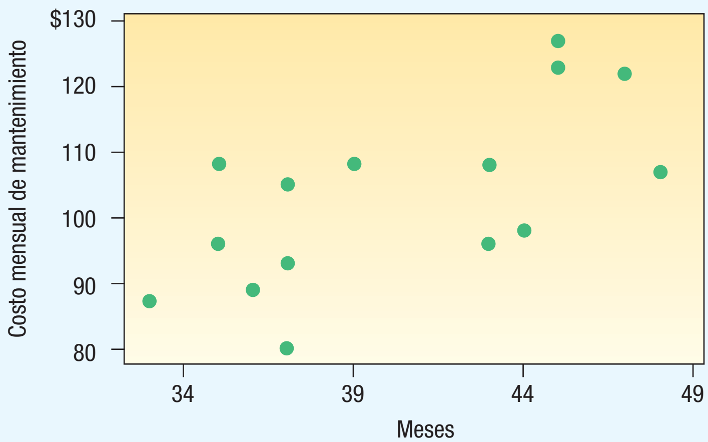
38. En la siguiente lista se registra la cantidad de comisiones que ganaron el mes anterior los ocho miembros del personal de ventas de Best Electronics. Determine el coeficiente de sesgo utilizando ambos métodos. *Sugerencia:* el uso de una hoja de cálculo agilizará las operaciones.

980.9	1 036.5	1 099.5	1 153.9	1 409.0	1 456.4	1 718.4	1 721.2

39. En la siguiente tabla se registra la cantidad de robos de automóviles en una ciudad grande la semana pasada. Determine el coeficiente de sesgo utilizando ambos métodos. *Sugerencia:* el uso de una hoja de cálculo agilizará las operaciones.

3	12	13	7	8	3	8

40. El gerente de servicios de información de Wilkin Investigations, una empresa privada, estudia la relación entre el tiempo de uso (en meses) de una máquina compuesta de impresora, copiadora y fax, y su costo de mantenimiento mensual. El gerente elaboró el siguiente diagrama sobre una muestra de 15 máquinas. ¿Qué puede concluir el gerente acerca de la relación entre las variables?



Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

41. Una compañía de seguros de automóvil arrojó la siguiente información relacionada con la edad de un conductor y el número de accidentes registrados el año previo. Diseñe un diagrama de dispersión con los datos y redacte un breve resumen.

Edad	Accidentes	Edad	Accidentes
16	4	23	0
24	2	27	1
18	5	32	1
17	4	22	3

42. Wendy's ofrece ocho diferentes condimentos (mostaza, catsup, cebolla, mayonesa, pepinillos, lechuga, tomate y guarnición) para hamburguesas. El administrador de una de las tiendas recogió la información que aparece enseguida relativa al número de condimentos que se pidieron y el grupo de edad de los clientes. ¿Qué puede concluir respecto de la información? ¿Quién tiende a ordenar la mayor o la menor cantidad de condimentos?

Cantidad de condimentos	Edad			
	Menos de 18	De 18 hasta 40	De 40 hasta 60	60 o mayores
0	12	18	24	52
1	21	76	50	30
2	39	52	40	12
3 o más	71	87	47	28

43. En la siguiente lista se muestra el número de trabajadores empleados y desempleados de 20 años o mayores, de acuerdo con su género en Estados Unidos.

Género	Número de trabajadores (miles)	
	Empleados	Desempleados
Hombres	70 415	4 209
Mujeres	61 402	3 314

a. ¿Cuántos trabajadores se registraron?
b. ¿Qué porcentaje de trabajadores estaban desempleados?
c. Compare el porcentaje de desempleados en el caso de hombres y mujeres.

EJERCICIOS DE LA BASE DE DATOS

Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**.

44. Consulte los datos sobre Real Estate, que contienen información acerca de las casas que se vendieron en Goodyear, Arizona, el año anterior. Prepare un reporte sobre los precios de venta de las casas. Asegúrese de responder en su informe las siguientes preguntas:
a. Elabore un diagrama de caja. Estime el primer y tercer cuartiles. ¿Hay datos atípicos?
b. Desarrolle un diagrama de dispersión con el precio en el eje vertical y el tamaño de la casa en el horizontal. ¿Le parece que hay alguna relación entre ambas variables? ¿La relación es directa o inversa?
c. Elabore un diagrama de dispersión con el precio en el eje vertical y la distancia al centro de la ciudad en el horizontal. ¿Parece que hay alguna relación entre ambas variables? ¿La relación es directa o inversa?

45. Consulte los datos sobre béisbol 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012.
a. En la base de datos, la variable "construido (*built*)" es el año en que el estadio se construyó. Usando esta variable, cree una nueva: "edad", restando el valor de la variable "construido" del año vigente para cada equipo. Diseñe un diagrama de caja ¿Hay datos atípicos? Si los hay, ¿cuáles estadios serían atípicos?
b. Seleccione la variable relacionada con el salario del equipo y diseñe un diagrama de caja. ¿Hay datos atípicos? ¿Cuáles son los cuartiles? Redacte un breve resumen de su análisis. ¿Cómo se comparan los salarios de los Yanquis de Nueva York con los otros equipos?

c. Trace un diagrama de dispersión en cuyo eje vertical se indique el número de juegos ganados y la nómina del equipo, en el eje horizontal. ¿Cuáles son sus conclusiones?

d. Seleccione la variable "juegos ganados (*wins*)". Trace un diagrama de puntos. ¿Qué conclusiones puede obten er a partir de esta gráfica?

46. Consulte los datos de los autobuses del Distrito Escolar Buena.

a. Refiérase a la variable "costo de mantenimiento". Desarrolle un diagrama de caja. ¿Cuáles son el primer y tercer cuartiles? ¿Hay datos atípicos?

b. Determine el costo mediano de mantenimiento. Basándose en la mediana, desarrolle una tabla de contingencias en donde el fabricante sea una variable y en la otra se indique si el costo de mantenimiento estuvo por arriba o por debajo de la mediana. ¿Cuáles son sus conclusiones?

REPASO DE LOS CAPÍTULOS 1 a 4

Esta sección constituye un repaso de los conceptos y términos más importantes que estructuran los capítulos 1 a 4. El capítulo 1 inició con una descripción del significado y objetivo de la estadística. Enseguida se describieron los diferentes tipos de variables y los cuatro niveles de medición. El capítulo 2 se centró en la descripción de un conjunto de observaciones y la forma en la que se organizaban en una distribución de frecuencias, y en la representación de la distribución de frecuencias mediante histogramas o polígonos de frecuencias. El capítulo 3 comenzó con la descripción de medidas de ubicación, como la media, la media ponderada, la mediana, la media geométrica y la moda. Este capítulo también incluyó las medidas de dispersión o propagación. En esta sección se estudiaron el rango, la desviación media, la varianza y la desviación estándar. El capítulo 4 incluyó diversas técnicas de graficación, como los diagramas de puntos, los diagramas de caja y los diagramas de dispersión. También se abordó el coeficiente de sesgo, que indica la falta de simetría que puede existir en un conjunto de datos.

A lo largo de esta sección se destacó la importancia del software estadístico, como Excel y Minitab. En estos capítulos muchas capturas de pantalla demostraron la rapidez y eficacia con la que se puede organizar un conjunto de datos en una distribución de frecuencias; también expusieron cómo calcular diversas medidas de ubicación o de variación y la información que se presenta de forma gráfica.

PROBLEMAS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

1. Una muestra de 50 fondos depositados en la cuenta de cheques miniatura del First Federal Savings Bank reveló las siguientes cantidades:

$124	$ 14	$150	$289	$ 52	$156	$203	$ 82	$ 27	$248
39	52	103	58	136	249	110	298	251	157
186	107	142	185	75	202	119	219	156	78
116	152	206	117	52	299	58	153	219	148
145	187	165	147	158	146	185	186	149	140

Utilice un paquete de software estadístico como Excel o Minitab para contestar las siguientes preguntas.

a. Determine la media, la mediana y la desviación estándar.

b. Precise el primer y tercer cuartiles.

c. Desarrolle un diagrama de puntos. ¿Hay datos atípicos? ¿Las cantidades siguen una distribución simétrica o están sesgadas? Sustente su respuesta.

d. Organice la distribución de fondos en una distribución de frecuencia.

e. Redacte un breve resumen de los resultados que obtuvo en los puntos anteriores.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2. A continuación se presenta una lista de los 44 presidentes de Estados Unidos y sus edades cuando comenzaron sus respectivos periodos.

Número	Nombre	Edad	Número	Nombre	Edad
1	Washington	57	7	Jackson	61
2	J. Adams	61	8	Van Buren	54
3	Jefferson	57	9	W. H. Harrison	68
4	Madison	57	10	Tyler	51
5	Monroe	58	11	Polk	49
6	J. Q. Adams	57	12	Taylor	64

(*continúa*)

(*continuación*)

Número	Nombre	Edad	Número	Nombre	Edad
13	Fillmore	50	29	Harding	55
14	Pierce	48	30	Coolidge	51
15	Buchanan	65	31	Hoover	54
16	Lincoln	52	32	F. D. Roosevelt	51
17	A. Johnson	56	33	Truman	60
18	Grant	46	34	Eisenhower	62
19	Hayes	54	35	Kennedy	43
20	Garfield	49	36	L. B. Johnson	55
21	Arthur	50	37	Nixon	56
22	Cleveland	47	38	Ford	61
23	B. Harrison	55	39	Carter	52
24	Cleveland	55	40	Reagan	69
25	McKinley	54	41	G. H. W. Bush	64
26	T. Roosevelt	42	42	Clinton	46
27	Taft	51	43	G. W. Bush	54
28	Wilson	56	44	Obama	47

Utilice un paquete de software estadístico como Excel o Minitab para contestar las siguientes preguntas.

a. Determine la media, la mediana y la desviación estándar.
b. Precise el primer y tercer cuartiles.
c. Desarrolle un diagrama de puntos. ¿Hay datos atípicos? ¿Las cantidades siguen una distribución simétrica o están sesgadas? Sustente su respuesta.
d. Organice la distribución de fondos en una distribución de frecuencia.
e. Redacte un breve resumen de los resultados que obtuvo en los puntos anteriores.

3. A continuación se enlista el ingreso *per capita* de los 50 estados y el distrito de Columbia.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Estado	Cantidad	Estado	Cantidad
Alabama	$30 894	Montana	$30 790
Alaska	38 138	Nebraska	34 440
Arizona	31 936	Nevada	38 994
Arkansas	28 473	New Hampshire	39 753
California	39 626	New Jersey	46 763
Colorado	39 491	New Mexico	29 929
Connecticut	50 762	New York	44 027
Delaware	39 131	North Carolina	32 247
DC	57 746	North Dakota	32 763
Florida	36 720	Ohio	33 320
Georgia	32 095	Oklahoma	32 391
Hawaii	37 023	Oregon	33 299
Idaho	29 920	Pennsylvania	36 825
Illinois	38 409	Rhode Island	37 523
Indiana	32 288	South Carolina	29 767
Iowa	33 038	South Dakota	32 030
Kansas	34 799	Tennessee	32 172
Kentucky	29 729	Texas	35 166
Louisiana	31 821	Utah	29 406
Maine	32 095	Vermont	34 871
Maryland	43 788	Virginia	39 540
Massachusetts	46 299	Washington	38 212
Michigan	33 788	West Virginia	28 206
Minnesota	38 859	Wisconsin	34 405
Mississippi	27 028	Wyoming	40 655
Missouri	32 789		

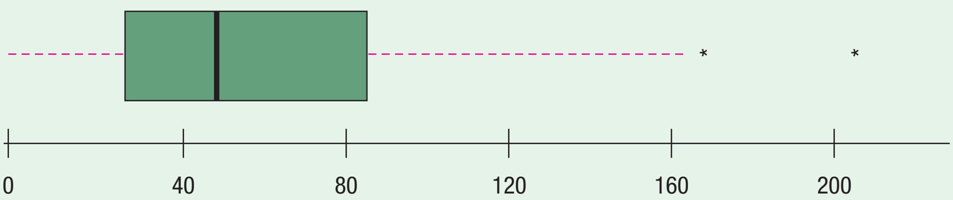
Utilice un paquete de software estadístico como Excel o Minitab para contestar las siguientes preguntas.

a. Determine la media, la mediana y la desviación estándar.
b. Precise el primer y tercer cuartiles.
c. Desarrolle un diagrama de puntos. ¿Hay datos atípicos? ¿Las cantidades siguen una distribución simétrica o están sesgadas? Sustente su respuesta.
d. Organice la distribución de fondos en una distribución de frecuencia.
e. Redacte un breve resumen de los resultados que obtuvo en los anteriores.

4. Una muestra de 12 casas que se vendieron la semana pasada en St. Paul, Minnesota, reveló la información que aparece enseguida. Trace un diagrama de dispersión. ¿Es posible concluir que, conforme las dimensiones (expresadas en miles de pies cuadrados) de la casa aumentan, el precio de venta (en miles de dólares) también se incrementa?

Dimensiones de la casa (miles de pies cuadrados)	Precio de venta (miles de dólares)	Dimensiones de la casa (miles de pies cuadrados)	Precio de venta (miles de dólares)
1.4	100	1.3	110
1.3	110	0.8	85
1.2	105	1.2	105
1.1	120	0.9	75
1.4	80	1.1	70
1.0	105	1.1	95

5. Consulte el siguiente diagrama:



a. ¿Cuál es el nombre de la gráfica?
b. ¿Cuál es la mediana y los valores del primer y tercer cuartiles?
c. ¿Es la distribución positivamente sesgada? Indique cómo lo sabe.
d. ¿Hay datos atípicos? Si es el caso, estime los valores.
e. ¿Puede determinar el número de observaciones en el estudio?

CASOS

A. Century National Bank

El siguiente caso aparecerá en las subsecuentes secciones de repaso. Suponga que usted trabaja en el departamento de planeación del Century National Bank y le reporta a la señora Lamberg. Usted necesita hacer un análisis de datos y preparar un breve informe escrito. Recuerde que el señor Selig es el presidente del banco, de modo que usted querrá asegurarse de que su informe sea completo y exacto. El apéndice A.6 contiene una copia de los datos.

Century National Bank cuenta con oficinas en diversas ciudades de la región central y el sureste de Estados Unidos. Al señor Dan Selig, presidente y director ejecutivo, le gustaría conocer las características de sus clientes con cuentas de cheques. ¿Cuál es el saldo típico de estos?

¿Cuántos servicios bancarios más utilizan los clientes con cuentas de cheques? ¿Utilizan el servicio de cajero automático y, de ser así, cuán a menudo? ¿Qué hay de las tarjetas de débito? ¿Quién las utiliza y con cuánta frecuencia?

Para comprender mejor a los clientes, el señor Selig pidió a la señora Wendy Lamberg, directora de planeación, que seleccionara una muestra de clientes y preparara un informe. Para comenzar, ella nombró un equipo de entre su personal. Usted es el jefe del equipo y el responsable de elaborar el informe. Elige una muestra aleatoria de 60 clientes. Además del saldo de cada cuenta al final del mes anterior, usted determina lo siguiente: 1) el número de transacciones en cajeros automáticos del mes previo; 2) la cantidad de servicios bancarios distintos (cuenta de ahorro, certificados de depósito, etc.) que utiliza el cliente; 3) si el cliente posee una tarjeta de débito (este es un servicio bancario relativamente nuevo respecto del cual los cargos se hacen directamente a la cuenta del cliente); 4) si se paga o no interés en la cuenta de cheques. La muestra incluye clientes de las sucursales de Cincinnati, Ohio; Atlanta, Georgia; Louisville, Kentucky y Erie, Pennsylvania.

1. Diseñe una gráfica o tabla en la que se representen los saldos de las cuentas de cheques. ¿Cuál es el saldo de un cliente típico? ¿Hay clientes con más de 2 000 dólares en sus cuentas? ¿Le parece que existe una diferencia en la distribución de las cuentas entre las cuatro sucursales? Los saldos tienden a acumularse en torno a un valor, ¿cuál es ese?
2. Determine la media y la mediana de los saldos de las cuentas de cheques. Compare la media y la mediana de los saldos de las cuatro sucursales. ¿Existe alguna diferencia entre las sucursales? Explique en su informe la diferencia entre la media y la mediana.
3. Determine el rango y la desviación estándar de los saldos de las cuentas de cheques. ¿Qué muestran el primer y tercer cuartiles? Determine el coeficiente de sesgo e indique lo que muestra. Como el señor Selig no maneja estadísticas diariamente, incluya una breve descripción e interpretación de la desviación estándar y otras medidas.

B. Wildcat Plumbing Supply, Inc.: ¿hay diferencias de género?

Wildcat Plumbing Supply ha dado servicios de plomería en el sur de Arizona durante más de 40 años. La compañía, que fue fundada por el señor Terrence St. Julian y hoy la dirige su hijo Cory, ha crecido de un puñado de empleados a más de 500. Cory está interesado en los diferentes puestos en la compañía en los que trabajan hombres y mujeres que llevan a cabo las mismas tareas, pero con diferente salario. Para investigar, recoge la información que aparece enseguida. Suponga que usted es un estudiante que lleva a cabo prácticas en el departamento de contabilidad y se le ha encomendado la tarea de redactar un informe que resuma la situación.

Salario anual (miles de dólares)	Mujer	Hombre
Menos de 30	2	0
30 hasta 40	3	1
40 hasta 50	17	4
50 hasta 60	17	24
60 hasta 70	8	21
70 hasta 80	3	7
80 o más	0	3

Para arrancar el proyecto, el señor Cory St. Julian organizó una junta con su personal, a la cual usted fue invitado. En esta junta se sugirió que usted calculara diversas medidas de ubicación, trazara diagramas (como una distribución de frecuencias acumulativas) y determinara los cuartiles tanto de hombres como de mujeres. Elabore los diagramas y redacte un informe que resuma los salarios anuales de los empleados de Wildcat Plumbing Supply. ¿Parece que hay diferencias de pago a partir del género?

C. Kimble Products: ¿hay alguna diferencia en el pago de comisiones?

En la junta nacional de ventas de enero, al director ejecutivo de Kimble Products se le cuestionó sobre la política de la compañía en lo que se refiere al pago de comisiones a sus representantes de ventas. La compañía vende artículos deportivos en dos mercados importantes. Tiene 40 representantes de ventas que se comunican directamente con una gran cantidad de clientes, como los departamentos de educación física de los principales institutos, universidades y franquicias de artículos deportivos profesionales. Además, 30 agentes de ventas representan a la compañía ante tiendas minoristas ubicadas en centros comerciales y grandes almacenes de descuento, como Kmart y Target.

Al llegar a las oficinas centrales, el director ejecutivo solicitó al gerente de ventas un informe en el que se compararan las comisiones que ganaron el año anterior las dos secciones del equipo de ventas. ¿Concluiría usted que existe alguna diferencia? El informe debe incluir información acerca de la tendencia central, así como sobre la dispersión en ambos grupos.

Comisiones que obtuvieron los representantes de ventas que atienden departamentos de deportes (en dólares)

354	87	1 676	1 187	69	3 202	680	39	1 683	1 106
883	3 140	299	2 197	175	159	1 105	434	615	149
1 168	278	579	7	357	252	1 602	2 321	4	392
416	427	1 738	526	13	1 604	249	557	635	527

Comisiones que obtuvieron los representantes de ventas que atienden grandes departamentos minoristas (en dólares)

1 116	681	1 294	12	754	1 206	1 448	870	944	1 255
1 213	1 291	719	934	1 313	1 083	899	850	886	1 556
886	1 315	1 858	1 262	1 338	1 066	807	1 244	758	918

TEST DE PRÁCTICAS

Existe un cuestionario de práctica al final de cada sección de revisión; este consta de dos partes. La primera contiene diversas preguntas objetivas, por lo general, con un espacio en blanco para la respuesta. La segunda consiste en problemas y ejercicios. En la mayoría de los casos, debería tomarle de 30 a 45 minutos completarlo. Los problemas requieren una calculadora. Verifique las soluciones en la sección de respuestas en la parte final del libro.

Parte 1: Preguntas objetivas

1. La ciencia de recolectar, organizar, presentar, analizar e interpretar los datos para ayudar a tomar decisiones eficaces se denomina ________. **1.** ________
2. Los métodos para organizar, resumir y presentar los datos de una manera informativa se llaman ________. **2.** ________
3. El grupo completo de individuos u objetos de interés, o las medidas que se obtienen de todos los individuos u objetos de interés se llama ________. **3.** ________

4. Mencione dos tipos de variables. **4.** ________________
____

5. ¿La cantidad de habitaciones en una casa es un ejemplo de variable discreta, continua o cualitativa? **5.** ________________

6. ¿De qué nivel de medición son un ejemplo los números en los jerseys de los jugadores de las Ligas Mayores de Béisbol? **6.** ________________

7. ¿Qué ejemplo de nivel de medición sería la clasificación de estudiantes por color de ojos? **7.** ________________

8. ¿A qué valor equivale siempre la suma de las diferencias entre cada valor y la media? **8.** ________________

9. Un grupo de datos contiene 70 observaciones. ¿Cuántas clases sugeriría usted para construir una distribución de frecuencias? **9.** ________________

10. ¿Qué porcentaje de los valores en un grupo de datos siempre es mayor que la mediana? **10.** ________________

11. El cuadrado de la desviación estándar es ________. **11.** ________________

12. ¿La desviación estándar asume un valor negativo cuando todos los valores son negativos, al menos la mitad de los valores son negativos o nunca? **12.** ________________

13. Considere la media, la mediana y el rango. ¿Cuál de los anteriores es el menos afectado por un dato atípico? **13.** ________________

Parte 2: Problemas

1. El índice de precios de valores Russell 2000 se incrementó en las siguientes cantidades los últimos tres años.

18%	4%	2%

¿Cuál es la media geométrica del incremento de los tres años? ________________

2. La siguiente información se refiere a los precios de venta, en miles de dólares, de casas que se vendieron en Warren, PA, durante 2014.

Precio de venta (miles de dólares)	Frecuencia
120.0 hasta 150.0	4
150.0 hasta 180.0	18
180.0 hasta 210.0	30
210.0 hasta 240.0	20
240.0 hasta 270.0	17
270.0 hasta 300.0	10
300.0 hasta 330.0	6

a. ¿Cuál es el intervalo de clase? ________________
b. ¿Cuántas casas se vendieron en 2014? ________________
c. ¿Cuántas casas se vendieron por menos de $210 000? ________________
d. ¿Cuál es la frecuencia relativa de la clase 210 hasta 240? ________________
e. ¿Cuál es el punto medio de la clase 150 hasta 180? ________________
f. ¿Entre cuáles dos cantidades está el rango de los precios de venta? ________________

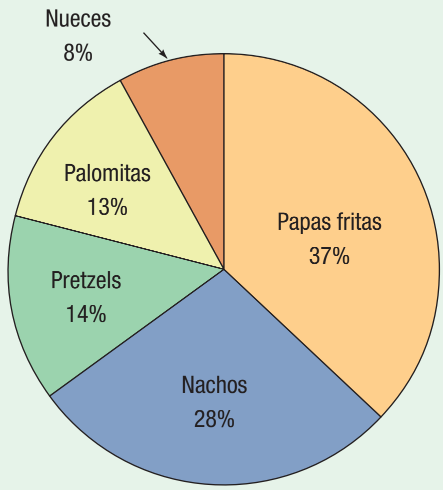
3. Una muestra de ocho estudiantes universitarios reveló que poseían el siguiente número de discos compactos.

52	76	64	79	80	74	66	69

a. ¿Cuál es el número medio de discos compactos? ________________
b. ¿Cuál es el número mediano de discos compactos? ________________
c. ¿Cuál es el cuadragésimo percentil? ________________
d. ¿Cuál es el rango del número de discos compactos? ________________
e. ¿Cuál es la desviación estándar del número de discos compactos? ________________

4. Un inversionista compró 200 acciones de Blair Company a 36 dólares cada una en julio de 2012, 300 acciones a 40 dólares cada una en septiembre de 2013, y 500 acciones a 50 dólares cada una en enero de 2014. ¿Cuál es su media ponderada del precio por acción? ________________

5. Durante el Súper Tazón de 2008 se consumieron 30 millones de libras de comida chatarra. Esta información se presenta en la siguiente gráfica.



a. ¿Cuál es el nombre que se le da a esta gráfica? ____________________

b. Estime, en millones de libras, la cantidad de papas fritas consumidas durante el juego. ____________________

c. ¿Cuál es la relación entre las papas fritas y las palomitas (el doble, la mitad, el triple o ninguna de las anteriores)? ____________________

d. ¿Qué porcentaje del total comprenden las papas fritas y los nachos? ____________________

5 Estudio de los conceptos de la probabilidad

ENCUESTAS RECIENTES revelaron que 60% de los turistas que viajaron a China visitaron la Ciudad Prohibida, el Templo del Cielo, la Gran Muralla y otros sitios históricos en Beijing o cerca de esta ciudad. Cuarenta por ciento visitó Xi'an y sus magníficos soldados, caballos y carrozas de terracota, que permanecieran enterrados durante más de 2000 años; 30% fue tanto a Beijing como a Xi'an. ¿Cuál es la probabilidad de que un turista haya visitado por lo menos uno de estos lugares? (vea el ejercicio 76, y el objetivo de aprendizaje OA5-3).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA5-1 Definir los términos *probabilidad, experimento, evento* y *resultado*.

OA5-2 Asignar probabilidades utilizando un enfoque clásico, empírico o subjetivo.

OA5-3 Calcular probabilidades mediante las *reglas de la adición*.

OA5-4 Calcular probabilidades mediante las *reglas de la multiplicación*.

OA5-5 Calcular probabilidades por medio de una tabla de contingencia.

OA5-6 Calcular probabilidades con base en el *teorema de Bayes*.

OA5-7 Determinar el número de resultados por medio del principio apropiado de conteo.

Introducción

Los capítulos 2, 3 y 4 se enfocan en la estadística descriptiva. En el capítulo 2 se organizaron las ganancias de 180 vehículos que vendió el Applewood Auto Group en una distribución de frecuencias, se muestran las ganancias más baja y más alta, así como el punto donde se presenta la concentración de datos. En el capítulo 3, mediante medidas numéricas de ubicación y dispersión, se definió una ganancia típica y se examinó la variación de la ganancia derivada de una venta; además, se describió la variación de las ganancias con medidas de dispersión como el rango y la desviación estándar. En el capítulo 4 se diseñaron diagramas y gráficas, como el diagrama de puntos y gráfica de dispersión, con el fin de presentar los datos de manera gráfica.

A la estadística descriptiva le concierne resumir datos recogidos de eventos pasados. Ahora se presenta la segunda faceta de la estadística, a saber, *el cálculo de la probabilidad de que algo ocurra en el futuro*; esta recibe el nombre de **inferencia estadística** o **estadística inferencial.**

Quien toma decisiones, pocas veces cuenta con la información completa para hacerlo. Por ejemplo:

- Toys and Things, un fabricante de juguetes y rompecabezas, creó un nuevo juego basado en una trivia deportiva y pretende saber si los fanáticos del deporte comprarán el juego. Slam Dunk y Home Run son dos de los nombres que se consideran. Para investigar, el presidente de la empresa decidió contratar a una firma de investigación de mercados. La firma seleccionó a 800 consumidores de la población y pidió a cada entrevistado su opinión acerca del nuevo juego y los nombres propuestos. La compañía estimará la proporción de la población que comprará el juego con base en los resultados de la muestra.
- El departamento de control de calidad de la fundidora Bethlehem Steel debe asegurar a la administración que el cable de un cuarto de pulgada que se fabrica tiene una fuerza de tensión aceptable. Es obvio que no se prueba la fuerza de tensión de todo el cable que se fabrica porque la prueba requiere tensar el cable hasta romperlo, es decir, lo destruye. De modo que se selecciona una muestra de 10 piezas y se prueban. A partir de los resultados del estudio, todo el cable que se fabrica se califica de aceptable o inaceptable.
- Otras preguntas que implican incertidumbre son: ¿debe suspenderse de inmediato la telenovela *Days of Our Lives*? ¿Será redituable un nuevo cereal con sabor a menta si se comercializa? ¿Charles Linden será elegido auditor del condado en Batavia County?

La inferencia estadística se vincula con las conclusiones relacionadas con una población con base en una muestra que se toma de ella (las poblaciones de los ejemplos anteriores son: todos los consumidores aficionados a las trivias deportivas; todos los cables de acero de un cuarto de pulgada; todos los televidentes que ven telenovelas; toda la gente que compra cereal para el desayuno, etc.).

Dada la incertidumbre existente en la toma de decisiones, es importante evaluar científicamente todos los riesgos implicados. La *teoría de la probabilidad,* a menudo conocida como la ciencia de la incertidumbre, resulta útil para hacer esta evaluación. Su aplicación permite a quien toma decisiones y posee información limitada analizar los riesgos y reducirlos al mínimo; por ejemplo, al lanzar al mercado un nuevo producto o aceptar un envío que quizá contenga partes defectuosas.

Puesto que los conceptos de la probabilidad son importantes en el campo de la inferencia estadística (tema que se analiza en el capítulo 8), en este capítulo se introduce el lenguaje básico de la probabilidad, el cual incluye términos como *experimento, evento, probabilidad subjetiva,* y *reglas de la adición y de la multiplicación.*

¿Qué es la probabilidad?

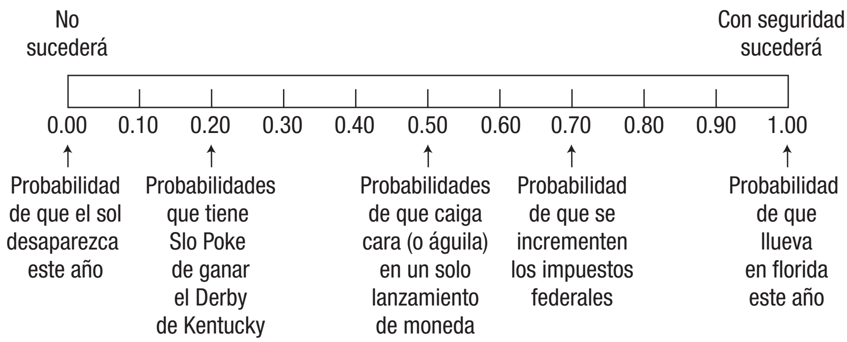
OA5-1
Definir los términos *probabilidad, experimento, evento* y *resultado.*

Sin duda, usted está familiarizado con términos como *probabilidad, azar* y *posibilidad*, los cuales con frecuencia se emplean de manera indistinta. El meteorólogo anuncia que hay 70% de probabilidad de lluvia para el domingo del Súper Tazón. Con base en una encuesta de consumidores que probaron una nueva pasta de dientes con sabor a plátano, la probabilidad de que sea un éxito financiero si se le comercializa es de 0.03 (esto significa que la posibilidad de que la pasta de dientes con sabor a plátano sea aceptada por el público es muy remota). ¿Qué es la probabilidad? En general es un valor numérico que describe la posibilidad de un suceso.

PROBABILIDAD Valor entre cero y uno, inclusive, que describe la posibilidad relativa (oportunidad o casualidad) de un evento.

Es común que una probabilidad se exprese en forma decimal, como 0.70, 0.27 o 0.50. No obstante, también se da en forma de fracción, como 7/10, 27/100 o 1/2. Se puede suponer cualquier número entre 0 y 1, inclusive. Si una compañía solo tiene cinco regiones de ventas, y el nombre o número de cada región se escribe en un trozo de papel, que se coloca en un sombrero, la probabilidad de seleccionar una de las cinco regiones es de 1. La probabilidad de sacar del sombrero un trozo de papel rotulado "Pittsburgh Steelers" es de 0. Por consiguiente, la probabilidad de 1 representa algo que seguramente sucederá, y la de 0, algo que no sucederá.

Cuanto más se aproxime una probabilidad a 0, más improbable es que el evento suceda. Cuanto más próxima se encuentre la probabilidad a 1, más seguro es que suceda. En el siguiente diagrama se muestra la relación e incluye algunas conjeturas personales. Sin embargo, usted podría seleccionar una probabilidad distinta de que Slo Poke gane el Derby de Kentucky o de que se incrementen los impuestos federales.



A veces, la probabilidad de un evento se expresa utilizando el término *chances* o *posibilidades*. Para explicar esto, alguien dice que los chances son "cinco a dos" de que un evento suceda. Esto significa que en un total de siete ensayos (5 + 2), el evento ocurrirá cinco veces y no sucederá dos veces. Utilizando los chances se puede calcular la probabilidad de que el evento ocurra como 5/(5 + 2) o 5/7. Así, si los chances o posibilidades a favor de un evento son x, la probabilidad del evento es $x/(x + y)$.

En el estudio de la probabilidad se utilizan tres palabras claves: **experimento**, **resultado** y **evento.** Dichos términos se emplean en el lenguaje de la vida cotidiana, pero en estadística adquieren significados específicos.

EXPERIMENTO Proceso que induce a que ocurra una y solo una de varias posibles observaciones.

Esta definición es más general que la empleada en las ciencias físicas, donde es fácil imaginar a alguien que manipula tubos de ensayo o microscopios. Respecto de la probabilidad, un experimento tiene dos o más posibles resultados y no se sabe cuál ocurrirá.

RESULTADO Ocurrencia particular de un experimento.

Por ejemplo, lanzar una moneda al aire constituye un experimento. Usted no sabe cuál será el resultado. Cuando se lanza a una moneda, un resultado particular es "cara". El resultado alternativo es "cruz". De manera similar, preguntar a 500 estudiantes universitarios si viajarían más de 200 kilómetros para asistir a un concierto de Mumford and Sons constituye un experimento, en el cual un posible resultado es que 273 estudiantes indiquen que sí lo harían. Otro es que 317 estudiantes irían al concierto. Todavía hay otro resultado, que 423 estudiantes respondan afirmativamente. Cuando se observan uno o más resultados en los experimentos, tenemos un evento.

EVENTO Conjunto de uno o más resultados de un experimento.

En la siguiente figura se presentan ejemplos para aclarar las definiciones de los términos *experimento*, *resultado* y *evento.*

En el caso del experimento del lanzamiento de un dado, hay seis posibles resultados, pero existen varios posibles eventos. Cuando se cuenta el número de miembros de la junta directiva de las compañías Fortune 500 que tienen más de 60 años de edad, el número posible de resultados varía de cero al total de miembros. Hay un número aún mayor de eventos posibles en este experimento.

		
Experimento	Lanzar un dado	Listado del número de miembros de la junta directiva de las compañías de *Fortune 500*, mayores de 60 años
Todos los posibles resultados	Se observa un 1 Se observa un 2 Se observa un 3 Se observa un 4 Se observa un 5 Se observa un 6	Ninguno tiene más de 60 Uno tiene más de 60 Dos tiene más de 60 ... 29 tiene más de 60 ... ... 48 tiene más de 60 ...
Algunos posibles eventos	Se observa un número par Se observa un número mayor que 4 Se observa un 3 o un número menor	Más de 13 tiene más de 60 Menos de 20 tiene más de 60

AUTOEVALUACIÓN 5-1

Video Games, Inc., creó recientemente un nuevo videojuego; entonces selecciona 80 jugadores veteranos para probar su facilidad de operación.

(a) ¿En qué consiste el experimento?
(b) ¿Cuál es uno de los posibles resultados?
(c) Suponga que 65 jugadores prueban el nuevo juego y dicen que les gusta. ¿Es 65 una probabilidad?
(d) La probabilidad de que el nuevo juego sea un éxito es de −1.0. Haga comentarios al respecto.
(e) Especifique un posible evento.

Enfoques para asignar probabilidades

OA5-2 Asignar probabilidades utilizando un enfoque clásico, empírico o subjetivo.

En esta sección se describen tres formas de asignar una probabilidad a un evento: clásica, empírica y subjetiva. Los métodos clásico y empírico son objetivos y se basan en datos e información. El método subjetivo se basa en la creencia o estimación de una persona acerca de la probabilidad de un evento.

Probabilidad clásica

La **probabilidad clásica** parte del supuesto de que los resultados de un experimento son *igualmente posibles.* De acuerdo con el punto de vista clásico, la probabilidad de un evento se calcula dividiendo el número de resultados favorables entre el número de posibles resultados:

PROBABILIDAD CLÁSICA

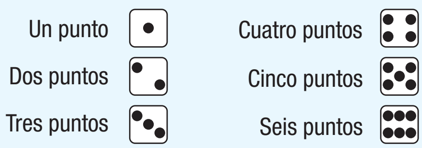
$$\text{Probabilidad de un evento} = \frac{\text{Número de resultados favorables}}{\text{Número total de posibles resultados}} \qquad \textbf{[5.1]}$$

EJEMPLO

Considere el experimento de lanzar un dado. ¿Cuál es la probabilidad del evento “cae un número par de puntos”?

SOLUCIÓN

Los posibles resultados son:



Hay tres resultados “favorables” (dos, cuatro y seis) en el conjunto de seis resultados igualmente posibles. Por consiguiente,

$$\text{Probabilidad de un número par} = \frac{3}{6} \begin{array}{l} \leftarrow \\ \leftarrow \end{array} \frac{\text{Número de resultados favorables}}{\text{Número total de posibles resultados}}$$
$$= 0.5$$

El concepto de conjuntos mutuamente excluyentes se presentó en el estudio de las distribuciones de frecuencias en el capítulo 2. Recuerde que se crean clases de tal manera que un evento particular se incluya en una sola de las clases y no haya superposición entre estas. Por lo tanto, solo uno de varios eventos puede presentarse en cierto momento.

MUTUAMENTE EXCLUYENTE El hecho de que un evento se presente significa que ninguno de los demás puede ocurrir al mismo tiempo.

La variable “género” da origen a resultados mutuamente excluyentes: hombre y mujer. Un empleado seleccionado al azar es hombre o mujer, pero no puede tener ambos géneros. Una pieza fabricada es aceptable o no lo es. La pieza no puede ser aceptable e inaceptable al mismo tiempo. En una muestra de piezas fabricadas, el evento de seleccionar una pieza no aceptable y el evento de seleccionar una pieza aceptable son mutuamente excluyentes.

Si un experimento incluye un conjunto de eventos con todo tipo de resultados posibles, como los eventos “un número par” y “un número impar” en el experimento del lanzamiento del dado, entonces el conjunto de eventos es **colectivamente exhaustivo.** En el experimento de lanzar un dado, cada resultado será par o impar. Por consiguiente, el conjunto es colectivamente exhaustivo.

COLECTIVAMENTE EXHAUSTIVO Por lo menos uno de los eventos debe ocurrir cuando se lleva a cabo un experimento.

Si el conjunto de eventos es colectivamente exhaustivo y los eventos son mutuamente excluyentes, la suma de las probabilidades es 1. En términos históricos, el enfoque clásico de la probabilidad se creó y aplicó en los siglos XVII y XVIII a los juegos de azar, como las cartas y los dados. Resulta innecesario llevar a cabo un experimento para determinar la probabilidad de un evento mediante el enfoque clásico porque el número total de resultados se sabe antes de realizar el experimento. Lanzar una moneda tiene dos posibles resultados; arrojar un dado tiene seis posibles resultados. Por lógica, es posible determinar la probabilidad de sacar una cruz al lanzar una moneda o tres caras al lanzar tres monedas.

El enfoque clásico de la probabilidad también puede aplicarse a la lotería. En Carolina del Sur, uno de los juegos de Education Lottery es "Pick 3". Para concursar, una persona compra un billete y selecciona tres números entre 0 y 9. Una vez a la semana se seleccionan tres números en forma aleatoria de una máquina que hace girar tres contenedores, cada uno de los cuales contiene bolas numeradas de 0 a 9. Una forma de ganar consiste en acertar los números, así como el orden de estos. Dado que hay 1 000 posibles resultados (000 a 999), la probabilidad de ganar con un número de tres dígitos es de 0.001, o 1 en 1 000.

Probabilidad empírica

La **probabilidad empírica** o **frecuencia relativa,** el segundo tipo de probabilidad, se basa en el número de veces que ocurre el evento como proporción del número de intentos conocidos.

PROBABILIDAD EMPÍRICA La probabilidad de un evento representa una fracción de los sucesos similares en el pasado.

La fórmula para determinar la probabilidad empírica es:

$$\text{Probabilidad empírica} = \frac{\text{Número de veces que el evento ocurre}}{\text{Número total de observaciones}}$$

El enfoque empírico de la probabilidad se basa en la llamada *ley de los grandes números.* La clave para determinar probabilidades de forma empírica consiste en que una mayor cantidad de observaciones proporcionarán un cálculo más preciso de la probabilidad.

LEY DE LOS GRANDES NÚMEROS En una gran cantidad de intentos, la probabilidad empírica de un evento se aproximará a su probabilidad real.

Para explicar la ley de los grandes números, suponga que lanza una moneda común. El resultado de cada lanzamiento es cara o cruz. Si lanza la moneda una sola vez, la probabilidad empírica de las caras es cero o uno. Si lanza la moneda una gran cantidad de veces, la probabilidad del resultado de las caras se aproximará a 0.5. En la siguiente tabla se muestran los resultados de un experimento en el que se lanza una moneda 1, 10, 50, 100, 500, 1 000 y 10 000 veces, y, enseguida, se calcula la frecuencia relativa de las caras. Observe que conforme se incrementa el número de intentos, la probabilidad empírica de que salga una cara se aproxima a 0.5, el cual es su valor de acuerdo con el enfoque clásico de la probabilidad.

Número de ensayos	Número de caras	Frecuencia relativa de las caras
1	0	.00
10	3	.30
50	26	.52
100	52	.52
500	236	.472
1 000	494	.494
10 000	5 027	.5027

¿Qué queda demostrado? A partir de la definición clásica de probabilidad, la posibilidad de obtener una cara en un solo lanzamiento de una moneda común es de 0.5. Según el enfoque empírico de la frecuencia relativa de la probabilidad, la de cada evento se aproxima al mismo valor determinado de acuerdo con la definición clásica de probabilidad.

Este razonamiento permite emplear el enfoque empírico o el de frecuencia relativa para determinar una probabilidad. He aquí algunos ejemplos.

- El semestre anterior, 80 estudiantes se registraron para cursar Estadística administrativa 101 en la Scandia University. Doce estudiantes obtuvieron A. Con base en dicha información, y de acuerdo con la regla empírica de la probabilidad, la posibilidad calculada de que un estudiante obtenga una A es de 0.15.
- Kobe Bryant, jugador de Los Angeles Lakers logró 381 de 451 intentos de tiro libre durante la temporada 2011-2012 de la NBA. De acuerdo con la regla empírica de la probabilidad, las posibilidades de lograr su siguiente intento de tiro son de 0.845.

Las compañías de seguros de vida confían en datos similares a los anteriores para determinar la aceptabilidad de un solicitante, así como la prima que se le va a cobrar. Las tablas de mortalidad incluyen una lista de las probabilidades de que una persona de determinada edad fallezca en el siguiente año. Por ejemplo, la probabilidad de que una mujer de 20 años de edad fallezca en el próximo año es de 0.0015.

El concepto empírico se ilustra con el siguiente ejemplo.

EJEMPLO

El 1 de febrero de 2003 explotó el trasbordador espacial Columbia. Este fue el segundo desastre en 113 misiones espaciales de la NASA. Con base en esta información, ¿cuál es la probabilidad de que una futura misión concluya con éxito?

SOLUCIÓN

Para simplificar, utilice letras o números. *P* representa la probabilidad y, en este caso, $P(A)$ representa la probabilidad de que una futura misión concluya con éxito.

$$\text{Probabilidad de un vuelo exitoso} = \frac{\text{Número de vuelos exitosos}}{\text{Número total de vuelos}}$$

$$P(A) = \frac{111}{113} = 0.98$$

Este resultado sirve como aproximación de la probabilidad. En otras palabras, por experiencia, la probabilidad de que una futura misión del trasbordador espacial concluya con éxito es de 0.98.

Probabilidad subjetiva

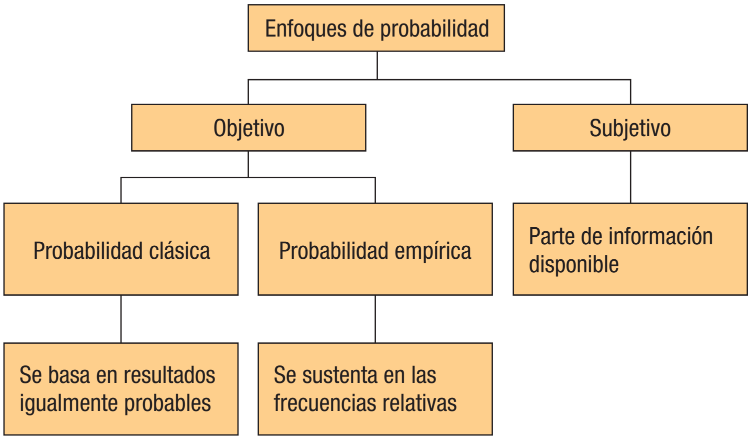
Si se cuenta con poca o ninguna experiencia o información con la cual sustentar la probabilidad, es posible aproximarla en forma subjetiva. En esencia, esto significa que un individuo evalúa las opiniones e información disponibles y luego calcula o asigna la probabilidad. Esta se denomina adecuadamente **probabilidad subjetiva.**

CONCEPTO SUBJETIVO DE PROBABILIDAD Posibilidad (probabilidad) de un evento en particular que asigna un individuo a partir de cualquier información disponible.

Algunos ejemplos de probabilidad subjetiva son:

1. Calcular la posibilidad de que los Patriotas de Nueva Inglaterra jueguen el Súper Tazón el año que viene.
2. Estimar la posibilidad de que una persona se vea involucrada en un accidente automovilístico durante los próximos 12 meses.
3. Calcular la posibilidad de que el déficit presupuestario de Estados Unidos se reduzca a la mitad en los siguientes 10 años.

En la gráfica 5.1 de la página siguiente, se resumen los diferentes tipos de probabilidad. Un enunciado probabilístico siempre asigna una posibilidad a un evento que no ha ocurrido aún. Por supuesto, hay un amplio grado de incertidumbre en este tipo de probabilidad, la cual se basa, principalmente, en el conocimiento que posee el individuo del proceso que estudia. Por ejemplo, dado el amplio conocimiento que un individuo tiene acerca del lanzamiento de dados, puede establecer que la probabilidad de que aparezca un punto en el lanzamiento de un dado no cargado es de un sexto. Sin embargo, la experiencia respecto de la aceptación del mercado de un nuevo producto que no ha sido probado es escasa. Por ejemplo, aun cuando la directora de investigación de mercado prueba un producto recién creado en 40 tiendas minoristas y establece que existe 70% de posibilidades de que el producto genere ventas por más de un millón de unidades, posee un conocimiento limitado sobre cómo reaccionarán los consumidores cuando se comercialice en todo el país. En ambos casos (el de la persona que lanza un dado y en el que se prueba un nuevo producto), el individuo asigna un valor probabilístico a un evento de interés, y solo existe una diferencia: la confianza del pronosticador en la precisión de la aproximación. No obstante, prescindiendo del punto de vista, se aplicarán las mismas leyes de la probabilidad (que se exponen en las siguientes secciones).



GRÁFICA 5.1 Resumen de enfoques de la probabilidad

AUTOEVALUACIÓN 5-2

1. Se selecciona al azar una carta de una baraja convencional de 52 cartas. ¿Cuál es la probabilidad de que la carta resulte ser una reina? ¿Qué enfoque de la probabilidad empleó para responder la pregunta?
2. El Center for Child Care publica información sobre 539 niños, así como el estado civil de sus padres. Hay 333 casados, 182 divorciados y 24 viudos. ¿Cuál es la probabilidad de que un niño elegido al azar tenga un padre divorciado? ¿Qué enfoque utilizó?
3. ¿Cuál es la probabilidad de que usted tenga un millón de dólares ahorrado para cuando se retire? ¿Qué enfoque de la probabilidad utilizó para responder la pregunta?

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

1. Hay personas que apoyan la reducción de los impuestos federales con el fin de incrementar los gastos del consumidor, aunque otros están en contra. Se seleccionan dos personas y se registran sus opiniones. Si ninguna está indecisa, elabore una lista de los posibles resultados.
2. Un inspector de control de calidad selecciona una pieza para probarla. Luego, la declara aceptable, reparable o chatarra. Entonces prueba otra pieza. Elabore una lista de los posibles resultados de este experimento relacionado con dos piezas.
3. Una encuesta de 34 estudiantes en la Wall College of Business mostró que estos tienen las siguientes especialidades:

Contabilidad	10
Finanzas	5
Economía	3
Administración	6
Marketing	10

Suponga que elige a un estudiante y observa su especialidad.

a. ¿Cuál es la probabilidad de que el estudiante tenga una especialidad en administración?
b. ¿Qué concepto de probabilidad utilizó para hacer este cálculo?

4. Una compañía grande debe contratar un nuevo presidente. El consejo directivo prepara una lista final de cinco candidatos, todos con las mismas cualidades. Dos de ellos son miembros de un grupo minoritario. Para evitar que el prejuicio influya en el momento de elegir al presidente, la compañía decide elegirlo por sorteo.
   a. ¿Cuál es la probabilidad de que uno de los candidatos que pertenece a un grupo minoritario sea contratado?
   b. ¿Qué concepto de probabilidad utilizó para hacer este cálculo?
5. En cada uno de los siguientes casos, indique si se utilizó la probabilidad clásica, empírica o subjetiva.
   a. Un jugador de béisbol consigue 30 hits en 100 turnos al bate. La probabilidad de que consiga un hit en su siguiente turno es de 0.3.
   b. Para estudiar problemas ambientales se forma un comité de estudiantes con siete miembros. ¿Cuál es la probabilidad de que cualquier de los siete sea elegido vocero del equipo?
   c. Usted compra uno de los cinco millones de boletos vendidos por la lotería de Canadá. ¿Cuáles son las posibilidades de que gane el premio de un millón de dólares?
   d. La probabilidad de un terremoto al norte de California en los próximos 10 años es de 0.80.
6. Una empresa promoverá a dos empleados de un grupo de seis hombres y tres mujeres.
   a. Elabore una lista de los posibles resultados.
   b. ¿Qué concepto de probabilidad utilizaría al hacer estos cálculos?
7. Se eligió una muestra de 40 ejecutivos de la industria del petróleo para someter a prueba un cuestionario. Una pregunta relacionada con cuestiones ambientales requería que se respondiera *sí* o *no.*
   a. ¿En qué consiste el experimento?
   b. Indique un posible evento.
   c. Diez de los 40 ejecutivos respondieron que sí. Con base en estas respuestas de la muestra, ¿cuál es la probabilidad de que un ejecutivo de la industria del petróleo responda de manera afirmativa?
   d. ¿Qué concepto de probabilidad se ilustra?
   e. ¿Los posibles resultados son igualmente probables y mutuamente excluyentes?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

8. Una muestra de 2 000 conductores con licencia reveló la siguiente cantidad de violaciones al límite de velocidad.

Cantidad de violaciones	Cantidad de conductores
0	1 910
1	46
2	18
3	12
4	9
5 o más	5
Total	2 000

   a. ¿En qué consiste el experimento?
   b. Indique un posible evento.
   c. ¿Cuál es la probabilidad de que un conductor haya cometido dos violaciones al límite de velocidad?
   d. ¿Qué concepto de probabilidad se ilustra?
9. Los clientes del Bank of America seleccionan su propio número de identificación personal de tres dígitos (NIP) para emplearlo en los cajeros automáticos.
   a. Considere ésto un experimento y haga una lista de cuatro posibles resultados.
   b. ¿Cuál es la probabilidad de que el señor Jones y la señora Smith seleccionen el mismo NIP?
   c. ¿Qué concepto de probabilidad utilizó en la respuesta *b?*
10. Un inversionista compra 100 acciones de AT&T y registra los cambios de precio diariamente.
   a. Elabore una lista de los posibles eventos de este experimento.
   b. ¿Qué concepto de probabilidad utilizó en el punto anterior?

OA5-3
Calcular probabilidades mediante las reglas de la adición.

Reglas de adición para calcular probabilidades

Existen dos reglas de la adición: la regla especial y la regla general. He aquí la primera.

Regla especial de la adición

Cuando se aplica la **regla especial de la adición,** los eventos deben ser *mutuamente excluyentes* (recuerde que esto significa que cuando un evento ocurre, ninguno de los demás eventos puede ocurrir al mismo tiempo). Un ejemplo de este tipo de eventos en el experimento del lanzamiento de un dado son los eventos "un número cuatro o mayor" y "un número dos o menor". Si el resultado se encuentra en el primer grupo {4, 5 y 6}, entonces no puede estar en el segundo grupo {1 y 2}. Otro ejemplo consiste en que un producto proveniente de la línea de montaje no puede estar defectuoso y en buen estado al mismo tiempo.

Si dos eventos *A* y *B* son mutuamente excluyentes, la regla especial de la adición establece que la probabilidad de que ocurra uno *u* otro es igual a la suma de sus probabilidades. Esta regla se expresa mediante la siguiente fórmula:

REGLA ESPECIAL DE LA ADICIÓN $$P(A \text{ o } B) = P(A) + P(B) \quad [5.2]$$

En el caso de los tres eventos mutuamente excluyentes designados *A, B* y *C,* la regla se expresa de la siguiente manera:

$$P(A \text{ o } B \text{ o } C) = P(A) + P(B) + P(C)$$

Un ejemplo ayudará a entender los detalles.

EJEMPLO

Una máquina automática llena bolsas de plástico con una combinación de frijoles, brócoli y otras verduras. La mayoría de estas contiene el peso correcto, aunque, como consecuencia de la variación del tamaño del frijol y de algunas verduras, un paquete podría pesar menos o más. Una revisión de 4 000 paquetes que se llenaron el mes previo arrojó los siguientes datos:

Peso	Evento	Número de paquetes	Probabilidad de que ocurra el evento	
Menos peso	*A*	100	.025	← $\frac{100}{4\,000}$
Peso satisfactorio	*B*	3 600	.900	
Más peso	*C*	300	.075	
		4 000	1.000	

¿Cuál es la probabilidad de que un paquete en particular pese menos o más?

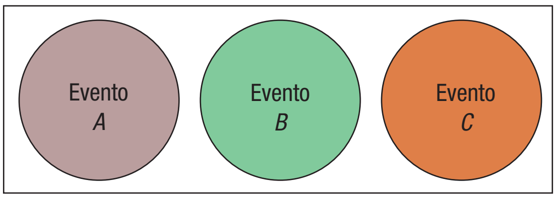
SOLUCIÓN

El resultado "pesa menos" es el evento *A*; el resultado "pesa más" es el evento *C.* Al aplicar la regla especial de la adición se tiene:

$$P(A \text{ o } C) = P(A) + P(C) = 0.025 + 0.075 = 0.10$$

Observe que los eventos son mutuamente excluyentes, lo cual significa que un paquete de verduras mixtas no puede pesar menos, tener el peso satisfactorio y pesar más al mismo tiempo. Estos también son colectivamente exhaustivos; es decir, que un paquete seleccionado debe pesar menos, tener un peso satisfactorio o pesar más.

El lógico inglés J. Venn (1834-1923) creó un diagrama para representar de manera gráfica el resultado de un experimento. El concepto de *eventos mutuamente excluyentes,* así como de otras reglas para combinar probabilidades, se ilustra mediante este recurso. Para construir un diagrama de Venn, primero se encierra un espacio de forma rectangular, el cual representa el total de posibles resultados. Así, un evento se representa por medio de un área circular que se dibuja dentro del rectángulo, la cual corresponde a la probabilidad del evento. En el diagrama de Venn de la derecha se ilustra el concepto de *eventos mutuamente excluyentes* (los eventos no se superponen). En el diagrama suponga que los eventos *A, B* y *C* son igualmente probables.



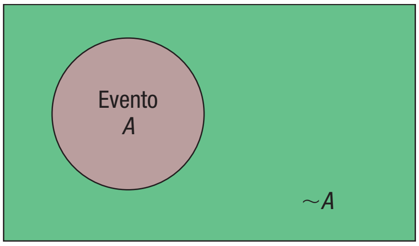
Regla del complemento

La probabilidad de que una bolsa de verduras mixtas seleccionadas pese menos, $P(A)$, más la probabilidad de que no sea una bolsa con menos peso, $P(\sim A)$, que se lee *no A,* deber ser, por lógica, igual a 1. Esto se escribe:

$$P(A) + P(\sim A) = 1$$

Esta expresión puede reformularse:

REGLA DEL COMPLEMENTO $\quad P(A) = 1 - P(\sim A) \quad$ **[5.3]**



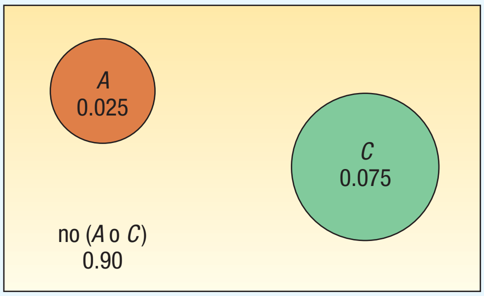
La **regla del complemento** se emplea para determinar la probabilidad de que un evento ocurra restando de 1 la probabilidad de un evento que no ha ocurrido. Esta regla es útil porque a veces es más fácil calcular la probabilidad de que un evento suceda determinando la probabilidad de que no suceda y restando el resultado de 1. Observe que los eventos A y $\sim A$ son mutuamente excluyentes y colectivamente exhaustivos. Por consiguiente, las probabilidades de A y $\sim A$ suman 1. En un diagrama de Venn la regla del complemento se ilustra a la izquierda.

EJEMPLO

Refiriéndonos al ejemplo anterior, la probabilidad de que una bolsa de verduras mixtas pese menos es de 0.025 y la probabilidad de que pese más es de 0.075. Aplique la regla del complemento para demostrar que la probabilidad de una bolsa con un peso satisfactorio es de 0.900. Muestre la solución en un diagrama de Venn.

SOLUCIÓN

La probabilidad de que la bolsa no tenga un peso satisfactorio es igual a la suma de la probabilidad de tener mayor peso más la de tener menos. Es decir, $P(A \text{ o } C) = P(A) + P(C) = 0.025 + 0.075 = 0.100$. La bolsa tiene un peso satisfactorio si no tiene menos ni más peso; así que $P(B) = 1 - [P(A) + P(C)] = 1 - [0.025 + 0.075] = 0.900$. El diagrama de Venn que representa este caso es el siguiente:



AUTOEVALUACIÓN 5-3

Se va a encuestar a una muestra de empleados de Worldwide Enterprises sobre un nuevo plan de cuidado de la salud. Los empleados se clasifican como se muestra en la tabla de abajo.

(a) Determine la probabilidad de que la primera persona elegida:
 (i) sea de mantenimiento o secretaria,
 (ii) no sea de administración.

(b) Dibuje un diagrama de Venn que ilustre sus respuestas al punto anterior.

(c) ¿Los eventos del punto (a), inciso (*i*) son complementarios, mutuamente excluyentes o ambos?

Clasificación	Evento	Número de empleados
Supervisores	*A*	120
Mantenimiento	*B*	50
Producción	*C*	1 460
Administración	*D*	302
Secretarias	*E*	68

Regla general de la adición

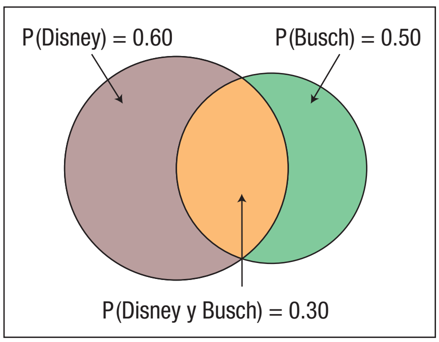
Los resultados de un experimento pueden no ser mutuamente excluyentes. Por ejemplo, suponga que la Florida Tourist Commission seleccionó una muestra de 200 turistas que visitaron el estado durante el año. La encuesta reveló que 120 fueron a Disney World y 100 a Busch Gardens, cerca de Tampa. ¿Cuál es la probabilidad de que una persona seleccionada haya visitado Disney World o Busch Gardens? Si se emplea la regla especial de la adición, la probabilidad de seleccionar un turista que haya ido a Disney World es de 0.60, que se determina mediante la división 120/200. De manera similar, la probabilidad de que un turista haya ido a Busch Gardens es de 0.50. La suma de estas probabilidades es de 1.10. Sin embargo, la probabilidad no puede ser mayor que 1. La explicación es que muchos turistas visitaron ambas atracciones turísticas y se les contó dos veces. Una revisión de las respuestas de la encuesta reveló que 60 de los 200 encuestados visitó, en realidad, ambas atracciones turísticas.

Para responder cuál es la probabilidad de elegir a una persona que haya visitado Disney World o Busch Gardens, 1) sume la probabilidad de que un turista haya visitado Disney World y la probabilidad de que haya visitado Busch Gardens; y 2) reste la probabilidad de que haya visitado ambas atracciones turísticas. Por consiguiente:

$$\begin{aligned} P(\text{Disney o Busch}) &= P(\text{Disney}) + P(\text{Busch}) - P(\text{tanto Disney como Busch}) \\ &= 0.60 + 0.50 - 0.30 = 0.80 \end{aligned}$$

Cuando dos eventos ocurren al mismo tiempo, se habla de **probabilidad conjunta.** El que un turista visite ambas atracciones turísticas (0.30) es un ejemplo de probabilidad conjunta.

En el siguiente diagrama de Venn se muestran dos eventos que no son mutuamente excluyentes. Ambos se superponen para ilustrar el evento conjunto (que algunas personas hayan visitado ambas atracciones).



PROBABILIDAD CONJUNTA Mide la posibilidad de que dos o más eventos sucedan simultáneamente.

ESTADÍSTICA EN ACCIÓN

Si usted desea llamar la atención en la siguiente reunión a la que asista, diga que cree que por lo menos dos personas presentes nacieron en la misma fecha; es decir, el mismo día, pero no necesariamente el mismo año. Si hay 30 personas en la sala, la probabilidad de que las fechas se dupliquen es de 0.706. Si hay 60 personas en la sala, la probabilidad de que por lo menos dos personas compartan la misma fecha de cumpleaños es de 0.994. Si solo hay 23 personas, las probabilidades son iguales, es decir, 0.50. *Sugerencia:* para calcularlo, determine la probabilidad de que todos hayan nacido en distintos días y aplique la regla del complemento. Inténtelo en clase.

Así, la regla general de adición, utilizada para calcular la probabilidad de dos eventos que no son mutuamente excluyentes, es

REGLA GENERAL DE LA ADICIÓN $P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$ **[5.4]**

En el caso de la expresión $P(A \text{ o } B)$, la letra *o* sugiere que puede ocurrir A o puede ocurrir B. Esto también incluye la posibilidad de que A y B ocurran. Tal uso de *o* a veces se denomina **inclusivo.** También es posible escribir $P(A \text{ o } B \text{ o ambos})$ para hacer hincapié en el hecho de que la unión de dos eventos incluye la intersección de A y B.

Al comparar las reglas general y especial de la adición, la diferencia que importa consiste en determinar si los eventos son mutuamente excluyentes. Si lo son, entonces la probabilidad conjunta $P(A \text{ y } B)$ es 0 y se podría aplicar la regla especial de la adición. De lo contrario, tome en cuenta la probabilidad conjunta y aplique la regla general de la adición.

EJEMPLO

¿Cuál es la probabilidad de que una carta escogida al azar de una baraja convencional sea rey o de corazones?

SOLUCIÓN

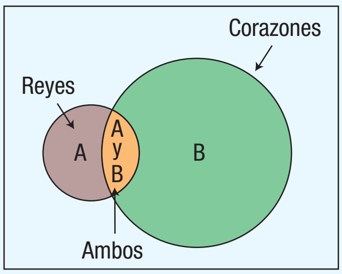
Quizá se sienta tentado a sumar la probabilidad de sacar un rey y la probabilidad de sacar una carta de corazones. Sin embargo, este enfoque crea problemas. Al hacerlo así, el rey de corazones se contará con los reyes y con los corazones. De esta manera, si suma la probabilidad de sacar un rey (hay 4 en una baraja de 52 cartas) a la probabilidad de sacar un corazón (hay 13 en una baraja de 52 cartas) 17 de 52 cartas cumplen con el requisito, pero ha contado dos veces el rey de corazones. Necesita restar una carta de las 17, de tal manera que el rey de corazones solo se cuente una vez. Por lo tanto, hay 16 cartas que son corazones o reyes. Así que la probabilidad es de 16/52 = 0.3077.

Carta	Probabilidad	Explicación
Rey	$P(A) = 4/52$	4 reyes en una baraja de 52 cartas
Corazones	$P(B) = 13/52$	13 corazones en una baraja de 52 cartas
Rey de corazones	$P(A \text{ y } B) = 1/52$	1 rey de corazones en una baraja de 52 cartas

De acuerdo con la fórmula [5.4]:

$$\begin{aligned} P(A \text{ o } B) &= P(A) + P(B) - P(A \text{ y } B) \\ &= 4/52 + 13/52 + 1/52 \\ &= 16/52, \text{ o } 0.3077 \end{aligned}$$

Un diagrama de Venn representa estos resultados, que no son mutuamente excluyentes.



AUTOEVALUACIÓN 5-4

Cada año se llevan a cabo exámenes físicos de rutina como parte de un programa de servicios de salud para los empleados de General Concrete, Inc. Se descubrió que 8% de los empleados requieren calzado ortopédico; 15% necesitan tratamiento dental mayor y 3% tanto zapatos ortopédicos como tratamiento dental mayor.

(a) ¿Cuál es la probabilidad de que un empleado elegido de forma aleatoria requiera zapatos ortopédicos o tratamiento dental mayor?
(b) Muestre esta situación en forma de diagrama de Venn.

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

11. Los eventos *A* y *B* son mutuamente excluyentes. Suponga que $P(A) = 0.30$ y $P(B) = 0.20$. ¿Cuál es la probabilidad de que *A* o *B* ocurran? ¿Cuál es la probabilidad de que ni *A* ni *B* sucedan?

12. Los eventos *X* y *Y* son mutuamente excluyentes. Si $P(X) = 0.05$ y $P(Y) = 0.02$. ¿Cuál es la probabilidad de que *X* o *Y* ocurran? ¿Cuál es la probabilidad de que ni *X* ni *Y* sucedan?

13. Un estudio de 200 empresas de publicidad reveló los siguientes ingresos después de impuestos:

Ingreso después de impuestos	Número de empresas
Menos de $1 millón	102
De $1 millón a $20 millones	61
$20 millones o más	37

a. ¿Cuál es la probabilidad de que una empresa de publicidad seleccionada al azar tenga un ingreso después de impuestos menor a un millón de dólares?

b. ¿Cuál es la probabilidad de que una empresa de publicidad seleccionada al azar tenga un ingreso después de impuestos entre uno y 20 millones de dólares o un ingreso de 20 millones de dólares o más? ¿Qué regla de probabilidad aplicó?

14. El presidente de la junta directiva afirma: “Hay 50% de posibilidades de que esta compañía obtenga utilidades; 30% de que termine sin pérdidas ni ganancias y 20% de que pierda dinero durante el próximo trimestre”.

a. Aplique una de las reglas de la adición para determinar la probabilidad de que la compañía no pierda dinero el siguiente trimestre.

b. Aplique la regla del complemento para determinar la probabilidad de que no pierda dinero el próximo trimestre.

15. Suponga que la probabilidad de que saque una A en esta clase es de 0.25 y que la probabilidad de obtener una B es de 0.50. ¿Cuál es la probabilidad de que su calificación sea mayor que C?

16. Se lanzan al aire dos monedas. Si *A* es el evento “dos caras” y *B* es el evento “dos cruces”, ¿*A* y *B* son mutuamente excluyentes? ¿Son complementos?

17. Las probabilidades de los eventos *A* y *B* son 0.20 y 0.30, respectivamente. La probabilidad de que *A* y *B* ocurran es de 0.15. ¿Cuál es la probabilidad de que *A* o *B* ocurran?

18. Sean $P(X) = 0.55$ y $P(Y) = 0.35$. Suponga que la probabilidad de que ambos ocurran es de 0.20. ¿Cuál es la probabilidad de que *X* o *Y* ocurran?

19. Suponga que los dos eventos *A* y *B* son mutuamente excluyentes. ¿Cuál es la probabilidad de que se presenten de forma conjunta?

20. Un estudiante toma dos cursos, uno de historia y otro de matemáticas. La probabilidad de pasar el curso de historia es de 0.60, la de aprobar matemáticas es de 0.70 y la de pasar ambos es de 0.50. ¿Cuál es la probabilidad de acreditar por lo menos uno?

21. El acuario de Sea Critters Depot contiene 140 peces espada. De estos, 80 son verdes (44 hembras y 36 machos), y 60 son anaranjados (36 hembras y 24 machos). Se captura uno al azar en este acuario.

a. ¿Cuál es la probabilidad de que sea verde?

b. ¿Cuál es la probabilidad de que sea un macho?

c. ¿Cuál es la probabilidad de que sea un macho verde?

d. ¿Cuál es la probabilidad de que sea macho o verde?

22. Un estudio llevado a cabo por el National Service Park reveló que 50% de los vacacionistas que se dirigen a la región de las Montañas Rocallosas visitan el parque de Yellowstone; 40%, el de los Tetons, y 35%, ambos lugares.

a. ¿Cuál es la probabilidad de que un vacacionista visite por lo menos una de estas atracciones?

b. ¿Qué nombre recibe la probabilidad de 0.35?

c. ¿Los eventos son mutuamente excluyentes? Explique su respuesta.

Reglas de la multiplicación

OA5-4
Calcular probabilidades mediante las *reglas de la multiplicación*.

En esta sección se estudian las reglas para calcular la probabilidad de que la ocurrencia de dos eventos sea simultánea; es decir, su probabilidad conjunta. Por ejemplo, 16% de las declaraciones de impuestos de 2012 fueron preparadas por H&R Block, y 75% mostraron un reembolso. ¿Cuál es la posibilidad de que la forma fiscal de una persona haya sido preparada por H&R Block, y que esa persona reciba un reembolso? Los diagramas de Venn ilustran este hecho como la intersección de dos eventos. Para determinar la probabilidad de dos eventos que se presentan simultáneamente se emplea la regla de la multiplicación; de la cual, hay dos tipos: especial y general.

Regla especial de la multiplicación

Esta regla requiere que dos eventos, A y B, sean independientes, y lo son si el hecho de que uno ocurra no altera la probabilidad de que el otro suceda.

INDEPENDENCIA Si un evento ocurre, no tiene ningún efecto sobre la probabilidad de que otro evento acontezca.

Una forma de entender la independencia consiste en suponer que los eventos A y B ocurren en diferentes tiempos. Por ejemplo, cuando el evento B ocurre después del evento A, ¿influye A en la probabilidad de que el evento B ocurra? Si la respuesta es no, entonces A y B son eventos independientes. Para ilustrar la independencia, suponga que se lanzan al aire dos monedas. El resultado del lanzamiento de una moneda (cara o cruz) no se altera por el resultado de cualquier moneda lanzada previamente.

En el caso de dos eventos independientes A y B, la probabilidad de que A y B ocurran se determina multiplicando las dos probabilidades, tal es la **regla especial de la multiplicación**, cuya expresión simbólica es la siguiente:

REGLA ESPECIAL DE LA MULTIPLICACIÓN $P(A \text{ y } B) = P(A)P(B)$ **[5.5]**

En el caso de tres eventos independientes, A, B y C, la regla especial de la multiplicación que se utiliza para determinar la probabilidad de que los tres eventos ocurran es:

$$P(A \text{ y } B \text{ y } C) = P(A)P(B)P(C)$$

EJEMPLO

Una encuesta que llevó a cabo la American Automobile Association (AAA) reveló que el año anterior 60% de sus miembros hicieron reservaciones en líneas aéreas. Dos de ellos fueron seleccionados al azar. ¿Cuál es la probabilidad de que ambos hicieran reservaciones el año previo?

SOLUCIÓN

La probabilidad de que el primero haya hecho una reservación el año pasado es de 0.60, que se expresa como $P(R_1) = 0.60$, en la que R_1 representa el hecho de que el primer miembro hizo una reservación. La probabilidad de que el segundo miembro elegido haya hecho una reservación es la misma, así que $P(R_2) = 0.60$. Como el número de miembros de la AAA es muy grande, se supone que R_1 y R_2 son independientes. En consecuencia, de acuerdo con la fórmula [5.5], la probabilidad de que ambos hayan hecho una reservación es de 0.36, que se calcula de la siguiente manera:

$$P(R_1 \text{ y } R_2) = P(R_1)P(R_2) = (0.60)(0.60) = 0.36$$

Todos los posibles resultados pueden representarse como se muestra a continuación. Aquí, R significa que se hizo la reservación y $\sim R$, que no se hizo.

Con las probabilidades y la regla del complemento se calcula la probabilidad conjunta de cada resultado. Por ejemplo, la probabilidad de que ningún miembro haga una reservación es de 0.16. Además, la probabilidad de que el primero y el segundo miembros (regla especial de la adición) hagan una reservación es de 0.48 (0.24 + 0.24). También se puede observar que los resultados son mutuamente excluyentes y colectivamente exhaustivos. Por lo tanto, las probabilidades suman 1.

Resultados		Probabilidad conjunta	
R_1	R_2	(.60)(.60) =	.36
R_1	$\sim R_2$	(.60)(.40) =	.24
$\sim R_1$	R_2	(.40)(.60) =	.24
$\sim R_1$	$\sim R_2$	(.40)(.40) =	.16
Total			1.00

AUTOEVALUACIÓN 5-5

Por experiencia, Teton Tire sabe que la probabilidad de que una llanta XB-70 rinda 60 000 millas antes de quedar lisa o falle es de 0.95. A cualquier llanta que no dure las 60 000 millas se le hacen arreglos. Usted adquiere cuatro llantas XB-70. ¿Cuál es la probabilidad de que las cuatro llantas tengan una duración de 60 000 millas?

Regla general de la multiplicación

Si dos eventos no son independientes, se dice que son **dependientes.** Con el fin de ilustrar el concepto de dependencia, suponga que hay 10 latas de refresco en un refrigerador, 7 de los cuales son normales y 3 dietéticos. Se saca una lata del refrigerador. La probabilidad de que sea una lata de refresco dietético es de 3/10, y la probabilidad de que sea una lata de refresco normal es de 7/10. Luego, se elige una segunda lata del refrigerador sin devolver la primera. La probabilidad de que la segunda lata sea de refresco dietético depende de que la primera lo haya sido o no. La probabilidad de que la segunda lata sea de refresco dietético es:

> 2/9, si la primera bebida es dietética (solo quedan dos latas de refresco dietético en el refrigerador).
> 3/9, si la primera lata elegida es normal (los tres refrescos aún están en el refrigerador).

La fracción 2/9 (o 3/9) es **probabilidad condicional** porque su valor se encuentra condicionado (o depende) del hecho de que un refresco regular o dietético haya sido el primero en ser seleccionado del refrigerador.

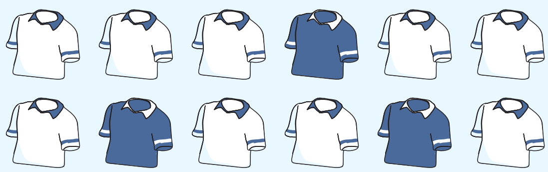
PROBABILIDAD CONDICIONAL Probabilidad de que un evento en particular ocurra, dado que otro evento haya acontecido.

En la regla general de la multiplicación se requiere la probabilidad condicional para calcular la probabilidad conjunta de dos eventos que no son independientes. Para dos eventos, *A* y *B* que no son independientes, la probabilidad condicional se representa como $P(B|A)$, y se expresa como la probabilidad de *B* dada *A*. O la probabilidad de *B* es condicional a la ocurrencia y efecto del evento *A*. Simbólicamente, la regla general de la multiplicación para dos eventos que no son independientes es:

REGLA GENERAL DE LA MULTIPLICACIÓN $$P(A \text{ y } B) = P(A)P(B|A) \quad \textbf{[5.6]}$$

EJEMPLO

Un golfista tiene 12 camisas en su clóset. Suponga que 9 son blancas y las demás, azules. Como se viste de noche, simplemente toma una camisa y se la pone. Juega golf dos veces seguidas y no lava las camisas usadas ni las regresa al clóset. ¿Cuál es la probabilidad de que las dos camisas elegidas sean blancas?



SOLUCIÓN

El evento que se relaciona con el hecho de que la primera camisa seleccionada sea blanca es W_1. La probabilidad es $P(W_1) = 9/12$ porque 9 de cada 12 camisas son blancas. El evento de que la segunda camisa seleccionada también sea blanca se identifica con W_2. La probabilidad condicional relacionada con el hecho de que la segunda camisa seleccionada sea blanca, dado que la primera camisa seleccionada es blanca también, es $P(W_2|W_1) = 8/11$. ¿A qué se debe esto? A que después de seleccionar la primera camisa, quedan 11 camisas en el clóset y 8 de estas son blancas. Para determinar la probabilidad de que se elijan dos camisas blancas se aplica la fórmula [5.6]:

$$P(W_1 \text{ y } W_2) = P(W_1)P(W_2|W_1) = \left(\frac{9}{12}\right)\left(\frac{8}{11}\right) = 0.55$$

Por consiguiente, la probabilidad de seleccionar dos camisas, y que ambas sean de color blanco, es de 0.55.

Es posible ampliar la regla general de la multiplicación para que incluya más de dos eventos. En el caso de los tres eventos, *A*, *B* y *C*, la fórmula es:

$$P(A \text{ y } B \text{ y } C) = P(A)P(B|A)P(C|A \text{ y } B)$$

En el caso del ejemplo de la camisa de golf, la probabilidad de elegir tres camisas blancas sin reemplazo es:

$$P(W_1 \text{ y } W_2 \text{ y } W_3) = P(W_1)P(W_2|W_1)P(W_3|W_1 \text{ y } W_2) = \left(\frac{9}{12}\right)\left(\frac{8}{11}\right)\left(\frac{7}{10}\right) = 0.38$$

Así, la probabilidad de seleccionar tres camisas sin reemplazo, todas las cuales sean blancas, es de 0.38.

AUTOEVALUACIÓN 5-6

La junta directiva de Tarbell Industries consta de ocho hombres y cuatro mujeres. Un comité de cuatro miembros será elegido al azar para llevar a cabo una búsqueda, en todo el país, del nuevo presidente de la compañía.

(a) ¿Cuál es la probabilidad de que los cuatro miembros del comité de búsqueda sean mujeres?
(b) ¿Cuál es la probabilidad de que los cuatro miembros del comité de búsqueda sean hombres?
(c) ¿Las probabilidades de los eventos descritos en los puntos anteriores suman 1? Explique su respuesta.

OA5-5 Calcular probabilidades por medio de una tabla de contingencia.

Tablas de contingencia

A menudo, los resultados de una encuesta se registran en una tabla de dos direcciones y se utilizan para determinar diversas probabilidades. Ya se ha descrito esta idea a partir de la sección "Tablas de contingencia" del capítulo 4. Para recordarlo: una tabla de dos direcciones es una tabla de contingencia.

TABLA DE CONTINGENCIA Se utiliza para clasificar observaciones de una muestra de acuerdo con dos o más características identificables.

Una tabla de contingencia consiste en una tabulación cruzada que resume simultáneamente dos variables de interés, así como la relación entre estas. El nivel de medición puede ser nominal. He aquí algunos ejemplos.

- Se preguntó a 150 adultos su género y la cantidad de cuentas de Facebook que usan. En la siguiente tabla se resumen los resultados.

	Género		
Cuentas de Facebook	**Hombres**	**Mujeres**	**Total**
0	20	40	60
1	40	30	70
2 o más	10	10	20
Total	70	80	150

ESTADÍSTICA EN ACCIÓN

En el año 2000, George W. Bush ganó la presidencia de Estados Unidos por un mínimo margen. Surgieron muchas historias sobre las elecciones, algunas de las cuales hablaban de irregularidades en las votaciones y otras dieron lugar a interesantes preguntas. En una elección local de Michigan resultó un empate entre dos candidatos para un puesto de elección. Para

(continúa)

- La American Coffee Producers Association proporciona la siguiente información sobre la edad y la cantidad de café que se consumió en un mes.

	Consumo de café			
Edad (años)	**Bajo**	**Moderado**	**Alto**	**Total**
Menos de 30	36	32	24	92
30 hasta 40	18	30	27	75
40 hasta 50	10	24	20	54
50 o más	26	24	29	79
Total	90	110	100	300

De acuerdo con esta tabla, cada uno de los 300 entrevistados se clasifica según dos criterios: 1) edad y 2) cantidad de café que consume.

En el siguiente ejemplo se muestra la forma en que las reglas de adición y multiplicación se emplean en tablas de contingencias.

(*continuación*)

resolver el empate, los candidatos sacaron una hoja de papel de una caja que contenía dos hojas, una rotulada *Ganador*, y otra sin marcar. Para determinar qué candidato sacaría primero el papel, los funcionarios electorales lanzaron una moneda al aire. El ganador del lanzamiento también sacó el papel del ganador. Ahora bien, ¿era realmente necesario lanzar una moneda al aire? No, porque los dos eventos son independientes. Ganar en el lanzamiento de la moneda no altera la probabilidad de que cualquiera de los candidatos saque la hoja con el nombre del ganador.

EJEMPLO

El mes anterior, la Asociación Nacional de Administradores de Salas Cinematográficas realizó una encuesta entre 500 adultos seleccionados al azar. La encuesta preguntaba a las personas su edad y el número de veces que habían visto una película en un cine. Los resultados se resumen en la tabla 5.1

TABLA 5.1 Número de películas vistas por mes y por edad

	Edad			
Películas por mes	**Menos de 30** B_1	**30 hasta 60** B_2	**60 o más** B_3	**Total**
0 A_1	15	50	10	75
1 o 2 A_2	25	100	75	200
3, 4 o 5 A_3	55	60	60	175
6 o más A_4	5	15	30	50
Total	100	225	175	500

La asociación está interesada en entender las probabilidades de que un adulto vaya a ver una película al cine, especialmente en el caso de adultos mayores de 60 años. Esta información es útil para tomar decisiones con respecto a los descuentos en boletos y concesiones para los mayores. Determine la probabilidad de seleccionar un adulto que vio:

1. Seis o más películas por mes.
2. Dos o menos películas por mes.
3. Seis o más películas por mes **o** tiene 60 años o más.
4. Seis o más películas por mes **dado que** la persona tiene 60 años o más.
5. Seis o más películas por mes **y** tiene 60 años o más.

y establezca la:

6. Independencia del número de películas por mes que fueron vistas, y la edad del adulto.

SOLUCIÓN

La tabla 5.1 es de contingencias. En este tipo de tabla se clasifica a un individuo o a un objeto de acuerdo con dos criterios; en este ejemplo, un adulto de la muestra se clasifica por edad y por el número de películas que vio en el cine por mes. Las reglas de la adición (fórmulas [5.2] y [5.4]) y las de la multiplicación (fórmulas [5.5] y [5.6]) permiten responder varias preguntas de probabilidad con base en la tabla de contingencia.

1. Para encontrar la probabilidad de que un adulto seleccionado al azar haya visto seis o más películas por mes, enfóquese en la línea "6 o más" (también nombrada A_4) en la tabla 5.1. La tabla muestra que 50 de los 500 adultos están en esta clase. Utilizando el enfoque empírico, la probabilidad se calcula así:

$$P(6 \text{ o más}) = P(A_4) = \frac{50}{500} = 0.10$$

Esta probabilidad indica que 10% de los adultos vieron seis o más películas por mes.

2. Para determinar la probabilidad de seleccionar al azar un adulto que vio dos o menos películas por mes, deben combinarse dos resultados: no ver ninguna película por mes y ver una o dos películas por mes. Ambos resultados son mutuamente excluyentes; es decir, una persona solo puede clasificarse como que no vio películas, o bien vio una o dos películas por mes; por tanto, se utiliza la regla especial de la adición (fórmula [5.2]) sumando las posibilidades de no ver ninguna película y ver una o dos:

$$P[(\text{ver } 0) \text{ o } (\text{ver } 1 \text{ o } 2) = P(A_1) + P(A_2) = \left(\frac{75}{500} + \frac{200}{500}\right) = 0.55$$

Por lo tanto, 55% de los adultos de la muestra vio dos o menos películas por mes.

3. Para determinar la probabilidad de seleccionar al azar a un adulto que vio "6 o más" películas por mes o cuya edad es "60 o más", de nuevo se usan las reglas de la adición. Sin embargo, en este caso los resultados **no son** mutuamente excluyentes. ¿Por qué? Porque una persona puede ver más de seis películas al mes, tener 60 años o más, o ambas. De manera que ambos grupos no son mutuamente excluyentes porque es posible que una persona pueda quedar en los dos. Para determinar esta probabilidad se utiliza la regla general de la adición (fórmula [5.4]):

$$P[(6 \text{ o más}) \text{ o } (60 \text{ o más})] = P(A_4) + P(B_3) - P(A_4 \text{ y } B_3) = \left(\frac{50}{500} + \frac{175}{500} - \frac{30}{500}\right) = 0.39$$

Por lo tanto, 39% de los adultos tienen 60 años o más, ven seis o más películas al mes, o ambos.

4. Para determinar la probabilidad de seleccionar a una persona que ve seis o más películas por mes dado que la persona tiene 60 años o más, enfóquese solo en la columna rotulada B_3 en la tabla 5.1 Es decir, solo interesan los 175 adultos que tienen 60 años o más. De estos 175 adultos, 30 vieron seis películas o más. Utilizando la regla general de la multiplicación (fórmula [5.6]):

$$P[(6 \text{ o más}) \text{ dado que } (60 \text{ o más})] = P(A_4|B_3) = \frac{30}{175} = 0.17$$

De los 500 adultos, 17% de los que tienen 60 años o más vieron seis o más películas al mes. Esto recibe el nombre de probabilidad condicional porque está basada en la "condición" de tener 60 años o más. Recuerde que en el punto 1, 10% de todos los adultos vieron seis películas o más por mes; aquí se ve que 17% de quienes tienen 60 años o más ven películas. Esta es una información valiosa para los administradores de los cines con respecto a las características de sus clientes.

5. La probabilidad de que una persona viera seis o más películas y tenga 60 años o más se basa en dos condiciones que deben darse forzosamente. Esto es, que ambos resultados, "6 o más" (A_4) y "60 o más" (B_3) deben ocurrir conjuntamente. Para encontrar esta probabilidad conjunta se utiliza la regla especial de la multiplicación (fórmula [5.6]).

$$P[(6 \text{ o más}) \text{ y } (60 \text{ o más})] = P(A_4 \text{ y } B_3) = P(A_4)\,P(B_3|A_4)$$

Para calcular la probabilidad conjunta, calcule primero la probabilidad simple del primer resultado, A_4, seleccionar al azar una persona que vea seis o más películas. Para encontrar la probabilidad, refiérase a la fila A_4 en la tabla 5.1. De los 500 adultos, 50 vieron seis o más películas. Así que $P(A_4) = 50/500$.

Enseguida, calcule la probabilidad condicional $P(B_3|A_4)$. Esta es la probabilidad de seleccionar un adulto que tiene 60 años o más dado que la persona haya visto 6 o más películas. La probabilidad condicional es:

$$P[(60 \text{ o más}) \text{ dado que } (60 \text{ o más})] = P(B_3|A_4) = 30/50$$

Utilizando ambas probabilidades, la probabilidad conjunta de que un adulto vea seis o más películas y tenga 60 años o más es:

$$P[(6 \text{ o más}) \text{ y } (60 \text{ o más})] = P(A_4 \text{ y } B_3) = P(A_4)\,P(B_3|A_4) = (50/500)(30/50) = 0.06$$

Basándonos en la información de la muestra de la tabla 5.1, la probabilidad de que un adulto tenga más de 60 años y vea seis películas o más es 6%. Es importante saber que ese 6% es relativo a los 500 adultos.

¿Existe otra forma de determinar esta probabilidad conjunta sin usar la fórmula de la regla especial de la multiplicación? Sí la hay. Mire directamente a la célula donde la fila A_4, (6 o más), y la columna B_3, (60 o más), se intersectan. Hay 30 adultos en esta celda que reúnen ambos criterios, así que $P(A_4 \text{ y } B_3) = 30/500 = 0.06$. Es el mismo resultado que calculando con la fórmula.

6. ¿Son interdependientes los eventos? Es posible responder a esta pregunta con la ayuda de los resultados del punto 4, en el que se encontró que la probabilidad de seleccionar a un adulto de 60 años o más, dado que vio seis o más películas, era 0.17. Si la edad no es un factor en la asistencia al cine entonces se espera que la probabilidad de que una persona que tenga 30 años o menos y haya visto seis o más películas también sea 17%. Esto es, ambas probabilidades condicionales son iguales. La probabilidad de que un adulto vea seis o más películas por mes dado que tiene menos de 30 años es:

$$P[(6 \text{ o más}) \text{ dado que (menos de 30)}] = \frac{5}{100} = 0.05$$

Como ambas probabilidades no son iguales, el número de películas vistas y la edad no son independientes. Dicho de otro modo, en el caso de los 500 adultos, la edad está relacionada con el número de películas vistas. En el capítulo 15 se investiga el concepto de independencia con mayor detalle.

AUTOEVALUACIÓN 5-7

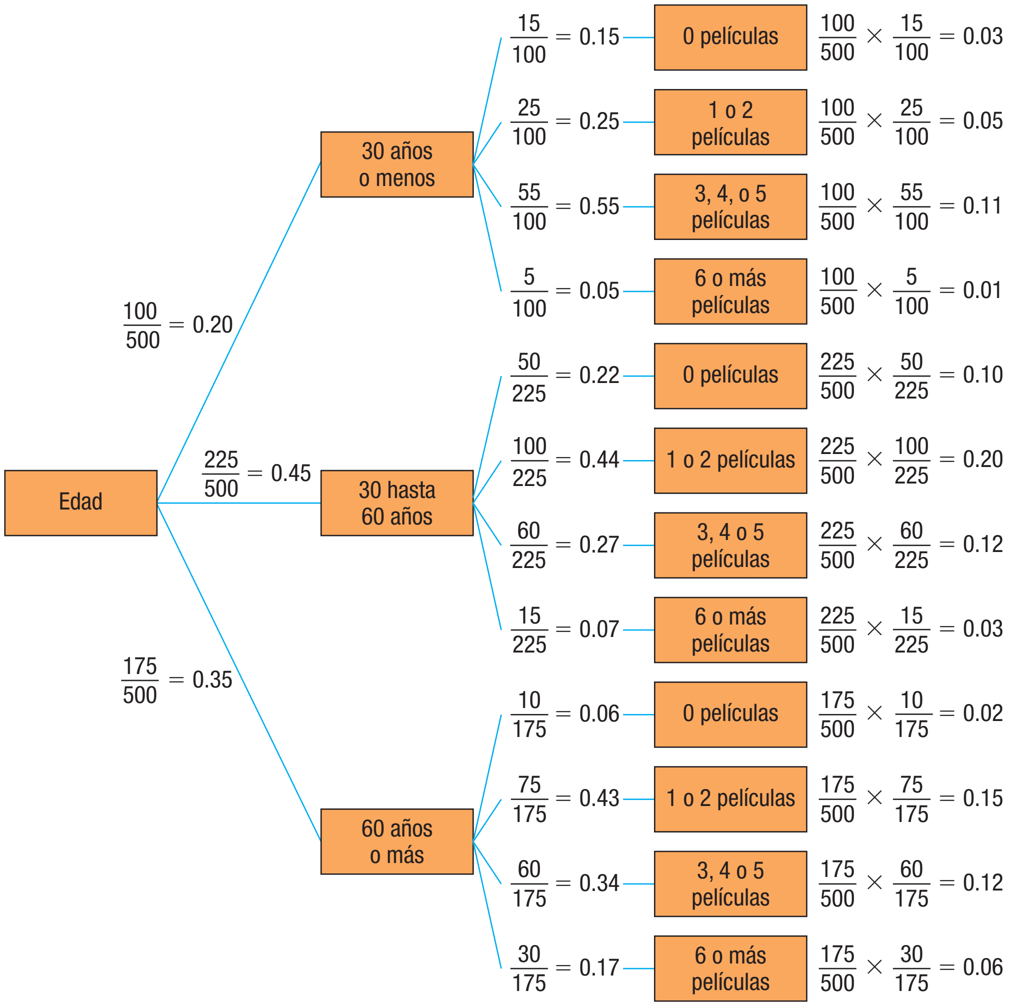
Consulte la tabla 5.1 para calcular las probabilidades de seleccionar a un adulto que tenga:
(a) De 30 hasta 60 años.
(b) Menos de 60 años.
(c) Menos de 30 años o no fue al cine.
(d) Menos de 30 años y no fue al cine.

Diagramas de árbol

El **diagrama de árbol** es una gráfica útil para organizar y calcular probabilidades para problemas similares al ejemplo previo. Este tipo de problema implica varias etapas y cada una se ilustra con la rama del árbol. Las ramas del árbol se etiquetan con las probabilidades. Se utiliza la información de la tabla 5.1 para mostrar la construcción de un diagrama de árbol.

1. Para construir un diagrama de árbol, comience dibujando un punto grueso a la izquierda para representar la raíz del árbol (vea la gráfica 5.2 en la página siguiente).
2. Hay tres ramas principales que salen de la raíz. En la rama superior se representa el evento de un adulto que tiene menos de 30 años. La rama se etiqueta con la probabilidad $P(B_1) = 100/500$. En la siguiente rama se representa el resultado de los adultos que tienen de 30 hasta 60 años, y se etiqueta con la probabilidad $P(B_2) = 225/500$. La rama restante se etiqueta $P(B_3) = 175/500$.
3. De cada una de las ramas principales *salen* cuatro ramas, las cuales representan las cuatro categorías de películas vistas por mes: 0; 1 o 2; 3, 4 o 5; y 6 o más. Las ramas superiores del árbol representan la probabilidad condicional de un adulto que no vio ninguna película dado que tiene menos de 30 años. Estas se escriben $P(A_1)|B_1)$, $P(A_2)|B_1)$, $P(A_3)|B_1)$ y $P(A_4)|B_1)$, donde $P(A_1)|B_1)$, donde A_1 se refiere a no ver películas; A_2, a ver 1 o 2 películas por mes; A_3, a ver 3, 4 o 5 películas por mes; y A_4, a ver 6 o más películas por mes. Para la rama superior del árbol, estas probabilidades son 15/100, 25/100, 55/100 y 5/100. Las probabilidades condicionales se escriben en forma similar en el resto de las ramas.
4. Por último, se determinan las diversas probabilidades conjuntas. Para las ramas superiores, los eventos son: un adulto no ve películas en el mes y tiene 30 años o menos; un adulto ve 1 o 2 películas y tiene 30 años o menos; y un adulto ve 3, 4 o 5 películas al mes y tiene 30 años o menos; y un adulto ve 6 películas o más y tiene 30 años o menos. Estas probabilidades conjuntas se muestran en el lado derecho de la gráfica 5.2. Para explicar, la probabilidad conjunta de que un adulto seleccionado al azar tenga menos de 30 años y no vea películas durante el mes es:

$$P(B_1 \text{ y } A_1) = P(B_1)P(A_1 \mid B_1) = \left(\frac{100}{500}\right)\left(\frac{15}{100}\right) = 0.03$$



GRÁFICA 5.2 Diagrama de árbol que muestra la edad y el número de películas vistas

En los tres diagramas se resumen todas las probabilidades basándose en la tabla de contingencia 5.1. Por ejemplo, las probabilidades condicionales muestran que el grupo "60 o más" tiene el porcentaje más alto (17%) de quienes ven seis o más películas al mes. El grupo "30 a 60" tiene el porcentaje más alto (22%) de no ver películas al mes. Con base en las probabilidades conjuntas, 20% de los adultos encuestados ven una o dos películas por mes y tienen 30 hasta 60 años. Por tanto, se pueden hacer muchas observaciones mediante la información presentada en un diagrama de árbol.

AUTOEVALUACIÓN 5-8

Considere una encuesta a algunos consumidores relacionada con la cantidad relativa de visitas que hacen a una tienda Sears (con frecuencia, en ocasiones o nunca) y con el hecho de que la tienda se ubique en un centro comercial (sí y no). Cuando las variables son de escala nominal, tal como estos datos, por lo general los resultados se resumen en una tabla de contingencia.

	Centro comercial cerrado		
Visitas	**Sí**	**No**	**Total**
Con frecuencia	60	20	80
En ocasiones	25	35	60
Nunca	5	50	55
	90	105	195

Determine la probabilidad de seleccionar a un comprador que visitó una tienda Sears:

(a) Con frecuencia.
(b) Ubicada en un centro comercial cerrado.
(c) Ubicada en un centro comercial cerrado o lo hizo con frecuencia.
(d) Con frecuencia, dado que se ubicaba en un centro comercial cerrado.

Además:

(e) ¿Las variables "cantidad de visitas" y "centro comercial cerrado" son independientes?
(f) ¿Cuál es la probabilidad de elegir un comprador que visitó con frecuencia una tienda Sears ubicada en un centro comercial cerrado?
(g) Trace un diagrama de árbol y determine las diversas probabilidades conjuntas.

EJERCICIOS

23. Suponga que $P(A) = 0.40$ y $P(B|A) = 0.30$. ¿Cuál es la probabilidad conjunta de A y B?

24. Suponga que $P(X_1) = 0.75$ y $P(Y_2|X_1) = 0.40$. ¿Cuál es la probabilidad conjunta de X_1 y Y_2?

25. Un banco local informa que 80% de sus clientes tienen cuenta de cheques; 60% tienen cuenta de ahorros y 50% cuentan con ambas. Si se elige un cliente al azar, ¿cuál es la probabilidad de que el cliente tenga una cuenta de cheques o de ahorros? ¿Cuál es la probabilidad de que el cliente no tenga cuenta de cheques ni de ahorros?

26. All Seasons Plumbing tiene dos camiones de servicio que se descomponen con frecuencia. Considere estas probabilidades: que el primer camión esté disponible: 0.75; que el segundo esté disponible: 0.50; que ambos lo estén: 0.30. ¿Cuál es la probabilidad de que ningún camión se encuentre disponible?

27. Observe la siguiente tabla.

	Primer evento			
Segundo evento	A_1	A_2	A_3	**Total**
B_1	2	1	3	6
B_2	1	2	1	4
Total	3	3	4	10

a. Determine $P(A_1)$.
b. Estime $P(B_1|A_2)$.
c. Aproxime $P(B_2 \text{ y } A_3)$.

28. Clean-brush Products envió por accidente tres cepillos dentales eléctricos defectuosos a una farmacia, además de 17 sin defectos.
a. ¿Cuál es la probabilidad de que los primeros dos cepillos eléctricos vendidos sean devueltos a la farmacia por estar defectuosos?
b. ¿Cuál es la probabilidad de que los primeros dos cepillos eléctricos vendidos no estén defectuosos?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

29. Cada vendedor de Puchett, Sheets, and Hogan Insurance Agency recibe una calificación (debajo del promedio, promedio y por encima del promedio) en lo que se refiere a sus habilidades en ventas. A cada vendedor también se le califica por su potencial para progresar (regular, bueno o excelente). En la siguiente tabla se muestra una clasificación cruzada de estas características de los 500 empleados.

	Potencial para progresar		
Habilidades en ventas	**Regular**	**Bueno**	**Excelente**
Debajo del promedio	16	12	22
Promedio	45	60	45
Por encima del promedio	93	72	135

a. ¿Qué nombre recibe esta tabla?
b. ¿Cuál es la probabilidad de que una persona elegida al azar tenga habilidad para las ventas por encima del promedio y excelente potencial para progresar?
c. Construya un diagrama de árbol que muestre las probabilidades, las probabilidades condicionales y las probabilidades conjuntas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

30. Un inversionista cuenta con tres acciones ordinarias independientes entre sí y con las mismas probabilidades de: 1) incrementar su valor; 2) bajar su valor; 3) permanecer con el mismo valor. Elabore una lista de los posibles resultados de este experimento. Calcule la probabilidad de que por lo menos dos de las acciones aumenten de valor.

31. Una encuesta a 545 estudiantes de educación superior planteó las siguientes preguntas: "¿Cuál es tu deporte de invierno favorito?", y "¿A qué tipo de institución asistes?". Los resultados se resumen en la siguiente tabla.

	Deporte de invierno favorito			
Tipo de institución	**Snowboarding**	**Esquí**	**Patinaje sobre hielo**	**Total**
Escuela técnica	68	41	46	155
Universidad	84	56	70	210
Facultad de posgrado	59	74	47	180
Total	211	171	163	545

Utilizando a estos 545 estudiantes como muestra, se seleccionó uno al azar para el estudio.

a. ¿Cuál es la probabilidad de elegir a un estudiante cuyo deporte favorito sea el esquí?
b. ¿Cuál es la probabilidad de elegir a un estudiante de una escuela técnica?
c. Si el estudiante elegido está en una universidad, ¿cuál es la probabilidad de que prefiera el patinaje sobre hielo?
d. Si el estudiante elegido prefiere el snowboarding, ¿cuál es la probabilidad de que asista a una escuela técnica?
e. Si se selecciona un estudiante de posgrado, ¿cuál es la probabilidad de que prefiera el esquí o el patinaje sobre hielo?

32. Si pregunta a tres extraños las fechas de sus cumpleaños, ¿cuál es la probabilidad de que a) todos haya nacido el miércoles; b) todos hayan nacido en diferentes días de la semana; c) todos hayan nacido el sábado?

OA5-6
Calcular probabilidades con base en el teorema de Bayes.

Teorema de Bayes

En el siglo XVIII, el reverendo Thomas Bayes, un ministro presbiteriano inglés, planteó esta pregunta: "¿Dios realmente existe?". Dado su interés en las matemáticas, intentó crear una fórmula para llegar a la probabilidad de que Dios existiera con base en la evidencia disponible en la Tierra. Más tarde, Pierre-Simon Laplace perfeccionó el trabajo de Bayes y le dio el nombre de teorema de Bayes. De una forma entendible, el **teorema de Bayes** es el siguiente:

TEOREMA DE BAYES

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \quad \textbf{[5.7]}$$

En la fórmula [5.7] asuma que los eventos A_1 y A_2 son mutuamente excluyentes y colectivamente exhaustivos, y A_i se refiere a cualquiera de ambos eventos. De ahí que en este caso A_1 y A_2 sean complementos. El significado de los símbolos utilizados se ilustra en el siguiente ejemplo.

Suponga que 5% de la población de Umen, un país ficticio del Tercer Mundo, tiene una enfermedad propia del país. Sea A_1 el evento "padece la enfermedad" y A_2 el evento "no padece la enfermedad". Por lo tanto, si selecciona al azar a una persona de Umen, la probabilidad de que el individuo elegido padezca la enfermedad es de 0.05 o $P(A_1) = 0.05$. Esta probabilidad, $P(A_1) = P(\text{padece la enfermedad}) = 0.05$, recibe el nombre de **probabilidad *a priori*.** Se le da este nombre porque la probabilidad se asigna antes de obtener los datos empíricos.

PROBABILIDAD *A PRIORI* La que está basada en el nivel de información actual.

Por ende, la probabilidad *a priori* de que una persona no padezca la enfermedad es de 0.95, o $P(A_2) = 0.95$, que se calcula mediante la resta $1 - 0.05$.

Existe una técnica de diagnóstico para detectar la enfermedad, pero no es muy precisa. Sea B el evento "la prueba revela la presencia de la enfermedad". Suponga que la evidencia histórica muestra que si una persona padece realmente la enfermedad, la probabilidad de que la prueba indique su presencia es de 0.90. De acuerdo con las definiciones de probabilidad condicional que se establecieron en el capítulo, dicho enunciado se expresa de la siguiente manera:

ESTADÍSTICA EN ACCIÓN

Un estudio reciente de la National Collegiate Athletic Association (NCAA) informó que de los 150 000 muchachos que cursan el último año de la escuela secundaria y juegan en su equipo de basquetbol, 64 formarían un equipo profesional. En otras palabras, las posibilidades de que un jugador de basquetbol del último año de la escuela secundaria integre dicho equipo son de 1 en 2 344. De acuerdo con el mismo estudio:

1. Las posibilidades de que un jugador de basquetbol del último año de la escuela secundaria juegue en

(*continúa*)

$$P(B|A_1) = 0.90$$

Suponga la probabilidad de que la prueba indique la presencia de la enfermedad en una persona que en realidad no la padece es de 0.15.

$$P(B|A_2) = 0.15$$

Elija al azar a una persona de Umen y aplique la prueba. Los resultados indican que la enfermedad está presente. ¿Cuál es la probabilidad de que la persona en realidad la padezca? Lo que desea saber, en forma simbólica, es $P(A_1|B)$, que se interpreta de la siguiente manera: *P*(padece la enfermedad | la prueba resulta positiva). La probabilidad $P(A_1|B)$ recibe el nombre de **probabilidad *a posteriori***.

PROBABILIDAD *A POSTERIORI* La que se revisa a partir de información adicional.

Con la ayuda del teorema de Bayes (fórmula [5.7]) es posible determinar la probabilidad a posteriori:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$

$$= \frac{(0.05)(0.90)}{(0.05)(0.90) + (0.95)(0.15)} = \frac{0.0450}{0.1875} = 0.24$$

Así que la probabilidad de que una persona padezca la enfermedad, dado que la prueba fue positiva, es de 0.24. ¿Cómo se interpreta el resultado? Si selecciona al azar a una persona de la población, la probabilidad de que esté enferma es de 0.05. Si se le somete a la prueba y resulta positiva, la probabilidad de que la persona padezca realmente la enfermedad se incrementa cinco veces, de 0.05 a 0.24.

En el problema anterior solo había dos eventos mutuamente excluyentes y colectivamente exhaustivos: A_1 y A_2. Si hay n eventos $A_1, A_2, \ldots, A_n$, el teorema de Bayes (fórmula [5.7]) se transforma en

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

Con la notación anterior, los cálculos del problema de Umen se resumen en la siguiente tabla:

Evento, A_i	Probabilidad *a priori*, $P(A_i)$	Probabilidad condicional, $P(B \mid A_i)$	Probabilidad conjunta, $P(A_i$ y $B)$	Probabilidad *a posteriori*, $P(A_i \mid B)$
Padece la enfermedad, A_1	0.05	0.90	0.0450	0.0450/.1875 = 0.24
No padece la enfermedad, A_2	0.95	0.15	0.1425	0.1425/.1875 = 0.76
			$P(B)$ = 0.1875	1.00

He aquí otro ejemplo del teorema de Bayes.

EJEMPLO

Un fabricante de teléfonos celulares compra un microchip en particular, denominado *LS-24*, a tres proveedores: Hall Electronics, Schuller Sales y Crawford Components. Del total de piezas, 30% lo adquiere de Hall Electronics; 20%, de Schuller Sales y el restante 50%, de Crawford Components. El fabricante cuenta con amplios historiales sobre los tres proveedores y conoce los porcentajes de defectos de los dispositivos de cada proveedor: 3%, en el caso de Hall Electronics; 5% en el de Schuller Sales y 4% en el de Crawford Components.

Cuando el fabricante recibe el material, lo lleva directamente a un depósito y no lo inspecciona ni lo identifica con el nombre del proveedor. Un trabajador selecciona un microchip para instalarlo y lo encuentra defectuoso. ¿Cuál es la probabilidad de que lo haya fabricado Schuller Sales?

(*continuación*)

alguna universidad son de casi 1 en 40.

2. Las posibilidades de que el mismo chico juegue basquetbol universitario como estudiante del último año de la universidad son de 1 en 60.
3. Si usted juega basquetbol como estudiante del último año de la universidad, sus posibilidades de formar parte de un equipo profesional son de casi 1 en 37.5.

SOLUCIÓN

Como primer paso, resuma parte de la información incluida en el enunciado del problema.

- Hay tres eventos mutuamente excluyentes y colectivamente exhaustivos, es decir, tres proveedores:

  A_1: el LS-24 se le compró a Hall Electronics;
  A_2: el LS-24 se le compró a Schuller Sales;
  A_3: el LS-24 se le compró a Crawford Components.

- Las probabilidades *a priori* son:

  $P(A_1) = 0.30$ La probabilidad de que Hall Electronics haya fabricado microchip.
  $P(A_2) = 0.20$ La probabilidad de que haya sido Schuller Sales.
  $P(A_3) = 0.50$ La probabilidad de que haya sido Crawford Components.

- La información adicional es la siguiente:

  B_1: el LS-24 tiene defectos; o
  B_2: no los tiene.

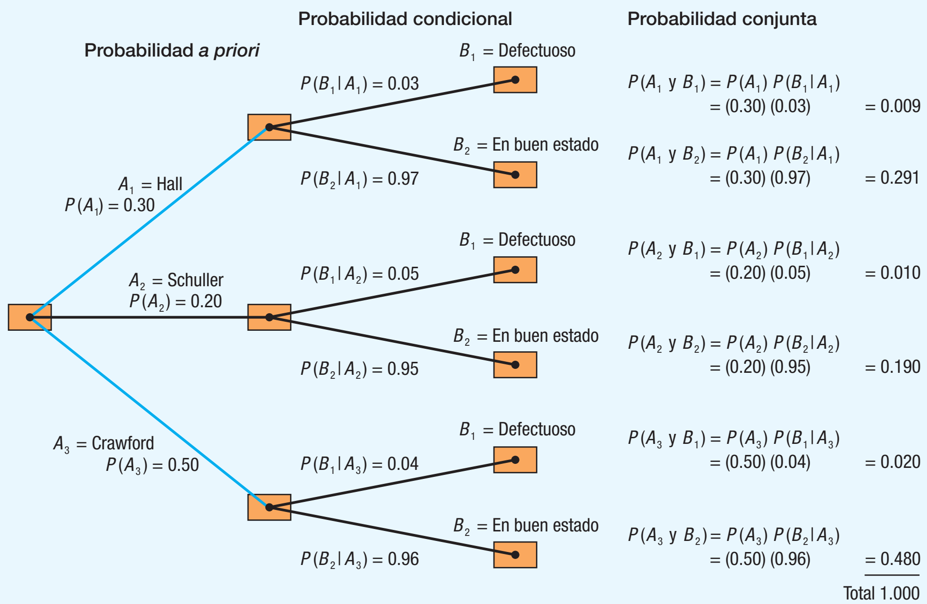
- Se dan las siguientes probabilidades condicionales.

  $P(B_1|A_1) = 0.03$ La probabilidad de que un microchip LS-24 fabricado por Hall Electronics esté defectuoso.
  $P(B_1|A_2) = 0.05$ La probabilidad de que un microchip LS-24 fabricado por Schuller Sales esté defectuoso.
  $P(B_1|A_3) = 0.04$ La probabilidad de que un microchip LS-24 fabricado por Crawford Components esté defectuoso.

- Se selecciona uno de los dispositivos del depósito. Como el fabricante no los identificó, no está seguro de cuál proveedor los fabricó. Desea determinar la probabilidad de que Schuller Sales haya sido. La probabilidad se expresa como $P(A_2|B_1)$.

Observe el registro de calidad de Schuller. Es el peor de los tres proveedores. Ahora que ha encontrado una pieza defectuosa, sospecha que $P(A_2|B_1)$ es mayor que $P(A_2)$. Es decir, la probabilidad revisada es mayor que 0.20. Pero, ¿cuán mayor? El teorema de Bayes ofrece la respuesta. Como primer paso considere el diagrama de árbol de la gráfica 5.3.

Los eventos son dependientes, así que la probabilidad *a priori* de la primera rama se multiplica por la probabilidad condicional de la segunda para obtener la probabilidad conjunta; la cual se observa en la última columna de la gráfica 5.3. Para construir el diagrama de árbol de la gráfica 5.3, se empleó una sucesión de etapas que iban del proveedor hasta determinar si el microchip tenía defectos o no.



GRÁFICA 5.3 Diagrama de árbol del problema de la fabricación de un teléfono celular

Lo que necesita hacer es invertir el proceso. Esto es, en lugar de desplazarse de izquierda a derecha en la gráfica 5.3, necesita hacerlo de derecha a izquierda. Tiene un microchip defectuoso, y quiere determinar la probabilidad de que se le haya comprado a Schuller Sales. ¿Cómo se consigue este objetivo? Primero considere las probabilidades conjuntas como frecuencias relativas de entre 1 000 casos. Por ejemplo, la posibilidad de que Hall Electronics haya fabricado un dispositivo defectuoso es de 0.009. Así que de 1 000 casos es de esperar 9 piezas defectuosas de este fabricante. Observe que en 39 de 1 000 casos se encontrará un microchip defectuoso, lo cual se calcula sumando 9 + 10 + 20; de este total, Schuller and Sales fabricó 10. Por consiguiente, la probabilidad de que el defectuoso se le haya comprado a este proveedor es de 10/39 = 0.2564. Ha determinado la probabilidad revisada de $P(A_2|B_1)$. Antes de encontrar el dispositivo defectuoso, la probabilidad de habérselo comprado a Schuller Sales era de 0.20. Esta posibilidad se ha incrementado a 0.2564.

Esta información se resume en la siguiente tabla:

Evento, A_i	**Probabilidad *a priori*, $P(A_i)$**	**Probabilidad condicional, $P(B_1\|A_i)$**	**Probabilidad conjunta, $P(A_i \text{ y } B_1)$**	**Probabilidad *a posteriori*, $P(A_i\|B_1)$**
Hall	0.30	0.03	0.009	0.009/0.039 = 0.2308
Schuller	0.20	0.05	0.010	0.010/0.039 = 0.2564
Crawford	0.50	0.04	0.020	0.020/0.039 = 0.5128
			$P(B_1)$ = 0.039	1.0000

La probabilidad de que el microchip LS-24 defectuoso provenga de Schuller Sales puede determinarse formalmente mediante el teorema de Bayes. Calcule $P(A_2|B_1)$, en la que A_2 se refiere a este proveedor y B_1, al hecho de que el dispositivo estaba defectuoso:

$$P(A_2|B_1) = \frac{P(A_2)P(B_1|A_2)}{P(A_1)P(B_1|A_1) + P(A_2)P(B_1|A_2) + P(A_3)P(B_1|A_3)}$$

$$= \frac{(0.20)(0.05)}{(0.30)(0.03) + (0.20)(0.05) + (0.50)(0.04)} = \frac{0.010}{0.039} = 0.2564$$

Es el mismo resultado que se obtuvo en la gráfica 5.3 y en la tabla de probabilidad condicional.

AUTOEVALUACIÓN 5-9

Considere el ejemplo anterior junto con la solución.

(a) Diseñe una fórmula para determinar la probabilidad de que la pieza seleccionada provenga de Crawford Components, dado que se trataba de un microchip en buenas condiciones.

(b) Calcule la probabilidad con el teorema de Bayes.

EJERCICIOS

33. $P(A_1) = 0.60$, $P(A_2) = 0.40$, $P(B_1|A_2) = 0.05$, y $P(B_1|A_2) = 0.10$. Aplique el teorema de Bayes para determinar $P(A_1|B_1)$.

34. $P(A_1) = 0.20$, $P(A_2) = 0.40$, $P(A_3) = 0.40$, $P(B_1|A_1) = 0.25$, $P(B_1|A_2) = 0.05$ y $P(B_1|A_3) = 0.10$. Aplique el teorema de Bayes para determinar $P(A_3|B_1)$.

35. El equipo de béisbol los Gatos Salvajes de Ludlow, un equipo de las Ligas Menores perteneciente a los Indios de Cleveland, juega 70% de sus partidos por la noche y 30%, de día. El equipo gana 50% de los juegos nocturnos y 90% de los diurnos. De acuerdo con el periódico de hoy, ganaron ayer. ¿Cuál es la probabilidad de que el partido se haya jugado de noche?

36. La doctora Stallter ha enseñado estadística básica durante varios años. Ella sabe que 80% de los estudiantes terminará los problemas asignados. También, que entre quienes hacen sus tareas, 90% pasará el curso. Entre los que no hacen su tarea, 60% pasará el curso. Mike Fishbaugh cursó estadística el semestre previo con la doctora Stallter y pasó. ¿Cuál es la probabilidad de que haya terminado sus tareas?

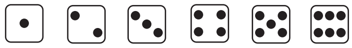
37. El departamento de crédito de Lion's Department Store en Anaheim, California, informó que 30% de las ventas se paga con efectivo; 30%, con tarjeta de crédito y 40%, con tarjeta de débito. Veinte por ciento de las compras con efectivo, 90% de las compras con tarjeta de crédito y 60% de las compras con tarjeta de débito son por más de 50 dólares. La señora Tina Stevens acaba de comprar un vestido nuevo que le costó 120 dólares. ¿Cuál es la probabilidad de que haya pagado en efectivo?

38. Una cuarta parte de los residentes de Burning Ridge Estates deja las puertas de sus cocheras abiertas cuando sale de su hogar. El jefe de la policía de la localidad calcula que a 5% de las cocheras les robarán algo, pero solo a 1% de las cocheras con puertas cerradas les robarán algo. Si roban una cochera, ¿cuál es la probabilidad de que las puestas hayan estado abiertas?

OA5-7
Determinar el número de resultados por medio del principio apropiado de conteo.

Principios de conteo

Si la cantidad de posibles resultados de un experimento es pequeña, resulta relativamente fácil contarlos. Por ejemplo, existen seis posibles resultados del lanzamiento de un dado, a saber:



Sin embargo, si hay un número muy grande de resultados, tal como el número de caras y cruces en un experimento con 10 lanzamientos de una moneda, sería tedioso contar todas las posibilidades. Todos podrían ser caras, una cruz y nueve caras, dos caras y ocho cruces, y así sucesivamente. Para facilitar la cuenta se analizarán tres fórmulas para contar: la **fórmula de la multiplicación** (no confundir con la *regla* de la multiplicación descrita en el capítulo), la **fórmula de las permutaciones** y la **fórmula de de las combinaciones.**

Fórmula de la multiplicación

Primero, la fórmula de la multiplicación.

FÓRMULA DE LA MULTIPLICACIÓN Si hay m formas de hacer una cosa y n formas de hacer otra, hay $m \times n$ formas de hacer ambas.

En términos de la fórmula:

FÓRMULA DE LA MULTIPLICACIÓN Número total de disposiciones $= (m)(n)$ **[5.8]**

Esta fórmula se puede extender a más de dos eventos. En el caso de tres eventos m, n y o:

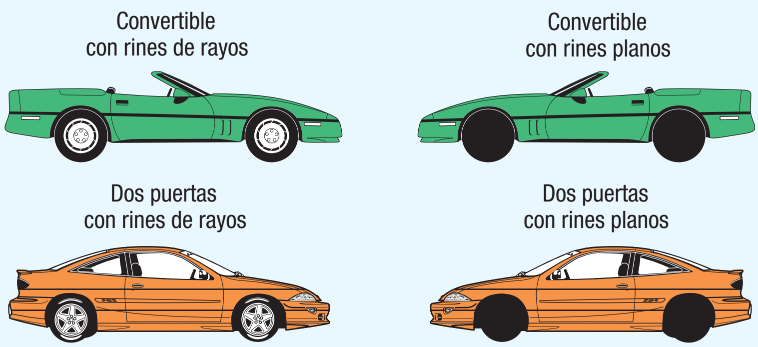
$$\text{Número total de disposiciones} = (m)(n)(o)$$

EJEMPLO

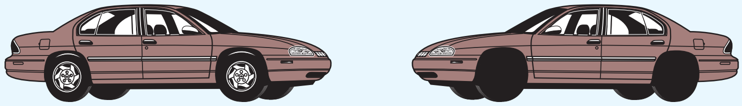
Un distribuidor de automóviles quiere anunciar que por 29 999 dólares usted puede comprar un vehículo convertible, un sedán de dos puertas o un modelo de cuatro puertas, y elegir entre rines de rayos o planos. Basándose en el número de modelos con rines planos, ¿cuántos tipos distintos de vehículos se pueden ofrecer?

SOLUCIÓN

Por supuesto, el distribuidor podría determinar el número total de disposiciones haciendo un diagrama y contando. Hay seis.



Cuatro puertas con rines de rayos | Cuatro puertas con rines planos



Mediante la fórmula de la multiplicación se verifica el resultado (en cuyo caso m es el número de modelos y n, el tipo de rin). De acuerdo con la fórmula [5.8]:

$$\text{Número total de posibles disposiciones} = (m)(n) = (3)(2) = 6$$

No resultó difícil contar todas las posibles combinaciones de modelos y rines en este ejemplo. Sin embargo, suponga que el distribuidor decide ofrecer ocho modelos y seis tipos de rines. Resultaría tedioso representar y contar todas las posibles alternativas. Más bien, se puede aplicar la fórmula de la multiplicación. En este caso, hay $(m)(n) = (8)(6) = 48$ posibles disposiciones.

Observe que en las aplicaciones anteriores de la fórmula de la multiplicación había *dos o más agrupamientos de los cuales usted hizo selecciones.* El distribuidor, por ejemplo, ofreció una variedad de modelos y rines para elegir. Si un constructor de casas ofrece cuatro diferentes estilos de exteriores y tres modelos de interiores, se aplicaría la fórmula de la multiplicación para determinar cuántas combinaciones existen. Hay 12 posibilidades.

AUTOEVALUACIÓN 5-10

1. Women's Shopping Network vende suéteres y pantalones para dama en su canal de televisión por cable. La mercancía se ofrece en colores coordinados. Si los suéteres se encuentran disponibles en cinco colores y los pantalones en cuatro, ¿cuántos diferentes conjuntos se pueden anunciar?
2. Pioneer fabrica tres modelos de receptores estereofónicos, dos de reproductores MP3, cuatro de bocinas y tres de carruseles de CD. Cuando se venden juntos, los cuatro tipos de componentes forman un *sistema.* ¿Cuántos diferentes sistemas puede ofrecer la empresa de electrónica?

Fórmula de las permutaciones

La fórmula de la multiplicación se aplica para determinar el número de posibles disposiciones de dos o más grupos. En contraste, la **fórmula de las permutaciones** se aplica para determinar el número posible de disposiciones cuando solo hay un grupo de objetos. He aquí algunos ejemplos de esta clase de problemas.

- Tres piezas electrónicas (un transistor, un LED y un sintetizador) se van a montar en una unidad que se conecta a un aparato de televisión. Estas se pueden montar en cualquier orden. La pregunta es: ¿de cuántas formas pueden montarse?
- Un operador de máquinas debe llevar a cabo cuatro verificaciones de seguridad antes de arrancar su máquina. No importa el orden en que realice las verificaciones. ¿De cuántas formas puede hacerlas?

Un orden para el primer ejemplo sería: primero el transistor, en seguida el LED y en tercer lugar el sintetizador. A esta distribución se le conoce como **permutación.**

PERMUTACIÓN Cualquier distribución de r objetos seleccionados de un solo grupo de n posibles objetos.

Observe que las distribuciones *a b c* y *b a c* son permutaciones diferentes. La fórmula para contar el número total de estas es:

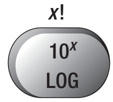
FÓRMULA DE LAS PERMUTACIONES $${}_nP_r = \frac{n!}{(n-r)!} \qquad \textbf{[5.9]}$$

donde:

n representa el total de objetos;
r representa el total de objetos seleccionados.

Antes de resolver los dos problemas planteados, considere que en las permutaciones y las combinaciones (que se plantean en breve) se emplea la notación denominada *n factorial.* Esta se representa como $n!$ y significa el producto de $n(n - 1)(n - 2)(n - 3) \cdots (1)$. Por ejemplo, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$.

Muchas calculadoras tienen la tecla $x!$, que ejecuta el cálculo y le ahorra mucho tiempo. Por ejemplo, la calculadora Texas Instrument TI-36X tiene la siguiente tecla:



Es la “tercera función”, así que revise el manual de usuario o internet para leer las instrucciones.

La notación factorial se puede eliminar cuando los mismos números aparecen tanto en el numerador como en el denominador, como se muestra a continuación:

$$\frac{6!3!}{4!} = \frac{6 \cdot 5 \cdot \cancel{4} \cdot \cancel{3} \cdot \cancel{2} \cdot \cancel{1}(3 \cdot 2 \cdot 1)}{\cancel{4} \cdot \cancel{3} \cdot 2 \cdot \cancel{1}} = 180$$

Por definición, cero factorial, que se escribe $0!$, es igual a uno. Es decir, $0! = 1$.

EJEMPLO

Respecto del grupo de tres piezas electrónicas que se van a montar en cualquier orden, ¿de cuántas formas puede hacerse?

SOLUCIÓN

Hay tres piezas electrónicas que van a montarse, así que $n = 3$. Como las tres se van a insertar en la unidad conectable, $r = 3$. De acuerdo con la fórmula [5.9], el resultado es:

$${}_nP_r = \frac{n!}{(n - r)!} = \frac{3!}{(3 - 3)!} = \frac{3!}{0!} = \frac{3!}{1} = 6$$

Es posible verificar el número de permutaciones obtenidas con la fórmula de las permutaciones. Determine cuántos *espacios* hay que llenar y las posibilidades para cada *espacio.* En el problema de las tres piezas electrónicas hay un sitio en la unidad conectable para cada pieza. Hay tres posibilidades para el primer lugar, dos para el segundo (una se ha agotado) y una para el tercero:

$$(3)(2)(1) = 6 \text{ permutaciones.}$$

Las seis formas en que las tres piezas electrónicas, representadas con las letras *A, B, C,* se pueden ordenar, son:

ABC	*BAC*	*CAB*	*ACB*	*BCA*	*CBA*

En el ejemplo anterior, se seleccionan y distribuyen todos los objetos, es decir que $n = r$. En muchos casos, solo se seleccionan algunos objetos y se ordenan tomándolos de entre los n posibles. En el siguiente ejemplo se explican los detalles de este caso.

EJEMPLO

Betts Machine Shop, Inc., cuenta con ocho tornos, aunque solo hay tres espacios disponibles en el área de producción para las máquinas. ¿De cuántas maneras se pueden distribuir las ocho máquinas en los tres espacios disponibles?

SOLUCIÓN

Hay ocho posibilidades para el primer espacio disponible en el área de producción, siete para el segundo (una se ha agotado) y seis para el tercero. Por consiguiente:

$$(8)(7)(6) = 336$$

es decir, hay un total de 336 diferentes distribuciones posibles. Este resultado también podría obtenerse aplicando la fórmula [5.9]. Si $n = 8$ máquinas y $r = 3$ espacios disponibles, la fórmula da como resultado

$$_nP_r = \frac{n!}{(n-r)!} = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{(8)(7)(6)\cancel{5!}}{\cancel{5!}} = 336$$

Fórmula de las combinaciones

Si el orden de los objetos seleccionados *no* es importante, cualquier selección se denomina **combinación.** La fórmula para contar el número de r combinaciones de objetos de un conjunto de n objetos es:

FÓRMULA DE LAS COMBINACIONES $$_nC_r = \frac{n!}{r!(n-r)!}$$ **[5.10]**

Por ejemplo, si los ejecutivos Able, Baker y Chauncy van a ser elegidos para formar un comité de negociación de una fusión, solo existe una posible combinación con estos tres; el comité formado por Able, Baker y Chauncy es el mismo comité que el que forman Baker, Chauncy y Able. De acuerdo con la fórmula de las combinaciones:

$$_nC_r = \frac{n!}{r!(n-r)!} = \frac{3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1(1)} = 1$$

EJEMPLO

La sala cinematográfica Grand 16 utiliza equipos de tres empleados para trabajar en la dulcería cada noche. Hay siete empleados disponibles. ¿Cuántos equipos diferentes pueden programarse para cubrir el turno?

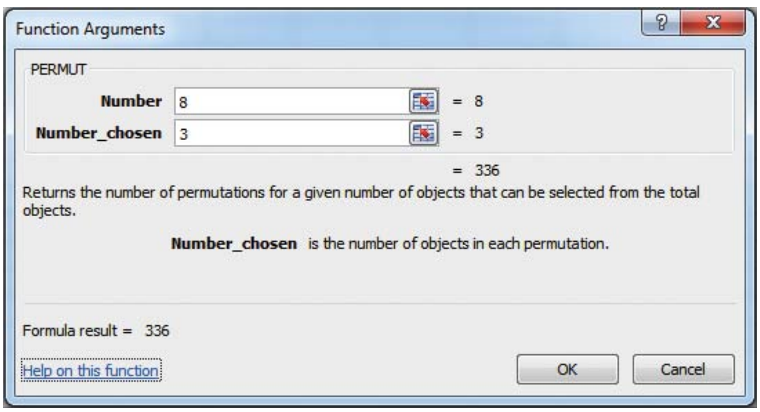
SOLUCIÓN

De acuerdo con la fórmula [5.10], hay 35 combinaciones, determinadas por

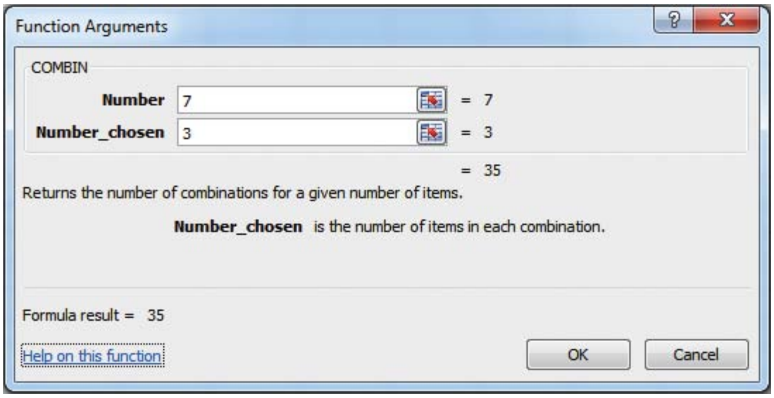
$$_7C_3 = \frac{n!}{r!(n-r)!} = \frac{7!}{3!(7-3)!} = \frac{7!}{3!4!} = 35$$

Los siete empleados, en grupos de tres, crearían la posibilidad de 35 equipos diferentes.

Cuando el número de permutaciones o combinaciones es grande, los cálculos son laboriosos; pero el software de las computadoras y las calculadoras de mano tienen *funciones* para calcularlo. A continuación aparece una salida de Excel que contiene la ubicación de los ocho tornos en el área de producción de Betts Machine Shop, Inc. Hay un total de 336 distribuciones.



Enseguida aparece la salida del número de equipos en la sala cinematográfica Grand 16. Se eligen tres empleados de entre los siete disponibles.



AUTOEVALUACIÓN 5-11

1. Un músico piensa escribir una escala basada solo en cinco cuerdas: B bemol, C, D, E y G. Sin embargo, solo tres de las cinco cuerdas se van a utilizar en sucesión, por ejemplo: C, B bemol y E. No se permiten repeticiones como B bemol, B bemol y E.
   (a) ¿Cuántas permutaciones de las cinco cuerdas, tomadas de tres en tres, son posibles?
   (b) De acuerdo con la fórmula [5.9], ¿cuántas permutaciones son posibles?
2. Los números dígitos se van a emplear en grupos de códigos de cuatro para identificar una prenda. El código 1083 podría identificar una blusa azul, talla mediana; el grupo de código 2031 podría identificar unos pantalones talla 18, etcétera. No están permitidas las repeticiones de números; es decir, el mismo número no se puede utilizar dos veces (o más) en una sucesión completa. Por ejemplo, 2256, 2562 o 5559 no estarían permitidos. ¿Cuántos diferentes grupos de códigos se pueden asignar?
3. En el ejemplo relacionado con la sala cinematográfica Grand 16 hubo 35 posibles equipos de tres a partir de siete empleados.
   (a) Aplique la fórmula [5.10] para demostrar que esto es verdadero.
   (b) El gerente de la sala desea planificar sus equipos para la dulcería con equipos de cinco empleados durante los fines de semana para atender a más clientes. A partir de siete empleados, ¿cuántos equipos de cinco son posibles?
4. En un juego de lotería se seleccionan al azar tres números de una tómbola de bolas numeradas del 1 al 50.
   (a) ¿Cuántas permutaciones son posibles?
   (b) ¿Cuántas combinaciones son posibles?

EJERCICIOS

39. Resuelva las siguientes operaciones:
- **a.** 40!/35!
- **b.** ${}_7P_4$
- **c.** ${}_5C_2$

40. Resuelva las siguientes operaciones:
- **a.** 20!/17!
- **b.** ${}_9P_3$
- **c.** ${}_7C_2$

41. Un encuestador seleccionó en forma aleatoria a cuatro de 10 personas disponibles. ¿Cuántos diferentes grupos son posibles?

42. Un número telefónico consta de siete dígitos, los primeros tres representan el enlace. ¿Cuántos números telefónicos son posibles con el enlace 537?

43. Una compañía de entregas rápidas debe incluir cinco ciudades en su ruta. ¿Cuántas diferentes rutas se pueden formar suponiendo que no importa el orden en que se incluyen las ciudades?

44. Una representante de la Environmental Protection Agency (EPA) piensa seleccionar muestras de 10 terrenos. El director tiene 15 terrenos de los cuales la representante puede recoger las muestras. ¿Cuántas diferentes muestras son posibles?

45. Un encuestador nacional formula 15 preguntas diseñadas para medir el desempeño del presidente de Estados Unidos. El encuestador seleccionará 10 de las preguntas. ¿Cuántas distribuciones de estas preguntas se pueden formar tomando en cuenta el orden?

46. Una compañía va a crear tres nuevas divisiones. Para dirigir cada una de ellas hay siete gerentes elegibles. ¿De cuántas formas se podrían elegir a los tres nuevos directores? *Sugerencia:* asuma que la asignación de la división sí hace diferencia.

RESUMEN DEL CAPÍTULO

I. Una probabilidad es un valor entre 0 y 1, inclusive, que representa las posibilidades de que cierto evento ocurra.
 A. Un experimento es la observación de alguna actividad o el acto de tomar una medida.
 B. Un resultado es una consecuencia particular de un experimento.
 C. Un evento es la colección de uno o más resultados de un experimento.

II. Existen tres definiciones de probabilidad.
 A. La definición clásica se aplica cuando un experimento generará *n* resultados igualmente posibles.
 B. La definición empírica se emplea cuando el número de veces que ocurre un evento se divide entre el número de observaciones.
 C. Una probabilidad subjetiva se basa en cualquier información disponible.

III. Dos eventos son mutuamente excluyentes si, como consecuencia de que uno de los dos sucede, el otro no puede ocurrir.

IV. Los eventos son independientes si el hecho de que un evento suceda no influye en la probabilidad de que el otro ocurra.

V. Las reglas de la adición se refieren a la probabilidad de que cualquiera de dos o más eventos puedan ocurrir.
 A. La regla especial de la adición se aplica cuando los eventos son mutuamente excluyentes.

$$P(A \text{ o } B) = P(A) + P(B) \qquad \textbf{[5.2]}$$

 B. La regla general de la adición se aplica cuando los eventos no son mutuamente excluyentes.

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) \qquad \textbf{[5.4]}$$

 C. La regla del complemento se utiliza para determinar la probabilidad de un evento restando de 1 la probabilidad de que el evento no suceda.

$$P(A) = 1 - P(\sim A) \qquad \textbf{[5.3]}$$

VI. Las reglas de la multiplicación se aplican cuando dos o más eventos ocurren simultáneamente.
 A. La regla especial de la multiplicación se refiere a eventos que son independientes.

$$P(A \text{ y } B) = P(A)P(B) \qquad \textbf{[5.5]}$$

 B. La regla general de la multiplicación se aplica en eventos que no son independientes.

$$P(A \text{ y } B) = P(A)P(B|A) \qquad \textbf{[5.6]}$$

 C. Una probabilidad conjunta es la de que dos o más eventos sucedan al mismo tiempo.
 D. Una probabilidad condicional es la de que un evento suceda, dado que otro evento ha sucedido.
 E. El teorema de Bayes es un método que consiste en revisar una probabilidad, dado que se tiene información adicional. En el caso de dos eventos mutuamente excluyentes y colectivamente exhaustivos,

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \qquad \textbf{[5.7]}$$

VII. Existen tres reglas de conteo útiles para determinar la cantidad de resultados de un experimento.
 A. La regla de la multiplicación establece que si hay *m* formas de que un evento suceda y *n* formas de que otro pueda suceder, entonces hay *mn* formas en que ambos eventos pueden suceder.

$$\text{Número de disposiciones} = (m)(n) \qquad \textbf{[5.8]}$$

 B. Una permutación es un arreglo en el que el orden de los objetos seleccionados de un conjunto específico es importante.

$${}_nP_r = \frac{n!}{(n-r)!} \qquad \textbf{[5.9]}$$

 C. Una combinación es un arreglo en el que el orden de los objetos seleccionados de un conjunto específico no es importante.

$${}_nC_r = \frac{n!}{r!(n-r)!} \qquad \textbf{[5.10]}$$

ESTADÍSTICA EN ACCIÓN

Las estadísticas gubernamentales muestran que hay alrededor de 1.7 muertes provocadas por accidentes automovilísticos por cada 100 millones de millas recorridas. Si usted maneja una milla a la tienda para comprar un billete de lotería y enseguida regresa a casa, habrá recorrido dos millas. Por consiguiente, la probabilidad de que usted se una a este grupo de estadísticas en sus siguientes dos millas de viaje redondo es de 2 × 1.7/100 000 000 = 0.000 000 037. Esto también se expresa como "una en 29 411 765". Por tanto, si usted maneja a la tienda a comprar su boleto, la probabilidad de morir (o matar a alguien) es más de cuatro veces la de sacarse la lotería: una en 120 526 770. **http://www.durangobill.com/PowerballOdds.htlm**

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
$P(A)$	Probabilidad de *A*	*P de A*
$P(\sim A)$	Probabilidad de no *A*	*P de no A*
$P(A \text{ y } B)$	Probabilidad de *A* y *B*	*P de A y B*
$P(A \text{ o } B)$	Probabilidad de *A* o *B*	*P de A o B*
$P(A\|B)$	Probabilidad de *A* dado que *B* ocurrió	*P de A, dado B*
${}_nP_r$	Permutación de *n* elementos seleccionados *r* a la vez	*Pnr*
${}_nC_r$	Combinación de *n* elementos seleccionados *r* a la vez	*Cnr*

EJERCICIOS DEL CAPÍTULO

47. El departamento de investigación de mercados de Pepsico planea realizar una encuesta entre adolescentes sobre un refresco recién creado. A cada uno de ellos se le va a pedir compararlo con su refresco favorito.
a. ¿En qué consiste el experimento?
b. ¿Cuál es un posible evento?

48. El número de veces que ocurrió un evento en el pasado se divide entre la cantidad de veces que ocurre. ¿Cómo se llama este enfoque de la probabilidad?

49. La probabilidad de que la causa y la cura de todo tipo de cáncer se descubran antes del año 2020 es de 0.20. ¿Qué enfoque de la probabilidad ilustra este enunciado?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

50. Berdine's Chicken Factory posee varias tiendas en el área del Hilton Head, Carolina del Sur. El propietario va a entrevistar a los candidatos para el puesto de mesero, y le gustaría tener información referente a la propina que estos esperan ganar por cuenta (o nota). Un estudio de 500 cuentas recientes indicó que el mesero ganaba las siguientes propinas por turno de 8 horas.

Propina	Número
\$0 hasta \$ 20	200
20 hasta 50	100
50 hasta 100	75
100 hasta 200	75
200 o más	50
Total	500

a. ¿Cuál es la probabilidad de recibir una propina de 200 dólares o más?
b. ¿Se consideran mutuamente excluyentes las categorías "\$0 hasta \$20", "\$20 hasta \$50", etcétera?
c. Si las probabilidades relacionadas con cada resultado se sumaran, ¿cuál sería el total?
d. ¿Cuál es la probabilidad de que una propina sea hasta de 50 dólares?
e. ¿Cuál es la probabilidad de que una propina sea inferior a 200 dólares?

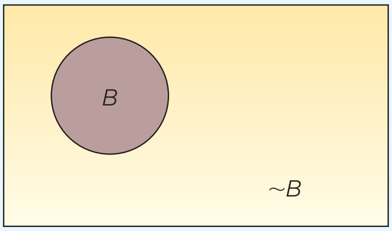
51. Ganar en todas las carreras "Triple Corona" se considera la mayor hazaña de un caballo de carreras de pedigrí. Después de un exitoso Derby de Kentucky, Corn on the Cob es un gran favorito (dos a uno) para ganar las apuestas de Preakness.
a. Si Corn on the Cob también es favorito dos a uno para ganar las apuestas de Belmont, ¿cuál es la probabilidad de que gane la Triple Corona?
b. ¿Cuáles tendrían que ser sus oportunidades para las apuestas de Preakness para que sea una "apuesta segura" para ganar la Triple Corona?

52. La primera carta de una baraja de 52 naipes es un rey.
a. Si lo regresa a la baraja, ¿cuál es la probabilidad de sacar un rey en la segunda selección?
b. Si no lo regresa a la baraja, ¿cuál es la probabilidad de sacar un rey en la segunda selección?
c. ¿Cuál es la probabilidad de seleccionar un par de reyes en los dos primeros intentos (suponiendo que el primer rey no fue reemplazado?

53. Armco, un fabricante de sistemas de semáforos, descubrió que, en las pruebas de vida acelerada, 95% de los sistemas recién desarrollados duraban tres años antes de descomponerse al cambiar de señal.
a. Si una ciudad compra cuatro de estos sistemas, ¿cuál es la probabilidad de que los cuatro sistemas funcionen adecuadamente durante tres años por lo menos?

b. ¿Qué regla de la probabilidad se ejemplifica en este caso?
c. Representando los cuatro sistemas con letras, escriba una ecuación para demostrar cómo llegó a la respuesta del punto *a.*

54. Observe el siguiente dibujo.



a. ¿Qué nombre recibe el dibujo?
b. ¿Qué regla de la probabilidad se ilustra?
c. *B* representa el evento que se refiere a la selección de una familia que recibe prestaciones sociales. ¿A qué es igual $P(B) + P(\sim B)$?

55. En un programa de empleados que realizan prácticas de gerencia en Claremont Enterprises, 80% de ellos son mujeres y 20%, hombres. De las mujeres, 90% fue a la universidad, así como 78% de los hombres.
a. Se elige al azar a uno de estos empleados. ¿Cuál es la probabilidad de que sea una mujer que no asistió a la universidad?
b. ¿El género y la asistencia a la universidad son independientes? ¿Por qué?
c. Construya un diagrama de árbol que muestre las probabilidades condicionales y probabilidades conjuntas.
d. ¿Las probabilidades conjuntas suman 1.00? ¿Por qué?

56. Suponga que la probabilidad de que cualquier vuelo de Delta Airlines llegue 15 minutos después de la hora programada es de 0.90. Se selecciona al azar un vuelo de Delta en cuatro días diferentes.
a. ¿Cuál es la probabilidad de que los cuatro vuelos seleccionados lleguen tarde?
b. ¿Cuál es la probabilidad de que ninguno de los vuelos seleccionados llegue tarde?
c. ¿Cuál es la probabilidad de que por lo menos uno de los vuelos seleccionados no llegue tarde?

57. Kiddie Carts International tiene 100 empleados. De estos, 57 son trabajadores por hora, 40 son supervisores, 2 son secretarias y el otro empleado es el presidente. Suponga que selecciona un empleado.
a. ¿Cuál es la probabilidad de que el empleado seleccionado sea un trabajador por hora?
b. ¿Cuál es la probabilidad de que el empleado seleccionado sea un trabajador por hora o un supervisor?
c. Respecto del punto *b,* ¿estos eventos son mutuamente excluyentes?
d. ¿Cuál es la probabilidad de que el empleado seleccionado no sea trabajador por hora ni supervisor?

58. Buster Posey, de los Gigantes de San Francisco, tuvo el promedio de bateo más alto en la temporada 2012 de las Ligas Mayores de Béisbol. Su promedio fue de 0.336. Suponga que la probabilidad de conectar un *hit* es de 0.336 en cada turno al bate, y que durante un partido batea tres veces.
a. ¿Qué tipo de probabilidad constituye este ejemplo?
b. ¿Cuál es la probabilidad de conectar tres hits en un juego?
c. ¿Cuál es la probabilidad de que no conecte ningún hit en un juego?
d. ¿Cuál es la probabilidad de conectar por lo menos un hit?

59. Quedan cuatro equipos femeninos deportivos en una competencia de eliminatorias en un torneo de basquetbol. Un equipo resulta favorecido en el pronóstico del marcador de la semifinal por probabilidades de dos a uno; uno de los equipos de la otra semifinal, por probabilidades de tres a uno. Determina la probabilidad de que:
a. Ambos equipos ganen sus juegos.
b. Ninguno de los equipos gane su juego.
c. Cuando menos uno de los equipos gane su juego.

60. Hay tres claves etiquetadas como "doble diario" en el programa de juegos *Jeopardy*. Participan tres concursantes igualmente aptos. Determine la probabilidad de que:
a. Un solo concursante encuentre los tres "doble diario".
b. El retador se lleve todos los "doble diario".
c. Cada uno de los concursantes elija precisamente un "doble diario".

61. Brooks Insurance, Inc., pretende ofrecer seguros de vida a hombres de 60 años por internet. Las tablas de mortalidad indican que la probabilidad de que un hombre de esa edad sobreviva otro año es de 0.98. Si el seguro se ofrece a cinco hombres de 60 años:

a. ¿Cuál es la probabilidad de que los cinco hombres sobrevivan?
b. ¿Cuál es la probabilidad de que por lo menos uno no sobreviva?

62. De las casas construidas en el área de Quail Creek, 40% tienen sistema de seguridad. Se seleccionan tres casas al azar.
a. ¿Cuál es la probabilidad de que las tres cuenten con sistema de seguridad?
b. ¿Cuál es la probabilidad de que ninguna cuente con sistema de seguridad?
c. ¿Cuál es la probabilidad de que por lo menos una cuente con sistema de seguridad?
d. ¿Los eventos son dependientes o independientes?

63. Repase el ejercicio anterior, pero suponga que hay 10 casas en el área de Quail Creek y cuatro de ellas cuentan con sistema de seguridad. Se eligen tres casas al azar.
a. ¿Cuál es la probabilidad de que las tres cuenten con sistema de seguridad?
b. ¿Cuál es la probabilidad de que ninguna cuente con sistema de seguridad?
c. ¿Cuál es la probabilidad de que por lo menos una cuente con sistema de seguridad?
d. ¿Los eventos son dependientes o independientes?

64. Hay 20 familias viviendo en el Willbrook Farms Development. De ellas, 10 elaboraron sus propias declaraciones de impuestos del año anterior, 7 la encargaron a un profesional de la localidad y las 3 restantes las encargaron a H&R Block.
a. ¿Cuál es la probabilidad de seleccionar a una familia que haya preparado su propia declaración?
b. ¿Cuál es la probabilidad de seleccionar a dos familias que hayan preparado sus propias declaraciones?
c. ¿Cuál es la probabilidad de seleccionar a tres familias que hayan preparado sus propias declaraciones?
d. ¿Cuál es la probabilidad de seleccionar a dos familias que no presentaron sus declaraciones mediante H&R Block?

65. La junta directiva de Saner Automatic Door Company consta de 12 miembros, 3 de los cuales son mujeres. Para redactar un nuevo manual relacionado con la política y procedimientos de la compañía, se elige al azar un comité de 3 miembros de la junta directiva para llevar a cabo la redacción.
a. ¿Cuál es la probabilidad de que todos los miembros del comité sean hombres?
b. ¿Cuál es la probabilidad de que por lo menos un miembro del comité sea mujer?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

66. Una encuesta reciente publicada en *BloombergBusinessWeek* aborda el tema de los salarios de los directores ejecutivos de grandes compañías y si los accionistas ganan o pierden dinero.

	Director ejecutivo con un salario mayor que $1 000 000	Director ejecutivo con un salario menor que $1 000 000	Total
Los accionistas ganaron dinero	2	11	13
Los accionistas perdieron dinero	4	3	7
Total	6	14	20

Se selecciona al azar una compañía de la lista de 20 estudiadas. Determine la probabilidad de que:
a. El director ejecutivo gane más de un millón de dólares.
b. El director ejecutivo gane más de un millón de dólares o los accionistas pierdan dinero.
c. El director ejecutivo gane más de un millón de dólares dado que los accionistas pierden dinero.
d. Se seleccionen dos directores ejecutivos que ganen más de un millón de dólares?

67. Althoff and Roll, una empresa de inversiones de Augusta, Georgia, se anuncia con frecuencia en el *Augusta Morning Gazette,* el periódico de la región. El personal de marketing del *Gazette* calcula que 60% del mercado potencial de Althoff and Roll leyó el periódico y que 85% de quienes lo leyeron recuerdan la publicidad de Althoff and Roll.
a. ¿Qué porcentaje del mercado potencial de la compañía ve y recuerda el anuncio?
b. ¿Qué porcentaje del mercado potencial de la compañía ve el anuncio pero no lo recuerda?

68. Una compañía de internet localizada en Carolina del Sur tiene boletos de temporada para los juegos de basquetbol de Los Angeles Lakers. Su presidente siempre invita a uno de los cuatro vicepresidentes al juego, y afirma que selecciona a la persona al azar. Uno de ellos no ha sido invitado para ir a ninguno de los últimos cinco juegos en casa de los Lakers. ¿Cuál es la probabilidad de que ello se deba al azar?

69. Un proveedor minorista de computadoras compró un lote de 1 000 discos CD-R e intentó formatearlos para una aplicación particular. Había 857 discos compactos en perfectas condiciones, 112 se podían utilizar, aunque tenían sectores en malas condiciones y el resto era inservible.
a. ¿Cuál es la probabilidad de que un CD seleccionado no se encuentre en perfecto estado?
b. Si el disco no se encuentra en perfectas condiciones, ¿cuál es la probabilidad de que no se le pueda utilizar?

70. Un inversionista compró 100 acciones de Fifth Third Bank y 100 de Santee Electric Cooperative. La probabilidad de que las acciones del banco incrementen su valor en un año es de 0.70. La probabilidad de que las utilidades de la compañía eléctrica se incrementen en el mismo periodo es de 0.60.
a. ¿Cuál es la probabilidad de que ambos lotes de acciones aumenten de precio durante el periodo?
b. ¿Cuál es la probabilidad de que las acciones del banco incrementen su precio, aunque las utilidades no lo hagan?
c. ¿Cuál es la probabilidad de que por lo menos uno de los lotes aumente de precio?

71. Flashner Marketing Research, Inc., se especializa en evaluar las posibles tiendas de ropa para dama en centros comerciales. Al Flashner, el presidente, informa que califica las posibles tiendas como buenas, regulares y malas. Los registros de evaluaciones previas muestran que 60% de las veces los candidatos fueron evaluados como buenos; 30% de las veces, como regulares; y 10% de las ocasiones, como malos. De quienes fueron calificados como buenos, 80% hicieron mejoras el primer año; de los que fueron calificados como regulares, 60% lo hicieron; y de los que fueron mal evaluados, 20% mejoraron sus instalaciones el primer año. Connie´s Apparel fue uno de los clientes de Flashner e hizo mejoras el año anterior. ¿Cuál es la probabilidad de que se le haya dado originalmente una mala calificación?

72. Se recibieron de la fábrica dos cajas de camisas para caballero Old Navy. La primera caja contenía 25 camisas polo y 15 camisas Super-T; la segunda, 30 camisas polo y 10 camisas Super-T. Una de las cajas se seleccionó al azar y se extrajo una camisa, también en forma aleatoria, para revisarla. La camisa era polo. Dada esta información, ¿cuál es la probabilidad de que provenga de la primera caja?

73. En la compra de una pizza grande en Tony´s Pizza, el cliente recibe un cupón que puede raspar para ver si tiene premio. La probabilidad de ganar un refresco es 0.10, y la de ganar una pizza grande es 0.02. Usted tiene planes de almorzar mañana en Tony's. Determine la probabilidad de que usted:
a. Gane una pizza grande o un refresco.
b. No gane nada.
c. No gane nada en tres visitas consecutivas.
d. Gane por lo menos algo en sus siguientes tres visitas.

74. Para el juego diario de la lotería en Illinois, los participantes seleccionan tres opciones de entre los números dígitos. No pueden seleccionar un número más de una vez, así que un billete ganador podría ser, por ejemplo, 307, pero no 337. Al comprar un billete se puede seleccionar un conjunto de números. Los ganadores se anuncian en televisión todas las noches.
a. ¿Cuántos son los diferentes resultados (números de tres dígitos) posibles?
b. Si compra un billete, ¿cuál es la probabilidad de ganar?
c. Suponga que compra tres boletos y selecciona un número diferente para cada uno. ¿Cuál es la probabilidad de que no gane con cualquiera de los boletos?

75. Hace varios años, Wendy´s Hamburgers anunció que hay 256 diferentes formas de pedir una hamburguesa. Es posible elegir entre cualquiera de las siguientes combinaciones: mostaza, cátsup, cebolla, pepinillos, tomate, salsa, mayonesa y lechuga. ¿Es correcto el anuncio? Explique la forma en la que llegó a la respuesta.

76. Encuestas recientes revelaron que 60% de los turistas que viajaron a China visitaron la Ciudad Prohibida, el Templo del Cielo, la Gran Muralla y otros sitios históricos en Beijing o cerca de esta ciudad. Cuarenta por ciento visitó Xi'an y con sus magníficos soldados, caballos y carrozas de terracota, que permanecieron enterrados desde hace más de 2 000 años; 30% fue tanto a Beijing como a Xi'an. ¿Cuál es la probabilidad de que un turista haya visitado por lo menos uno de estos lugares?

77. Considere una nueva goma de mascar que ayuda a quienes desean dejar de fumar. Si 60% de la gente que la mastica tiene éxito en dejar de fumar, ¿cuál es la probabilidad de que en un grupo de cuatro fumadores que mascan la goma por lo menos uno deje el cigarro?

78. Reynolds Construction Company está de acuerdo en no construir casas *iguales* en una nueva subdivisión. Se ofrecen cinco diseños de exterior a los posibles compradores. La constructora ha uniformado tres planos de interior que pueden incorporarse a cualquiera de los cinco modelos de exteriores. ¿Cuántos planos de exterior e interior se pueden ofrecer a los posibles compradores?

79. A un nuevo modelo de automóvil deportivo le fallan los frenos 15% del tiempo y 5% tiene un mecanismo de dirección defectuoso. Suponga –y espere- que estos problemas se presenten de manera independiente. Si ocurre uno u otro problema, el automóvil recibe el nombre de *limón.* Si ambos problemas se presentan, el automóvil se denomina *riesgo.* Su profesor compró uno de estos automóviles el día de ayer. Determine la probabilidad de que sea:
a. Un limón.
b. Un riesgo.

80. En el estado de Maryland, las placas tienen tres números seguidos de tres letras. ¿Cuántas diferentes placas son posibles?

81. Hay cuatro candidatos para el cargo de director ejecutivo de Dalton Enterprises. Tres de los solicitantes tiene más de 60 años de edad. Dos son mujeres, de las cuales solo una rebasa esa edad.

a. ¿Cuál es la probabilidad de que un candidato tenga más de 60 años y sea mujer?
b. Si el candidato es hombre, ¿cuál es la probabilidad de que tenga menos de 60 años?
c. Si el individuo tiene más de 60 años, ¿cuál es la probabilidad de que sea mujer?

82. Tim Beckie es propietario de Bleckie Investment and Real Estate Company. La empresa recientemente compró cuatro terrenos en Holly Farms Estates y seis en Newburg Woods. Los terrenos son igual de atractivos y se venden casi al mismo precio.
a. ¿Cuál es la probabilidad de que los siguientes dos terrenos que se vendan se ubiquen en Newburg Woods?
b. ¿Cuál es la probabilidad de que por lo menos uno de los siguientes cuatro que se vendan se ubique en Holly Farms?
c. ¿Estos eventos son independientes o dependientes?

83. La contraseña de una computadora consta de cuatro caracteres; los cuales pueden ser una de las 26 letras del alfabeto. Cada carácter se puede incluir más de una vez. ¿Cuántas diferentes contraseñas puede haber?

84. Una caja con 24 latas contiene una lata contaminada. Tres latas se van a elegir al azar para probarlas.
a. ¿Cuántas diferentes combinaciones de tres latas podrían seleccionarse?
b. ¿Cuál es la probabilidad de que la lata contaminada se seleccione para la prueba?

85. El acertijo de un periódico presenta un problema de comparación. Los nombres de los 10 presidentes de Estados Unidos aparecen en una columna, y los vicepresidentes se colocan en la segunda, en lista aleatoria. En el acertijo se pide al lector que ponga en correspondencia a cada presidente con su vicepresidente. Si usted realiza las correspondencias al azar, ¿cuántas son posibles? ¿Cuál es la probabilidad de que las 10 sean correctas?

86. Dos componentes, *A* y *B*, operan en serie y ambos deben trabajar para que el sistema funcione. Suponga que estos son independientes. ¿Cuál es la probabilidad de que el sistema funcione en estas condiciones? La probabilidad de que *A* funcione es de 0.90, igual que la de *B*.

87. Horwege Electronics, Inc., compra tubos de televisión a cuatro proveedores. Tyson Wholesale proporciona 20% de los tubos; Fuji Importers, 30%; Kirkpatricks, 25%; y Parts, Inc., 25%. Tyson Wholesale normalmente tiene la mejor calidad porque solo 3% de sus tubos llegan defectuosos; en cuanto a Fuji Importers, 4% de sus tubos tienen defectos; 7% de los de Kirkpatricks y 6.5% de los de Parts, Inc., también los tienen.
a. ¿Cuál es el porcentaje total de tubos defectuosos?
b. Se descubrió un tubo de televisión defectuoso en el último envío. ¿Cuál es la probabilidad de que provenga de Tyson Wholesale?

88. ABC Auto Insurance clasifica a cada conductor, según su riesgo, en bueno, medio o malo. Quienes solicitan un seguro caen dentro de estos tres grupos en porcentajes de 30%, 50% y 20%, respectivamente. La probabilidad de que un *buen* conductor tenga un accidente es de 0.01; la probabilidad de un conductor de riesgo *medio* es de 0.03 y la probabilidad de que un *mal* conductor tenga un accidente es de 0.10. La compañía le vende al señor Brophy una póliza de seguro y él tiene un accidente. Determine la probabilidad de que el señor Brophy sea:
a. Un *buen* conductor.
b. Un conductor de riesgo *medio*.
c. Un *mal* conductor.

89. Usted hace un viaje aéreo que involucra tomar tres vuelos independientes. Si hay 80% de probabilidades de que cada etapa específica del viaje se realice a tiempo, ¿cuál es la probabilidad de que los tres vuelos lleguen a tiempo?

90. La probabilidad de que un servidor de red D-Link se caiga es de 0.05. Si usted tiene tres servidores independientes, ¿cuál es la probabilidad de que al menos uno de ellos sea funcional?

91. Samsung fabrica 22% de todas las pantallas de cristal líquido (LCD). ¿Cuál es la probabilidad de que en un conjunto de tres compras independientes de pantallas, cuando menos una sea Samsung?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

92. Consulte los datos sobre Real Estate, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.
a. Distribuya los datos en una tabla que muestre el número de casas con alberca frente al número de casas sin alberca en cada uno de los cinco municipios. Si selecciona una casa al azar, calcule las siguientes probabilidades:
1. La casa se localiza en el primer municipio o tiene alberca.

2. Dado que la casa se encuentra en el tercer municipio, que tenga alberca.
3. Tiene alberca y se localiza en el tercer municipio.

b. Distribuya los datos en una tabla que muestre el número de casas con cochera frente a las que no la tienen en cada uno de los cinco municipios. Si elige una casa al azar, calcule las siguientes probabilidades.
1. Tiene cochera.
2. Si se localiza en el quinto municipio, que no tenga cochera.
3. Tiene cochera y se localiza en el tercer municipio.
4. No tiene cochera o se localiza en el segundo municipio.

93. Consulte los datos sobre béisbol 2012, que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Establezca tres variables:

- Divida a los equipos en dos grupos, los que tuvieron una temporada ganadora y los que no. Es decir, cree una variable para contar los equipos que ganaron 81 juegos o más y los que ganaron 80 juegos o menos.
- Cree una nueva variable para la asistencia, con tres categorías: una asistencia inferior a 2.0 millones; una de 2.0 hasta 3.0 millones y una de 3.0 millones o más.
- Cree una variable que muestre cuáles equipos jugaron en un estadio de menos de 15 años de antigüedad, contra uno que tiene 15 años o más.

Siga las siguientes instrucciones:

a. Elabore una tabla que muestre el número de equipos que ganaron en la temporada frente a los que perdieron de acuerdo con las tres categorías de asistencia. Si selecciona un equipo al azar, calcule las siguientes probabilidades:
1. Tener una temporada ganadora.
2. Tener una temporada ganadora o contar con una asistencia de más de 3.0 millones.
3. Dada una asistencia de más de 3.0 millones, tener una temporada ganadora.
4. Tener una temporada perdedora y contar con una asistencia de menos de 2.0 millones.

b. Elabore una tabla que muestre el número de equipos que tuvieron una temporada ganadora contra los que jugaron en estadios antiguos o nuevos. Si selecciona un equipo al azar, calcule las siguientes probabilidades:
1. Seleccionar un equipo con una temporada ganadora.
2. Seleccionar un equipo con un récord ganador que haya jugado en un estadio nuevo.
3. El equipo tuvo un récord ganador o jugó en un estadio nuevo.

94. Consulte los datos de los autobuses del Distrito Escolar Buena. Establezca una variable que divida la edad de estos en tres grupos: nuevos (menos de cinco años), medios (mayores de cinco años pero menores de 10) y viejos (10 o más). El costo mediano de mantenimiento es de 456 dólares. Basándose en este valor, cree una variable para aquellos que están por debajo de la mediana (bajo mantenimiento) y los que están por encima de la mediana (alto mantenimiento). Finalmente, desarrolle una tabla que muestre la relación entre el costo de mantenimiento y la edad del autobús.

a. ¿Qué porcentaje de los autobuses son nuevos?
b. ¿Qué porcentaje de los nuevos autobuses tienen un bajo mantenimiento?
c. ¿Qué porcentaje de los viejos autobuses tienen alto mantenimiento?
d. ¿El costo de mantenimiento parece estar relacionado con la edad del autobús? *Sugerencia:* compare el costo de mantenimiento de los viejos autobuses con el costo de los nuevos. ¿Concluiría usted que el costo de mantenimiento es independiente de la edad?

6 Distribuciones discretas de probabilidad

ESTADÍSTICAS RECIENTES SUGIEREN que 15% de quienes visitan un sitio web de ventas de menudeo realiza una compra. Un minorista desea verificar esta afirmación. Para hacerlo, seleccionó una muestra de 16 "visitas" de su sitio y descubrió que en realidad cuatro realizaron una compra. ¿Cuál es la probabilidad de que exactamente esa cantidad de personas realicen una compra? ¿Cuántas compras debe esperar? ¿Cuál es la probabilidad de que cuatro o más "visitas" terminen en compra? (vea el ejercicio 49 y el objetivo de aprendizaje OA6-4).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA6-1 Identificar las características de una distribución de probabilidad.

OA6-2 Distinguir entre una variable aleatoria discreta y una continua.

OA6-3 Calcular la media, la varianza y la desviación estándar de una distribución de probabilidad discreta.

OA6-4 Explicar los supuestos de la distribución binomial y aplicarlos en el cálculo de probabilidades.

OA6-5 Explicar los supuestos de la distribución hipergeométrica y aplicarlos en el cálculo de probabilidades.

OA6-6 Explicar los supuestos de la distribución de Poisson y aplicarlos en el cálculo de probabilidades.

Introducción

Los capítulos 2 a 4 se dedicaron al estudio de la estadística descriptiva: datos en bruto organizados en una distribución de frecuencias, la cual se representa en tablas, gráficas y diagramas. Asimismo, se calculó una medida de ubicación —como la media aritmética, la mediana o la moda— para localizar un valor típico cercano al centro de la distribución. Mediante el rango y la desviación estándar se describió la dispersión de los datos. Dichos capítulos se centran en describir *algo que sucedió.*

A partir del capítulo 5, el tema cambia: ahora se analiza *algo que posiblemente suceda.* Esta faceta de la estadística recibe el nombre de *estadística inferencial.* El objetivo consiste en hacer inferencias (afirmaciones) sobre una población con base en determinada cantidad de observaciones, denominadas *muestras,* que se seleccionan de la población. En el capítulo 5 se estableció que una probabilidad es un valor entre cero y uno, inclusive, y se analizó la forma en que las probabilidades pueden combinarse de acuerdo con las reglas de la adición y la multiplicación.

En este capítulo se comienza el estudio de las **distribuciones de probabilidad.** Una distribución de probabilidad proporciona toda la gama de valores que pueden presentarse en un experimento. Es similar a una distribución de frecuencias relativas, pero, en lugar de referirse al pasado, describe la probabilidad de que un evento se presente en el futuro. Por ejemplo, si un fabricante de medicamentos afirma que cierto tratamiento permitirá que 80% de la población baje de peso, la agencia de protección al consumidor quizá someta a prueba el tratamiento con una muestra de seis personas. Si la afirmación del fabricante es cierta, es *casi imposible* tener un resultado en el que nadie en la muestra pierda peso y es *muy probable* que 5 de cada 6 lo pierdan.

En este capítulo se examinan la media, la varianza y la desviación estándar de una distribución de probabilidad, así como tres distribuciones de probabilidad que se presentan con frecuencia: binomial, hipergeométrica y de Poisson.

¿Qué es una distribución de probabilidad?

OA6-1 Identificar las características de una distribución de probabilidad.

Una distribución de probabilidad muestra los posibles resultados de un experimento y la probabilidad de que cada uno se presente.

DISTRIBUCIÓN DE PROBABILIDAD Lista de todos los resultados de un experimento y la probabilidad asociada a cada uno de ellos.

A continuación se mencionan las principales características de una distribución de probabilidad.

CARACTERÍSTICAS DE UNA DISTRIBUCIÓN DE PROBABILIDAD

1. La probabilidad de un resultado en particular se encuentra entre 0 y 1, inclusive.
2. Los resultados son eventos mutuamente excluyentes.
3. La lista es exhaustiva. Por lo tanto, la suma de las probabilidades de los diversos eventos es igual a 1.

¿Cómo generar una distribución de probabilidad? Esto se ilustra mediante el siguiente ejemplo.

EJEMPLO

Suponga que le interesa el número de caras que aparecen en tres lanzamientos de una moneda. Tal es el experimento. Los resultados posibles son: cero caras, una cara, dos caras y tres caras. ¿Cuál es la distribución de probabilidad del número de caras?

SOLUCIÓN

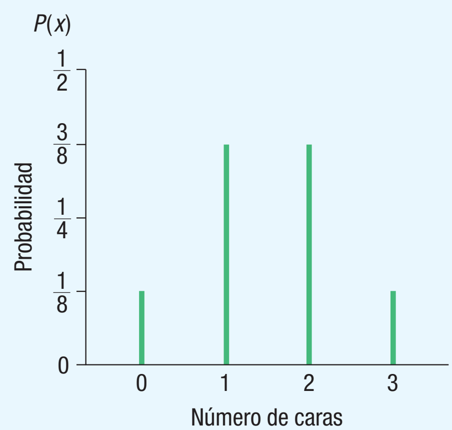
Hay ocho resultados posibles. En el primer lanzamiento puede aparecer una cara, una cruz en el segundo y otra cruz en el tercero, o puede obtener cruz, cruz y cara, en ese orden. Para obtener el conteo de resultados aplique la fórmula de la multiplicación [5.8]: (2)(2)(2), es decir, ocho posibles resultados. Estos se listan en seguida.

Resultado posible	Lanzamiento de la moneda Primero	Segundo	Tercero	Número de caras
1	Cruz	Cruz	Cruz	0
2	Cruz	Cruz	Cara	1
3	Cruz	Cara	Cruz	1
4	Cruz	Cara	Cara	2
5	Cara	Cruz	Cruz	1
6	Cara	Cruz	Cara	2
7	Cara	Cara	Cruz	2
8	Cara	Cara	Cara	3

Observe que el resultado "cero caras" ocurre solo una vez; "una cara", tres veces; "dos caras", tres veces, y "tres caras", una sola vez. Es decir, "cero caras" se presentó una de ocho veces. Por consiguiente, la probabilidad de cero caras es de un octavo; la probabilidad de una cara es de tres octavos, y así sucesivamente. La distribución de probabilidad se muestra en la tabla 6.1. Como uno de estos resultados debe suceder, el total de probabilidades de todos los eventos posibles es 1.000. Esto siempre se cumple. La gráfica 6.1 contiene la misma información.

TABLA 6.1 Distribución de probabilidad de los eventos relativos a cero, una, dos y tres caras en tres lanzamientos de una moneda

Número de caras, x	Probabilidad del resultado, $P(x)$
0	$\frac{1}{8} = 0.125$
1	$\frac{3}{8} = 0.375$
2	$\frac{3}{8} = 0.375$
3	$\frac{1}{8} = 0.125$
Total	$\frac{8}{8} = 1.000$



GRÁFICA 6.1 Presentación gráfica del número de caras que resultan de tres lanzamientos de una moneda y la probabilidad correspondiente

Consulte el ejemplo del lanzamiento de una moneda de la tabla 6.1. La probabilidad de x se representa como $P(x)$. De esta manera, la probabilidad de cero caras es $P(0 \text{ caras}) = 0.125$, y la probabilidad de una cara es $P(1 \text{ cara}) = 0.375$, etcétera. La suma de estas probabilidades mutuamente excluyentes es 1; es decir, de acuerdo con la tabla 6.1, $0.125 + 0.375 + 0.375 + 0.125 = 1.00$.

AUTOEVALUACIÓN 6-1

Los posibles resultados de un experimento que implica el lanzamiento de un dado son: uno, dos, tres, cuatro, cinco y seis.

(a) Elabore la distribución de probabilidad para el número de posibles resultados.
(b) Represente gráficamente la distribución de probabilidad.
(c) ¿Cuál es la suma de las probabilidades?

Variables aleatorias

OA6-2 Distinguir entre una variable aleatoria discreta y una continua.

En cualquier experimento aleatorio, los resultados se presentan al azar; así, a este se le denomina *variable aleatoria.* Por ejemplo, lanzar un dado constituye un experimento: puede ocurrir cualquiera de los seis resultados posibles. Algunos experimentos dan origen a resultados de índole cuantitativa (como dólares, peso o número de niños); otros generan resultados de naturaleza cualitativa (como el color o la afiliación religiosa). Cada valor de la variable aleatoria se relaciona con una probabilidad que indica la posibilidad de un resultado determinado. He aquí unos cuantos ejemplos que aclaran el concepto de **variable aleatoria.**

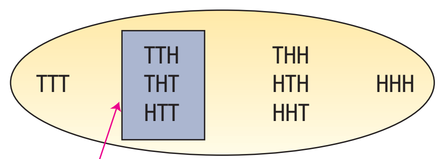
- Si cuenta el número de empleados ausentes en el turno matutino del lunes, el número puede ser 0, 1, 2, 3,… El número de ausencias es una variable aleatoria.
- Usted tiene cuatro lingotes de acero, cuyos pesos pueden ser de 2 492 libras, 2 497 libras, 2 506 libras, etcétera. El peso es una variable aleatoria.
- Otras variables aleatorias pueden ser la cantidad de focos defectuosos producidos por hora en Cleveland Company, Inc.; el grado escolar (9°, 10°, 11° o 12°) de los miembros del equipo femenino de basquetbol de la secundaria St. James; el número de corredores del maratón de Nueva York de 2013 y la cantidad diaria de conductores multados por conducir bajo la influencia del alcohol en Brazoria County, Texas, el mes previo.

VARIABLE ALEATORIA Cantidad que resulta de un experimento que, por azar, puede adoptar diferentes valores.

En el siguiente diagrama se ilustran los términos *experimento, resultado, evento* y *variable aleatoria.* Primero, en el caso del experimento en el que se lanza una moneda tres veces, hay ocho posibles resultados. En este experimento nos interesa la posibilidad de que salga una cara en tres lanzamientos. La variable aleatoria es el número de caras; y se desea saber la probabilidad del evento que tiene una variable aleatoria igual a 1. El resultado es P(1 cara en 3 lanzamientos) = 0.375.

Una variable aleatoria puede ser *discreta* o *continua.*

Posibles *resultados* de tres lanzamientos de moneda (T = cruz; H = cara)



Ocurre el *evento* {una cara} y la *variable aleatoria* $x = 1$.

Variable aleatoria discreta

Una variable aleatoria discreta adopta solo cierto número de valores separados. Por ejemplo, el Banco de las Carolinas cuenta el número de tarjetas de crédito pertenecientes a un grupo de clientes. Los datos se resumen en la tabla de frecuencias de la derecha.

En esta tabla de frecuencias, el número de tarjetas es la variable aleatoria discreta.

Número de tarjetas de crédito	Porcentaje
0	3%
1	10%
2	18%
3	21%
4 o más	48%
Total	100%

VARIABLE ALEATORIA DISCRETA Variable aleatoria que solo adopta valores claramente separados.

A veces, una variable aleatoria discreta asume valores fraccionarios o decimales. Estos deben estar separados por cierta distancia. Por ejemplo, las calificaciones de los jueces por destreza técnica y formas artísticas en una competencia de patinaje artístico son valores decimales, como 7.2, 8.9 y 9.7. Dichos valores son discretos, pues hay una distancia entre las calificaciones, como 8.3 y 8.4. Una calificación no puede tener un valor de 8.34 o de 8.347.

Variable aleatoria continua

Por otra parte, una variable aleatoria puede ser continua. Si se mide algo, como el ancho de una recámara, la estatura de una persona o la presión de la llanta de un automóvil, se trata de una *variable aleatoria continua.* Es posible suponer una infinidad de valores, con ciertas limitaciones. Por ejemplo:

- Los tiempos de los vuelos comerciales entre Atlanta y Los Ángeles son de 4.67 horas, 5.13 horas, etcétera. La variable aleatoria es el tiempo medido en horas.
- La presión, medida en libras por pulgada cuadrada (psi), de un nuevo neumático Chevy Trailblazer puede ser de 32.78 psi, 31.62 psi, 33.07 psi, etcétera. En otras palabras, es razonable que se presente cualquier valor entre 28 y 35. La variable aleatoria es la presión de la llanta.

Como ocurre con las variables aleatorias discretas, la probabilidad de una variable aleatoria continua puede resumirse con una **distribución de probabilidad.** Así que, ¿cuál es la diferencia entre una distribución de probabilidad y una variable aleatoria? Una variable aleatoria representa el resultado particular de un experimento, mientras que una distribución de probabilidad representa todos los posibles resultados, así como la probabilidad correspondiente.

Las herramientas que se utilizan y las interpretaciones probabilísticas son diferentes en el caso de distribuciones de probabilidades discretas y continuas. Este capítulo se limita al análisis e interpretación de distribuciones discretas. En el siguiente capítulo se estudian las distribuciones continuas. ¿Cuál diría que es la diferencia entre ambos tipos de distribuciones? Por lo general, una distribución discreta es el resultado de contar algo, como:

- El número de caras que se presentan en tres lanzamientos de una moneda.
- El número de estudiantes que obtienen A en clase.
- El número de empleados de producción que se ausentaron este día durante el segundo turno.
- El número de comerciales de 30 segundos que se transmiten hoy en la NBC de las 8:00 a las 11:00 p. m.

Las distribuciones continuas son el resultado de algún tipo de medición, como:

- La duración de cada canción en el último álbum de Linkin Park.
- El peso de cada estudiante de un curso.
- La temperatura ambiente en el momento en que usted lee este libro.
- La edad en años de cada uno de los 161 000 empleados de la Ford Motor Company.

OA6-3
Calcular la media, la varianza y la desviación estándar de una distribución de probabilidad discreta.

Media, varianza y desviación estándar de una distribución de probabilidad discreta

En el capítulo 3 se estudiaron las medidas de ubicación y la variación de una distribución de frecuencias. La media indica la posición central de los datos, y la varianza describe la dispersión de estos. De forma similar, una distribución de probabilidad queda resumida por su media y su varianza. La media de una distribución de frecuencias se identifica mediante la letra minúscula griega mu (μ), y la desviación estándar, con sigma (σ).

Media

La media constituye un valor típico para representar la posición central de una distribución de probabilidad. También es el valor promedio de la variable aleatoria. La media de una distribución de probabilidad también recibe el nombre de **valor esperado.** Se trata de un promedio ponderado en el que los posibles valores de una variable aleatoria se ponderan con sus correspondientes probabilidades de ocurrir.

La media de una distribución de probabilidad discreta se calcula con la fórmula:

MEDIA DE UNA DISTRIBUCIÓN DE PROBABILIDAD $\mu = \Sigma[xP(x)]$ **[6.1]**

donde $P(x)$ es la probabilidad de un valor particular x. En otras palabras, cada valor x se multiplica por sus probabilidades de ocurrir y enseguida se suman los productos.

Varianza y desviación estándar

La media constituye un valor típico para resumir una distribución de probabilidad discreta. Sin embargo, no describe el grado de dispersión (variación) en una distribución. La varianza sí lo hace. La fórmula de la varianza de una distribución de probabilidad es:

VARIANZA DE UNA DISTRIBUCIÓN DE PROBABILIDAD $\sigma^2 = \Sigma[(x - \mu)^2 P(x)]$ **[6.2]**

Los pasos para el cálculo son los siguientes:

1. La media se resta de cada valor de la variable aleatoria y la diferencia se eleva al cuadrado.
2. Cada diferencia al cuadrado se multiplica por su probabilidad.
3. Se suman los productos resultantes para obtener la varianza.

La desviación estándar, σ, se determina al extraer la raíz cuadrada positiva de σ^2; es decir, $\sigma = \sqrt{\sigma^2}$.

En el siguiente ejemplo se explican los detalles del cálculo e interpretación de la media y la desviación estándar de una distribución de probabilidad.

EJEMPLO

John Ragsdale vende automóviles nuevos en Pelican Ford. Por lo general, John vende la mayor cantidad de automóviles el sábado, así que desarrolló la siguiente distribución de probabilidades, en la cual se muestra la cantidad de automóviles que espera vender un sábado determinado.

Cantidad de automóviles vendidos, *x*	Probabilidad, *P*(*x*)
0	0.1
1	0.2
2	0.3
3	0.3
4	0.1
Total	1.0

1. ¿De qué tipo de distribución se trata?
2. ¿Cuántos automóviles espera vender John un sábado normal?
3. ¿Cuál es la varianza de la distribución?

SOLUCIÓN

1. Se trata de una distribución de probabilidad discreta de la variable aleatoria "número de automóviles vendidos". Observe que John solo espera vender cierto rango de automóviles; no espera vender cinco automóviles, ni 50. Además, no puede vender medio vehículo. Solo puede vender 0, 1, 2, 3 o 4 automóviles. Asimismo, los resultados son mutuamente excluyentes: no puede vender un total de tres y cuatro automóviles el mismo sábado.
2. La media de la cantidad de automóviles vendidos se calcula al multiplicar el número de automóviles que vendió por la probabilidad de vender dicho número, y sumar los productos de acuerdo con la fórmula [6.1]:

$$\begin{aligned}\mu &= \Sigma[xP(x)]\\ &= 0(0.1) + 1(0.2) + 2(0.3) + 3(0.3) + 4(0.1)\\ &= 2.1\end{aligned}$$

Estos cálculos se resumen en la siguiente tabla.

Número de automóviles vendidos, *x*	Probabilidad, *P*(*x*)	*x* · *P*(*x*)
0	0.1	0.0
1	0.2	0.2
2	0.3	0.6
3	0.3	0.9
4	0.1	0.4
Total	1.0	μ = 2.1

¿Cómo se interpreta una media de 2.1? Este valor indica que, a lo largo de una gran cantidad de sábados, John Ragsdale espera vender un promedio de 2.1 automóviles por día. Por supuesto, no es posible vender *exactamente* 2.1 automóviles un sábado en particular. Sin embargo, el valor esperado se utiliza para predecir la media aritmética de la cantidad de automóviles vendidos a largo plazo. Por ejemplo, si John trabaja 50 sábados en un año, puede esperar vender (50)(2.1) o 105 automóviles solo durante los sábados. Por consiguiente, a veces la media recibe el nombre de *valor esperado.*

3. En la siguiente tabla se ilustran los pasos para calcular la varianza utilizando la fórmula [6.2]. Las primeras dos columnas repiten la distribución de probabilidad. En la tercera columna se resta la media de cada valor de la variable aleatoria. En la cuarta se elevan al cuadrado las diferencias de la tercera columna. En la quinta se multiplica cada diferencia al cuadrado de la cuarta columna por la probabilidad correspondiente. La varianza es la suma de los valores de la quinta columna.

Número de automóviles vendidos, x	**Probabilidad, $P(x)$**	**$(x - \mu)$**	**$(x - \mu)^2$**	**$(x - \mu)^2P(x)$**
0	.1	0 − 2.1	4.41	0.441
1	.2	1 − 2.1	1.21	0.242
2	.3	2 − 2.1	0.01	0.003
3	.3	3 − 2.1	0.81	0.243
4	.1	4 − 2.1	3.61	0.361
				$\sigma^2 = 1.290$

Recuerde que la desviación estándar, σ, es la raíz cuadrada positiva de la varianza. En este ejemplo es $\sqrt{\sigma^2} = \sqrt{1.290} = 1.136$ automóviles. ¿Cómo se interpreta una desviación estándar de 1.136 automóviles? Si la vendedora Rita Kirsch también vendió un promedio de 2.1 automóviles los sábados y la desviación estándar de sus ventas fue de 1.91 automóviles, se concluye que hay más variabilidad en las ventas sabatinas de Kirsch que en las de Ragsdale (pues $1.91 > 1.136$).

AUTOEVALUACIÓN 6-2

Pizza Palace ofrece tres tamaños de refresco de cola. El mediano cuesta 0.80 dólares; el mediano, 0.90 y el grande, 1.20; 30% de las bebidas vendidas son de tamaño chico; 50%, mediano, y 20%, grande. Cree una distribución de probabilidad para la variable aleatoria “precio” y responda las siguientes preguntas.

(a) ¿Se trata de una distribución de probabilidad discreta? Indique por qué sí o por qué no.
(b) Calcule el precio promedio que se cobra por refresco de cola.
(c) ¿Cuál es la varianza de la cantidad que se cobra por un refresco de cola? ¿Cuál es la desviación estándar?

EJERCICIOS

1. Calcule la media y la varianza de la siguiente distribución de probabilidad discreta.

x	$P(x)$
0	0.2
1	0.4
2	0.3
3	0.1

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2. Calcule la media y la varianza de la siguiente distribución de probabilidad discreta.

x	$P(x)$
2	0.5
8	0.3
10	0.2

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

3. Calcule la media y la varianza de la siguiente distribución de probabilidad.

x	$P(x)$
5	0.1
10	0.3
15	0.2
20	0.4

4. ¿Cuáles de las siguientes variables aleatorias son discretas y cuáles son continuas?
- **a.** El número de cuentas nuevas conseguidas por un vendedor en un año.
- **b.** El tiempo que transcurre entre la llegada de cada cliente a un cajero automático.
- **c.** El número de clientes en la estética Big Nick.
- **d.** La cantidad de combustible que contiene el tanque de gasolina de su automóvil.
- **e.** La cantidad de miembros del jurado pertenecientes a una minoría.
- **f.** La temperatura ambiente el día de hoy.

5. La información que sigue representa el número de llamadas diarias al servicio de emergencia que el servicio voluntario de ambulancias de Walterboro, Carolina del Sur, durante los últimos 50 días. En otras palabras, hubo 22 días en los que se realizaron dos llamadas de emergencia, y nueve días en los que se realizaron tres llamadas de emergencia.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Número de llamadas	Frecuencia
0	8
1	10
2	22
3	9
4	1
Total	50

- **a.** Convierta la información del número de llamadas en una distribución de probabilidad.
- **b.** ¿Es un ejemplo de distribución de probabilidad discreta o continua?
- **c.** ¿Cuál es la media de la cantidad de llamadas de emergencia al día?
- **d.** ¿Cuál es la desviación estándar de la cantidad de llamadas diarias?

6. El director de admisiones de la Universidad Kinzua, en Nueva Escocia, estimó la distribución de admisiones de estudiantes para el segundo semestre con base en la experiencia de años pasados. ¿Cuál es el número de admisiones esperado para el segundo semestre? Calcule la varianza y la desviación estándar del número de admisiones.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Admisiones	Probabilidad
1 000	0.6
1 200	0.3
1 500	0.1

7. Belk Department Store tiene una venta especial este fin de semana. Los clientes que registren cargos por compras superiores a 50 dólares en su tarjeta de crédito de Belk recibirán una tarjeta especial de la lotería de la empresa. El cliente raspará la tarjeta para descubrir la cantidad que se descontará del total de compras. A continuación se muestra tanto el premio como el porcentaje de veces que se deducirá del total de las compras.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Suma de premios	Probabilidad
$ 10	0.50
25	0.40
50	0.08
100	0.02

- **a.** ¿Cuál es la cantidad media deducida de la compra total?
- **b.** ¿Cuál es la desviación estándar de la cantidad deducida del total de las compras?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

8. La Downtown Parking Authority, de Tampa, Florida, reportó los siguientes datos de una muestra de 250 clientes relacionados con el número de horas que se estacionan y las cantidades que pagan.

Número de horas	Frecuencia	Pago
1	20	$ 3.00
2	38	6.00
3	53	9.00
4	45	12.00
5	40	14.00
6	13	16.00
7	5	18.00
8	36	20.00
	250	

a. Convierta la información del número de horas de estacionamiento en una distribución de probabilidad. ¿Es una distribución de probabilidad discreta o continua?
b. Determine la media y la desviación estándar de la cantidad de horas de estacionamiento. ¿Qué respondería si se le pregunta por el número de horas que se estaciona un cliente normal?
c. Calcule la media y la desviación estándar del pago.

OA6-4
Explicar los supuestos de la distribución binomial y aplicarlos en el cálculo de probabilidades.

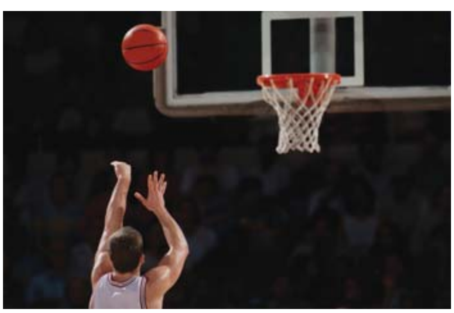
Distribución de probabilidad binomial

La **distribución de probabilidad binomial** es de probabilidad discreta, y se presenta con mucha frecuencia. Hay cuatro requisitos para describir los resultados experimentales con una distribución binomial. El primero es que solo haya dos posibles resultados en determinado ensayo del experimento. He aquí dos ejemplos: en una prueba, el enunciado en una pregunta de cierto o falso solo se puede responder de dos maneras; por otro lado, en un resort, la supervisora de ama de llaves revisa el trabajo de un empleado y lo evalúa como aceptable o inaceptable. Una característica clave de ambos resultados es que son mutuamente excluyentes. Esto significa que la respuesta a una pregunta como la del primer ejemplo debe ser ya sea *cierto* o *falso*, pero no puede ser cierta y falsa a la vez. Otro ejemplo es el resultado de una llamada de ventas; el cliente compra el producto o no, pero la venta no puede arrojar ambos resultados. Con frecuencia los dos posibles resultados se clasifican como "éxito" y "fracaso". Sin embargo, esta distinción no implica que un resultado sea bueno y el otro sea malo, solo que son dos resultados mutuamente excluyentes.

El segundo requisito de la distribución binomial es que la variable aleatoria es el número de éxitos en el número total de ensayos. Por ejemplo, se lanza una moneda cinco veces y se cuenta el número de veces que aparece una cara; se seleccionan 10 trabajadores aleatoriamente y se registra cuántos tienen más de 50 años; o se eligen al azar 20 cajas de Raisin Bran de Kellogg y se identifica cuántas pesan más de lo que indica el paquete. En cada ejemplo, se cuenta el número de éxitos de un número fijo de ensayos.

Un tercer requisito es que la probabilidad de éxito sea la misma para cada ensayo. Tres ejemplos son:

- En una prueba con 10 preguntas de cierto o falso se sabe que hay 10 ensayos, y la probabilidad de adivinar correctamente la respuesta de cada uno de ellos es de 0.5 o para una prueba de 20 preguntas de opción múltiple con cuatro opciones y solo una respuesta correcta, se sabe que hay 20 ensayos y que la probabilidad de adivinar al azar la respuesta correcta para cada uno es de 0.25.



- Bones Albaugh es un jugador universitario de basquetbol de la primera división, que acierta 70% de sus tiros libres. Si tiene cinco oportunidades de tiro en el juego de esta noche, la probabilidad de tener éxito en cada intento es de 0.70.
- En una encuesta reciente, 18% de los adultos entrevistados indicó que una barra de Snickers era su dulce favorito. Se elige una muestra de 15 adultos y se les pregunta cuál es su dulce favorito. La probabilidad de que uno responda "una barra de Snickers" es de 0.18.

El último requisito de una distribución de probabilidad binomial consiste en que cada ensayo sea *independiente* de cualquier otro; esto implica que no deben existir patrones en los ensayos; es decir, que el resultado de un ensayo particular no influya en el resultado de otro. Dos ejemplos de lo anterior son:

- Una familia joven tiene dos hijos, ambos varones. La probabilidad de que el tercer hijo sea varón sigue siendo 0.50. Es decir, el género del tercer hijo es independiente de los otros dos.
- Suponga que 20% de los pacientes atendidos en la sala de urgencias del Waccamaw Hospital no tiene seguro médico. Si el segundo paciente atendido durante el turno vespertino no tiene seguro, eso no afecta la probabilidad de que el tercero, el décimo o cualquiera de los otros pacientes cuente o no con dicho beneficio.

EXPERIMENTO DE PROBABILIDAD BINOMIAL

1. El resultado de cada ensayo de un experimento se clasifica en una de dos categorías mutuamente excluyentes: éxito o fracaso.
2. La variable aleatoria permite contar el número de éxitos en una cantidad fija de ensayos.
3. La probabilidad de éxito y fracaso es la misma en cada ensayo.
4. Los ensayos son independientes, lo cual significa que el resultado de un ensayo no influye en el resultado del otro.

¿Cómo se calcula una probabilidad binomial?

Para construir una probabilidad binomial particular se necesita: 1) el número de ensayos y 2) la probabilidad de éxito de cada ensayo. Por ejemplo, si la Hannah Landscaping Company planta hoy 10 pinos Norfolk, sabiendo que 90% de estos árboles sobrevivirá, se puede calcular la probabilidad binomial de que exactamente ocho árboles sobrevivan. En este caso, el número de ensayos son los 10 árboles, la probabilidad de éxito es 0.90 y el número de éxitos es 8. De hecho, se puede calcular una probabilidad binomial para cualquier número de éxitos de 0 a 10 árboles sobrevivientes.

Una probabilidad binomial se calcula mediante la fórmula:

FÓRMULA DE LA PROBABILIDAD BINOMIAL $$P(x) = {}_nC_x = \pi^x(1 - \pi)^{n-x} \qquad [6.3]$$

donde:

C es el símbolo de combinación;
n es el número de ensayos;
x es la variable aleatoria definida como el número de éxitos;
π es la probabilidad de éxito en cada ensayo.

La letra griega π (pi) representa un parámetro de población binomial. No la confunda con la constante matemática 3.1416.

EJEMPLO

US Airways tiene cinco vuelos diarios de Pittsburgh al aeropuerto regional de Bradford, Pennsylvania. Suponga que la probabilidad de que cualquier vuelo llegue tarde es de 0.20. ¿Cuál es la probabilidad de que ningún vuelo llegue tarde hoy? ¿Cuál es la probabilidad de que exactamente uno de los vuelos llegue tarde hoy?

SOLUCIÓN

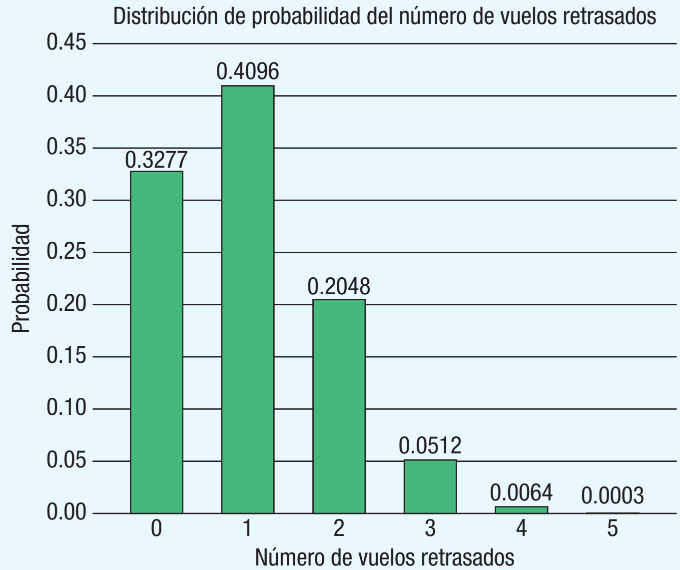
Aplique la fórmula [6.3]. La probabilidad de que un vuelo llegue tarde es de 0.20, así, $\pi = 0.20$. Hay cinco vuelos, por lo que $n = 5$, y x (la variable aleatoria) se refiere al número de éxitos. En este caso, un éxito consiste en que un avión llegue tarde. Como no hay demoras en las llegadas, $x = 0$.

$$\begin{aligned} P(0) &= {}_nC_x(\pi)^x(1 - \pi)^{n-x} \\ &= {}_5C_0(0.20)^0(1 - 0.20)^{5-0} = (1)(1)(0.3277) = 0.3277 \end{aligned}$$

La probabilidad de que exactamente uno de los cinco vuelos llegue tarde es de 0.4096, que se calcula de la siguiente manera:

$$P(1) = {}_nC_x(\pi)^x(1 - \pi)^{n-x}$$
$$= {}_5C_1(0.20)^1(1 - 0.20)^{5-1} = (5)(0.20)(0.4096) = 0.4096$$

La distribución de probabilidad binomial completa con $\pi = 0.20$ y $n = 5$ se muestra en la siguiente gráfica de barras. Observe que la probabilidad de que exactamente tres vuelos lleguen tarde es de 0.0512, y al analizar el diagrama de barras se concluye que la distribución del número de llegadas demoradas tiene un sesgo positivo.



La media (π) y la varianza (σ^2) de una distribución binomial se calculan con la siguiente fórmula, fácil y rápida:

MEDIA DE UNA DISTRIBUCIÓN BINOMIAL $\mu = n\pi$ **[6.4]**

VARIANZA DE UNA DISTRIBUCIÓN BINOMIAL $\sigma^2 = n\pi(1 - \pi)$ **[6.5]**

Por ejemplo, respecto del número de vuelos retrasados, recuerde que $\pi = 0.20$ y $n = 5$. Por lo tanto:

$$\mu = n\pi = 5(0.20)) = 1.0$$
$$\sigma^2 = n\pi(1 - \pi) = 5(0.20)(1 - 0.20) = 0.80$$

La media de 1.0 y la varianza de 0.80 se verifican con las fórmulas [6.1] y [6.2]. La distribución de probabilidad del resultado de la gráfica de barras superior, así como los detalles de los cálculos, se muestran a continuación.

Número de vuelos retrasados, x	$P(x)$	$xP(x)$	$x - \mu$	$(x - \mu)^2$	$(x - \mu)^2P(x)$
0	0.3277	0.0000	−1	1	0.3277
1	0.4096	0.4096	0	0	0
2	0.2048	0.4096	1	1	0.2048
3	0.0512	0.1536	2	4	0.2048
4	0.0064	0.0256	3	9	0.0576
5	0.0003	0.0015	4	16	0.0048
		$\mu = 1.0000$			$\sigma^2 = 0.7997$

Tablas de probabilidad binomial

Con la fórmula [6.3] se construye una distribución de probabilidad binomial para cualesquiera valores de n y π. Sin embargo, si n es grande, los cálculos consumen más tiempo. Por conveniencia, las tablas del apéndice B.1 muestran el resultado de la aplicación de la fórmula en el caso de varios valores de n y π. En la tabla 6.2 se muestra parte del apéndice B.1 para $n = 6$ y diversos valores de π.

TABLA 6.2 Probabilidades binomiales para $n = 6$ y valores seleccionados de π

					$n = 6$ Probabilidad						
$x\backslash\pi$	**0.05**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**0.95**
0	**0.735**	0.531	0.262	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000
1	**0.232**	0.354	0.393	0.303	0.187	0.094	0.037	0.010	0.002	0.000	0.000
2	**0.031**	0.098	0.246	0.324	0.311	0.234	0.138	0.060	0.015	0.001	0.000
3	**0.002**	0.015	0.082	0.185	0.276	0.313	0.276	0.185	0.082	0.015	0.002
4	**0.000**	0.001	0.015	0.060	0.138	0.234	0.311	0.324	0.246	0.098	0.031
5	**0.000**	0.000	0.002	0.010	0.037	0.094	0.187	0.303	0.393	0.354	0.232
6	**0.000**	0.000	0.000	0.001	0.004	0.016	0.047	0.118	0.262	0.531	0.735

EJEMPLO

En el sudoeste, 5% de todas las llamadas por teléfono celular se interrumpen. ¿Cuál es la probabilidad de que, en seis llamadas seleccionadas al azar, ninguna se interrumpa? ¿Exactamente una? ¿Exactamente dos? ¿Exactamente tres? ¿Exactamente cuatro? ¿Exactamente cinco? ¿Exactamente seis de seis?

SOLUCIÓN

Las condiciones binomiales se cumplen: a) hay solo dos posibles resultados (una llamada determinada se interrumpe o no); b) existe una cantidad fija de ensayos (6); c) hay una probabilidad constante de éxito (0.05); d) los ensayos son independientes.

Consulte la tabla 6.2 y localice la probabilidad de que exactamente cero llamadas se interrumpan. Descienda por el margen izquierdo hasta llegar al valor 0 de x. Ahora siga por la horizontal hasta la columna con un encabezado p de 0.05 para determinar la probabilidad. Esta es de 0.735 (los valores en la tabla 6.2 se redondearon a tres lugares decimales).

La probabilidad de que exactamente una llamada se interrumpa en una muestra de seis es de 0.232. La distribución de probabilidad completa de $n = 6$ y $p = 0.05$ es la siguiente:

Número de llamadas interrumpidas, x	Probabilidad de que ocurra, $P(x)$	Número de llamadas interrumpidas, x	Probabilidad de que ocurra, $P(x)$
0	0.735	4	0.000
1	0.232	5	0.000
2	0.031	6	0.000
3	0.002		

Por supuesto, existe una ligera posibilidad de obtener cinco llamadas interrumpidas en seis selecciones aleatorias. Esta es de 0.00000178, que se determina al sustituir los valores adecuados en la fórmula binomial:

$$P(5) = {}_6C_5(0.5)^5(0.95)^1 = (6)(0.05)^5(0.95) = 0.00000178$$

En el caso de seis de seis, la probabilidad exacta es de 0.000000016. Por consiguiente, la probabilidad de seleccionar cinco o seis llamadas interrumpidas de una muestra de seis es muy pequeña.

Es posible calcular la media o valor esperado de la distribución del número de llamadas interrumpidas:

$$\mu = n\pi = (6)(0.05)) = 0.30$$

$$\sigma^2 = n\pi(1 - \pi) = 6(0.05)(0.95) = 0.285$$

AUTOEVALUACIÓN 6-3

De los empleados de la planta de General Mills en Laskey Road, 95% recibe su sueldo bimestral por medio de transferencias de fondos electrónicos. Este mecanismo también recibe el nombre de *depósito directo.* Suponga que selecciona una muestra aleatoria de siete empleados.

(a) ¿Esta situación cumple los supuestos de la distribución binomial?
(b) ¿Cuál es la probabilidad de que a los siete empleados se les haga un depósito directo?
(c) Aplique la fórmula [6.3] para determinar la probabilidad exacta de que a cuatro de los siete empleados de la muestra se les haga un depósito directo.
d) Utilice Excel para verificar sus respuestas a los puntos *b* y *c.*

	A	B
1	Success	Probability
2	0	0.0230
3	1	0.0910
4	2	0.1754
5	3	0.2198
6	4	0.2011
7	5	0.1432
8	6	0.0826
9	7	0.0397
10	8	0.0162
11	9	0.0057
12	10	0.0017
13	11	0.0005
14	12	0.0001
15	13	0.0000
16	14	0.0000
17	15	0.0000

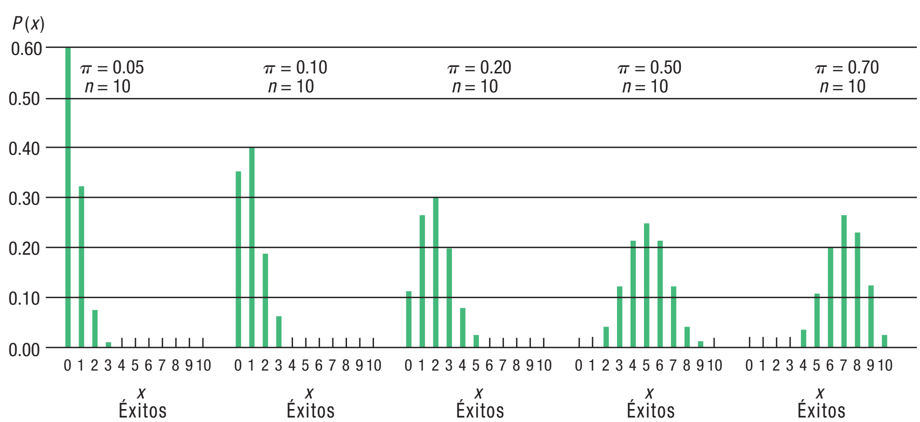
El apéndice B.1 es limitado; ofrece probabilidades para *n* valores de 1 a 15, y para valores de π de 0.05, 0.10, [...] 0.90 y 0.95. Un programa de software puede generar las probabilidades de un número específico de éxitos, dados *n* y π. En la salida Excel que aparece a la izquierda se muestra la probabilidad cuando $n = 40$ y $\pi = 0.09$. Observe que el número de éxitos se detiene en 15, pues las probabilidades de 16 a 40 se aproximan mucho a 0. Las instrucciones se detallan en la sección "Comandos de software" en el apéndice C.

Es preciso mencionar otras cuestiones adicionales relacionadas con la distribución de probabilidad binomial.

1. Si *n* permanece igual y π se incrementa de 0.05 a 0.95, la forma de la distribución cambia. Observe la tabla 6.3 y la gráfica 6.2; las probabilidades de que π sea 0.05 presentan un sesgo positivo. Conforme π se aproxima a 0.50, la distribución se torna más simétrica. A medida que π sea mayor a 0.50 *y* se aproxime a 0.95, la distribución de probabilidad adquiere un sesgo negativo. En la tabla 6.3 se destacan las probabilidades de $n = 10$ y valores de π de 0.05, 0.10, 0.20, 0.50 y 0.70. Las gráficas de estas distribuciones de probabilidad se muestran en la gráfica 6.2.

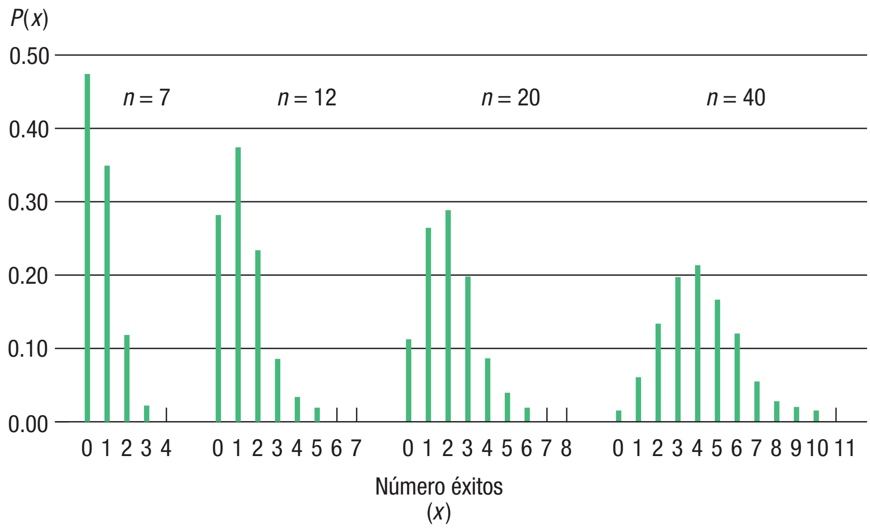
TABLA 6.3 Probabilidad de 0, 1, 2, ... éxitos para valores de $\pi = 0.05, 0.10, 0.20, 0.50$ y 0.70 con una *n* de 10

$x \backslash \pi$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000
1	0.315	0.387	0.268	0.121	0.040	0.010	0.002	0.000	0.000	0.000	0.000
2	0.075	0.194	0.302	0.233	0.121	0.044	0.011	0.001	0.000	0.000	0.000
3	0.010	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001	0.000	0.000
4	0.001	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006	0.000	0.000
5	0.000	0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001	0.000
6	0.000	0.000	0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011	0.001
7	0.000	0.000	0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057	0.010
8	0.000	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.302	0.194	0.075
9	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.268	0.387	0.315
10	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.107	0.349	0.599



GRÁFICA 6.2 Representación gráfica de la distribución de probabilidad binomial para valores de $\pi = 0.05, 0.10, 0.20, 0.50$ y 0.70 con una *n* de 10

2. Si π (la probabilidad de éxito) conserva el mismo valor, pero n aumenta, la forma de la distribución binomial se torna más simétrica. En la gráfica 6.3 se muestra el caso en el que π permanece constante en 0.10, pero n se incrementa de 7 a 40.



GRÁFICA 6.3 Representación gráfica de la distribución de probabilidad binomial para valores π de 0.10 y n de 7, 12, 20 y 40

EJERCICIOS

9. En una situación binomial, $n = 4$ y $\pi = 0.25$. Determine las probabilidades de los siguientes eventos usando la fórmula binomial.
   a. $x = 2$
   b. $x = 3$
10. En una situación binomial, $n = 5$ y $\pi = 0.40$. Determine las probabilidades de los siguientes eventos usando la fórmula binomial.
   a. $x = 1$
   b. $x = 2$
11. Suponga una distribución binomial en la que $n = 3$ y $\pi = 0.60$.
   a. Consulte el apéndice B.1 y elabore una lista de probabilidades para valores de $x = 0, 1, 2$ y 3.
   b. Determine la media y la desviación estándar de la distribución a partir de las definiciones generales de las fórmulas [6.1] y [6.2].
12. Suponga que existe una distribución binomial en la que $n = 5$ y $\pi = 0.30$.
   a. Consulte el apéndice B.1 y elabore una lista de probabilidades para valores de $x = 0, 1, 2$ y 3.
   b. Determine la media y la desviación estándar de la distribución a partir de las definiciones generales de las fórmulas [6.1] y [6.2].
13. Un estudio de la American Society of Investors descubrió que 30% de inversionistas particulares había utilizado un agente de descuentos. Considere una muestra aleatoria de nueve personas y determine la probabilidad de que:
   a. Exactamente dos personas hayan utilizado un agente de descuentos.
   b. Exactamente cuatro personas hayan recurrido a él.
   c. Ninguna persona lo haya empleado.
14. El servicio postal de Estados Unidos informa que 95% de la correspondencia de primera clase dentro de la misma ciudad se entrega en un periodo de dos días a partir del momento en que se envía. Se enviaron seis cartas de forma aleatoria a diferentes lugares.
   a. ¿Cuál es la probabilidad de que las seis lleguen en un plazo de dos días?
   b. ¿Cuál es la probabilidad de que exactamente cinco lleguen en un plazo de dos días?
   c. Determine la media del número de cartas que llegarán en un plazo de dos días.
   d. Calcule la varianza y la desviación estándar del número de cartas que llegarán en un plazo de dos días.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

15. Las normas de la industria sugieren que 10% de los vehículos nuevos requiere un servicio de garantía durante el primer año. El día de ayer, Jones Nissan, de Sumter, Carolina del Sur, vendió 12 automóviles marca Nissan.
   a. ¿Cuál es la probabilidad de que ninguno de estos vehículos requiera servicio de garantía?
   b. ¿Cuál es la probabilidad de que exactamente uno requiera servicio de garantía?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

c. Determine la probabilidad de que exactamente dos requieran servicio de garantía.
d. Calcule la media y la desviación estándar de esta distribución de probabilidad.

16. Un agente de telemarketing hace seis llamadas por hora y es capaz de hacer una venta con 30% de sus contactos. Para las siguientes dos horas, determine:
a. La probabilidad de realizar exactamente cuatro ventas.
b. La probabilidad de no realizar ninguna venta.
c. La probabilidad de hacer exactamente dos ventas.
d. La media de la cantidad de ventas durante un periodo de dos horas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

17. Una encuesta reciente de la American Accounting Association reveló que 23% de los estudiantes graduados en contabilidad elige la contaduría pública. Suponga que elige una muestra de 15 estudiantes recién graduados.
a. ¿Cuál es la probabilidad de que dos hayan elegido contaduría pública?
b. ¿Cuál es la probabilidad de que cinco hayan elegido contaduría pública?
c. ¿Cuántos graduados esperaría que eligieran contaduría pública?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

18. Se reporta que 16% de los hogares estadounidenses utilizan exclusivamente un teléfono celular como servicio telefónico. En una muestra de ocho hogares, encuentre:
a. La probabilidad de que ninguno use un celular como su servicio exclusivo.
b. La probabilidad de que exactamente cinco usen solo el celular.
c. El número medio de hogares que solo usan el celular.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Distribuciones de probabilidad binomial acumulada

Tal vez desee conocer la probabilidad de adivinar la respuesta a seis *o más* preguntas de verdadero o falso de un total de 10. O quizás esté interesado en la probabilidad de *seleccionar,* en forma aleatoria, *menos de dos* artículos defectuosos en la producción de la hora anterior. En estos casos usted necesita distribuciones de frecuencia acumulada similares a las del capítulo 2. En el siguiente ejemplo se ilustra este hecho.

EJEMPLO

Un estudio del departamento de transporte de Illinois concluyó que 76.2% de quienes ocupaban los asientos delanteros de los vehículos utilizaba cinturón de seguridad. Esto significa que los dos ocupantes de la parte delantera lo hacían. Suponga que decide comparar la información con el uso actual que se da al cinturón de seguridad. Seleccione una muestra de 12 vehículos.

1. ¿Cuál es la probabilidad que los ocupantes de la parte delantera en exactamente siete de los 12 vehículos seleccionados utilicen cinturones de seguridad?
2. ¿Cuál es la probabilidad que los ocupantes de la parte delantera de por lo menos siete de los 12 vehículos utilicen cinturón de seguridad?

SOLUCIÓN

Esta situación satisface los requisitos binomiales.

- En un vehículo en particular, ambos ocupantes de la parte delantera utilizan cinturón de seguridad o no lo hacen. Solo hay dos posibles resultados.
- Existe una cantidad fija de ensayos, 12 en este caso, pues se verifica esa cantidad de vehículos.
- La probabilidad de un "éxito" (los ocupantes utilizan cinturón de seguridad) es la misma en un vehículo u otro: 76.2%.
- Los ensayos son independientes. El que en el cuarto vehículo seleccionado todos los ocupantes utilicen cinturón de seguridad no influye en los resultados del quinto vehículo, o en los del décimo.

Aplique la fórmula [6.3] para determinar la probabilidad de que los ocupantes de *exactamente* siete vehículos de la muestra utilicen cinturón de seguridad. En este caso, $n = 12$ y $\pi = 0.762$.

$$P(x = 7) = {}_{12}C_7(0.762)^7(1 - 0.762)^{12-7} = 792(0.149171)(0.000764) = 0.0902$$

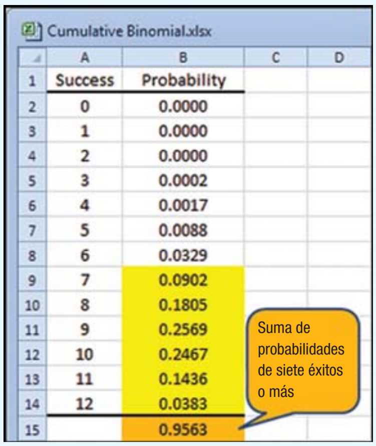
De esta manera, se concluye que la probabilidad de que los ocupantes de exactamente siete de los 12 vehículos de la muestra utilicen cinturones de seguridad es de aproximadamente 9%. Como se hizo en esta ecuación, con frecuencia se emplea una barra "|" para dar a entender "dado que". Así, en esta ecuación se trata de conocer la probabilidad de que x sea igual a 7 "dado que el número de ensayos es de 12 y la probabilidad de un éxito es de 0.762".

Para determinar la probabilidad de que los ocupantes de siete o más de los vehículos utilicen su cinturón de seguridad, aplique la fórmula [6.3] de este capítulo, así como la regla especial de la adición del capítulo anterior (vea fórmula [5.2]).

Como los eventos son mutuamente excluyentes (lo cual significa que una muestra de 12 vehículos no puede tener un *total* de 7 ni, al mismo tiempo, un *total* de 8 vehículos en que los ocupantes utilizan cinturón de seguridad), se determina la probabilidad de que en siete de ellos los ocupantes utilicen cinturón de seguridad; la probabilidad de que en ocho de los vehículos los ocupantes utilicen cinturones de seguridad y, así sucesivamente, hasta la probabilidad de que en todos los vehículos de la muestra los ocupantes utilicen cinturón de seguridad. La probabilidad de cada uno de estos resultados se suma enseguida.

$$\begin{aligned} P(x \geq 7) &= P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10) + P(x = 11) + P(x = 12) \\ &= 0.0902 + 0.1805 + 0.2569 + 0.2467 + 0.1436 + 0.0383 \\ &= 0.9562 \end{aligned}$$

De esta manera, la probabilidad de seleccionar 12 automóviles y hallar que los ocupantes de siete o más vehículos utilizan cinturón de seguridad es de 0.9562. Esta información se muestra en la siguiente hoja de cálculo de Excel. Existe una pequeña diferencia en la respuesta con software como consecuencia del redondeo. Los comandos de Excel son similares a los que se indican en los "Comandos de software" en el apéndice C.


Cumulative Binomial.xlsx

	A	B	C	D
1	Success	Probability		
2	0	0.0000		
3	1	0.0000		
4	2	0.0000		
5	3	0.0002		
6	4	0.0017		
7	5	0.0088		
8	6	0.0329		
9	7	0.0902		
10	8	0.1805		
11	9	0.2569		
12	10	0.2467		
13	11	0.1436		
14	12	0.0383		
15		0.9563		

AUTOEVALUACIÓN 6-4

Un estudio reciente reveló que 40% de las mujeres en el área metropolitana de San Diego que trabaja tiempo completo también hace trabajo voluntario en la comunidad. Suponga que se eligen al azar ocho mujeres del área de San Diego.

(a) ¿Cuáles son los valores para n y π?
(b) ¿Cuál es la probabilidad de que exactamente tres mujeres hagan trabajo voluntario?
(c) ¿Cuál es la probabilidad de que al menos una mujer haga trabajo voluntario?

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

19. En una distribución binomial, $n = 8$ y $\pi = 0.30$. Determine las probabilidades de los siguientes eventos.
- **a.** $x = 2$
- **b.** $x \leq 2$ (la probabilidad de que x sea igual o menor que 2).
- **c.** $x \geq 3$ (la probabilidad de que x sea igual o mayor que 3).

20. En una distribución binomial, $n = 12$ y $\pi = 0.60$. Determine las probabilidades de los siguientes eventos.
- **a.** $x = 5$
- **b.** $x \leq 5$
- **c.** $x \geq 5$

21. En un estudio reciente se descubrió que 90% de las familias de Estados Unidos tiene televisores de pantalla grande. Considere una muestra de nueve familias y determine la probabilidad de que:

a. Las nueve tengan televisores de pantalla grande.
b. Menos de cinco tengan televisores de pantalla grande.
c. Más de cinco tengan televisores de pantalla grande.
d. Al menos siete tengan televisores de pantalla grande.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

22. Un fabricante de marcos para ventanas sabe, por experiencia, que 5% de la producción tendrá algún tipo de defecto menor que requerirá reparación. Considere una muestra de 20 marcos y determine la probabilidad de que:
a. Ninguno requiera reparación.
b. Por lo menos uno requiera reparación.
c. Más de dos requieran reparación.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

23. La rapidez con que las compañías de servicios resuelven problemas es de suma importancia. Georgetown Telephone Company afirma que es capaz de resolver 70% de los problemas de los clientes el mismo día en que se reportan. Suponga que los 15 casos que se reportaron el día de hoy son representativos de todas las quejas.
a. ¿Cuántos problemas esperaría que se resolvieran este día? ¿Cuál es la desviación estándar?
b. ¿Cuál es la probabilidad de que 10 problemas se resuelvan hoy?
c. ¿Cuál es la probabilidad de que 10 u 11 problemas se resuelvan hoy?
d. ¿Cuál es la probabilidad de que más de 10 problemas se resuelvan hoy?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

24. Se afirma que 80% de los autos que se aproximan a una caseta individual de peaje en Nueva Jersey están equipados con un *transponder E-ZPass*. Encuentre la probabilidad de que en una muestra de seis autos:
a. Todos lo tengan.
b. Cuando menos tres lo tengan.
c. Ninguno lo tenga.

OA6-5
Explicar los supuestos de la distribución hipergeométrica y aplicarlos para calcular probabilidades.

Distribución de probabilidad hipergeométrica

Para aplicar una distribución binomial, la probabilidad de un éxito debe permanecer igual en cada ensayo. Por ejemplo, la probabilidad de adivinar la respuesta correcta a una pregunta de verdadero o falso es de 0.50. Esta es igual para cada pregunta de un examen. Asimismo, suponga que 40% de los electores registrados en un distrito electoral es republicano. Si se seleccionan al azar 27 de los votantes registrados, la probabilidad de elegir a un republicano en la primera selección es de 0.40. La posibilidad de elegir a un republicano en la siguiente selección es la misma, tomando en cuenta que el muestreo incluye *reemplazo,* lo cual significa que la persona elegida vuelve a la población antes de escoger a la que sigue.

No obstante, la mayor parte del muestreo se realiza *sin reemplazos.* Por lo tanto, si la población es pequeña, la probabilidad de cada observación cambia. Por ejemplo, si la población consta de 20 elementos, la probabilidad de seleccionar un elemento de ella es de 1/20. Si el muestreo se realiza sin reemplazos, únicamente quedan 19 elementos después de la primera selección; la probabilidad de elegir un elemento en la segunda selección es de solo 1/19. En la tercera selección, la probabilidad es de 1/18,y así sucesivamente. Esto supone que la población es **finita;** es decir, se conoce el número de sus elementos, el cual es relativamente reducido. Ejemplos de poblaciones finitas son los 2 842 republicanos de un distrito electoral, las 9 241 solicitudes para la escuela de medicina y las 18 Dakota 4×4 Crew Cab 2014 actualmente en existencia en Helfman Dodge Chrysler Jeep en Houston, Texas.

Recuerde que uno de los criterios relacionados con la distribución binomial estriba en que la probabilidad de éxito debe permanecer igual en todos los ensayos. Como la probabilidad de éxito no es la misma en todos estos cuando se realiza un muestreo sin reemplazo en una población relativamente pequeña, no debe aplicarse la distribución binomial. En lugar de esta se aplica la **distribución hipergeométrica.** Por lo tanto, 1) si se selecciona una muestra de una población finita sin reemplazo y 2) si el tamaño de la muestra n es mayor que 5% del tamaño de la población, se aplica la distribución hipergeométrica para determinar la probabilidad de un número específico de éxitos o fracasos. Esto resulta especialmente apropiado cuando el tamaño de la población es pequeño.

La fórmula de la distribución de probabilidad hipergeométrica es la siguiente:

DISTRIBUCIÓN HIPERGEOMÉTRICA

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n} \qquad \textbf{[6.6]}$$

donde:

N representa el tamaño de la población;
S es el número de éxitos en la población;
x es el número de éxitos en la muestra; este puede asumir los valores 0, 1, 2, 3…;
n es el tamaño de la muestra o el número de ensayos;
C es el símbolo de combinación.

En resumen, una distribución de probabilidad hipergeométrica tiene las siguientes características:

DISTRIBUCIÓN DE PROBABILIDAD HIPERGEOMÉTRICA

1. Los resultados de cada ensayo de un experimento se clasifican en dos categorías exclusivas: éxito o fracaso.
2. La variable aleatoria es la cantidad de éxitos de un número fijo de ensayos.
3. Los ensayos *no son independientes*.
4. Los muestreos se realizan con una población finita sin reemplazo y $n/N > 0.05$. Por lo tanto, la probabilidad de éxito cambia en cada ensayo.

En el siguiente ejemplo se ilustran los detalles para determinar una probabilidad con la distribución de probabilidad hipergeométrica.

EJEMPLO

Play Time Toys, Inc., tiene 50 empleados en el departamento de ensamblado. Solo 40 de ellos pertenecen al sindicato. Se eligen al azar cinco empleados para formar un comité que hablará con la empresa sobre los horarios de inicio de los turnos. ¿Cuál es la probabilidad de que cuatro de los cinco empleados elegidos para formar parte del comité pertenezcan al sindicato?

SOLUCIÓN

En este caso, la población consiste en los 50 empleados del departamento de ensamblado. Solo se puede elegir una vez a un empleado para formar parte del comité. De ahí que el muestreo se lleve a cabo sin reemplazo. Por lo tanto, en cada ensayo cambia la probabilidad de elegir a un empleado sindicalizado. La distribución hipergeométrica es adecuada para determinar la probabilidad. En este problema,

N es igual a 50 (el número de empleados);
S es igual a 40 (el número de empleados sindicalizados);
x es igual a 4 (el número de empleados sindicalizados elegidos);
n es igual a 5 (el número de empleados elegidos).

Se desea calcular la probabilidad de que cuatro de los cinco miembros del comité sean sindicalizados. Al sustituir estos valores en la fórmula [6.6] se obtiene:

$$P(4) = \frac{({}_{40}C_4)({}_{50-40}C_{5-4})}{{}_{50}C_5} = \frac{(91\ 390)(10)}{2\ 118\ 760} = 0.431$$

Por consiguiente, la probabilidad de elegir al azar a cinco trabajadores de ensamblado de los 50 trabajadores y encontrar que cuatro de ellos son sindicalizados es de 0.431.

En la tabla 6.4 se muestran las probabilidades hipergeométricas de encontrar 0, 1, 2, 3, 4 y 5 empleados sindicalizados en el comité.

Con el fin de comparar ambas distribuciones de probabilidad, en la tabla 6.5 (página siguiente) se muestran las probabilidades hipergeométricas y binomiales del ejemplo de Play Time Toys, Inc. Como 40 de los 50 empleados del departamento de ensamblado están sindicalizados, se establece que $\pi = 0.80$ para la distribución binomial. Las probabilidades binomiales de la tabla 6.5 provienen de la distribución binomial con $n = 5$ y $\pi = 0.80$.

TABLA 6.4 Probabilidades hipergeométricas ($n = 5$, $N = 50$ y $S = 40$) del número de empleados sindicalizados en el comité

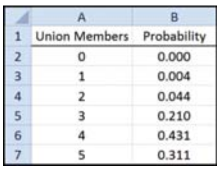
Miembros del sindicato	Probabilidad
0	0.000
1	0.004
2	0.044
3	0.210
4	0.431
5	0.311
	1.000

TABLA 6.5 Probabilidades hipergeométrica y binomial del departamento de ensamble de Play Time Toys, Inc.

Número de miembros sindicalizados en el comité	Probabilidad hipergeométrica, $P(x)$	Probabilidad binomial ($n = 5$ y $\pi = 0.80$)
0	0.000	0.000
1	0.004	0.006
2	0.044	0.051
3	0.210	0.205
4	0.431	0.410
5	0.311	0.328
	1.000	1.000

Cuando no es posible satisfacer alguno de los requisitos binomiales de una probabilidad constante de éxito, se debe recurrir a la distribución de probabilidad hipergeométrica. No obstante, según se indica en la tabla 6.5, es posible, en ciertas condiciones, emplear los resultados de la distribución binomial para calcular la distribución hipergeométrica. Esto conduce a la siguiente regla de oro:

Si los elementos seleccionados no se regresan a la población, se puede aplicar la distribución binomial para calcular la distribución hipergeométrica cuando $n < 0.05N$. Es decir, basta la distribución binomial si el tamaño de la muestra es menor que 5% de la población.



	A	B
1	Union Members	Probability
2	0	0.000
3	1	0.004
4	2	0.044
5	3	0.210
6	4	0.431
7	5	0.311

En Excel es posible generar una distribución hipergeométrica. Observe la pantalla a la izquierda. Los pasos pertinentes se incluyen en la sección "Comandos de software" en el apéndice C al final del libro.

AUTOEVALUACIÓN 6-5

Horwege Discount Brokers hace planes para contratar este año a cinco analistas financieros. Hay un grupo de 12 candidatos aprobados, y George Horwege, el propietario, decide elegir al azar a quienes va a contratar. De los solicitantes aprobados, ocho son hombres y cuatro son mujeres. ¿Cuál es la probabilidad de que tres de los cinco contratados sean hombres?

EJERCICIOS

25. Una población consta de 10 elementos, seis de los cuales se encuentran defectuosos. En una muestra de tres elementos, ¿cuál es la probabilidad de que exactamente dos sean defectuosos? Suponga que las muestras se toman sin reemplazo.

26. Una población consta de 15 elementos, cuatro de los cuales son aceptables. En una muestra de cuatro elementos, ¿cuál es la probabilidad de que exactamente tres sean aceptables? Suponga que las muestras se toman sin reemplazo.

27. La sucursal del Banco Nacional de Wyoming en Riverton tiene 10 préstamos hipotecarios mayores a un millón de dólares. De estos 10 préstamos, tres son "por debajo del agua". Un préstamo es "por debajo del agua" si la cantidad prestada es mayor al valor de la propiedad. El ejecutivo de préstamos decidió seleccionar al azar dos de estos préstamos para determinar si se ajustaban a los estándares bancarios. ¿Cuál es la probabilidad de que ninguno de los préstamos seleccionados sea "por debajo del agua"?

28. El departamento de sistemas de computación cuenta con ocho profesores, de los cuales seis son titulares. La doctora Vonder, directora, desea formar un comité de tres profesores del departamento para que revisen el plan de estudios. Si selecciona el comité al azar:
  a. ¿Cuál es la probabilidad de que todos los miembros del comité sean titulares?
  b. ¿Cuál es la probabilidad de que por lo menos un miembro del comité no sea titular? (*sugerencia*: aplique la regla del complemento para responder esta pregunta).

29. Keith's Florists tiene 15 camiones de entrega, cuya labor principal es entregar flores y arreglos florales en la zona de Greenville, Carolina del Sur. De todos sus camiones, seis presentan problemas con los frenos. En forma aleatoria se seleccionó una muestra de cinco camiones. ¿Cuál es la probabilidad de que dos de los camiones probados presenten frenos defectuosos?

30. El juego de Lotto, patrocinado por la Comisión de la Lotería de Louisiana, otorga el premio mayor a un concursante que haga coincidir seis de los posibles números. Suponga que hay 40 pelotas de ping-pong numeradas del 1 al 40. Cada número aparece una sola vez y las pelotas ganadoras se seleccionan sin reemplazo.

a. La comisión informa que la probabilidad de que coincidan todos los números es de 1 en 3 838 380. ¿Qué significa esto en términos de probabilidad?
b. Aplique la fórmula de la distribución de probabilidad hipergeométrica para determinar esta probabilidad: la comisión de la lotería también otorga un premio si un concursante hace coincidir cuatro o cinco de los seis números ganadores. *Sugerencia*: divida los 40 números en dos grupos (ganadores y no ganadores).
c. Calcule la probabilidad, de nuevo con la fórmula de la distribución de probabilidad hipergeométrica, para hacer coincidir cuatro de los seis números ganadores.
d. Calcule la probabilidad de que coincidan cinco de los seis números ganadores.

Distribución de probabilidad de Poisson

OA6-6
Explicar los supuestos de la distribución de Poisson y aplicarlos para calcular probabilidades.

La **distribución de probabilidad de Poisson** describe el número de veces que se presenta un evento durante un intervalo específico; el cual puede ser de tiempo, distancia, área o volumen.

La distribución se basa en dos supuestos: el primero consiste en que la probabilidad es proporcional a la longitud del intervalo; el segundo, en que los intervalos son independientes. En otras palabras, cuanto más grande sea el intervalo, mayor será la probabilidad; además, el número de veces que se presenta un evento en un intervalo no influye en el resto de los intervalos. La distribución también constituye una forma restrictiva de la distribución binomial cuando la probabilidad de un éxito es muy pequeña y n es grande. En general, a esta se le conoce con el nombre de *ley de eventos improbables,* lo cual significa que la probabilidad, p, de que ocurra un evento en particular es muy pequeña. La distribución de Poisson es una distribución de probabilidad discreta porque se genera contando.

La distribución de probabilidad de Poisson posee tres características:

EXPERIMENTO DE PROBABILIDAD DE POISSON

1. La variable aleatoria es el número de veces que ocurre un evento durante un intervalo definido.
2. La probabilidad de que ocurra el evento es proporcional al tamaño del intervalo.
3. Los intervalos no se superponen y son independientes.

Esta distribución de probabilidad posee muchas aplicaciones. Se le utiliza como modelo para describir la distribución de errores en una entrada de datos, el número de rayones y otras imperfecciones en las cabinas de automóviles recién pintados, el número de partes defectuosas en envíos, el número de clientes que esperan mesa en un restaurante o que esperan entrar en una de las atracciones de Disney World, y el número de accidentes en la carretera federal I-75 en un periodo de tres meses.

En la siguiente fórmula se describe matemáticamente la distribución de Poisson:

DISTRIBUCIÓN DE POISSON $$P(x) = \frac{\mu^x e^{-\mu}}{x!} \qquad \textbf{[6.7]}$$

donde:

μ (mu) es la media de la cantidad de veces (éxitos) que se presenta un evento en un intervalo particular;
e es la constante 2.71828 (base del sistema de logaritmos neperianos);
x es el número de veces que se presenta un evento;
$P(x)$ es la probabilidad de un valor específico de x.

La media de número de éxitos, μ, puede determinarse con $n\pi$; en este caso, n es el número total de ensayos, y π, la probabilidad de éxito.

MEDIA DE UNA DISTRIBUCIÓN DE POISSON $$\mu = n\pi \qquad \textbf{[6.8]}$$

La varianza de Poisson también es igual a su media. Por ejemplo, la probabilidad de que un cheque cobrado en un banco sea rechazado es de 0.0003, y se cobran 10 000 cheques; así, la media y la

ESTADÍSTICA EN ACCIÓN

Cerca del final de la Segunda Guerra Mundial, los alemanes crearon bombas propulsadas por cohetes que lanzaron hacia la ciudad de Londres. El comando militar aliado no sabía si estas se lanzaban de forma aleatoria o si tenían un objetivo. Con el fin de averiguarlo, se dividió la ciudad de Londres en 586 regiones cuadradas y se registró la distribución de los bombarderos en cada región cuadrada de la siguiente manera:

Bombardeos	0	1	2	3	4	5
Regiones	229	221	93	35	7	1

Con el fin de interpretar estos datos, en la tabla se señala que 229 regiones no fueron bombardeadas. Siete regiones fueron atacadas cuatro veces. De acuerdo con la distribución de Poisson, con una media de 0.93 bombardeos por

(*continúa*)

(*continuación*)
región, se obtiene la siguiente cantidad esperada de bombardeos:

Bombardeos					
0	1	2	3	4	5 o más
Regiones					
231.2	215.0	100.0	31.0	7.2	1.6

Puesto que la cantidad real de bombardeos se aproxima a la cantidad esperada, el comando militar concluyó que las bombas caían de forma aleatoria. Los alemanes no habían creado una bomba con un dispositivo para dar en un blanco determinado.

varianza del número de cheques rebotados es de 3.0, que se determina mediante la operación $\mu = n\pi = 10\,000(0.0003) = 3.0$.

Recuerde que, en el caso de una distribución binomial, existe una cantidad fija de ensayos. Por ejemplo, en una prueba de opción múltiple de cuatro preguntas, solo puede haber cero, uno, dos, tres o cuatro éxitos (respuestas correctas). Sin embargo, la variable aleatoria, *x,* en el caso de una distribución de Poisson puede adoptar una *infinidad de valores;* es decir, 0, 1, 2, 3, 4, 5, etcétera. No obstante, *las probabilidades se tornan muy bajas después de las primeras veces que se presenta un evento* (éxitos).

Para ejemplificar el cálculo de la distribución de Poisson, suponga que pocas veces se pierde equipaje en Northeast Airlines. En la mayoría de los vuelos no se pierden maletas; en algunos se pierde una; en unos cuantos se pierden dos; pocas veces se pierden tres, etc. Suponga que una muestra aleatoria de 1 000 vuelos arroja un total de 300 maletas perdidas. De esta manera, la media aritmética del número de maletas perdidas por vuelo es de 0.3, que se calcula al dividir 300 entre 1 000. Si el número de maletas perdidas por vuelo se rige por una distribución de Poisson con $\mu = 0.3$, las diversas probabilidades se calculan con la fórmula [6.7]:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

Por ejemplo, la probabilidad de que no se pierda ninguna maleta es la siguiente:

$$P(0) = \frac{(0.3)^0(e^{-0.3})}{0!} = 0.7408$$

En otras palabras, en 74% de los vuelos no habrá maletas perdidas. La probabilidad de que se pierda exactamente una maleta es:

$$P(1) = \frac{(0.3)^1(e^{-0.3})}{1!} = 0.2222$$

Por consiguiente, se espera que se pierda exactamente una maleta en 22% de los vuelos.

Las probabilidades de Poisson también se pueden consultar en el apéndice B.2.

EJEMPLO

De acuerdo con el ejemplo anterior, el número de maletas se rige por una distribución de Poisson con una media de 0.3. Consulte el apéndice B.2 para determinar la probabilidad de que ninguna maleta se pierda en un vuelo. ¿Cuál es la probabilidad de que se pierda exactamente una maleta en un vuelo? ¿En qué momento debe sospechar el supervisor que en un vuelo se están perdiendo demasiadas maletas?.

SOLUCIÓN

Parte del apéndice B.2 se reproduce en la tabla 6.6. Para determinar la probabilidad de que ninguna maleta se pierda, se localiza la columna con el encabezado "0.3" y se desciende por dicha columna hasta el renglón señalado con "0". La probabilidad de que no haya maletas perdidas es de 0.7408. La probabilidad de que se pierda una maleta es 0.2222, y se muestra en el siguiente renglón de la tabla, en la misma columna. La probabilidad de que se pierdan dos maletas es de 0.0333, renglón

TABLA 6.6 Tabla de Poisson para diversos valores de μ (del apéndice B.5)

	μ								
x	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

inferior; en el caso de tres maletas perdidas, la probabilidad es de 0.0033; y en el de cuatro maletas perdidas, de 0.0003. Por consiguiente, a un supervisor no debería sorprenderle que se pierda una maleta, pero debería esperar ver con menos frecuencia más de una maleta perdida.

Estas probabilidades también se determinan con Excel. Los comandos para calcular las probabilidades de Poisson están en el apéndice C.

Poisson Probability Distribution.xlsx

	A	B
1	Success	Probability
2	0	0.7408
3	1	0.2222
4	2	0.0333
5	3	0.0033
6	4	0.0003
7	5	0.0000
8	6	0.0000
9	7	0.0000

Se ha mencionado que la distribución de probabilidad de Poisson constituye una forma restrictiva de la distribución binomial. Es decir, se puede calcular una probabilidad binomial con la de Poisson.

La distribución de probabilidad de Poisson se caracteriza por el número de veces que se presenta un evento durante un intervalo. Algunos ejemplos son:

- El número de palabras mal escritas por página en un periódico.
- La cantidad de llamadas por hora que recibe Dyson Vacuum Cleaner Company.
- El número de vehículos que vende por día Hyatt Buick GMC, en Durham, Carolina del Norte.
- La cantidad de anotaciones en un encuentro de futbol colegial.

En cada uno de estos ejemplos existe algún tipo de intervalo: palabras mal escritas por página, llamadas por hora, vehículos vendidos por día o anotaciones por partido.

En el ejemplo anterior (cantidad de maletas perdidas en cada vuelo), el intervalo es un "vuelo". Se conocía la media del número de maletas perdidas por vuelo, pero no el número de pasajeros ni la probabilidad de que se perdiera una maleta. Se sospechó que el número de pasajeros era lo bastante grande y que la probabilidad de que un pasajero perdiera su maleta era baja. En el ejemplo siguiente se aplicó la distribución de Poisson para calcular una probabilidad binomial cuando n (el número de ensayos) es grande y π (la probabilidad de un éxito), pequeña.

EJEMPLO

Coastal Insurance Company asegura propiedades frente a la playa a lo largo de Virginia, Carolina del Norte y del Sur, y las costas de Georgia; el cálculo aproximado es que, cualquier año, la probabilidad de que un huracán de categoría III (vientos sostenidos de más de 110 millas por hora) o mayor azote una región de la costa (la isla de St. Simons, Georgia, por ejemplo) es de 0.05. Si un dueño de casa obtiene un crédito hipotecario de 30 años por una propiedad recién comprada en St. Simons, ¿cuáles son las posibilidades de que experimente por lo menos un huracán durante el periodo del crédito?

SOLUCIÓN

Para aplicar la distribución de probabilidad de Poisson se comienza por determinar la media o número esperado de tormentas que se ajustan al criterio y que azotan St. Simons durante el periodo de 30 años. Es decir,

$$\mu = n\pi = 30(0.05)) = 1.5$$

donde:

n es el número de años, 30 en este caso;
π es la probabilidad de que toque tierra un huracán que se ajuste al criterio;
μ es la media o número esperado de tormentas en un periodo de 30 años.

Para determinar la probabilidad de que por lo menos una tormenta azote la isla de St. Simons, Georgia, primero calcule la probabilidad de que ninguna tormenta llegue a la costa y reste dicho valor de 1.

$$P(x \geq 1) = 1 - P(x = 0) = 1 - \frac{\mu^0 e^{-1.5}}{0!} = 1 - 0.2231 = 0.7769$$

Se concluye que las posibilidades de que un huracán de ese tipo azote la propiedad frente a la playa en St. Simons, durante el periodo de 30 años, mientras el crédito se encuentra vigente, son de 0.7769. En otras palabras, la probabilidad de que St. Simons sufra el azote de un huracán categoría III o mayor durante el periodo de 30 años es de un poco más de 75%.

Se debe insistir en que el intervalo, como antes se explicó, aún existe. Es decir, se espera que haya 1.5 tormentas en esa costa cada periodo de 30 años (el cual, es el intervalo).

En el caso anterior se utilizó la distribución de Poisson como aproximación de la binomial. Observe que se cumplieron las condiciones binomiales antes mencionadas.

- Solo hay dos posibles resultados: un huracán azota el área de St. Simons o no lo hace.
- Hay una cantidad fija de ensayos, en este caso, 30 años.
- Existe una probabilidad constante de éxito; es decir, la probabilidad de que un huracán azote la zona es de 0.05 por año.
- Los años son independientes. Es decir, el que una tormenta importante azota en el quinto año no influye en ningún otro año.

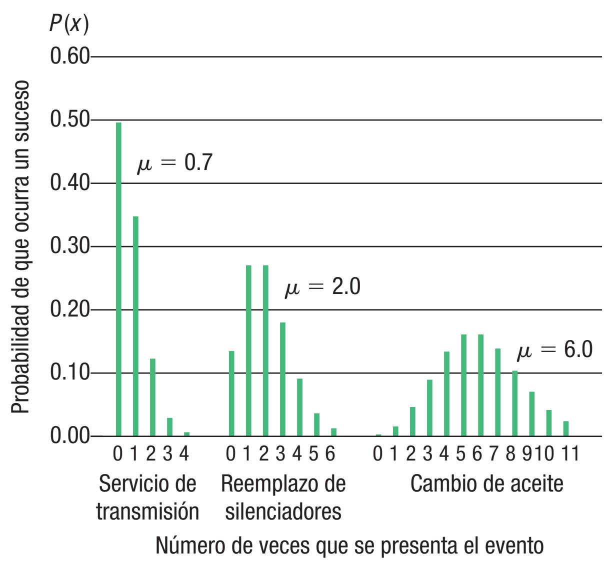
Para calcular la probabilidad de que por lo menos una tormenta azote el área en un periodo de 30 años aplique la distribución binomial:

$$P(x \geq 1) = 1 - P(x = 0) = 1 - [_{30}C_0(0.05)^0(0.95)^{30}] = 1 - [(1)(1)(0.2146)] = 0.7854$$

La probabilidad de que por lo menos un huracán azote el área de St. Simons durante el periodo de 30 años con la distribución binomial es de 0.7854.

¿Cuál es la respuesta correcta? ¿Por qué considerar el problema desde ambos puntos de vista? La respuesta que se obtiene con la distribución binomial es la más “técnicamente correcta”. La que se obtuvo con la distribución de Poisson puede tomarse como una aproximación de la binomial cuando n (el número de ensayos) es grande, y π (la probabilidad de un éxito), pequeña. Considere el problema desde ambas distribuciones para destacar la convergencia de las dos distribuciones discretas. En ocasiones, la aplicación de la distribución de Poisson permite una solución más rápida y, como puede ver, hay poca diferencia entre las respuestas. De hecho, conforme n aumenta y π disminuye, se reducen las diferencias entre ambas distribuciones.

La distribución de probabilidad de Poisson siempre tiene un sesgo positivo, y la variable aleatoria no posee límite superior específico. La distribución de Poisson en el caso de las maletas perdidas ($\mu = 0.3$) está muy sesgada. Conforme μ se incrementa, la distribución de Poisson adquiere más simetría. Por ejemplo, en la gráfica 6.4 se muestran las distribuciones del número de servicios



GRÁFICA 6.4 Distribuciones de probabilidad de Poisson con medias de 0.7, 2.0 y 6.0

de transmisión, reemplazos de silenciadores y cambios de aceite al día en Avellino's Auto Shop. Estas se ajustan a las distribuciones de Poisson con medias de 0.7, 2.0 y 6.0, respectivamente.

En resumen, la distribución de Poisson es en realidad una familia de distribuciones discretas, y todo lo que se requiere para construirla es la media del número de defectos, errores, etcétera, que se designan con μ.

AUTOEVALUACIÓN 6-6

A partir de las tablas actuariales, Washington Insurance Company determinó que la probabilidad de que un hombre de 25 años muera en el transcurso del próximo año es de 0.0002. Si la empresa vende de 4 000 pólizas a hombres de esa edad durante este año, ¿cuál es la probabilidad de que estos paguen exactamente una póliza?

EJERCICIOS

31. En una distribución de Poisson de $\mu = 0.4$.
  a. ¿Cuál es la probabilidad de que $x = 0$?
  b. ¿Cuál es la probabilidad de que $x > 0$?

32. En una distribución de Poisson de $\mu = 4$.
  a. ¿Cuál es la probabilidad de que $x = 2$?
  b. ¿Cuál es la probabilidad de que $x \leq 2$?
  c. ¿Cuál es la probabilidad de que $x > 2$?

33. La señorita Bergen es ejecutiva del Coastal Bank and Trust. A partir de sus años de experiencia, calcula que la probabilidad de que un solicitante no pague un préstamo inicial es de 0.025. El mes anterior realizó 40 préstamos.
  a. ¿Cuál es la probabilidad de que no se paguen tres préstamos?
  b. ¿Cuál es la probabilidad de que por lo menos no se paguen tres préstamos?

34. Un promedio de dos automóviles por minuto llegan a la salida de Elkhart de la autopista de Indiana. La distribución de llegadas se aproxima a una distribución de Poisson.
  a. ¿Cuál es la probabilidad de que ningún automóvil llegue en un minuto?
  b. ¿Cuál es la probabilidad de que por lo menos llegue un automóvil en un minuto?

35. Se calcula que 0.5% de quienes se comunican al departamento de servicio al cliente de Dell, Inc., escuchará un tono de línea ocupada. ¿Cuál es la probabilidad de que de las 1 200 personas que se comunicaron hoy, por lo menos cinco hayan escuchado un tono de línea ocupada?

36. En el pasado, las escuelas del condado de Los Ángeles cerraron un promedio de tres días cada año por emergencias climáticas. ¿Cuál es la probabilidad de que las escuelas de dicho condado cierren cuatro días el próximo año?

RESUMEN DEL CAPÍTULO

I. Una variable aleatoria es un valor numérico determinado por el resultado de un experimento.

II. Una distribución de probabilidad es una lista de posibles resultados de un experimento y la probabilidad asociada con cada resultado.
  A. Una distribución de probabilidad discreta solo puede adoptar ciertos valores. Las principales características son:
    1. La suma de las probabilidades es 1.00.
    2. La probabilidad de un resultado se encuentra entre 0.00 y 1.00.
    3. Los resultados son mutuamente excluyentes.
  B. Una distribución continua puede adoptar una infinidad de valores dentro de un rango específico.

III. La media y la varianza de una distribución de probabilidad se calculan de la siguiente manera:
  A. La media es igual a:

$$\mu = \Sigma[xP(x)] \qquad \textbf{[6.1]}$$

  B. La varianza es igual a:

$$\sigma^2 = \Sigma[(x - \mu)^2 P(x)] \qquad \textbf{[6.2]}$$

IV. La distribución binomial posee las siguientes características:
  A. Cada resultado se clasifica en una de dos categorías mutuamente excluyentes.
  B. La distribución es resultado de la cuenta del número de éxitos en una cantidad fija de ensayos.
  C. La probabilidad de un éxito es la misma de un ensayo al siguiente.
  D. Cada ensayo es independiente.

E. Una probabilidad binomial se determina de la siguiente manera:

$$P(x) = {}_nC_x \pi^x(1 - \pi)^{n-x} \quad \textbf{[6.3]}$$

F. La media se calcula de la siguiente manera:

$$\mu = n\pi \quad \textbf{[6.4]}$$

G. La varianza es

$$\sigma^2 = n\pi(1 - \pi) \quad \textbf{[6.5]}$$

V. La distribución hipergeométrica posee las siguientes características:

A. Solo hay dos posibles resultados.
B. La probabilidad de un éxito no es la misma en cada ensayo.
C. La distribución resulta de contar el número de éxitos en una cantidad fija de ensayos.
D. Se le utiliza cuando se toman muestras sin reemplazo de una población finita.
E. Una probabilidad hipergeométrica se calcula a partir de la siguiente ecuación:

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n} \quad \textbf{[6.6]}$$

VI. La distribución de Poisson posee las siguientes características:

A. Describe el número de veces que se presenta un evento en un intervalo específico.
B. La probabilidad de un "éxito" es proporcional a la longitud del intervalo.
C. Los intervalos que no se superponen son independientes.
D. Es una forma restrictiva de la distribución binomial, en la que n es grande y π es pequeña.
E. La probabilidad de Poisson se determina a partir de la siguiente ecuación:

$$\mu = n\pi \quad \textbf{[6.7]}$$

F. La media y la varianza son:

$$\sigma^2 = n\pi \quad \textbf{[6.8]}$$

EJERCICIOS DEL CAPÍTULO

37. ¿Cuál es la diferencia entre una variable aleatoria y una distribución de probabilidad?

38. Indique si la variable aleatoria en cada uno de los siguientes enunciados es discreta o continua.

a. El tiempo de espera para un corte de cabello.
b. El total de automóviles que rebasa un corredor cada mañana.
c. El número de hits de un equipo femenil de softbol de preparatoria.
d. La cantidad de pacientes atendidos en el South Strand Medical Center entre las seis y diez de la noche, cada noche.
e. La distancia que recorrió en su automóvil con el último tanque de gasolina.
f. El número de clientes del Wendy's de Oak Street que utilizaron las instalaciones.
g. La distancia entre Gainesville, Florida, y todas las ciudades de Florida con una población de por lo menos 50 000 habitantes.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

39. Una inversión producirá 1 000, 2 000 y 5 000 dólares a fin de año. Las probabilidades de estos valores son de 0.25, 0.60 y 0.15, respectivamente. Determine la media y la varianza del valor de la inversión.

40. El siguiente anuncio se colocó en la tienda de golf en el campo de Myrtle Beach, Carolina del Sur.

A los miembros del Club de Golf Blackmoor

La tienda de golf está organizando la rifa de un palo
TaylorMade R9 10.5° (con valor de 300 dólares).
Cada boleto cuesta cinco dólares.
Solo se venderán 80 boletos.
¡Venga a la tienda de golf por los suyos!

John Underpar compra un boleto.

a. ¿Cuáles son los posibles resultados monetarios para el señor Underpar?
b. ¿Cuáles son las probabilidades de los posibles resultados?
c. Resuma el "experimento" del señor Underpar como una distribución de probabilidad.
d. ¿Cuál es la media o el valor esperado de la distribución de probabilidad? Explique su resultado.
e. Si se venden los 80 boletos, ¿cuál es la ganancia esperada para el Club?

41. Croissant Bakery, Inc., ofrece pasteles con decorados especiales para cumpleaños, bodas y otras ocasiones. La pastelería también tiene pasteles normales. En la siguiente tabla se incluye el número total de pasteles vendidos por día, así como la probabilidad correspondiente. Calcule la media, la varianza y la desviación estándar del número de pasteles vendidos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Cantidad de pasteles vendidos en un día	Probabilidad
12	0.25
13	0.40
14	0.25
15	0.10

42. Abajo se muestran los premios de la lotería Powerball y sus correspondientes pronósticos y probabilidades de ocurrencia. El precio del boleto es de un dólar. Encuentre la media y la desviación estándar del premio. *Sugerencia*: no olvide incluir el costo del boleto y su correspondiente probabilidad.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Divisiones	Premios	Pronósticos	Probabilidad
Five plus Powerball	$50 000 000	146 107 962	0.000000006844
Match 5	200 000	3 563 609	0.000000280614
Four plus Powerball	10 000	584 432	0.000001711060
Match 4	100	14 255	0.000070145903
Three plus Powerball	100	11 927	0.000083836351
Match 3	7	291	0.003424657534
Two plus Powerball	7	745	0.001340482574
One plus Powerball	4	127	0.007812500000
Zero plus Powerball	3	69	0.014285714286

43. En una encuesta reciente, 35% indicó que el helado de chocolate era su favorito. Suponga que se selecciona una muestra de diez personas para preguntarles cuál es su sabor favorito de helado.
a. ¿Cuántas personas de la muestra esperaría usted que mencionen *chocolate*?
b. ¿Cuál es la probabilidad de que exactamente cuatro personas incluidas en la muestra digan *chocolate*?
c. ¿Cuál es la probabilidad de que cuatro o más respondan *chocolate*?

44. Suponga una comunidad del suroeste de Estados Unidos con 30% de habitantes hispanoparlantes. Se acusó a uno de ellos de asesinar a un estadounidense que no hablaba español. De los primeros 12 posibles jurados, solo dos son estadounidenses hispanoparlantes y 10 no lo son. El abogado de la defensa se opone a la elección del jurado, pues dice que habrá prejuicio contra su cliente. El fiscal no está de acuerdo y arguye que la probabilidad de esta composición del jurado es frecuente. Calcule la probabilidad y explique los supuestos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

45. Un auditor de Health Maintenance Services of Georgia informa que 40% de los asegurados de 55 años de edad o mayores utiliza la póliza durante el año. Se seleccionan al azar 15 asegurados de los registros de la compañía.
a. ¿Cuántos asegurados cree que utilizaron la póliza el año anterior?
b. ¿Cuál es la probabilidad de que 10 de los asegurados seleccionados hayan utilizado la póliza el año previo?
c. ¿Cuál es la probabilidad de que 10 o más de los asegurados seleccionados hayan utilizado la póliza el año anterior?
d. ¿Cuál es la probabilidad de que más de 10 de los asegurados seleccionados hayan utilizado la póliza el año previo?

46. Tire and Auto Supply contempla hacer una división de 2 a 1 de las acciones. Antes de realizar la transacción, por lo menos dos terceras partes de los 1 200 accionistas de la compañía deben aprobar la oferta. Para evaluar la probabilidad de que la oferta se apruebe, el director de finanzas eligió una muestra de 18 accionistas. Contactó a cada uno y comprobó que 14 aprobaron la propuesta. ¿Cuál es la posibilidad de este evento, si dos terceras partes de los accionistas dan su aprobación?

47. Un estudio federal informó que 7.5% de la fuerza laboral de Estados Unidos tiene problemas con las drogas. Una oficial antidrogas del estado de Indiana decidió investigar esta afirmación y seleccionó una muestra de 20 trabajadores:
a. ¿Cuántos cree que presenten problemas de adicción? ¿Cuál es la desviación estándar?
b. ¿Cuál es la probabilidad de que *ninguno* de los trabajadores de la muestra manifieste problemas de adicción?

c. ¿Cuál es la probabilidad de que *por* lo *menos uno* de los trabajadores de la muestra presente tales problemas?

48. Un banco de Hawái informa que 7% de sus clientes con tarjeta de crédito dejará de pagar en algún momento. La sucursal de Hilo envió el día de hoy 12 nuevas tarjetas.
   a. ¿Cuántos de los nuevos tarjetahabientes cree que dejarán de pagar? ¿Cuál es la desviación estándar?
   b. ¿Cuál es la probabilidad de que *ninguno* de los tarjetahabientes deje de pagar?
   c. ¿Cuál es la probabilidad de que *por lo menos uno* deje de pagar?

49. Estadísticas recientes sugieren que 15% de las personas que visitan un sitio de ventas de menudeo en la web compra algo. Un minorista desea verificar esta afirmación. Para hacerlo, seleccionó una muestra de 16 "visitas" de su sitio y descubrió que cuatro compraron algo.
   a. ¿Cuál es la probabilidad de que exactamente cuatro realicen una compra?
   b. ¿Cuántas compras deben esperarse?
   c. ¿Cuál es la probabilidad de que cuatro o más "visitas" terminen en compra?

50. En el capítulo 19 se estudia la *muestra de aceptación.* El muestreo de aceptación se utiliza para supervisar la calidad de la materia prima que entra. Suponga que un comprador de componentes electrónicos permite que 1% de los componentes se encuentren defectuosos. Para garantizar la calidad de las partes que entran, por lo general se toman 20 partes como muestra y se permite una parte defectuosa.
   a. ¿Cuál es la probabilidad de aceptar un lote con 1% de partes defectuosas?
   b. Si la calidad del lote que ingresa en realidad fue de 2%, ¿cuál es la probabilidad de que se acepte?
   c. Si la calidad del lote que ingresa en realidad fue de 5%, ¿cuál es la probabilidad de que se acepte?

51. Recientemente, Unilever, Inc., desarrolló un nuevo jabón líquido para el cuerpo con aroma a jengibre. Su investigación muestra que 30% de los hombres aprobó la nueva fragancia. Para ahondar en el estudio, el grupo de investigación de mercado de Unilever seleccionó aleatoriamente a 15 hombres y les preguntó si les gustaba la fragancia. ¿Cuál es la probabilidad de que a seis o más hombres les guste el aroma de jengibre en un jabón líquido para cuerpo?

52. La doctora Richmond, psicóloga, estudia el hábito de ver televisión durante el día entre estudiantes de preparatoria. Ella cree que 45% de los estudiantes de preparatoria ve telenovelas por la tarde. Para investigar un poco más, elige una muestra de 10.
   a. Elabore una distribución de probabilidad del número de estudiantes de la muestra que ven telenovelas.
   b. Determine la media y la desviación estándar de esta distribución.
   c. ¿Cuál es la probabilidad de encontrar que exactamente cuatro ven telenovelas?
   d. ¿Cuál es la probabilidad de que menos de la mitad de los estudiantes elegidos vean telenovelas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

53. Un estudio reciente llevado a cabo por Penn, Shone, and Borland para LastMinute.com reveló que 52% de quienes viajan por negocios planea sus viajes menos de dos semanas antes de partir. El estudio se va a repetir en un área que abarca tres estados con una muestra de 12 ejecutivos que viajan por negocios de manera frecuente.
   a. Elabore una distribución de probabilidad del número de ejecutivos que planean sus viajes a dos semanas de partir.
   b. Determine la media y la desviación estándar de esta distribución.
   c. ¿Cuál es la probabilidad de que exactamente cinco de los 12 ejecutivos planeen sus viajes dos semanas antes de partir?
   d. ¿Cuál es la probabilidad de que cinco o más de los 12 ejecutivos seleccionados planeen sus viajes dos semanas antes de partir?

54. El IRS estudia la categoría de las contribuciones para la beneficencia. Se seleccionó una muestra de 25 declaraciones de parejas jóvenes de entre 20 y 35 años de edad con un ingreso bruto de más de 100 000 dólares. De estas 25 declaraciones, cinco incluían contribuciones de beneficencia de más de 1 000 dólares. Suponga que cuatro de ellas se seleccionan para practicarles una auditoría completa.
   a. Explique por qué resulta adecuada la distribución hipergeométrica.
   b. ¿Cuál es la probabilidad de que exactamente una de las cuatro declaraciones auditadas tuviera deducciones de beneficencia de más de 1 000 dólares?
   c. ¿Cuál es la probabilidad de que por lo menos una de las cuatro declaraciones auditadas tuvieran deducciones de beneficencia de más de 1 000 dólares?

55. El despacho de abogados Hagel and Hagel se localiza en el centro de Cincinnati. La empresa tiene 10 socios; 7 viven en Ohio y 3 en el norte de Kentucky. La señora Wendy Hagel, gerente, desea nombrar un comité de tres socios que estudien la posibilidad de mudar el despacho al norte de Kentucky. El comité se selecciona al azar de entre los socios. Establezca la probabilidad de que:
   a. Un miembro del comité viva en el norte de Kentucky y los otros en Ohio.
   b. Por lo menos un miembro del comité viva en el norte de Kentucky.

56. Topten es una fuente líder de productos ecológicos eficientes. Su lista de los siete principales vehículos en términos de rendimiento de combustible para 2014 incluye a tres autos Honda.

a. Determine la distribución de probabilidad del número de autos Honda en una muestra de dos autos elegidos entre los siete más económicos.

b. ¿Cuál es la posibilidad de que en la muestra de dos haya por lo menos un Honda?

57. El cargo de jefe de la policía en la ciudad de Corry, Pennsylvania, se encuentra vacante. Un comité de búsqueda, integrado por los residentes de esa población tiene la responsabilidad de recomendarle al alcalde de la ciudad a alguien para el puesto. Hay 12 candidatos, cuatro de los cuales son mujeres o miembros de una minoría. El comité decide entrevistar a los 12 candidatos. Primero seleccionaron al azar a cuatro candidatos para entrevistarlos el primer día, ninguno de los cuales resultó ser mujer ni miembro de una minoría. En una de sus columnas editoriales, el periódico local *Corry Press* sugiere que hay discriminación. ¿Cuál es la probabilidad de que así sea?

58. En la lista siguiente se muestra la población por estado de los 15 con mayor cantidad de habitantes. Asimismo, se incluye información sobre el hecho de que un límite del estado está en el golfo de México, el océano Atlántico o el océano Pacífico (línea costera).

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Rango	Estado	Población	Línea costera
1	California	36 553 215	Sí
2	Texas	23 904 380	Sí
3	Nueva York	19 297 729	Sí
4	Florida	18 251 243	Sí
5	Illinois	12 852 548	No
6	Pennsylvania	12 432 792	No
7	Ohio	11 466 917	No
8	Michigan	10 071 822	No
9	Georgia	9 544 750	Sí
10	Carolina del Norte	9 061 032	Sí
11	Nueva Jersey	8 685 920	Sí
12	Virginia	7 712 091	Sí
13	Washington	6 468 424	Sí
14	Massachusetts	6 449 755	Sí
15	Indiana	6 345 289	No

Observe que cinco de los 15 estados no tienen costa. Suponga que se seleccionan tres estados al azar. Determine la probabilidad de que:

a. Ninguno de los estados seleccionados tenga costa.

b. Exactamente un estado tenga costa.

c. Por lo menos un estado seleccionado tenga costa.

59. Las ventas de automóviles Lexus en la zona de Detroit se rigen por una distribución de Poisson con una media de tres al día.

a. ¿Cuál es la probabilidad de que ningún Lexus se venda determinado día?

b. ¿Cuál es la probabilidad de que durante cinco días consecutivos se venda por lo menos un Lexus?

60. Suponga que 1.5% de las antenas de los nuevos teléfonos celulares Nokia tiene defectos. Considere una muestra aleatoria de 200 antenas y calcule las siguientes probabilidades:

a. Ninguna de las antenas tiene defectos.

b. Tres o más antenas tienen defectos.

61. Un estudio relacionado con las filas de las cajas registradoras en Safeway Supermarket, en el área de South Strand, reveló que durante los fines de semana, entre las 16:00 y las 19:00 horas, hay un promedio de cuatro clientes en la fila de espera. Determine la probabilidad de que al visitar Safeway en este horario encuentre lo siguiente:

a. Ningún cliente en la fila.

b. Cuatro clientes en la fila de espera.

c. Cuatro o menos clientes en la fila.

d. Cuatro o más clientes en espera.

62. Un estudio interno llevado a cabo por el departamento de servicios tecnológicos de Lahey Electronics reveló que los empleados de la compañía reciben un promedio de dos correos electrónicos no relacionados con el trabajo por hora. Suponga que la recepción de estos correos obedece aproximadamente a una distribución de Poisson.

a. ¿Cuál es la probabilidad de que Linda Lahey, presidenta de la compañía, haya recibido exactamente un correo no relacionado con el trabajo entre las 16:00 y las 17:00 horas de ayer?

b. ¿Cuál es la probabilidad de que haya recibido cinco o más correos de este tipo durante ese horario?

c. ¿Cuál es la probabilidad de que no haya recibido correos no relacionados con el trabajo en ese horario?

63. Los informes recientes relacionados con el crimen indican que cada minuto ocurren 3.1 robos de vehículos motorizados en Estados Unidos. Suponga que la distribución de los robos por minuto se puede aproximar por medio de una distribución de probabilidad de Poisson.

a. Calcule la probabilidad de que ocurran exactamente *cuatro* robos en un minuto.

b. ¿Cuál es la probabilidad de que *no* haya robos en un minuto?

c. ¿Cuál es la probabilidad de que *por lo menos* haya un robo en un minuto?

64. Los recientes y difíciles tiempos económicos causaron un incremento en el índice de exclusión de hipotecas. Las estadísticas del Penn Bank and Trust Company revelan que su índice de exclusión mensual de hipotecas es ahora de un préstamo entre cada 136. El mes anterior, el banco aprobó 300 préstamos.

a. ¿Cuántas exclusiones esperaría usted que hubiera tenido el banco el mes previo?

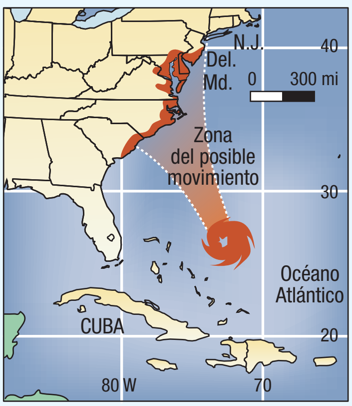
b. ¿Cuál es la probabilidad de que haya exactamente dos exclusiones?

c. ¿Cuál es la probabilidad de que haya exactamente una exclusión?

65. La National Aeronautics and Space Administration (NASA) ha sufrido dos desastres. El Challenger estalló en el océano Atlántico en 1986 y el Columbia estalló al este de Texas en 2003. Ha habido un total de 113 misiones espaciales. Suponga que los errores se siguen presentando con la misma razón y considere las siguientes 23 misiones. ¿Cuál es la probabilidad de que se presenten exactamente dos fallas? ¿Cuál es la probabilidad de que no se presenten fallas?

66. De acuerdo con la "teoría de enero", si el mercado accionario sube durante ese mes, seguirá haciéndolo durante el resto del año. Si no sube, no lo hará a lo largo del año. De acuerdo con un artículo de *The Wall Street Journal*, esta teoría se mantuvo vigente 29 de los últimos 34 años. Suponga que la teoría es falsa; es decir, la probabilidad de que el mercado suba o baje es de 0.50. ¿Cuál es la probabilidad de que esto suceda por casualidad? (es posible que requiera un paquete de software, como Excel o Minitab).

67. Durante la segunda ronda del torneo abierto de golf de 1989 en Estados Unidos, cuatro jugadores registraron un hoyo en uno al jugar el sexto hoyo. Se calcula que la posibilidad de que un jugador profesional de golf registre un hoyo en uno es de 3 708 a 1; por lo tanto, la probabilidad es de 1/3 709. Ese día participaron 155 jugadores de golf en la segunda ronda. Calcule la probabilidad de que cuatro jugadores de golf registren un hoyo en uno al jugar el sexto hoyo.



LA TORMENTA CONTINÚA
Posición : **27.8 N, 71.4 W**
Movimiento: **NNW a 8 mph**
Vientos constantes: **105 mph**
A las 23:00 horas del martes

Localización del huracán
Localización de la tormenta tropical

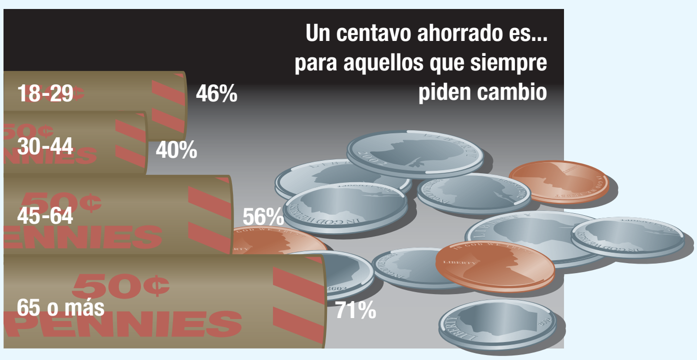
68. Suponga que el National Hurricane Center pronostica que los huracanes azotarán la zona afectada con un 0.95 de probabilidad. Responda las siguientes preguntas.

a. ¿De qué distribución de probabilidad se trata en este caso?

b. ¿Cuál es la probabilidad de que 10 huracanes toquen tierra en la zona afectada?

c. ¿Cuál es la probabilidad de que por lo menos 10 huracanes no lleguen a la zona afectada?

69. Un estudio reciente de CBS News informó que 67% de los adultos cree que el Departamento del Tesoro de Estados Unidos debe seguir acuñando monedas de un centavo.



Suponga que se selecciona una muestra de 15 adultos.

a. ¿Cuántos de los 15 adultos indicarían que el Departamento del Tesoro debe seguir acuñando monedas de un centavo? ¿Cuál es la desviación estándar?

b. ¿Cuál es la probabilidad de que *exactamente* ocho adultos indiquen que se deben seguir acuñando monedas de un centavo?

c. ¿Cuál es la probabilidad de que *por lo menos* ocho adultos indiquen que se deben seguir acuñando monedas de un centavo?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

70. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.

- **a.** Construya una distribución de probabilidad del número de habitaciones. Calcule la media y la desviación estándar de la distribución.
- **b.** Construya la distribución de probabilidad del número de baños. Calcule la media y la desviación estándar de la distribución.

71. Consulte los datos sobre Baseball 2012. Calcule la media de cuadrangulares por juego. Para hacerlo, encuentre primero la media de cuadrangulares por juego para 2012; después, divida este valor entre 162 (una temporada comprende 162 juegos). Enseguida multiplique por 2, dado que hay dos equipos en cada juego. Utilice la distribución de Poisson para estimar el número de cuadrangulares que se batearán en un juego. Encuentre la probabilidad de que:

- **a.** No haya cuadrangulares en un juego.
- **b.** Haya dos cuadrangulares en un juego.
- **c.** Haya cuando menos cuatro cuadrangulares en un juego.

7 Distribuciones de probabilidad continuas

LOS CRUCEROS DE LA LÍNEA Royal Viking informan que 80% de sus habitaciones se encuentra ocupado durante septiembre. En el caso de un crucero con 800 habitaciones, ¿cuál es la probabilidad de que 665 o más habitaciones se encuentren ocupadas en septiembre? (vea el ejercicio 60 y el objetivo de aprendizaje OA7-4).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA7-1 Describir la distribución de probabilidad uniforme y utilizarla para calcular probabilidades.

OA7-2 Describir las características de la distribución de probabilidad normal.

OA7-3 Describir la distribución de probabilidad normal estándar.

OA7-4 Aproximar la distribución de probabilidad binomial usando la distribución normal estándar para calcular probabilidades.

OA7-5 Describir la distribución de probabilidad exponencial y utilizarla para calcular probabilidades.

Introducción

En el capítulo 6 se inició el estudio de las distribuciones de probabilidad. Se consideraron tres distribuciones de probabilidad *discreta*: binomial, hipergeométrica y de Poisson; estas se basan en variables aleatorias discretas que solo adoptan valores claramente separados. Por ejemplo, si elige analizar 10 pequeñas empresas que iniciaron sus operaciones en 2010, la cantidad de empresas que todavía funcionan en 2013 puede ser 0, 1, 2, […], 10; no puede haber 3.7, 12 o −7. En este ejemplo, solo son posibles determinados resultados, los cuales, como se mencionó, se encuentran representados por valores separados con claridad. Además, el resultado se determina al contar el número de éxitos; en este caso, la cantidad de empresas del análisis que aún funcionan en 2013.

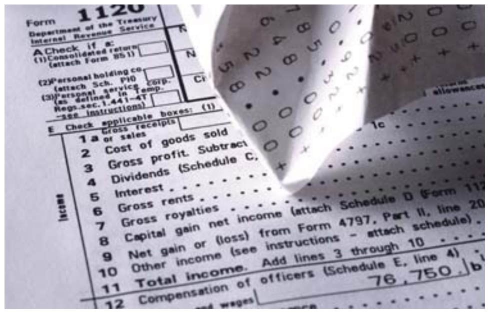
Ahora sigue el estudio de las distribuciones de probabilidad, explorando las *continuas.* Una distribución de probabilidad continua generalmente resulta de medir algo, como la distancia del dormitorio al salón de clases, el peso de un individuo o la cantidad de bonos que ganan los directores ejecutivos (*chief executive officers*, CEO). Como ejemplo, uno de los pescados que se vende en Dave's Inlet Fish Shack es el lenguado fresco. La distribución de la cantidad de lenguado que se vende tiene una media de cinco kilos por día y una desviación estándar de 1.5 kilos diarios. Esta distribución es continua porque Dave, el propietario, "mide" la cantidad de lenguado a diario. Es importante comprender que una variable aleatoria continua tiene un número infinito de valores dentro de un rango determinado. Así, en el caso de una variable aleatoria continua, la probabilidad es para un rango de valores. La probabilidad para un valor específico de una variable aleatoria continua es cero.

Este capítulo enseña cómo utilizar tres distribuciones continuas de probabilidad: uniforme, normal y exponencial.

La familia de distribuciones de probabilidad uniforme

OA7-1
Describir la distribución de probabilidad uniforme y utilizarla para calcular probabilidades.

La de probabilidad uniforme es, tal vez, la distribución más simple de una variable aleatoria continua. Esta tiene forma rectangular y está definida por valores mínimos y máximos. He aquí algunos ejemplos que se rigen por esta.



- Las ventas de gasolina en Kwik Fill, en Medina, Nueva York, siguen una distribución uniforme que varía entre 2 000 y 5 000 galones por día. La variable aleatoria es el número de galones vendidos dentro del intervalo entre estas cantidades.
- Los voluntarios de la Grand Strand Public Library elaboran formas para declaraciones de impuestos federales. El tiempo que tardan para confeccionar una forma 1040-EZ se rige por una distribución uniforme en el intervalo de 10 a 30 minutos. La variable aleatoria es la cantidad de minutos que emplean para llenar la forma, que puede asumir cualquier valor entre dicho intervalo.

En la gráfica 7.1 (página siguiente) se presenta una distribución uniforme. La forma de la distribución es rectangular y posee un valor mínimo *a* y un valor máximo *b.* Observe, asimismo, que la altura de la distribución es constante o uniforme para todos los valores entre *a* y *b.*

La media de una distribución uniforme se localiza a la mitad del intervalo entre los valores mínimo y máximo. Se calcula de la siguiente manera:

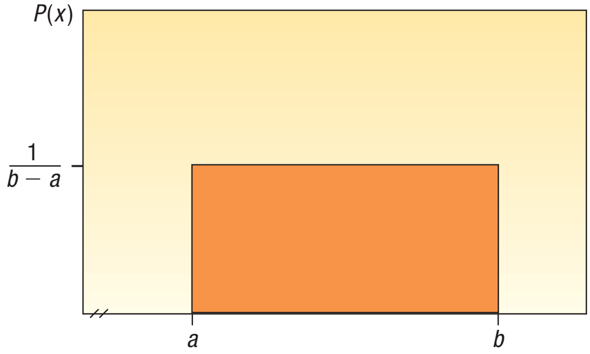
MEDIA DE LA DISTRIBUCIÓN UNIFORME

$$\mu = \frac{a + b}{2} \qquad [7.1]$$

La desviación estándar describe la dispersión de una distribución y, en la distribución uniforme, también se relaciona con el intervalo entre los valores máximo y mínimo.

DESVIACIÓN ESTÁNDAR DE LA DISTRIBUCIÓN UNIFORME

$$\sigma = \sqrt{\frac{(b - a)^2}{12}} \qquad [7.2]$$



GRÁFICA 7.1 Distribución uniforme continua

La ecuación de la distribución de probabilidad uniforme es:

DISTRIBUCIÓN UNIFORME $P(x) = \dfrac{1}{b - a}$ si $a \leq x \leq b$ y 0 en cualquier otro lugar **[7.3]**

Como se demostró en el capítulo 6, las distribuciones de probabilidad sirven para hacer afirmaciones relativas a los valores de una variable aleatoria. En el caso de distribuciones que describen una variable aleatoria continua, las áreas dentro de estas representan probabilidades. La forma rectangular de la distribución uniforme permite aplicar la fórmula del área de un rectángulo, la cual se determina al multiplicar su longitud por su altura. En el caso de la distribución uniforme, la altura del rectángulo es $P(x)$, que es $1/(b - a)$; mientras que la longitud de la base de la distribución es $b - a$. Observe que, si se multiplica la altura de la distribución por todo su intervalo para determinar el área, el resultado siempre es 1.00. En otras palabras, el área total dentro de una distribución de probabilidad continua es igual a 1.00. En general:

$$\text{Área} = (\text{altura})(\text{base}) = \frac{1}{(b - a)}(b - a) = 1.00$$

De este modo, si una distribución uniforme va de 10 a 15, la altura es 0.20, que se determina mediante $1/(15 - 10)$. La base es 5, que se calcula al restar $15 - 10$. El área total es:

$$\text{Área} = (\text{altura})(\text{base}) = \frac{1}{(15 - 10)}(15 - 10) = 1.00$$

En el siguiente ejemplo se ilustran las características de una distribución uniforme y la forma de calcular probabilidades por medio de esta.

EJEMPLO

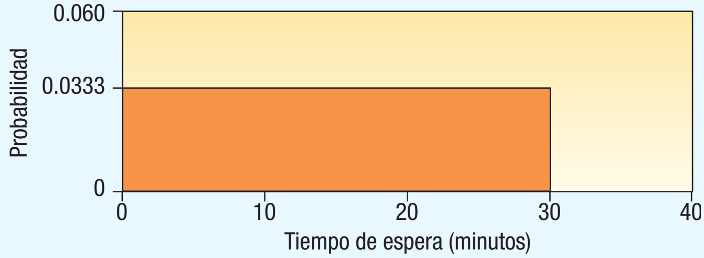
La Southwest Arizona State University proporciona servicio de transporte de autobús a los estudiantes mientras se encuentran en el recinto. Entre semana, un autobús llega a la parada de North Main Street y College Drive cada 30 minutos, desde las 6:00 hasta las 23:00. Los estudiantes llegan a la parada en tiempos aleatorios. El tiempo que espera un estudiante tiene una distribución uniforme de 0 a 30 minutos.

1. Trace una gráfica de la distribución.
2. Demuestre que el área de esta distribución uniforme es de 1.00.
3. ¿Cuánto tiempo esperará el autobús “normalmente” un estudiante? En otras palabras, ¿cuál es la media del tiempo de espera? ¿Cuál es la desviación estándar de los tiempos de espera?
4. ¿Cuál es la probabilidad de que un estudiante espere más de 25 minutos?
5. ¿Cuál es la probabilidad de que un estudiante espere entre 10 y 20 minutos?

SOLUCIÓN

En este caso, la variable aleatoria es el tiempo que espera un estudiante. El tiempo se mide en una escala continua, y los minutos de espera varían de 0 hasta 30.

1. En la gráfica 7.2 se muestra la distribución uniforme. La línea horizontal se traza a una altura de 0.0333, que se calcula mediante 1/(30 − 0). El intervalo de esta distribución es de 30 minutos.



GRÁFICA 7.2 Distribución de probabilidad uniforme de tiempos de espera de los estudiantes

2. El tiempo que los estudiantes deben esperar el autobús es uniforme a lo largo del intervalo de 0 a 30 minutos; así, en este caso, *a* es 0 y *b* es 30.

$$\text{Área} = (\text{altura})(\text{base}) = \frac{1}{(30-0)}(30-0) = 1.00$$

3. Para determinar la media, aplique la fórmula [7.1]:

$$\mu = \frac{(a+b)}{2} = \frac{(0+30)}{2} = 15$$

La media de la distribución es de 15 minutos; este es el tiempo de espera habitual del servicio de autobús.

Para determinar la desviación estándar de los tiempos de espera, aplique la fórmula [7.2]:

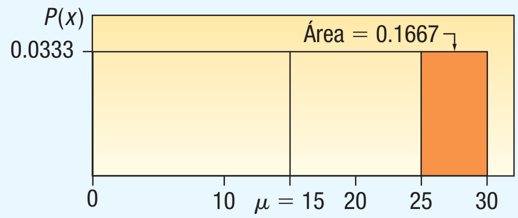
$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(30-0)^2}{12}} = 8.66$$

La desviación estándar de la distribución es de 8.66 minutos; tal es la variación de los tiempos de espera de los estudiantes.

4. El área dentro de la distribución en el intervalo de 25 a 30 minutos representa esta probabilidad en particular. De acuerdo con la fórmula del área:

$$P(25 < \text{tiempo de espera} < 30) = (\text{altura})(\text{base}) = \frac{1}{(30-0)}(5) = 0.1667$$

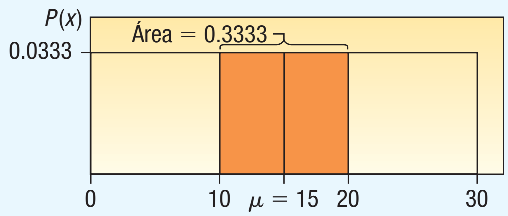
Así, la probabilidad de que un estudiante espere entre 25 y 30 minutos es 0.1667. Tal conclusión se ilustra en la siguiente gráfica:



5. El área dentro de la distribución en el intervalo de 10 a 20 minutos representa la probabilidad.

$$P(10 < \text{tiempo de espera} < 20) = (\text{altura})(\text{base}) = \frac{1}{(30-0)}(10) = 0.3333$$

Esta probabilidad se ilustra de la siguiente manera:



AUTOEVALUACIÓN 7-1

Los perros ovejeros australianos tienen una vida relativamente corta, pues su duración obedece a una distribución uniforme de entre 8 y 14 años.

(a) Trace la distribución uniforme. ¿Cuáles son los valores de la altura y de la base?
(b) Demuestre que el área total bajo la curva es de 1.00.
(c) Calcule la media y la desviación estándar de esta distribución.
(d) ¿Cuál es la probabilidad de que un perro en particular viva entre 10 y 14 años?
(e) ¿Cuál es la probabilidad de que un perro viva menos de nueve años?

EJERCICIOS

1. Una distribución uniforme se define en el intervalo de 6 a 10.
  a. ¿Cuáles son los valores de *a* y *b*?
  b. ¿Cuál es la media de esta distribución uniforme?
  c. ¿Cuál es la desviación estándar?
  d. Demuestre que el área total es de 1.00.
  e. Calcule la probabilidad de un valor mayor que siete.
  f. Calcule la probabilidad de un valor entre 7 y 9.

2. Una distribución uniforme se define en el intervalo de 2 a 5.
  a. ¿Cuáles son los valores de *a* y *b*?
  b. ¿Cuál es la media de esta distribución uniforme?
  c. ¿Cuál es la desviación estándar?
  d. Demuestre que el área total es de 1.00.
  e. Calcule la probabilidad de un valor mayor que 2.6.
  f. Calcule la probabilidad de un valor entre 2.9 y 3.7.

3. El precio de cierre de una acción común de Schnur Sporting Goods, Inc., está uniformemente distribuido entre 0 y 30 dólares por acción. Determine la probabilidad de que el precio de la acción sea:
  a. Mayor a 27 dólares.
  b. Menor o igual a 24 dólares.

4. De acuerdo con el Insurance Institute of America, una familia de cuatro miembros gasta entre 400 y 3 800 dólares anuales en toda clase de seguros. Suponga que el dinero que se gasta tiene una distribución uniforme entre estas cantidades.
  a. ¿Cuál es la media de la suma que se gasta en seguros?
  b. ¿Cuál es la desviación estándar de la suma que se gasta?
  c. Si elige una familia al azar, ¿cuál es la probabilidad de que gaste menos de 2 000 dólares anuales en seguros?
  d. ¿Cuál es la probabilidad de que una familia gaste más de 3 000 dólares anuales?

5. Las precipitaciones de abril en Flagstaff, Arizona, tienen una distribución uniforme de entre 0.5 y 3.00 pulgadas.
  a. ¿Cuáles son los valores de *a* y *b*?
  b. ¿Cuál es la precipitación media del mes? ¿Cuál es la desviación estándar?
  c. ¿Cuál es la probabilidad de que haya menos de una pulgada de precipitación en el mes?
  d. ¿Cuál es la probabilidad de que haya *exactamente* una pulgada de precipitación en el mes?
  e. ¿Cuál es la probabilidad de que haya más de 1.5 pulgadas de precipitación en el mes?

6. Los clientes con problemas técnicos en su conexión de internet pueden llamar a un número 01-800 para solicitar asistencia técnica. El técnico tarda entre 30 segundos y 10 minutos para resolver el problema. La distribución de este tiempo de asistencia tiene una distribución uniforme.
  a. ¿Cuáles son los valores (en minutos) de *a* y *b*?
  b. ¿Cuál es el tiempo medio que se requiere para resolver el problema? ¿Cuál es la desviación estándar del tiempo?
  c. ¿Qué porcentaje de problemas tardan más de cinco minutos en resolverse?
  d. Suponga que intenta determinar 50% de los tiempos de resolución de los problemas. ¿Cuáles son los puntos extremos de estos dos tiempos?

OA7-2
Describir las características de la distribución de probabilidad normal.

La familia de distribuciones de probabilidad normal

A continuación se estudia la distribución de probabilidad normal; la cual, a diferencia de la distribución uniforme (vea la fórmula [7.3]), tiene una fórmula muy compleja.

DISTRIBUCIÓN DE PROBABILIDAD NORMAL

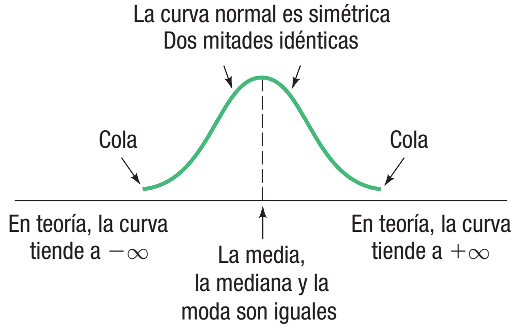
$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \qquad \textbf{[7.4]}$$

Sin embargo, no se preocupe por la complejidad de esta fórmula pues usted ya conoce varios de sus valores. Como se ha hecho hasta ahora, los símbolos μ y σ son la media y la desviación estándar. La letra griega π es una constante matemática natural, cuyo valor es aproximadamente 22/7 o 3.1416. Asimismo, la letra *e* también es una constante matemática que, además, es la base del sistema de logaritmos naturales y es igual a 2.718; *x* es el valor de una variable aleatoria continua. Así, una distribución normal se basa, o define, con su media y su desviación estándar.

No necesitará hacer cálculos con la fórmula [7.4]; más bien, requerirá la tabla que aparece en el apéndice B.3 para buscar diversas probabilidades. Estas también pueden calcularse utilizando las funciones de Excel, así como de otros programas estadísticos.

La distribución de probabilidad normal posee las siguientes características:

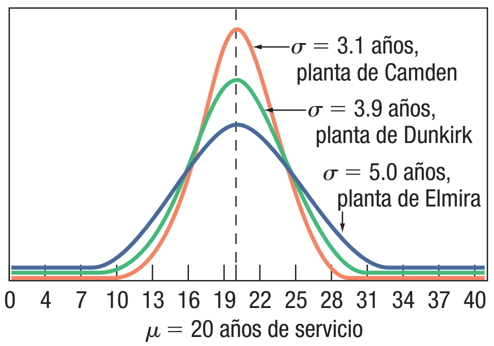
- Tiene **forma de campana** y posee una sola cima en el centro de la distribución. La media aritmética, la mediana y la moda son iguales, y se localizan en el centro de la distribución. El área total bajo la curva es de 1.00. La mitad del área bajo la curva normal se localiza a la derecha de este punto central, y la otra mitad, a la izquierda.
- Es **simétrica** respecto de la media. Si hace un corte vertical por el valor central a la curva normal, las dos mitades son imágenes especulares.
- Desciende suavemente en ambas direcciones del valor central. Es decir, la distribución es **asintótica.** La curva se aproxima más y más al eje *x,* sin tocarlo. En otras palabras, las colas de la curva se extienden indefinidamente en ambas direcciones.
- La localización de una distribución normal se determina a través de la media, μ. La dispersión (o propagación) de la distribución se determina por medio de la desviación estándar, σ.



GRÁFICA 7.3 Características de una distribución normal

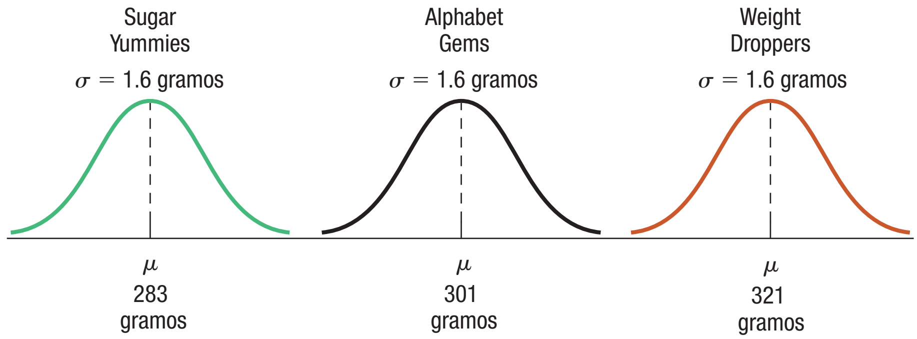
Estas características se muestran en la gráfica 7.3.

No existe una sola distribución de probabilidad normal, sino una "familia". Por ejemplo, en la gráfica 7.4 se comparan las distribuciones de probabilidad del tiempo de servicio de los empleados de tres diferentes plantas. En la planta de Camden, la media es de 20 años, y la desviación estándar, de 3.1 años. Existe otra distribución de probabilidad normal del tiempo de servicio en la planta de Dunkirk, donde $\mu = 20$ años y $\sigma = 3.9$ años. En la planta de Elmira, $\mu = 20$ años y $\sigma = 5.0$ años. Observe que las medias son las mismas, pero las desviaciones estándar difieren. A medida que la desviación estándar se reduce, la distribución se vuelve más estrecha y "picuda".

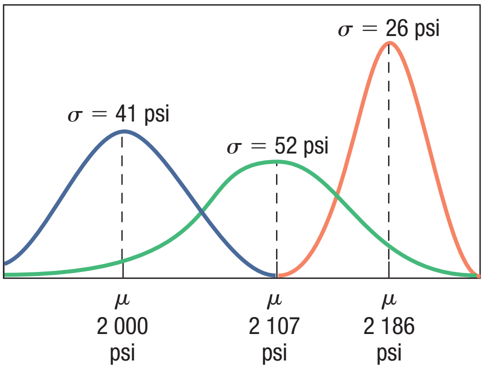


GRÁFICA 7.4 Distribución de probabilidad normal con medias iguales y desviaciones estándar diferentes

En la gráfica 7.5 se muestra la distribución de los pesos de las cajas de tres cereales diferentes. Los pesos tienen una distribución normal con diferentes medias y una desviación estándar idéntica.



GRÁFICA 7.5 Distribución de probabilidad normal con diferentes medias y desviación estándar igual



GRÁFICA 7.6 Distribuciones de probabilidad normal con medias y desviación estándar diferente

Por último, en la gráfica 7.6 se presentan tres distribuciones normales con media y desviaciones estándar diferentes. Estas muestran la distribución de fuerzas de tensión medidas en libras por pulgada cuadrada (psi) de tres clases de cables.

Recuerde que, en el capítulo 6, las distribuciones de probabilidad discreta muestran las posibilidades específicas de que ocurra un valor discreto. En el ejemplo de la sección "¿Cómo se calcula una probabilidad binomial?" se utiliza la distribución binomial para calcular la probabilidad de que ninguno de los cinco vuelos que llegan al Aeropuerto Regional Bradford de Pennsylvania esté retrasado.

En el caso de la distribución de probabilidad continua, las áreas bajo la curva definen probabilidades. El área total bajo la curva normal es de 1.0. Esto explica todos los posibles resultados. Como una distribución de probabilidad normal es simétrica, el área bajo la curva a tanto la izquierda de la media como a la derecha de esta es de 0.5. Aplique esta regla a la distribución de Sugar Yummies que se presenta en la gráfica 7.5. Es una distribución normal con una media de 283 gramos. Por consiguiente, la probabilidad de llenar una caja con más (o menos) de 283 gramos es de 0.5. También se puede determinar la probabilidad de que una caja pese entre 280 y 286 gramos. Sin embargo, para determinar esta probabilidad se necesita conocer la distribución de probabilidad normal estándar.

OA7-3
Describir la distribución de probabilidad normal estándar.

Distribución de probabilidad normal estándar

La cantidad de distribuciones normales es ilimitada, y cada una posee diferente media (μ), desviación estándar (σ), o ambas. Mientras que es posible proporcionar un número limitado de tablas de probabilidad de distribuciones discretas, como la binomial y la de Poisson, es impráctico elaborar tablas de una infinidad de distribuciones normales. Por fortuna, un miembro de la familia se utiliza para determinar las probabilidades de todas las distribuciones de probabilidad normal. Este representa la **distribución de probabilidad normal estándar**; la cual es única, pues tiene una media de 0 y una desviación estándar de 1.

Cualquier *distribución de probabilidad normal* puede convertirse en una *distribución de probabilidad normal estándar si* se resta la media de cada observación y esta diferencia se divide entre la desviación estándar. Los resultados reciben el nombre de **valores *z*** o **valores tipificados.**

VALOR *z* Distancia con signo entre un valor seleccionado, designado *x,* y la media (μ) dividida entre la desviación estándar, σ.

De esta manera, el valor *z* es la distancia de la media, medida en unidades de desviación estándar. La fórmula para esta conversión es:

VALOR NORMAL ESTÁNDAR
$$z = \frac{x - \mu}{\sigma} \qquad \textbf{[7.5]}$$

donde:

x es el valor de cualquier observación y medición;
μ es la media de la distribución;
σ es la desviación estándar de la distribución.

Según se observa en la definición anterior, un valor *z* expresa la distancia o diferencia entre un valor particular de *x* y la media aritmética en unidades de desviación estándar. Una vez que se estandarizan las observaciones de la distribución normal, los valores *z* se distribuyen normalmente con una media de 0 y una desviación estándar de 1. Así, la distribución *z* posee todas las características de cualquier distribución de probabilidad normal. Estas características aparecen en la gráfica 7.3 (página 189). En la tabla del apéndice B.1, que también se incluye al final del libro, se muestra una lista de las probabilidades de la distribución de probabilidad normal estándar. A continuación, se presenta una pequeña parte de esta tabla.

TABLA 7.1 Áreas bajo la curva normal

z	0.00	0.01	0.02	0.03	0.04	0.05	. . .
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	
⋮							

Aplicaciones de la distribución normal estándar

La distribución normal estándar es muy útil para determinar probabilidades para cualquier variable aleatoria normalmente distribuida. El procedimiento básico es encontrar el valor *z* de un valor particular de una variable aleatoria, basándose en la media y la desviación estándar de su distribución. Después, utilizando el valor *z*, es posible emplear la distribución normal estándar para encontrar varias probabilidades. En el siguiente ejemplo se describen los detalles de la aplicación.

ESTADÍSTICA EN ACCIÓN

Las aptitudes de un individuo dependen de una combinación de factores hereditarios y ambientales, cada uno de los cuales tiene más o menos la misma influencia. Por consiguiente, como en el caso de una distribución binomial con un gran número de ensayos, muchas habilidades y aptitudes tienen una distribución normal. Por ejemplo, las calificaciones en la prueba de razonamiento del examen SAT, que es la prueba más utilizada para las admisiones a las universidades en Estados Unidos. Las puntuaciones se basan en una distribución normal con una media de 1 500 y una desviación estándar de 300.

EJEMPLO

Los ingresos semanales de los supervisores de turno de la industria del vidrio se rigen por una distribución de probabilidad normal con una media de 1 000 dólares y una desviación estándar de 100 dólares. ¿Cuál es el valor *z* del ingreso *x* de un supervisor que percibe 1 100 dólares semanales? ¿Y el de un supervisor que gana 900 dólares semanales?

SOLUCIÓN

De acuerdo con la fórmula [7.5], los valores *z* de los valores *x* ($1 100 y $900) son:

Para $x = \$1\,100$

$$z = \frac{x - \mu}{\sigma} = \frac{\$1\,100 - \$1\,000}{\$100} = 1.00$$

Para $x = \$900$

$$z = \frac{x - \mu}{\sigma} = \frac{\$900 - \$1\,000}{\$100} = -1.00$$

El valor *z* de 1.00 indica que un ingreso semanal de 1 100 dólares está a una desviación estándar por encima de la media, y un valor *z* de −1.00 muestra que un ingreso de 900 dólares está a una desviación estándar por debajo de la media. Observe que ambos ingresos (1 100 y 900 dólares) se encuentran a la misma distancia (100 dólares) de la media.

AUTOEVALUACIÓN 7-2

Una encuesta aplicada en Estados Unidos recientemente concluyó que la persona típica consume 48 onzas de agua al día. Asuma que el consumo diario de agua sigue una distribución de probabilidad normal con una desviación estándar de 12.8 onzas.

(a) ¿Cuál es el valor *z* para una persona que consume 64 onzas de agua al día? Basado en este valor *z*, ¿cómo se compara esta persona con el promedio nacional?

(b) ¿Cuál es el valor *z* para una persona que consume 32 onzas de agua al día? Basado en este valor *z*, ¿cómo se compara esta persona con el promedio nacional?

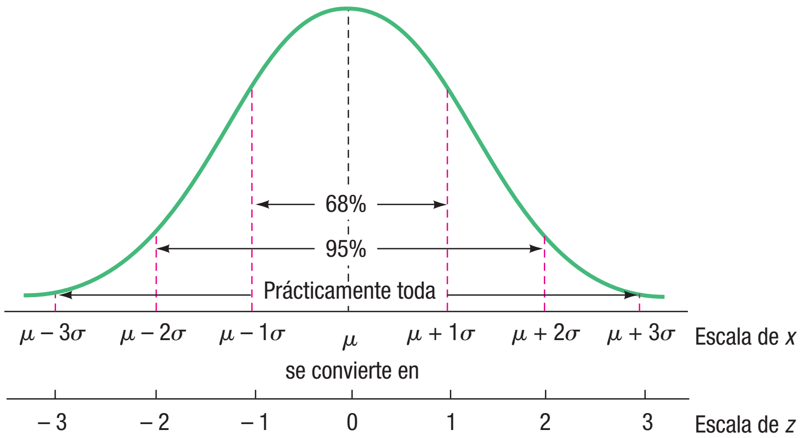
La regla empírica

La regla empírica se introdujo en el capítulo 3. Establece que si una variable aleatoria está normalmente distribuida, entonces:

1. Aproximadamente 68% de las observaciones caerán entre más y menos una desviación estándar de la media.
2. Aproximadamente 95% de las observaciones caerán entre más y menos dos desviaciones estándar de la media.
3. Prácticamente todas, o 99.7% de las observaciones caerán entre más y menos tres desviaciones estándar de la media.

Ahora bien; sabiendo cómo aplicar la distribución de probabilidad normal estándar, es posible verificar la regla empírica. Por ejemplo, una desviación estándar de la media es igual al valor *z* de 1.00. Al hablar de la tabla de probabilidad normal estándar, el valor *z* de 1.00 corresponde a una probabilidad de 0.3413. Por lo tanto, ¿qué porcentaje de las observaciones caerá entre más y menos una desviación estándar de la media? Se multiplica (2)(0.3413), lo que da 0.6826, o aproximadamente 68% de las observaciones están entre más y menos una desviación estándar de la media.

La regla empírica se resume en la siguiente gráfica.



La transformación de medidas en desviaciones normales estándar modifica la escala. Las conversiones también se muestran en la gráfica. Por ejemplo, $\mu + 1\sigma$ se convierte en un valor *z* de 1.00. Asimismo, $\mu - 2\sigma$ se transforma en un valor *z* de -2.00. Note que el centro de la distribución *z* es cero, lo cual indica que no hay desviación de la media, μ.

EJEMPLO

Como parte de su programa de control de calidad, la compañía Autolite Battery realiza pruebas acerca de la vida útil de las baterías. La vida media de una batería de celda alcalina D es de 19 horas. La vida útil de la batería se rige por una distribución normal con una desviación estándar de 1.2 horas. Responda las siguientes preguntas:

1. ¿Entre qué par de valores se localiza 68% de las baterías?
2. ¿Entre qué par de valores se localiza 95% de las baterías?
3. ¿Entre qué par de valores se localiza prácticamente la totalidad de las baterías?

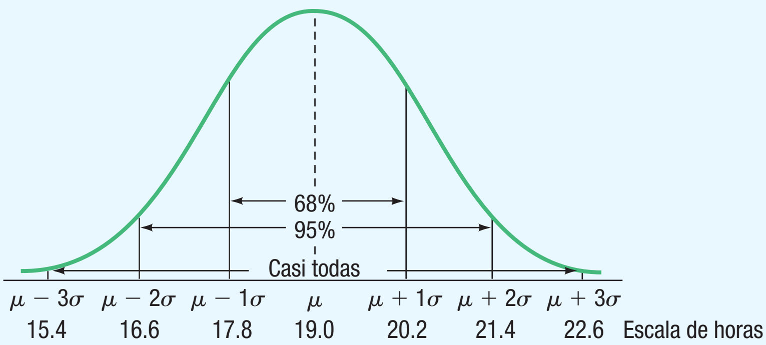
SOLUCIÓN

Aplique los resultados de la regla empírica en los siguientes planteamientos.

1. Alrededor de 68% de las baterías tiene una vida útil de entre 17.8 y 20.2 horas, lo cual se determina con el cálculo $19.0 \pm 1(1.2)$ horas.
2. Cerca de 95% de las baterías tiene una vida útil de entre 16.6 y 21.4 horas, lo cual se determina mediante $19.0 \pm 2(1.2)$ horas.

3. Casi todas las baterías tienen una vida útil de entre 15.4 y 22.6 horas, lo cual se determina por medio de 19.0 $\pm$ 3(1.2) horas.

Esta información se resume en la siguiente gráfica.



AUTOEVALUACIÓN 7-3

La distribución de los ingresos anuales de un grupo de empleados de mandos medios en Compton Plastics se aproxima a una distribución normal, con una media de 47 200 dólares y una desviación estándar de 800 dólares.

(a) ¿Entre qué par de valores se encuentran aproximadamente 68% de los ingresos?
(b) ¿Entre qué par de valores se encuentran aproximadamente 95% de los ingresos?
(c) ¿Entre qué par de valores se encuentran casi todos los ingresos?
(d) ¿Cuáles son los ingresos medio y modal?
(e) ¿Es simétrica la distribución de ingresos?

EJERCICIOS

7. Explique el significado del siguiente enunciado: "No existe solo una distribución de probabilidad normal, sino una 'familia'".

8. Enumere las características más importantes de una distribución de probabilidad normal.

9. La media de una distribución de probabilidad normal es de 500; la desviación estándar es de 10.
- **a.** ¿Entre qué par de valores se localiza alrededor de 68% del total de estos?
- **b.** ¿Entre qué par de valores se localiza alrededor de 95% del total de estos?
- **c.** ¿Entre qué par de valores se localizan casi todos estos?

10. La media de una distribución de probabilidad normal es de 60; la desviación estándar es de 5.
- **a.** ¿Alrededor de qué porcentaje de las observaciones se encuentra entre 55 y 65?
- **b.** ¿Cerca de qué porcentaje de las observaciones se encuentra entre 50 y 70?
- **c.** ¿Alrededor de qué porcentaje de las observaciones se encuentra entre 45 y 75?

11. La familia Kamp tiene gemelos, Rob y Rachel. Ambos se graduaron de la universidad hace dos años y actualmente cada uno gana 50 000 dólares anuales. Rachel trabaja en la industria de las ventas de menudeo, donde el salario medio de ejecutivos con menos de cinco años de experiencia es de 35 000 dólares, con una desviación estándar de 8 000 dólares. Rob es ingeniero. El salario medio de los ingenieros con menos de cinco años de experiencia es de 60 000 dólares, con una desviación estándar de 5 000 dólares. Calcule los valores z de Rob y Rachel. Comente los resultados.

12. Un artículo reciente que apareció en el *Cincinnati Enquirer* informó que el costo medio de la mano de obra para reparar una bomba de calefacción es de 90 dólares, con una desviación estándar de 22 dólares. Monte´s Plumbing and Heating Service terminó la reparación de dos bombas de calefacción por la mañana. El costo de la mano de obra de la primera bomba fue de 75 dólares, y de la segunda, de 100 dólares. Calcule los valores z de cada caso. Comente los resultados.

Determinación de áreas bajo la curva normal

La siguiente aplicación de la distribución normal estándar se relaciona con la determinación del área en una distribución normal entre la media y un valor elegido, que se identifica con x. En el siguiente ejemplo se ilustran los detalles.

EJEMPLO

En el ejemplo que se presenta en la sección “Aplicaciones de la distribución normal estándar” se reporta que el ingreso medio semanal de un supervisor de turno de la industria del vidrio tiene una distribución normal, con una media de 1 000 dólares y una desviación estándar de 100 dólares. Es decir, $\mu = \$1\,000$ y $\sigma = \$100$. ¿Cuál es la probabilidad de seleccionar a un supervisor cuyo ingreso semanal oscile entre 1 000 y 1 100 dólares?

SOLUCIÓN

Ya se ha convertido 1 100 dólares a un valor *z* de 1.00 mediante la fórmula [7.5] . Para repetir:

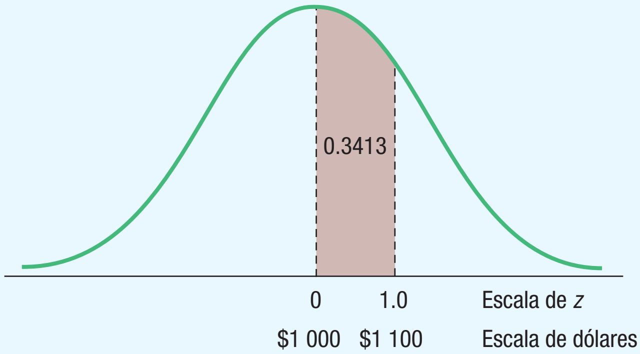
$$z = \frac{x - \mu}{\sigma} = \frac{\$1\,100 - \$1\,000}{\$100} = 1.00$$

La probabilidad asociada con un valor *z* de 1.00 se encuentra disponible en el apéndice B.3, una parte del cual se presenta a continuación. Para localizar la probabilidad, descienda por la columna izquierda hasta 1.0 y enseguida vaya a la columna con el encabezado 0.00. El valor es 0.3413.

z	0.00	0.01	0.02
⋮	⋮	⋮	⋮
0.7	0.2580	0.2611	0.2642
0.8	0.2881	0.2910	0.2939
0.9	0.3159	0.3186	0.3212
1.0	0.3413	0.3438	0.3461
1.1	0.3643	0.3665	0.3686
⋮	⋮	⋮	⋮

El área bajo la curva normal entre 1 000 y 1 100 dólares es de 0.3413. También puede decir que 34.13% de los supervisores de turno en la industria del vidrio gana entre 1 000 y 1 100 dólares semanales, o que la probabilidad de seleccionar a un supervisor cuyo ingreso oscile entre 1 000 y 1 100 dólares es de 0.3413.

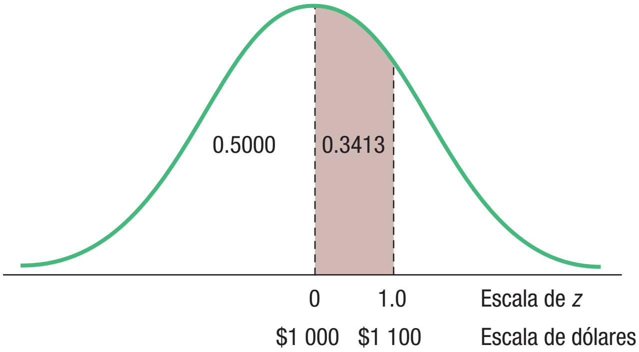
Esta información se resume en el siguiente diagrama.



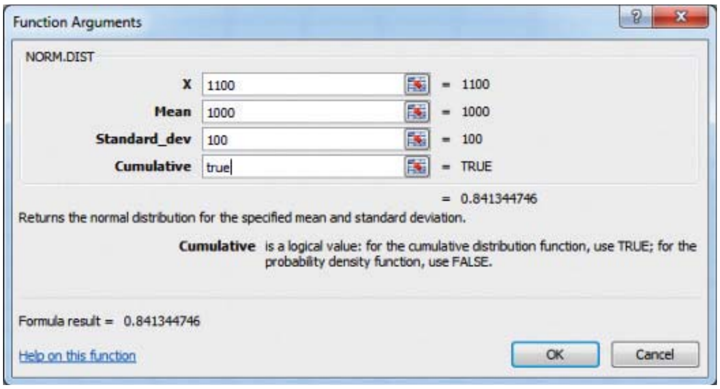
En el ejemplo anterior interesaba la probabilidad entre la media y un valor dado. Considere el siguiente cambio: en lugar de querer conocer la probabilidad de seleccionar al azar a un supervisor que gane entre 1 000 y 1 100 dólares, suponga que desea determinar la probabilidad de seleccionar a un supervisor que gane menos de 1 100 dólares. En notación probabilística, este enunciado se escribe como *P*(ingreso semanal $<$ \$1 100). El método de solución es el mismo. Determine la probabilidad de seleccionar a un supervisor que gane entre 1 000 dólares, la media y 1 100 dólares. Esta probabilidad es 0.3413. Enseguida, recuerde que la mitad del área, o probabilidad, se encuentra sobre la media, y la otra mitad, debajo de esta. En consecuencia, la probabilidad de seleccionar a un supervisor que gane menos de 1 000 dólares es de 0.5000. Por último, sume ambas probabilidades, de modo que $0.3413 + 0.5000 = 0.8413$. Alrededor de 84% de los supervisores de la industria del vidrio gana menos de 1 100 dólares mensuales (vea el diagrama de la página siguiente).

ESTADÍSTICA EN ACCIÓN

Muchos procesos, como llenar botellas de refresco o empacar fruta, tienen una distribución normal. Los fabricantes deben protegerse del llenado excesivo, así como del llenado incompleto. Si ponen demasiado en la lata o en la botella, regalan el producto; en caso contrario, el cliente quizá se sienta engañado y el gobierno cuestione la descripción que aparece en la etiqueta. A menudo se utilizan *gráficas de control*, con los límites trazados en tres desviaciones estándar por arriba y por debajo de la media, para supervisar esta clase de procesos de producción.



Excel calculará esta probabilidad. Los comandos que se requieren se encuentran en la sección "Comandos de software"**,** en el apéndice C. La respuesta es 0.8413, la misma que se calculó.



EJEMPLO

Consulte el ejemplo presentado en la sección "Aplicaciones de la distribución normal estándar" con respecto del ingreso semanal de los supervisores de turno en la industria del vidrio. La distribución de los ingresos semanales tiene una distribución de probabilidad normal, con una media de 1 000 dólares y una desviación estándar de 100 dólares. Determine la probabilidad de seleccionar a un supervisor de turno de la industria del vidrio cuyo ingreso:

1. Oscile entre 790 y 1 000 dólares.
2. Sea menor que 790 dólares.

SOLUCIÓN

Comience por localizar el valor z correspondiente a un ingreso semanal de 790 dólares. De acuerdo con la fórmula [7.5]:

$$z = \frac{x - \mu}{s} = \frac{\$790 - \$1\,000}{\$100} = -2.10$$

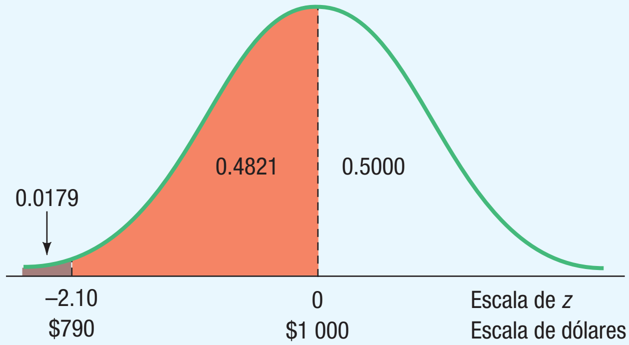
Consulte el apéndice B.3; siga hacia abajo por el margen izquierdo hasta la fila 2.1 y a lo largo de dicha fila, hasta la columna con el encabezado 0.00; ahí encontrará que el valor es de 0.4821. Así, el área bajo la curva normal estándar correspondiente a un valor z de 2.10 es de 0.4821. Sin embargo, como

z	0.00	0.01	0.02
⋮	⋮	⋮	⋮
2.0	0.4772	0.4778	0.4783
2.1	0.4821	0.4826	0.4830
2.2	0.4861	0.4864	0.4868
2.3	0.4893	0.4896	0.4898
⋮	⋮	⋮	⋮

la distribución normal es simétrica, el área entre 0 y un valor negativo de z es la misma que el área entre 0 y el correspondiente valor positivo de z. La probabilidad de localizar a un supervisor que gane entre 790 y 1 000 dólares es de 0.4821. En notación probabilística: $P(\$790 < \text{ingreso semanal} < \$1\,000) = 0.4821$.

La media divide la curva normal en dos mitades idénticas. El área bajo la mitad izquierda de la media es de 0.5000, y el área a la derecha es la misma. Como el área bajo la curva entre 790 y 1 000 dólares es 0.4821, el área debajo de 790 dólares es 0.0179, que se determina al restar 0.5000 − 0.4821. En notación probabilística: $P(\text{ingreso semanal} < \$790) = 0.0179$.

Esto significa que 48.21% de los supervisores tiene ingresos semanales que oscilan entre 790 y 1 000 dólares. Además, es previsible que 1.79% gane menos de 790 dólares a la semana. Esta información se resume en el siguiente diagrama.



AUTOEVALUACIÓN 7-4

La temperatura del café que vende Coffee Bean Cafe sigue una distribución de probabilidad normal, con una media de 150 grados Fahrenheit. La desviación estándar de esta distribución es de cinco grados.

(a) ¿Cuál es la probabilidad de que la temperatura del café esté entre los 150 y los 154 grados?
(b) ¿Cuál es la probabilidad de que la temperatura del café sea de más de 164 grados?

EJERCICIOS

13. Una población normal tiene una media de 20.0 y una desviación estándar de 4.0.
   a. Calcule el valor de z asociado con 25.0.
   b. ¿Qué proporción de la población se encuentra entre 20.0 y 25.0?
   c. ¿Qué proporción de la población es menor que 18.0?

14. Una población normal tiene una media de 12.2 y una desviación estándar de 2.5.
   a. Calcule el valor de z asociado con 14.3.
   b. ¿Qué proporción de la población se encuentra entre 12.2 y 14.3?
   c. ¿Qué proporción de la población es menor que 10.0?

15. Un estudio reciente con respecto a salarios por hora de integrantes de equipos de mantenimiento de las aerolíneas más importantes demostró que el salario medio por hora era de 20.50 dólares, con una desviación estándar de 3.50 dólares. Suponga que la distribución de los salarios por hora es una distribución de probabilidad normal. Se elige un integrante de un equipo al azar. Determine la probabilidad de que gane:
   a. Entre 20.50 y 24.00 dólares por hora.
   b. Más de 24.00 dólares por hora.
   c. Menos de 19.00 dólares por hora.

16. La media de una distribución de probabilidad normal es de 400 libras. La desviación estándar es de 10 libras.
   a. ¿Cuál es el área entre 415 libras y la media de 400 libras?
   b. ¿Cuál es el área entre la media y 395 libras?
   c. ¿Cuál es la probabilidad de seleccionar un valor al azar y descubrir que es menor a 395 libras?

Otra aplicación de la distribución normal se relaciona con la combinación de dos áreas o probabilidades. Una de estas se encuentra a la derecha de la media y la otra, a la izquierda.

EJEMPLO

Siguiendo con el ejemplo de la sección "Aplicaciones de la distribución normal estándar", utilizando los ingresos semanales de los supervisores de turno de la industria del vidrio, los ingresos semanales tienen una distribución de probabilidad normal, con una media de 1 000 dólares y una desviación estándar de 100 dólares. ¿Cuál es el área bajo esta curva normal, entre 840 y 1 200 dólares?

SOLUCIÓN

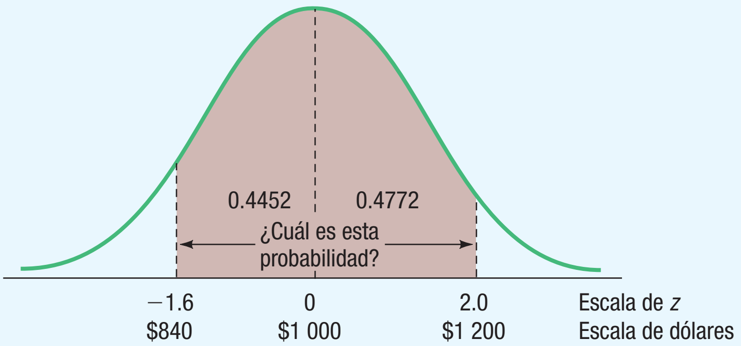
El problema se puede dividir en dos partes. En el caso del área entre 840 dólares y la media de 1 000 dólares:

$$z = \frac{\$840 - \$1\,000}{\$100} = \frac{-\$160}{\$100} = -1.60$$

En el del área entre la media de 1 000 y 1 200 dólares:

$$z = \frac{\$1\,200 - \$1\,000}{\$100} = \frac{\$200}{\$100} = 2.00$$

El área bajo la curva de un valor z de -1.60 es 0.4452 (apéndice B.3). El área bajo la curva de un valor z de 2.00 es 0.4772. Si suma ambas áreas: $0.4452 + 0.4772 = 0.9224$. Por consiguiente, la probabilidad de elegir un ingreso entre 840 y 1 200 dólares es de 0.9224. En notación probabilística: $P(\$840 < \text{ingreso semanal} < \$1\,200) = 0.4452 + 0.4772 = 0.9224$. Para resumir, 92.24% de los supervisores tiene un ingreso semanal de entre 840 y 1 200 dólares. Eso se muestra en el siguiente diagrama:



Otra aplicación de la distribución normal se relaciona con determinar el área entre valores del *mismo* lado de la media.

EJEMPLO

De regreso a la distribución del ingreso semanal de los supervisores de turno de la industria del vidrio ($\mu = \$1\,000$, $\sigma = \$100$), ¿cuál es el área bajo la curva normal entre 1 150 y 1 250 dólares?

SOLUCIÓN

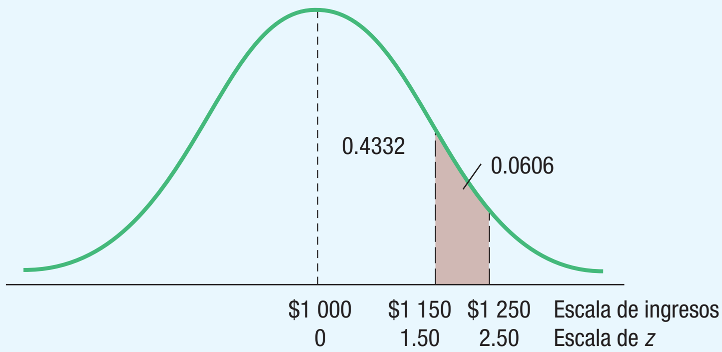
De nuevo, el caso se divide en dos partes, por lo que se aplica la fórmula [7.5]. Primero halle el valor z relacionado con un salario semanal de 1 250 dólares:

$$z = \frac{\$1\,250 - \$1\,000}{\$100} = 2.50$$

Enseguida determine el valor z de un salario semanal de 1 150 dólares:

$$z = \frac{\$1\,150 - \$1\,000}{\$100} = 1.50$$

De acuerdo con el apéndice B.3, el área relacionada con un valor z de 2.50 es de 0.4938. Así, la probabilidad de un salario semanal entre 1 000 y 1 250 dólares es de 0.4938. De manera similar, el área asociada con un valor z de 1.50 es 0.4332; de este modo, la probabilidad de un salario semanal entre 1 000 y 1 150 dólares es de 0.4332. La probabilidad de un salario semanal entre 1 150 y 1 250 dólares se calcula al restar el área asociada con un valor z de 1.50 (0.4332) de la probabilidad asociada con un valor z de 2.50 (0.4938). Por consiguiente, la probabilidad de un salario semanal entre 1 150 y 1 250 dólares es de 0.0606. En notación probabilística: $P(\$1\,150 < \text{ingreso semanal} < \$1\,250) = 0.4938 - 0.4332 = 0.0606$.



En síntesis, hay cuatro situaciones relacionadas con la determinación del área bajo la curva de la distribución de probabilidad normal estándar.

1. Para determinar el área entre 0 y z (o $-z$), se busca la probabilidad directamente en la tabla.
2. Para determinar el área más allá de z (o $-z$), se localiza la probabilidad de z en la tabla y se resta dicha probabilidad de 0.5000.
3. Para determinar el área entre dos puntos que se localizan en diferentes lados de la media, se determinan los valores z y se suman las probabilidades correspondientes.
4. Para determinar el área entre dos puntos que se localizan en el mismo lado de la media, se determinan los valores z y se resta la probabilidad menor de la mayor.

AUTOEVALUACIÓN 7-5

Retome la autoevaluación [7.4]. La temperatura del café que se vende en el Coffee Bean Cafe sigue una distribución de probabilidad normal, con una media de 150 grados Fahrenheit. La desviación estándar de esta distribución es cinco grados.

(a) ¿Cuál es la probabilidad de que la temperatura del café esté entre 146 y 156 grados?
(b) ¿Cuál es la probabilidad de que la temperatura del café sea superior a 156 grados pero inferior a 162 grados?

EJERCICIOS

17. Una distribución normal tiene una media de 50 y una desviación estándar de 4.0.
   a. Calcule la probabilidad de un valor localizado entre 44.0 y 55.0.
   b. Calcule la probabilidad de un valor mayor que 55.0.
   c. Calcule la probabilidad de un valor localizado entre 52.0 y 55.0.

18. Una población normal tiene una media de 80 y una desviación estándar de 14.0.
   a. Calcule la probabilidad de un valor localizado entre 75.0 y 90.0.
   b. Calcule la probabilidad de un valor de 75.0 o menor.
   c. Calcule la probabilidad de un valor localizado entre 55.0 y 70.0.

19. De acuerdo con el Internal Revenue Service (IRS), el reembolso medio de impuestos en 2013 fue de 3 000 dólares. Suponga que la desviación estándar es de 450 dólares y que las sumas devueltas tienen una distribución normal.
   a. ¿Qué porcentajes de reembolsos son superiores a 3 100 dólares?
   b. ¿Qué porcentajes de reembolsos son superiores a 3 100 dólares e inferiores a 3 500 dólares?
   c. ¿Qué porcentajes de reembolsos son superiores a 2 250 dólares e inferiores a 3 500 dólares?

20. La distribución del número de espectadores de *American Idol* sigue una distribución normal con una media de 29 millones, con una desviación estándar de cinco millones. Determine la probabilidad de que el programa de la próxima semana:
   a. Tenga entre 30 y 34 millones de espectadores.
   b. Tenga cuando menos 23 millones de espectadores.
   c. Sobrepase los 40 millones de espectadores.

21. WNAE, estación de AM dedicada a la transmisión de noticias, encuentra que el tiempo que los radioescuchas sintonizan la estación tiene una distribución normal. La media de la distribución es de 15.0 minutos, y la desviación estándar, de 3.5. Determine la probabilidad de que un radioescucha sintonice la estación:

a. Más de 20 minutos.
b. Durante 20 minutos o menos.
c. Entre 10 y 12 minutos.

22. Entre las ciudades de Estados Unidos con una población superior a 250 000 habitantes, la media del tiempo de viaje de ida al trabajo es de 24.3 minutos. El tiempo de viaje más largo pertenece a la ciudad de Nueva York, donde el tiempo medio es de 38.3 minutos. Suponga que la distribución de los tiempos de viaje en la ciudad de Nueva York tiene una distribución de probabilidad normal y la desviación estándar es de 7.5 minutos.

a. ¿Qué porcentaje de viajes en la ciudad de Nueva York consumen menos de 30 minutos?
b. ¿Qué porcentaje de viajes consumen entre 30 y 35 minutos?
c. ¿Qué porcentaje de viajes consumen entre 30 y 40 minutos?

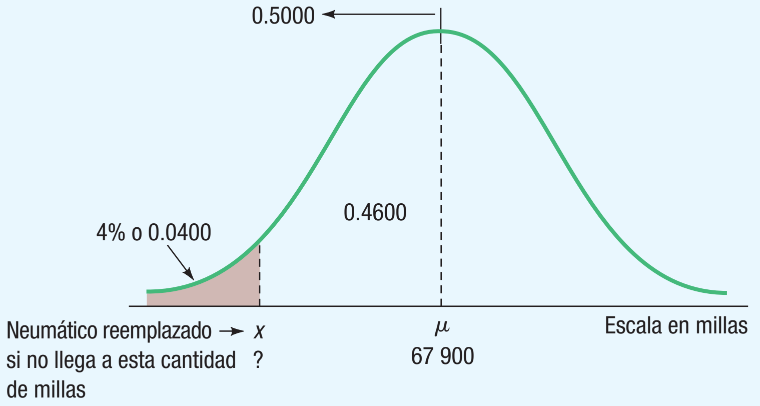
En los ejemplos anteriores se requiere determinar el porcentaje de eventos que se localiza entre dos observaciones, o el porcentaje que se encuentra por encima o por debajo de una observación *x.* Otra aplicación de la distribución normal se relaciona con el cálculo del valor de *x,* cuando se tiene el porcentaje por encima o por debajo de esta.

EJEMPLO

Layton Tire and Rubber Company pretende establecer como garantía la cantidad de millas que, como mínimo, recorrerá su nuevo neumático MX100. Algunas pruebas revelan que la media de millas es de 67 900, con una desviación estándar de 2 050, y una distribución de probabilidad normal. Layton desea determinar el número mínimo garantizado de millas, manera que no haya que sustituir más de 4% de los neumáticos. ¿Cuál es la cantidad mínima de millas que Layton debe garantizar?

SOLUCIÓN

En el siguiente diagrama se muestran las facetas del caso, en el que *x* representa el número mínimo garantizado de millas.



Al sustituir estos valores en la fórmula [7.5], se obtiene:

$$z = \frac{x - \mu}{\sigma} = \frac{x - 67\,900}{2\,050}$$

Observe que hay dos incógnitas: *z* y *x.* Para determinar *x,* primero calcule *z,* y después despeje *x.* Observe que el área por debajo de la curva normal a la izquierda de μ es de 0.5000. El área entre μ y *x* se determina al restar $0.5000 - 0.0400$. Enseguida consulte el apéndice B.3 y busque en la tabla el área más próxima a 0.4600; esta es 0.4599. Siga por los márgenes y lea el valor *z* de 1.75. Como este se encuentra a la izquierda de la media, en realidad es de -1.75. Estos pasos se ilustran en la tabla 7.2 (página siguiente).

TABLA 7.2 Áreas seleccionadas debajo de la curva normal

z …	0.03	0.04	0.05	0.06
⋮	⋮	⋮	⋮	⋮
1.5	0.4370	0.4382	0.4394	0.4406
1.6	0.4484	0.4495	0.4505	0.4515
1.7	0.4582	0.4591	0.4599	0.4608
1.8	0.4664	0.4671	0.4678	0.4686

Sabiendo que la distancia entre μ y x es de -1.75σ, o $z = -1.75$, se puede despejar x (número mínimo garantizado de millas):

$$z = \frac{x - 67\,900}{2\,050}$$

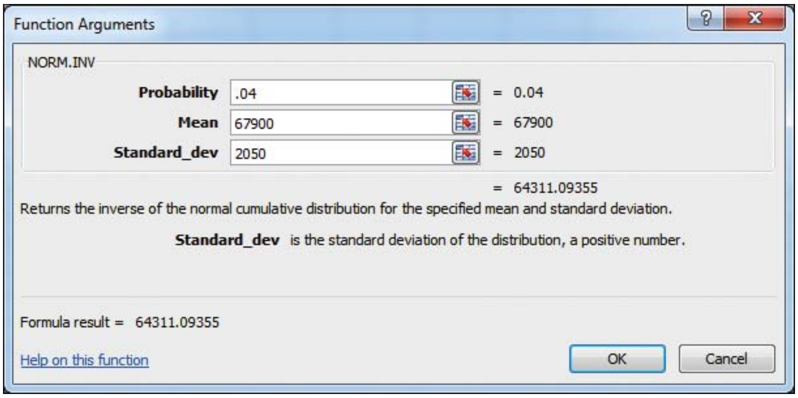
$$-1.75 = \frac{x - 67\,900}{2\,050}$$

$$-1.75(2\,050) = x - 67\,900$$

$$x = 67\,900 - 1.75(2\,050) = 64\,312$$

Por consiguiente, Layton puede anunciar que reemplazará de forma gratuita cualquier neumático que se desgaste antes de llegar a las 64 312 millas, y la empresa sabrá que solo 4% de los neumáticos se sustituirá de acuerdo con este plan.

Excel también puede encontrar el valor de la cantidad de millas. Vea la siguiente pantalla. Los comandos necesarios se dan en el apéndice C.



AUTOEVALUACIÓN 7-6

Un análisis de las calificaciones del examen final de un curso de introducción a la administración revela que tienen una distribución normal. La media de la distribución es 75, y la desviación estándar, 8. El profesor quiere recompensar con una A a los estudiantes cuyas calificaciones se encuentren dentro del 10% más alto. ¿Cuál es el punto de división de los estudiantes que merecen una A y los que merecen una B?

EJERCICIOS

23. Una distribución normal tiene una media de 50 y una desviación estándar de 4. Determine el valor por debajo del cual se presentará 95% del total de estos.

24. Una distribución normal tiene una media de 80 y una desviación estándar de 14. Determine el valor por encima del cual se presentará 80% del total de estos.

25. Suponga que el costo medio por hora de operación de un avión comercial se rige por una distribución normal, con una media de 2 100 dólares y una desviación estándar de 250 dólares. ¿Cuál es el costo de operación más bajo de 3% de los aviones?

26. La Prueba de Razonamiento SAT (antes conocida como Prueba de Aptitudes Escolares) es quizás la más amplia y la que más se utiliza para la admisión en las universidades de Estados Unidos. Las puntuaciones se basan en una distribución normal, con una media de 1 500 y una desviación estándar de 300. Clinton College desearía ofrecer una beca honorífica a aquellos estudiantes que obtengan puntuaciones que los coloquen en el 10% más alto. ¿Cuál es la puntuación mínima que se requiere para obtener la beca?

27. De acuerdo con una investigación de medios de comunicación, el estadounidense común escuchó 195 horas de música durante el año anterior. Este nivel se encuentra por debajo de las 290 horas de hace cuatro años. Dick Trythall es un gran aficionado de la música country y del oeste. Escucha música mientras trabaja en casa, lee y maneja su camión. Suponga que la cantidad de horas que escucha música tiene una distribución de probabilidad normal, con una desviación estándar de 8.5 horas.
   a. Si Dick se encuentra por encima de 1% en lo que se refiere al tiempo que escucha música, ¿cuántas horas al año escucha música?
   b. Suponga que la distribución de tiempos de hace cuatro años también tiene una distribución de probabilidad normal, con una desviación estándar de 8.5 horas. ¿Cuántas horas en realidad escucha música 1% de quienes menos lo hacen?

28. Según los datos más recientes disponibles, el costo medio anual para asistir a una universidad privada en Estados Unidos es de 26 889 dólares. Suponga que la distribución de los costos anuales se rigen por una distribución de probabilidad normal y que la desviación estándar es de 4 500 dólares. Del total de estudiantes de universidades privadas, 95% paga menos de cierta cantidad, ¿cuál es esta?

29. En teoría económica, una "tasa mínima de retorno" es, como su nombre lo indica, el retorno mínimo que una persona necesita antes de hacer una inversión. Una investigación revela que los retornos anuales de una clase especial de acciones comunes se distribuye de acuerdo con una distribución normal, con una media de 12% y una desviación estándar de 18%. Un corredor de bolsa desearía identificar una tasa mínima de retorno que esté por encima de ese valor en solo una de 20 acciones. ¿En cuánto debería establecer la tasa mínima de retorno?

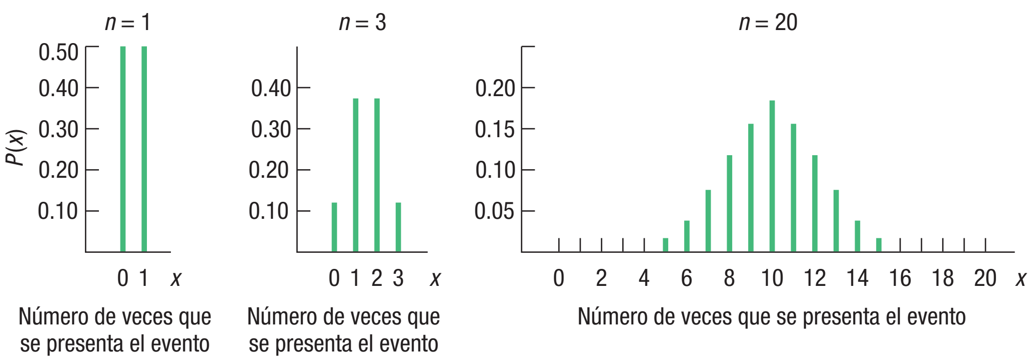
30. El fabricante de una impresora láser informa que la cantidad media de páginas que imprime un cartucho antes tener que reemplazarlo es de 12 200. La distribución de páginas impresas por cartucho se aproxima a la distribución de probabilidad normal, y la desviación estándar es de 820 páginas. El fabricante desea proporcionar lineamientos a los posibles clientes sobre el tiempo que deben esperar que les dure un cartucho. ¿Cuántas páginas por cartucho debe indicar el fabricante si desea tener 99% de certeza en todo momento?

Aproximación de la distribución normal a la binomial

OA7-4 Aproximar la distribución de probabilidad binomial usando la distribución normal estándar para calcular probabilidades.

En el capítulo 6 se describe la distribución de probabilidad binomial, la cual es discreta. En la tabla de probabilidades binomiales del apéndice B.1 se muestra una sucesión de una n de 1 a una n de 15. Si un problema implicaba una muestra de 60, generar una distribución binomial de una cantidad tan grande habría consumido demasiado tiempo. Un enfoque más eficiente consiste en aplicar la *aproximación de la distribución normal a la binomial.*

Es posible emplear la distribución normal (continua) en sustitución de la distribución binomial (discreta) en el caso de valores grandes de n, pues, conforme n se incrementa, una distribución binomial se aproxima cada vez más a una distribución normal.



GRÁFICA 7.7 Distribución binomial de una n de 1, 3 y 20, donde $\pi = 0.50$

En la gráfica 7.7 se describe el cambio de forma de una distribución binomial con $\pi = 0.50$, de una n de 3 a una n de 20. Observe cómo el caso en el que $n = 20$ se aproxima a la forma de la distribución normal. En otras palabras, compare el caso en el que $n = 20$ con la curva normal que se presenta en la gráfica 7.3 (página 189).

¿Cuándo se debe utilizar la aproximación normal? La distribución de probabilidad normal constituye una buena aproximación de la distribución de probabilidad binomial cuando $n\pi$ y $n(1 - \pi)$ tienen un valor mínimo de 5. Sin embargo, antes de aplicar la aproximación normal, debe estar seguro de que la distribución de interés es en verdad binomial. De acuerdo con el capítulo 6, se deben satisfacer cuatro criterios:

1. Solo existen dos resultados mutuamente excluyentes en un experimento: éxito o fracaso.
2. La distribución resulta del conteo del número de éxitos en una cantidad fija de ensayos.
3. La probabilidad de un éxito, π, es la misma de un ensayo a otro.
4. Cada ensayo es independiente.

Factor de corrección de continuidad

Para mostrar la aplicación de la aproximación de la distribución normal a la binomial, así como la necesidad de un factor de corrección, suponga que la administración del restaurante Santoni Pizza observa que 70% de sus nuevos clientes regresa a comer. ¿Cuál es la probabilidad de que 60 o más clientes regresen durante una semana en la que 80 nuevos clientes (primera vez) comieron en Santoni?

Observe que se cumplen las condiciones relacionadas con la distribución binomial: 1) solo hay dos posibles resultados: un cliente regresa o no lo hace; 2) es posible contar el número de éxitos, lo cual significa, por ejemplo, que 57 de los 80 clientes regresan; 3) los ensayos son independientes, es decir, si la persona número 34 regresa a comer por segunda vez, esto no influye en el hecho de que la persona 58 lo haga; 4) la probabilidad de que un cliente vuelva se mantiene en 0.70 para los 80 clientes.

Por consiguiente, la fórmula binomial [6.3] es aplicable.

$$P(x) = {}_nC_x(\pi)^x(1 - \pi)^{n-x}$$

Para determinar la probabilidad de que 60 o más clientes regresen al restaurante, primero necesita calcular la probabilidad de que regresen exactamente 60 clientes. Es decir:

$$P(x = 60) = {}_{80}C_{60}(0.70)^{60}(1 - 0.70)^{20} = 0.063$$

Enseguida, determine la probabilidad de que exactamente 61 clientes regresen. Es decir:

$$P(x = 61) = {}_{80}C_{61}(0.70)^{61}(1 - 0.70)^{19} = 0.048$$

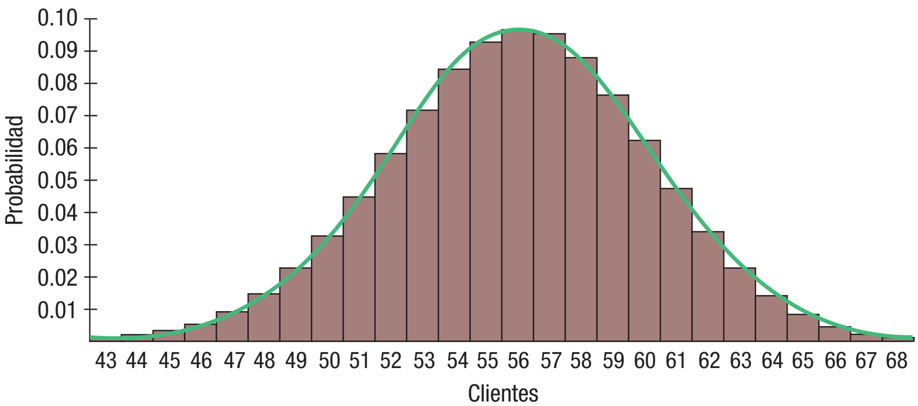
Continúe con el proceso hasta obtener la probabilidad de que regresen los 80 clientes. Por último, sume las probabilidades de 60 a 80. Resulta engorroso resolver este problema con este procedimiento. También se puede utilizar un paquete de software de computadora para determinar las diversas probabilidades. A continuación se incluye una lista de las probabilidades binomiales para $n = 80$ y $\pi = 0.70$, y x, el número de clientes que regresan, que oscila entre 43 y 68. La probabilidad

Número de clientes que regresan	Probabilidad	Número de clientes que regresan	Probabilidad
43	0.001	56	0.097
44	0.002	57	0.095
45	0.003	58	0.088
46	0.006	59	0.077
47	0.009	60	0.063
48	0.015	61	0.048
49	0.023	62	0.034
50	0.033	63	0.023
51	0.045	64	0.014
52	0.059	65	0.008
53	0.072	66	0.004
54	0.084	67	0.002
55	0.093	68	0.001

de que regrese cualquier cantidad de clientes inferior a 43 o superior a 68 es menor que 0.001. También es posible suponer que estas probabilidades son iguales a 0.000.

Se determina la probabilidad de que 60 o más clientes regresen al sumar $0.063 + 0.048 + \cdots + 0.001$, que equivale a 0.197. Sin embargo, al ver la siguiente gráfica se observa la similitud de esta distribución con una distribución normal. Todo lo que se necesita es "arreglar" las probabilidades discretas para obtener una distribución continua. Además, trabajar con una distribución normal implica hacer más cálculos que con la binomial.

El truco consiste en permitir que la probabilidad discreta de 56 clientes quede representada por un área bajo la curva continua entre 55.5 y 56.5; después, permitir que la probabilidad de los 57 clientes quede representada por un área entre 56.5 y 57.5, etcétera. Este enfoque es exactamente contrario al de redondear las cifras a un número entero.



Como la distribución normal sirve para determinar la probabilidad binomial de 60 o más éxitos, debe restar, en este caso, 0.5 de 60. El valor de 0.5 recibe el nombre de **factor de corrección de continuidad.** Este pequeño ajuste debe hacerse porque una distribución continua (la distribución normal) se utiliza para aproximar una distribución discreta (la distribución binomial). Al restar se obtiene $60 - 0.5 = 59.5$.

FACTOR DE CORRECCIÓN DE CONTINUIDAD Valor de 0.5 restado o sumado, según se requiera, a un valor seleccionado cuando una distribución de probabilidad discreta se aproxima por medio de una distribución de probabilidad continua.

Cómo aplicar el factor de corrección

Este se aplica en los siguientes cuatro casos:

1. Para la probabilidad de que *por lo menos* ocurra x, se utiliza el área *por encima de* $(x - 0.5)$.
2. Para la probabilidad de que ocurra más que x, se utiliza el área *por encima de* $(x + 0.5)$.
3. Para la probabilidad de que ocurra x o *menos,* se utiliza el área *debajo* de $(x + 0.5)$.
4. Para la probabilidad de que ocurra *menos que* x, se utiliza el área *debajo de* $(x - 0.5)$.

Para utilizar la distribución normal con el fin de aproximar la probabilidad de que regresen 60 o más clientes de los 80 que van a Santoni por primera vez, se sigue el siguiente procedimiento.

Paso 1. Establezca el valor z correspondiente a una x de 59.5 con la fórmula [7.5], y las fórmulas [6.4] y [6.5], de la media y la varianza de una distribución binomial:

$$\mu = n\pi = 80(0.70) = 56$$

$$\sigma^2 = n\pi(1 - \pi) = 80(0.70)(1 - 0.70) = 16.8$$

$$\sigma = \sqrt{16.8} = 4.10$$

$$z = \frac{x - \mu}{\sigma} = \frac{59.5 - 56}{4.10} = 0.85$$

Paso 2. Determine al área bajo la curva normal entre una π de 56 y una x de 59.5. Según el primer paso, el valor z correspondiente a 59.5 es de 0.85. Enseguida consulte el apén-

ESTADÍSTICA EN ACCIÓN

Muchas variables tienen una distribución normal aproximada, como las calificaciones del cociente intelectual, la esperanza de vida y la estatura en la edad adulta. Esto implica que casi todas las observaciones ocurrirán dentro de tres desviaciones estándar respecto de la media. Por otra parte, son poco frecuentes las que ocurren más allá de tres desviaciones estándar respecto de la media. Por ejemplo, la estatura media de un adulto de sexo masculino es de 68.2 pulgadas (casi 5 pies con 8 pulgadas), con una desviación estándar de 2.74. Esto significa que casi todos los hombres miden entre 60.0 pulgadas (5 pies) y 76.4 pulgadas (6 pies con 4 pulgadas) de estatura. LeBron James, jugador de basquetbol profesional de los Miami Heat, mide 80 pulgadas, o 6 pies con 8 pulgadas, lo cual rebasa las tres desviaciones estándar respecto de la media. La altura convencional de una puerta es de 6 pies con 8 pulgadas, y debe ser lo bastante alta para la mayoría de los hombres adultos, con excepción de una persona poco común, como LeBron James.

Otro ejemplo consiste en el hecho de que el asiento del conductor de la mayoría de los vehículos se encuentra colocado de manera que una persona que mida por lo menos 159 cm (62.5 pulgadas de estatura) se siente con comodidad. La distribución de estatura de mujeres adultas es más o menos una distribución

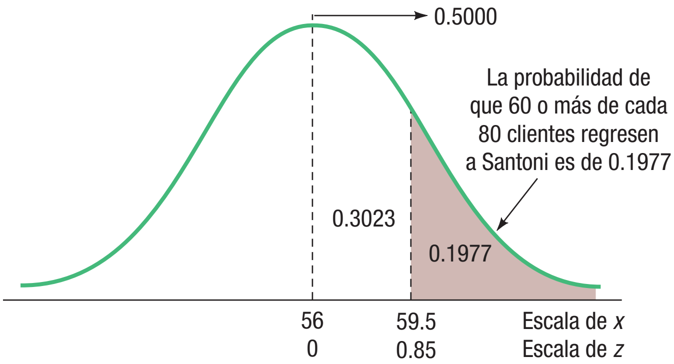
(*continúa*)

(*continuación*)

normal con una media de 161.5 y una desviación estándar de 6.3 cm. Por consiguiente, alrededor de 35% de las mujeres adultas no se sienta cómodamente en el asiento del conductor.

dice B.1, vaya hacia abajo del margen izquierdo hasta 0.8 y luego, en línea horizontal, hasta la columna con el encabezado 0.05. El área es de 0.3023.

Paso 3. Calcule el área más allá de 59.5, para restar 0.3023 de 0.5000 ($0.5000 - 0.3023 = 0.1977$). Por consiguiente, 0.1977 es la probabilidad de que regresen 60 o más clientes de los 80 que acuden por primera vez a Santoni. En notación probabilística: $P(\text{clientes} > 59.5) = 0.5000 - 0.3023 = 0.1977$. Las facetas de este problema se muestran en la siguiente gráfica:



Sin duda, usted estará de acuerdo en que utilizar la aproximación normal de la binomial constituye un método más eficaz para calcular la probabilidad de que regresen 60 o más clientes que acuden por primera vez. El resultado es comparable con el que se obtuvo en el "Factor de corrección de continuidad", donde se utilizó la distribución binomial. La probabilidad, al utilizar la distribución binomial, es de 0.197, mientras que con la aproximación normal es de 0.1977.

AUTOEVALUACIÓN 7-7

Un estudio de la compañía Great Southern Home Insurance reveló que en 80% de los robos que se reportaron, los bienes no fueron recuperados por los dueños.

(a) Durante un periodo en el que ocurrieron 200 robos, ¿cuál es la probabilidad de que los bienes robados no se recuperen en 170 o más casos?

(b) Durante un periodo en el que ocurrieron 200 robos, ¿cuál es la probabilidad de que no se recuperen los bienes robados en 150 o más casos?

EJERCICIOS

31. Suponga una distribución de probabilidad binomial con $n = 50$ y $\pi = 0.25$. Calcule lo siguiente:
 a. La media y la desviación estándar de la variable aleatoria.
 b. La probabilidad de que x sea 15 o mayor.
 c. La probabilidad de que x sea 10 o menor.

32. Suponga una distribución de probabilidad binomial con $n = 40$ y $\pi = 0.55$. Calcule lo siguiente:
 a. La media y la desviación estándar de la variable aleatoria.
 b. La probabilidad de que x sea 25 o mayor.
 c. La probabilidad de que x sea 15 o menor.
 d. La probabilidad de que x se encuentre entre 15 y 25, inclusive.

33. Dottie's Tax Service se especializa en declaraciones del impuesto sobre la renta de clientes profesionales, como médicos, dentistas, contadores y abogados. Una auditoría reciente de las declaraciones que elaboraba la empresa, que llevó a cabo el Internal Revenue Service (IRS) indicó que 5% de las declaraciones que había elaborado durante el año previo contenía errores. Considere que esta tasa continúa durante el año y Dottie´s elabora 60 declaraciones. Determine la probabilidad de que cometa errores en:
 a. Más de seis declaraciones.
 b. Por lo menos seis declaraciones.
 c. Exactamente seis declaraciones.

34. Shorty's Muffler anuncia que puede instalar un silenciador nuevo en 30 minutos o menos. No obstante, hace poco el departamento de estándares laborales de las oficinas centrales realizó un estudio y descubrió que 20% de los silenciadores no se instalaba en 30 minutos o menos. La sucursal Maumee instaló 50 silenciadores el mes anterior. Si el informe de la empresa es correcto:
 a. ¿Cuántas instalaciones de la sucursal Maumee se esperaría que tardaran más de 30 minutos?

b. ¿Cuál es la probabilidad de que ocho o menos instalaciones tarden más de 30 minutos?
c. ¿Cuál es la probabilidad de que exactamente ocho de las 50 instalaciones tarden más de 30 minutos?

35. Un estudio que realizó Taurus Health Club, famoso en Estados Unidos, reveló que 30% de sus nuevos miembros tiene un sobrepeso de 15 libras. Una campaña de promoción de membresías en un área metropolitana dio como resultado la captación de 500 nuevos miembros.
a. Se sugirió utilizar la aproximación normal de la distribución binomial para determinar la probabilidad de que 175 o más de los nuevos miembros tengan un sobrepeso de 15 libras. ¿Este problema es de naturaleza binomial? Explique.
b. ¿Cuál es la probabilidad de que 175 o más de los nuevos miembros tengan un sobrepeso de 15 libras?
c. ¿Cuál es la probabilidad de que 140 o más de los nuevos miembros tengan un sobrepeso de 15 libras?

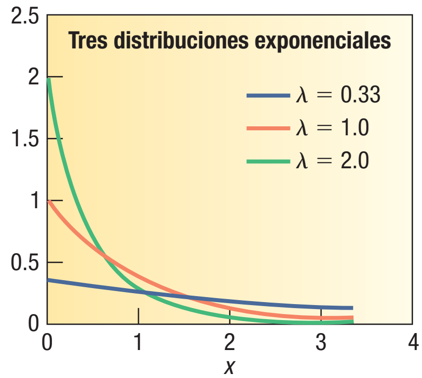
36. Un número reciente de *Bride Magazine* sugirió que las parejas que planean su boda deben esperar que dos terceras partes de las personas a las que envían invitación confirmen su asistencia. Rich y Stacy tienen planes de casarse este año y piensan enviar 197 invitaciones.
a. ¿Cuántos invitados se esperaría que aceptaran la invitación?
b. ¿Cuál es la desviación estándar?
c. ¿Cuál es la probabilidad de que 140 o más acepten la invitación?
d. ¿Cuál es la probabilidad de que exactamente 140 acepten la invitación?

La familia de distribuciones exponenciales

OA7-5
Describir la distribución de probabilidad exponencial y utilizarla para calcular probabilidades.

Hasta ahora, en este capítulo se han considerado dos distribuciones de probabilidad continua: uniforme y normal. La siguiente distribución continua que se explica es la exponencial. Por lo general, esta distribución de probabilidad continua describe los tiempos entre eventos que ocurren en secuencia. Las acciones suceden independientemente a un ritmo constante por unidad o duración de tiempo. Como el tiempo nunca es negativo, una variable aleatoria exponencial siempre será positiva. La distribución exponencial suele describir situaciones como:

- Los tiempos de servicio a clientes en el módulo de información de la Dallas Public Library.
- El tiempo entre las "entradas" en un sitio web.
- El tiempo de vida de un componente eléctrico.
- El tiempo que transcurre hasta que entra la siguiente llamada telefónica a un centro de servicio al cliente.



La distribución de probabilidad exponencial tiene un sesgo positivo. En esta característica difiere de las distribuciones uniforme y normal, que son simétricas. De hecho, la distribución es descrita por un solo parámetro, que se identifica como λ (se pronuncia "lambda"). Es común referirse a λ como el parámetro de "ritmo". En la siguiente gráfica se muestra el cambio en la forma de la distribución exponencial a medida que el valor de λ se modifica de 1/3 a de 1 a 2. Observe que conforme se reduce λ, la forma de la distribución cambia para volverse "menos sesgada".

Otra característica de la distribución exponencial es su estrecha relación con la distribución de Poisson, la cual es discreta y también tiene un solo parámetro, μ. La distribución de Poisson se describe en el capítulo 6 y también se trata de una distribución con sesgo positivo. Para explicar la relación entre la distribución de Poisson y las distribuciones exponenciales, suponga que el ritmo al que los clientes llegan a un restaurante familiar durante la cena es de seis por hora. Se utiliza la distribución de Poisson para determinar la probabilidad de que, en cualquier hora de la cena, lleguen dos clientes o siete, y así sucesivamente; así la distribución de Poisson tiene una media de seis. Pero suponga que en vez de estudiar el número de clientes *que llegan en una hora*, desea estudiar el tiempo que transcurre *entre cada llegada.* El tiempo entre llegadas es una distribución continua porque el tiempo se mide como una variable aleatoria continua. Si los clientes llegan a un ritmo de seis por hora, entonces es lógico que el tiempo medio o típico entre llegadas sea de 1/6 de hora o 10 minutos. Aquí es necesario ser consistentes con las unidades, de manera que considere 1/6 de hora. Así que, en general, si se sabe que los clientes llegan a cierto ritmo por hora, al que se llama μ, se espera que el tiempo

medio entre llegadas sea $1/\mu$. El parámetro de ritmo λ es igual a $1/\mu$. Por lo tanto, en este ejemplo, $\lambda = 1/6$.

En la gráfica de la distribución exponencial se comienza con el valor de λ cuando el valor de la variable aleatoria (x) es 0. La distribución se reduce de manera uniforme a medida que se desplaza a la derecha, con valores crecientes de x. En la fórmula [7.6] se describe la distribución de probabilidad exponencial con λ como parámetro de ritmo. Como ya se describió en la distribución de probabilidad de Poisson el capítulo 6, e es una constante matemática igual a 2.71828 que es la base del sistema logarítmico neperiano. Es una agradable sorpresa que tanto la media como la desviación estándar de la distribución de probabilidad exponencial sean iguales a $1/\lambda$.

DISTRIBUCIÓN EXPONENCIAL $$P(x) = \lambda e^{-\lambda x} \qquad \textbf{[7.6]}$$

En el caso de las distribuciones continuas, no se considera la probabilidad de que se presente un valor distinto. En vez de eso, las áreas o regiones que se muestran debajo de la gráfica de la distribución de probabilidades entre dos valores determinados dan la probabilidad de que la variable aleatoria esté en dicho intervalo. No se necesita una tabla de la distribución exponencial, como la que se incluye en el apéndice B.3, para la distribución normal. El área bajo la función de densidad exponencial se determina mediante una fórmula simple, y los cálculos que se requieren pueden realizarse con una calculadora de mano que tenga la tecla e^x. Con la mayoría de los softwares estadísticos también se pueden calcular las probabilidades exponenciales con solo ingresar λ, el parámetro de ritmo. La probabilidad de obtener un valor de llegada menor a uno en particular de x es:

ENCONTRAR LA PROBABILIDAD USANDO LA DISTRIBUCIÓN EXPONENCIAL $$P(\text{Tiempo de llegada} < x) = 1 - e^{-\lambda x} \qquad \textbf{[7.7]}$$

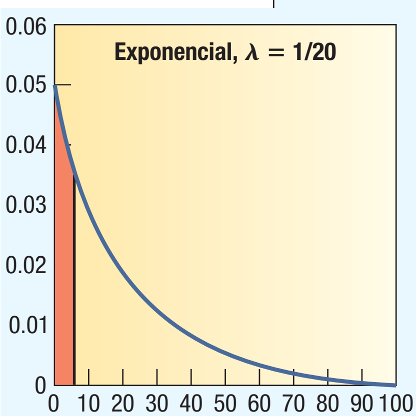
EJEMPLO

Las órdenes para pedidos de medicamentos por receta llegan a una farmacia virtual, de acuerdo con una distribución de probabilidad exponencial, a una media de una cada 20 segundos. Encuentre la probabilidad de que la siguiente orden llegue en menos de cinco segundos, o en más de 40.

SOLUCIÓN

Para comenzar, se determina el parámetro de ritmo λ, que en este caso es 1/20. Para encontrar la probabilidad, se inserta 1/20 en lugar de λ y 5 por x en la fórmula [7.7].

$$P(\text{Tiempo de llegada} < 5) = 1 - e^{-\frac{1}{20}(5)} = 1 - e^{-0.25} = 1 - 0.7788 = 0.2212$$



Se concluye que hay una probabilidad de 22% de que la siguiente orden llegue en menos de cinco segundos. La región se identifica como el área color marrón bajo la curva.

Los cálculos anteriores señalaron el área en la zona de la cola izquierda de la distribución exponencial como $\lambda = 1/20$, y el área entre 0 y 5 (es decir, el área que está por debajo de los cinco segundos). ¿Qué pasa si usted se interesa en el área de la cola derecha? Para encontrarla, use la regla del complemento. Vea la fórmula [5.3] en la sección 5.4 del capítulo 5. Dicho de otra forma, para encontrar la probabilidad de recibir la siguiente orden en más de 40 segundos se debe hallar la probabilidad de recibirla en menos de 40 segundos y restar el resultado de 1.00. Los pasos son:

1. Encuentre la probabilidad de recibir una orden *en menos* de 40 segundos.

$$P(\text{Llegada} < 40) = 1 - e^{-\frac{1}{20}(40)} = 1 - 0.1353 = 0.8647$$

2. Encuentre la probabilidad de recibir una orden *en más* de 40 segundos.

$$P(\text{Llegada} > 40) = 1 - P(\text{Llegada} < 40) = 1 - 0.8647 = 0.1353$$

Se concluye que la probabilidad de que pasen 40 segundos o más antes de recibir la siguiente orden en la farmacia virtual es de 13.5%.

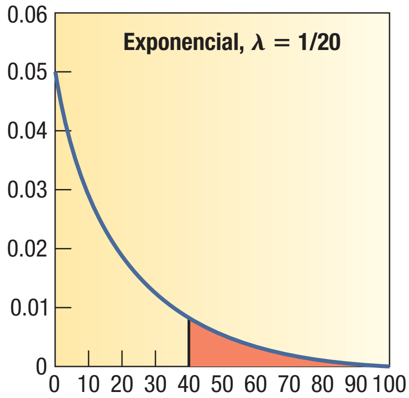
En el ejemplo anterior, al aplicar la distribución de probabilidad exponencial para calcular la probabilidad de que el tiempo de llegada sea mayor a 40 segundos, probablemente habrá observado que existe cierta redundancia. En general, si se desea encontrar la probabilidad de un tiempo mayor que algún valor x, como 40 en las ecuaciones anteriores, entonces:

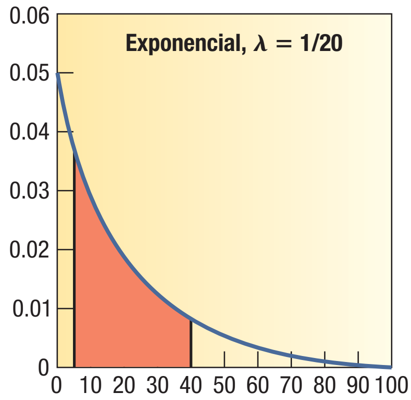
$$P(\text{Llegada} > x) = 1 - P(\text{Llegada} < x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

En otras palabras, reste la fórmula [7.7] de 1 para obtener el área en la cola derecha, la cual es $e^{-\lambda x}$. Por ello, la probabilidad de que pasen 40 segundos antes que llegue la siguiente orden se calcula directamente, sin la ayuda de la regla del complemento, de la siguiente forma:

$$P(\text{Llegada} > 40) = e^{-\frac{1}{20}(40)} = 0.1353$$

La gráfica de la izquierda muestra el resultado.





Si desea determinar la probabilidad de que pasen más de cinco segundos pero menos de 40 para que llegue la siguiente orden, use la fórmula [7.7] con un valor x de 40, y reste el valor de la fórmula [7.7], donde x es 5.

En símbolos, puede escribirlo así:

$$\begin{aligned} P(5 \le x \le 40) &= P(\text{Llegada} \le 40) - P(\text{Llegada} \le 5) \\ &= (1 - e^{-\frac{1}{20}(40)}) - (1 - e^{-\frac{1}{20}(5)}) = 0.8647 - 0.2212 = 0.6435 \end{aligned}$$

Se concluye que 64% del tiempo, el lapso entre órdenes oscilará entre cinco y 40 segundos (gráfica superior derecha).

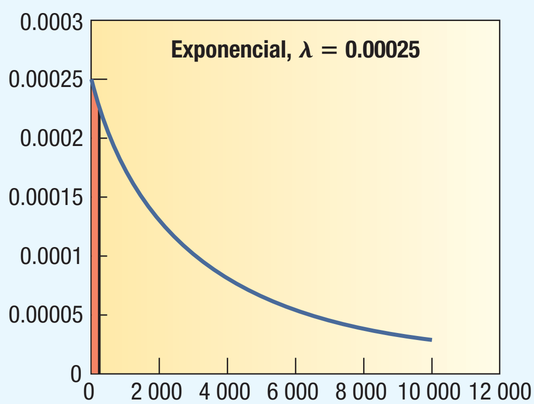
En los ejemplos anteriores se debe encontrar el porcentaje de valores ubicados entre dos cantidades, o el porcentaje de estos que está por encima o por debajo de uno en particular, x. También se puede utilizar la fórmula [7.7] "en reversa" para encontrar el valor de x cuando el porcentaje es superior o inferior a este. En el siguiente ejemplo se ilustra esta situación.

EJEMPLO

Compton Computers desea establecer una garantía mínima de tiempo de vida para su nueva unidad de fuente de poder. Las pruebas de calidad muestran que el tiempo de falla sigue una distribución exponencial con una media de 4 000 horas. Compton quiere un periodo de garantía en cuyo transcurso solo falle 5% de las fuentes de poder. ¿Qué valor debe establecer para el periodo de garantía?

SOLUCIÓN

Observe que 4 000 horas es una media y no un ritmo. Por lo tanto, es preciso establecer λ como 1/4 000, o 0.00025 fallas por hora. A la derecha se muestra un diagrama de la situación, donde x representa el tiempo de vida mínimo garantizado.



Utilice la fórmula [7.7] y, básicamente, trabaje hacia atrás para hallar la solución. En este caso, el parámetro de ritmo es 4 000 horas y se quiere que dicha área sea 0.05, tal como se muestra en el diagrama.

$$P(\text{Tiempo de llegada} < x) = 1 - e^{(-\lambda x)}$$

$$0.05 = 1 - e^{-\frac{1}{4\,000}(x)}$$

Enseguida, se resuelve la ecuación para x. Por lo tanto, se resta 1 de ambos lados de la ecuación y se multiplica por -1 para simplificar los signos. El resultado es:

$$0.95 = e^{-\frac{1}{4\,000}(x)}$$

El siguiente paso es tomar el logaritmo natural de ambos lados y resolverlo para x:

$$\ln(0.95) = -\frac{1}{4\,000}x$$

$$-(0.051293294) = -\frac{1}{4\,000}x$$

$$x = 205.17$$

En este caso, $x = 205.17$. De esta forma, Compton puede establecer el periodo de garantía en 205 horas, y esperar que alrededor de 5% de las fuentes de poder se devuelva.

AUTOEVALUACIÓN 7-8

El tiempo entre arribos de ambulancia en la sala de urgencias del Hospital Metodista sigue una distribución exponencial, con una media de 10 minutos.

(a) ¿Cuál es la probabilidad de que la próxima ambulancia llegue en 15 minutos o menos?
(b) ¿Cuál es la probabilidad de que la próxima ambulancia llegue en más de 25 minutos?
(c) ¿Cuál es la probabilidad de que la próxima ambulancia llegue en más de 15 minutos, pero menos de 25?
(d) Encuentre el 80o. percentil del tiempo entre los arribos de las ambulancias (esto significa que solo 20% de las corridas son más largas que este tiempo).

EJERCICIOS

37. Los tiempos de espera para recibir la comida después de hacer el pedido en la tienda Subway local siguen una distribución exponencial con una media de 60 segundos. Calcule la probabilidad de que un cliente espere:
a. Menos de 30 segundos.
b. Más de 120 segundos.
c. Entre 45 y 75 segundos.
d. Del total de clientes, 50% espera menos de cierta cantidad de segundos, ¿cuál es esta cantidad? ¿Cuál es la mediana?

38. El tiempo de vida de los televisores de plasma y LCD sigue una distribución exponencial con una media de 100 000 horas. Calcule la probabilidad de que un televisor:
a. Falle en menos de 10 000 horas.
b. Dure más de 120 000 horas.
c. Falle entre 60 000 y 100 000 horas de uso.
d. Encuentre el 90o. percentil. Del total de televisores, 10% dura más de cierta cantidad de horas, ¿cuántas son?

39. The Bureau of Labor Statitics realiza la encuesta llamada *American Time Use.* Esta mostró que el tiempo que se pasa en Estados Unidos utilizando una computadora para entretenimiento varía mucho según la edad. Los individuos de 75 años en adelante promediaron 0.3 horas (18 minutos) por día. Los de 15 a 19 años pasaron 1.0 horas al día. Considere que estos tiempos siguen una distribución exponencial y encuentre la proporción de cada grupo que pasa:
a. Menos de 15 minutos al día usando la computadora para entretenimiento.
b. Más de dos horas.
c. Entre 30 y 90 minutos.
d. Encuentre el 20o. percentil. Del total de individuos, 80% pasa más de una cantidad específica de tiempo, ¿cuál es esa cantidad?

40. El costo por artículo en el supermercado sigue una distribución exponencial. Hay muchos artículos baratos y pocos que son relativamente caros. El costo medio por artículo es de 3.50 dólares. Determine el porcentaje de artículos que cuestan:

a. Menos de un dólar.
b. Más de cuatro dólares.
c. Entre dos y tres dólares.
d. Encuentre el 40o. percentil. Del total de artículos del supermercado, 60% cuesta más un monto específico, ¿de cuánto dinero se trata?

RESUMEN DEL CAPÍTULO

I. La distribución uniforme es de probabilidad continua, y tiene las siguientes características:
  A. Su forma es rectangular.
  B. La media y la mediana son iguales.
  C. Su valor mínimo *a* y su valor máximo *b* la describen por completo.
  D. La siguiente ecuación de la región de *a* a *b* la describe:

$$P(x) = \frac{1}{b - a} \quad \textbf{[7.3]}$$

  E. La media y la desviación estándar de una distribución uniforme se calculan de la siguiente manera:

$$\mu = \frac{(a + b)}{2} \quad \textbf{[7.1]}$$

$$\sigma = \sqrt{\frac{(b - a)^2}{12}} \quad \textbf{[7.2]}$$

II. La distribución de probabilidad normal es continua, y tiene las siguientes características:
  A. Tiene forma de campana y posee una sola cima en el centro de la distribución.
  B. La distribución es simétrica.
  C. Es asintótica, lo cual significa que la curva se aproxima al eje *x* sin tocarlo jamás.
  D. Su media y su desviación estándar la describen por completo.
  E. Existe una familia de distribuciones de probabilidad normal.
    1. Se genera otra distribución de probabilidad normal cuando la media o la desviación estándar cambian.
    2. La distribución de probabilidad normal se describe por medio de la fórmula:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \quad \textbf{[7.4]}$$

III. La distribución de probabilidad normal estándar es un caso especial.
  A. Posee una media de 0 y una desviación estándar de 1.
  B. Toda distribución de probabilidad normal puede estandarizarse mediante la fórmula:

$$\mu = \frac{x - \mu}{\sigma} \quad \textbf{[7.5]}$$

  C. Al estandarizar una distribución de probabilidad normal, se indica la distancia de un valor de la media en unidades de desviación estándar.

IV. La distribución de probabilidad normal puede aproximarse a una distribución binomial en ciertas condiciones.
  A. $n\pi$ y $n(1 - \pi)$ deben tener un valor mínimo de 5.
    1. *n* es el número de observaciones.
    2. π es la probabilidad de un éxito.
  B. Las cuatro condiciones de una distribución de probabilidad binomial son:
    1. Solo hay dos posibles resultados.
    2. π permanece igual de un ensayo a otro.
    3. Los ensayos son independientes.
    4. La distribución es el resultado de la enumeración del número de éxitos en una cantidad fija de ensayos.
  C. La media y la varianza de una distribución binomial se calculan de la siguiente manera:

$$\mu = n\pi$$

$$\sigma^2 = n\pi(1 - \pi)$$

D. El factor de corrección de continuidad de 0.5 se emplea para extender el valor continuo de *x* media unidad en cualquier dirección. Esta corrección compensa la aproximación a una distribución discreta a través de una distribución continua.

V. La distribución de probabilidad exponencial describe los tiempos entre eventos que forman una secuencia.

A. Las acciones ocurren independientemente, a un ritmo constante por unidad o duración de tiempo.

B. La densidad de la probabilidad se calcula mediante la fórmula:

$$P(x) = \lambda e^{-\lambda x} \tag{7.6}$$

C. No es negativa, su sesgo es positivo, declina uniformemente hacia la derecha, y es asintótica.

D. El área bajo la curva se calcula mediante la fórmula:

$$P(\text{Tiempo de llegada} < x) = 1 - e^{-\lambda x} \tag{7.7}$$

E. Tanto la media como la desviación estándar son $1/\lambda$.

EJERCICIOS DEL CAPÍTULO

41. La cantidad de bebida de cola en una lata de 12 onzas tiene una distribución uniforme entre 11.96 onzas y 12.05 onzas.
- **a.** ¿Cuál es la cantidad media de bebida por lata?
- **b.** ¿Cuál es la desviación estándar de la cantidad de bebida por lata?
- **c.** ¿Cuál es la probabilidad de elegir una lata de bebida que contenga menos de 12 onzas?
- **d.** ¿Cuál es la probabilidad de elegir una lata de bebida que contenga más de 11.98 onzas?
- **e.** ¿Cuál es la probabilidad de elegir una lata de bebida que contenga más de 11 onzas?

42. Un tubo de pasta dental Listerine Control Tartar contiene 4.2 onzas. Conforme la gente utiliza la pasta, la cantidad que queda en cualquier tubo es aleatoria. Suponga que la cantidad de pasta restante en el tubo tiene una distribución uniforme. De acuerdo con estos datos, es posible determinar información relativa a la cantidad restante de un tubo de pasta dental sin invadir la privacidad de nadie.
- **a.** ¿Cuánta pasta esperaría que quedara en el tubo?
- **b.** ¿Cuál es la desviación estándar de la pasta que queda en el tubo?
- **c.** ¿Cuál es la probabilidad de que en el tubo queden menos de 3.0 onzas?
- **d.** ¿Cuál es la probabilidad de que en el tubo queden más de 1.5 onzas?

43. Muchas tiendas de menudeo ofrecen sus propias tarjetas de crédito. En el momento de hacer la solicitud de crédito, el cliente recibe 10% de descuento en su compra. El tiempo que se requiere para el proceso de la solicitud de crédito se rige por una distribución uniforme con tiempos que varían entre 4 y 10 minutos.
- **a.** ¿Cuál es el tiempo medio que dura el proceso de la solicitud?
- **b.** ¿Cuál es la desviación estándar del tiempo de proceso?
- **c.** ¿Cuál es la probabilidad de que una solicitud tarde menos de seis minutos?
- **d.** ¿Cuál es la probabilidad de que una solicitud tarde más de cinco minutos?

44. El tiempo que los huéspedes del hotel Grande Dunes, de Bahamas, esperan el ascensor tiene una distribución uniforme de entre cero y 3.5 minutos.
- **a.** Demuestre que el área bajo la curva es de 1.00.
- **b.** ¿Cuánto tiempo espera el cliente habitual el servicio de elevador?
- **c.** ¿Cuál es la desviación estándar del tiempo de espera?
- **d.** ¿Qué porcentaje de huéspedes espera menos de un minuto?
- **e.** ¿Qué porcentaje de huéspedes espera más de dos minutos?

45. Las ventas netas y el número de empleados de fabricantes de aluminio con características similares están organizados en una distribución de frecuencias. Ambos tienen distribuciones normales. La media de las ventas netas es de 180 millones de dólares, y la desviación estándar, de 25 millones de dólares. En el caso del número de empleados, la media es de 1 500, y la desviación estándar, de 120. Clarion Fabricators realizó ventas por 170 millones de dólares y tiene 1 850 empleados.
- **a.** Convierta las ventas y el número de empleados de Clarion en valores *z.*
- **b.** Localice los dos valores *z.*
- **c.** Compare las ventas de Clarion y su número de empleados con los de otros fabricantes.

46. El departamento de contabilidad de Weston Materials, Inc., fabricante de cocheras desmontables, indica que dos trabajadores de la construcción tardan una media de 32 horas, con una desviación estándar de dos horas, para armar el modelo Red Barn. Suponga que los tiempos de montaje tienen una distribución normal.

a. Determine los valores *z* de 29 y 34 horas. ¿Qué porcentaje de cocheras requiere entre 32 y 34 horas de armado?
b. ¿Qué porcentaje de cocheras requiere entre 29 y 34 horas de armado?
c. ¿Qué porcentaje de cocheras requiere 28.7 horas o menos de armado?
d. ¿Cuántas horas se requieren para armar 5% de las cocheras?

47. Un informe reciente indicaba que una familia común de cuatro miembros gasta 490 dólares al mes en alimentos. Suponga que la distribución de gastos de alimento de una familia de cuatro miembros sigue una distribución normal, con una media de 490 dólares y una desviación estándar de 90 dólares.
a. ¿Qué porcentaje de familias gasta más de 30 dólares y menos de 490 dólares en alimentos al mes?
b. ¿Qué porcentaje de familias gasta menos de 430 dólares al mes en alimentos?
c. ¿Qué porcentaje de familias gasta entre 430 y 600 dólares mensuales en alimentos?
d. ¿Qué porcentaje de familias gasta entre 500 y 600 dólares mensuales en alimentos?

48. Un estudio de llamadas telefónicas de larga distancia que se realizó en las oficinas centrales de General Electric, en Fairfield, Connecticut, demostró que las llamadas, en minutos, se rigen por una distribución de probabilidad normal. El lapso medio de tiempo por llamada fue de 4.2 minutos, con una desviación estándar de 0.60 minutos.
a. ¿Qué porcentaje de llamadas duró entre 4.2 y cinco minutos?
b. ¿Qué porcentaje de llamadas duró más de cinco minutos?
c. ¿Qué porcentaje de llamadas duró entre cinco y seis minutos?
d. ¿Qué porcentaje de llamadas duró entre cuatro y seis minutos?
e. Como parte de su informe al presidente, el director de comunicaciones desea informar la duración de 4% de las llamadas más largas. ¿Cuál es este tiempo?

49. Shaver Manufacturing, Inc., ofrece a sus empleados seguros de atención dental. Un estudio reciente realizado por el director de recursos humanos demuestra que el costo anual por empleado tuvo una distribución de probabilidad normal, con una media de 1 280 dólares y una desviación estándar de 420 dólares anuales.
a. ¿Qué porcentaje de empleados generó más de 1 500 dólares anuales de gastos dentales?
b. ¿Qué porcentaje de empleados generó entre 1 500 y 2 000 dólares anuales de gastos dentales?
c. Calcule el porcentaje que no generó gastos por atención dental.
d. ¿Cuál fue el costo de 10% de los empleados que generó gastos más altos por atención dental?

50. Las comisiones anuales que percibieron los representantes de ventas de Machine Products, Inc., fabricante de maquinaria ligera, tienen una distribución de probabilidad normal. El monto anual medio percibido es de 40 000 dólares, y la desviación estándar, de 5 000 dólares.
a. ¿Qué porcentaje de representantes de ventas percibe más de 42 000 dólares anuales?
b. ¿Qué porcentaje de representantes de ventas percibe entre 32 000 y 42 000 dólares anuales?
c. ¿Qué porcentaje de representantes de ventas percibe entre 32 000 y 35 000 dólares anuales?
d. El gerente desea gratificar a los representantes de ventas que perciben las comisiones más altas con un bono de 1 000 dólares. Lo puede conceder a 20% de ellos. ¿Cuál es el límite entre los que obtienen un bono y quienes no lo obtienen?

51. De acuerdo con el South Dakota Department of Health, la media de la cantidad de horas que se ve televisión a la semana es más alta entre mujeres adultas que entre hombres. Un estudio reciente mostró que las mujeres ven televisión un promedio de 34 horas a la semana, y los hombres, 29. Suponga que la distribución de horas que ven televisión tiene una distribución normal en ambos grupos, y que la desviación estándar entre las mujeres es de 4.5 horas, mientras que en los hombres es de 5.1 horas.
a. ¿Qué porcentaje de mujeres ve televisión menos de 40 horas a la semana?
b. ¿Qué porcentaje de hombres ve televisión más de 25 horas a la semana?
c. ¿Cuántas horas de televisión por semana ve 1% de las mujeres que pasan más tiempo en esta actividad? Encuentre el valor comparable en el caso de los hombres.

52. De acuerdo con un estudio del gobierno, la suma media que gastan cada año en lectura y entretenimiento los adultos de 25 a 34 años de edad es de 1 994 dólares. Suponga que la distribución de las sumas que se gastan tiene una distribución normal, con una desviación estándar de 450 dólares.
a. ¿Qué porcentaje de adultos gastó más de 2 500 dólares anuales en lectura y entretenimiento?
b. ¿Qué porcentaje gastó entre 2 500 y 3 000 dólares anuales en lectura y entretenimiento?
c. ¿Qué porcentaje gastó menos de 1 000 dólares anuales en lectura y entretenimiento?

53. La administración de Gordon Electronics piensa instituir un sistema de bonos para incrementar la producción. Con base en la experiencia previa, se sugiere pagar un bono sobre el 5% más alto de la producción. Los registros anteriores indican que la producción semanal tiene una distribución normal. La media de esta distribución es de 4 000 unidades a la semana, y la desviación estándar, de 60. Si el bono se paga sobre el 5% más alto de producción, ¿a partir de cuántas unidades se debe pagar?

54. Fast Service Truck Lines utiliza exclusivamente el Ford Super Duty F-750. La administración realizó un estudio acerca de los costos de mantenimiento y determinó que el número de millas que se recorrieron durante el año tenía una distribución normal. La media de la distribución fue de 60 000 millas, y la desviación estándar, de 2 000.

- **a.** ¿Qué porcentaje de los Ford Super Duty-750 registró en su bitácora 65 200 millas o más?
- **b.** ¿Qué porcentaje de los Ford Super Duty-750 registró en su bitácora más de 57 060 millas y menos de 58 280?
- **c.** ¿Qué porcentaje de los Ford Super Duty-750 recorrió 62 000 millas o menos durante el año?
- **d.** ¿Es razonable concluir que ninguno de los camiones recorrió más de 70 000 millas? Explique.

55. Best Electronics, Inc., promueve una política de devoluciones *sin complicaciones.* La cantidad de artículos devueltos al día tiene una distribución normal. La cantidad media de devoluciones de los clientes es de 10.3 al día, y la desviación estándar, de 2.25.

- **a.** ¿Qué porcentaje de días hay ocho o menos clientes que devuelven artículos?
- **b.** ¿Qué porcentaje de días hay entre 12 y 14 clientes que devuelven artículos?
- **c.** ¿Existe alguna probabilidad de que haya un día sin devoluciones?

56. Un informe noticioso reciente señala que 20% de los empleados le roba a la empresa cada año. Considere que una compañía tiene 50 empleados. Determine la probabilidad de que:

- **a.** Menos de 5 empleados roben.
- **b.** Más de 5 empleados roben.
- **c.** Exactamente 5 empleados roben.
- **d.** Más de cinco empleados y menos de 15 roben.

57. Como parte de su suplemento dominical dedicado a la salud, el diario *Orange County Register* informó que 64% de los varones estadounidenses mayores de 18 años considera que la nutrición es una prioridad en su vida. Suponga que se elige una muestra de 60 hombres. Establezca la probabilidad de que:

- **a.** Al menos 32 hombres consideren importante la nutrición.
- **b.** Al menos 44 hombres la consideren importante.
- **c.** Más de 32 y menos de 43 la consideren importante.
- **d.** Exactamente 44 hombres la consideren importante.

58. Se calcula que 10% de los alumnos que presentan la parte correspondiente a métodos cuantitativos del examen Certified Public Account (CPA) la reprobará. Este sábado presentarán el examen 60 estudiantes.

- **a.** ¿Cuántos esperaría que reprueben? ¿Cuál es la desviación estándar?
- **b.** ¿Cuál es la probabilidad de que reprueben exactamente dos estudiantes?
- **c.** ¿Cuál es la probabilidad de que reprueben por lo menos dos estudiantes?

59. La División de Tráfico de Georgetown, Carolina del Sur, informó que 40% de las persecuciones de automóviles da como resultado algún accidente grave o leve. Durante el mes en que ocurren 50 persecuciones de alta velocidad, ¿cuál es la probabilidad de que 25 o más terminen en un accidente grave o leve?

60. Del total de habitaciones de los cruceros de la línea Royal Viking, 80% se encuentra ocupado durante septiembre. En el caso de un crucero con 800 habitaciones, ¿cuál es la probabilidad de que 665 o más habitaciones se encuentren ocupadas ese mes?

61. El objetivo de los aeropuertos de Estados Unidos que tienen vuelos internacionales consiste en autorizarlos en un lapso de 45 minutos. Es decir, 95% de los vuelos se autoriza en un periodo de 45 minutos, y la autorización del 5% restante tarda más. Suponga, asimismo, que la distribución es aproximadamente normal.

- **a.** Si la desviación estándar del tiempo que se requiere para autorizar un vuelo internacional es de cinco minutos, ¿cuál es el tiempo medio para autorizar un vuelo?
- **b.** Suponga que la desviación estándar es de 10 minutos, no los cinco del inciso anterior. ¿Cuál es la nueva media?
- **c.** Un cliente tiene 30 minutos para abordar su limusina a partir del momento que aterriza su avión. Con una desviación estándar de 10 minutos, ¿cuál es la probabilidad de que cuente con tiempo suficiente para subir al vehículo?

62. Los fondos que despacha el cajero automático localizado cerca del área de pago en un centro comercial de Kroger, en Union, Kentucky, tienen una distribución de probabilidad normal con una media de 4 200 dólares al día y una desviación estándar de 720 dólares. La máquina se encuentra programada para notificar al banco más próximo si la cantidad que despacha el cajero es muy baja (menor que 2 500 dólares) o muy alta (más de 6 000 dólares).

- **a.** ¿Qué porcentaje de días se notificará al banco que la cantidad despachada es muy baja?
- **b.** ¿Qué porcentaje de días se notificará al banco que la cantidad despachada es muy alta?
- **c.** ¿Qué porcentaje de días no se notificará al banco la cantidad despachada?

63. Los pesos del jamón enlatado por la compañía Henline Ham tienen una distribución normal, con una media de 9.20 libras y una desviación estándar de 0.25 libras. En la etiqueta aparece un peso de 9.00 libras.
   a. ¿Qué proporción de latas pesa menos de la cantidad que señala la etiqueta?
   b. El propietario, Glen Henline, considera dos propuestas para reducir la proporción de latas con un peso menor al marcado en la etiqueta. Puede incrementar el peso medio a 9.25 y dejar igual la desviación estándar, o dejar el peso medio en 9.20 y reducir la desviación estándar de 0.25 libras a 0.15 libras. ¿Qué cambio le recomendaría?

64. El *Cincinnati Enquirer,* en su suplemento sabatino de negocios, informó que la cantidad media de horas trabajadas por semana por empleados de tiempo completo es de 43.9. El artículo indicó, además, que alrededor de una tercera parte de los empleados de tiempo completo trabaja menos de 40 horas a la semana.
   a. De acuerdo con esta información, y en el supuesto de que la cantidad de horas de trabajo tiene una distribución normal, ¿cuál es la desviación estándar de la cantidad de horas trabajadas?
   b. El artículo indicó incluso que 20% de los empleados de tiempo completo trabaja más de 49 horas a la semana. Determine la desviación estándar con esta información. ¿Son similares las dos aproximaciones de la desviación estándar? ¿Qué concluiría usted?

65. La mayoría de las rentas de automóviles por cuatro años abarcan hasta 60 000 millas. Si el arrendador rebasa esa cantidad, se aplica una sanción de 20 centavos por milla de renta. Suponga que la distribución de millas recorridas en rentas por cuatro años tiene una distribución normal. La media es de 52 000 millas, y la desviación estándar, de 5 000.
   a. ¿Qué porcentaje de rentas generará una sanción como consecuencia del exceso en millas?
   b. Si la compañía automotriz quisiera modificar los términos de arrendamiento de manera que 25 rentas rebasaen el límite de millas, ¿en qué punto debe establecerse el nuevo límite superior?
   c. Por definición, un automóvil de bajo recorrido de millas tiene cuatro años de uso y ha recorrido menos de 45 000 millas. ¿Qué porcentaje de automóviles devueltos se considera de bajo recorrido de millas?

66. El precio de las acciones del Banco de Florida al final de cada jornada de comercialización del año previo se rigió por una distribución normal. Suponga que durante el año hubo 240 jornadas de comercialización. El precio medio fue de 42.00 dólares por acción, y la desviación estándar, de 2.25 dólares.
   a. ¿Qué porcentaje de jornadas tuvo un precio superior a 45.00 dólares? ¿Cuántas jornadas calcularía usted?
   b. ¿Qué porcentaje de jornadas tuvo un precio que osciló entre 38.00 y 40.00 dólares?
   c. ¿Cuál fue el precio de las acciones que se mantuvo *más alto* 15% de las jornadas?

67. Las ventas anuales de novelas románticas tienen una distribución normal. Ahora bien, no se conoce la media ni la desviación estándar. Las ventas son superiores a 470 000 40% del tiempo, y 10%, superiores a 500 000. Determine la media y la desviación estándar.

68. Al establecer garantías en aparatos HDTV, el fabricante pretende establecer los límites de manera que pocos aparatos requieran reparación con cargo a él. Por otra parte, el periodo de garantía debe ser lo bastante prolongado para que la compra resulte atractiva para el comprador. La media del número de meses que abarca la garantía de un aparato HDTV es de 36.84, con una desviación estándar de 3.34 meses. ¿En qué punto deben establecerse los límites de garantía de manera que solo 10% de los aparatos HDTV requiera reparación con cargo al fabricante?

69. DeKorte Tele-Marketing, Inc., considera la compra de una máquina que selecciona aleatoriamente y marca números telefónicos en forma automática. La compañía realiza la mayoría de sus llamadas durante la tarde, así que las que se hacen a teléfonos comerciales son un desperdicio. El fabricante de la máquina argumenta que su programación reduce las llamadas a teléfonos comerciales a 15% del total. Para probar lo que dice, el director de compras de DeKorte programó la máquina para seleccionar una muestra de 150 números telefónicos. ¿Cuál es la probabilidad de que más de 30% de los números seleccionados sean comerciales, asumiendo que el argumento del fabricante es correcto?

70. Un detector de monóxido de carbono en el hogar de la familia Wheelock se activa una vez cada 200 días en promedio. Suponga que esta activación tiene una distribución exponencial. Determine la probabilidad de que:
   a. Haya una alarma dentro de los siguientes 60 días.
   b. Pasen cuando menos 400 días antes de la siguiente alarma.
   c. Pasen entre 150 y 250 días hasta la próxima alarma.
   d. Encuentre el tiempo mediano hasta la siguiente activación.

71. El “tiempo de arranque” (el lapso que transcurre entre la aparición de la pantalla del Bios hasta que el primer archivo se carga en Windows) de la computadora personal de Eric Mouser sigue una distribución exponencial, con una media de 27 segundos. Establezca la probabilidad de que este “arranque” requiera:

a. Menos de 15 segundos.
b. Más de seis segundos.
c. Entre 30 y 45 segundos.
d. ¿Cuál es el punto debajo del cual ocurre solo 10% de los “arranques”?

72. En Estados Unidos, el tiempo entre visitas a una sala de urgencias de un miembro de la población general sigue una distribución exponencial, con una media de 2.5 años. Determine la proporción de la población que:
a. Visitará una sala de urgencias dentro de los próximos seis meses.
b. No la visitará en los próximos seis años.
c. La visitará el siguiente año, pero no este.
d. Encuentre el primer y el tercer cuartiles de esta distribución.

73. Los tiempos entre fallas en una computadora personal siguen una distribución exponencial, con una media de 300 000 horas. Establezca la probabilidad de que:
a. Ocurra una falla en menos de 100 000 horas.
b. No haya fallas en las siguientes 500 000 horas.
c. La siguiente falla ocurra entre 200 000 y 350 000 horas.
d. ¿Cuáles son la media y la desviación estándar del tiempo entre fallas?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

74. Consulte los datos sobre Real Estate, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.
a. El precio de venta medio (en miles de dólares) de las casas se calculó en 221.10 dólares, con una desviación estándar de 47.11 dólares. Utilice la distribución normal para calcular el porcentaje de casas que se vende en más de 280 dólares. Compare con los resultados reales. ¿La distribución normal genera una buena aproximación de los resultados reales?
b. La distancia media desde el centro de la ciudad es de 14.629 millas, con una desviación estándar de 4.874 millas. Utilice la distribución normal para calcular la cantidad de casas que se ubican a 18 o más millas y a menos de 22 millas del centro de la ciudad. Compare con los resultados reales. ¿La distribución normal ofrece una buena aproximación de estos?

75. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012.
a. La asistencia media por equipo en la temporada fue de 2 495 millones, con una desviación estándar de 0.6421 millones. Utilice la distribución normal para calcular el número de equipos con asistencias superiores a 3.5 millones. Compare este resultado con el número real. Haga comentarios acerca de la exactitud del cálculo.
b. El salario medio por equipo fue de 98.023 millones de dólares, con una desviación estándar de 36.832 millones. Utilice la distribución normal para calcular el número de equipos con un salario superior a 50 millones de dólares. Compare este resultado con la cantidad real. Haga comentarios acerca de la exactitud de su aproximación.

76. Consulte los datos de los autobuses del Distrito Escolar Buena.
a. Refiérase a la variable “costo de mantenimiento”. El costo medio de mantenimiento del año previo fue de 450.29 dólares, con una desviación estándar de 53.69 dólares. Estime el número de autobuses con un costo de más de 50 dólares. Compare con el número real.
b. Refiérase a la variable “número de millas recorridas”. La media es 830.11 y la desviación estándar, de 42.19. Estime el número de autobuses que viajan más de 900 millas. Compare con el número del valor real.

REPASO DE LOS CAPÍTULOS 5 a 7

Los capítulos en esta sección estudian los métodos para hacer frente a la incertidumbre. En el capítulo 5 se describió el concepto de probabilidad. Una *probabilidad* es un valor entre 0 y 1, inclusive, que expresa la creencia de que un evento ocurrirá. También se vieron los métodos para calcular probabilidades utilizando las reglas de la adición y la multiplicación; se presentaron algunos principios de conteo, incluyendo las permutaciones y las combinaciones; y se describieron situaciones donde es posible utilizar el teorema de Bayes.

En el capítulo 6 se expusieron las distribuciones de probabilidad *discreta.* Las distribuciones de probabilidad son enumeraciones de todos los posibles resultados de un experimento y la probabilidad asociada con cada uno. Se describieron tres distribuciones de probabilidad discreta: la *distribución binomial,* la *distribución hipergeométrica* y la *distribución de Poisson.* Los requisitos para la distribución binomial son los siguientes: existen solo dos posibles resultados para cada ensayo; hay una probabilidad constante de éxito; existe un número fijo de ensayos y estos son independientes. La distribución binomial enlista las probabilidades para el número de éxitos en un número fijo de ensayos. La distribución hipergeométrica es similar a la binomial, pero la probabilidad de éxito no es constante, así que los ensayos no son independientes. La distribución de Poisson se distingue por una pequeña probabilidad de éxito en un gran número de ensayos y tiene las siguientes características: la variable aleatoria es el número de veces que ocurre un evento en un intervalo fijo; la probabilidad de éxito es proporcional al tamaño del intervalo y los intervalos son independientes y no se sobreponen.

En el capítulo 7 se describieron tres distribuciones de probabilidad continua: la *distribución de probabilidad uniforme,* la *distribución de probabilidad normal* y la *distribución exponencial.* La primera tiene una configuración rectangular y se describe por sus valores mínimo y máximo. La media y mediana son iguales y no tienen moda.

La segunda es la más usada y reportada. Sus principales características son: tiene forma de campana y es simétrica; está completamente descrita por su media y su desviación estándar y es asintótica, esto es, cae suavemente a ambos lados de su pico pero nunca llega a tocar el eje horizontal. Existe una familia de distribuciones normales, cada una con sus propias media y desviación estándar. Hay un número ilimitado de distribuciones de probabilidad normal.

Para encontrar las probabilidades de cualquier distribución de probabilidad normal es preciso convertirla en una *distribución de probabilidad normal estándar* calculando los *valores z*, los cuales miden la distancia entre *x* y la media en unidades de la desviación estándar. La distribución de probabilidad normal estándar tiene una media de 0 y una desviación estándar de 1. Resulta útil porque la probabilidad de cualquier evento a partir de una distribución de probabilidad normal puede calcularse mediante tablas de probabilidad normal estándar (vea el apéndice B.3).

La tercera describe el tiempo entre eventos que ocurren en secuencia. Estos suceden independientemente a un ritmo constante por unidad o duración de tiempo. La distribución de probabilidad exponencial tiene un sesgo positivo, con λ como el parámetro de “ritmo”. La media y la desviación estándar son iguales y son recíprocas de λ.

PROBLEMAS

1. Se dice que Proactine, un nuevo medicamento contra el acné, tiene 80% de eficacia: de cada 100 personas que se lo aplican, 80 muestran progresos significativos. Se aplica en el área afectada en un grupo de 15 personas. Determine la probabilidad de que:
   a. Las 15 muestren mejoras significativas.
   b. Menos de nueve muestren mejoras significativas.
   c. Doce o más personas muestren mejoras significativas.
2. Los clientes del Banco de Comercio de Idaho Falls, Idaho, tienen un rango de 0.005 de falla en el pago de sus préstamos de remodelación. El banco aprobó 400 créditos para mejoras menores de vivienda. Si aplica una distribución de Poisson al problema:
   a. ¿Cuál es la probabilidad de que ninguno de los 400 propietarios de vivienda se retrase en los pagos?
   b. ¿Cuántos de los 400 se espera que se retrasen?
   c. ¿Cuál es la probabilidad de que tres o más propietarios de vivienda se retrasen en el pago de los créditos?
3. Un estudio relacionado con la asistencia de aficionados a los partidos de basquetbol de la Universidad de Alabama reveló que la distribución de la asistencia es normal, con una media de 10 000 y una desviación estándar de 2 000.
   a. ¿Cuál es la probabilidad de que un partido registre una asistencia de 13 500 espectadores o más?
   b. ¿Qué porcentaje de partidos registra una asistencia de entre 8 000 y 11 500 aficionados?
   c. ¿Qué asistencia aproximada se registra en 10% de los partidos?
4. La compañía de seguros Daniel-James asegurará una plataforma marítima de producción de Mobil Oil contra pérdidas ocasionadas por el clima durante un año. El presidente de la aseguradora calcula las siguientes pérdidas (en millones de dólares) con las probabilidades correspondientes.

Monto de las pérdidas (millones de dólares)	Probabilidad de pérdida
0	0.98
40	0.016
300	0.004

a. ¿Cuál es el monto esperado que deberá pagar Daniel-James a Mobil por concepto de demandas?
b. ¿Cuál es la probabilidad de que Daniel-James pierda menos del monto esperado?
c. En caso de que Daniel-James sufra una pérdida, ¿cuál es la probabilidad de que sea de 300 millones de dólares?
d. Daniel-James fijó la prima anual en 2.0 millones de dólares. ¿Es una prima justa? ¿Cubrirá su riesgo?

5. La distribución de la cantidad de niños de edad escolar por familia en el área de Whitehall Estates, de Boise, Idaho, es la siguiente:

Número de niños	0	1	2	3	4
Porcentaje de familias	40	30	15	10	5

a. Determine la media y la desviación estándar del número de niños en edad escolar por familia en la región de Whitehall Estates.
b. Se planea construir una nueva escuela en la región de Whitehall Estates. Es necesario realizar un cálculo aproximado del número de niños en edad escolar. Hay 500 unidades familiares. ¿Cuántos niños habrá?
c. Se necesita información adicional de las familias que tienen niños exclusivamente. Convierta la información anterior de familias con niños. ¿Cuál es la media del número de niños en las familias con niños?

6. En la siguiente tabla se desglosan los miembros del 113o. Congreso de Estados Unidos por afiliación política.

	Partido		
	Demócratas	Republicanos	Total
Cámara	234	201	435
Senado	46	54	100
Total	280	255	535

a. Se elige al azar a un miembro del Congreso. ¿Cuál es la probabilidad de elegir a un republicano?
b. Si la persona elegida es miembro de la Cámara de Representantes, ¿cuál es la probabilidad de que sea republicano?
c. ¿Cuál es la probabilidad de elegir a un miembro de la Cámara de Representantes o a un demócrata?

CASOS

A. Century National Bank

Consulte los datos relativos a Century National Bank. ¿Es razonable que la distribución para verificar los saldos de las cuentas se aproxime a una distribución de probabilidad normal? Determine la media y la desviación estándar de una muestra de 60 clientes. Compare la distribución real con la teórica. Mencione algunos ejemplos específicos y haga comentarios acerca de sus conclusiones.

Divida los saldos de las cuentas en tres grupos de 20 cada uno, y coloque la tercera parte más pequeña en el primer grupo; la siguiente, en el segundo grupo y la parte con mayor saldo en el tercer grupo. Luego, elabore una tabla en la cual se incluya el número de cada una de las categorías de los saldos de las cuentas por sucursal. ¿Parece que las cuentas se relacionan con la sucursal correspondiente? Cite ejemplos o haga comentarios sobre sus conclusiones.

B. Auditor de elecciones

Algunos temas, como el incremento de los impuestos, la revocación de funcionarios electos o la expansión de los servicios públicos, pueden someterse a un referéndum si se recaban suficientes firmas válidas para apoyar la petición. Desafortunadamente, muchas personas firman la petición aunque no estén registradas en el distrito correspondiente, o lo hacen más de una vez.

Sara Ferguson, auditora de elecciones del condado de Venango, tiene que certificar la validez de las firmas antes de presentar la petición de manera oficial. No es de sorprender que su personal se encuentre agobiado de trabajo; así, ella piensa aplicar métodos estadísticos para dar validez a los documentos, los cuales contienen 200 firmas, en lugar de dar validez a cada una de estas. En una reunión profesional reciente, descubrió que, en algunas comunidades del estado, los funcionarios electorales verificaban apenas cinco firmas de cada página y rechazaban toda la página en caso de que dos o más fueran inválidas. Algunas personas piensan que cinco firmas pueden no ser suficientes para tomar una buena decisión. Sugieren verificar 10 firmas y rechazar la página si tres o más son inválidas.

Con el fin de investigar estos métodos, Sara pide a su personal que extraiga los resultados de la última elección y tome una muestra de 30 páginas. Sucede que el personal escogió 14 páginas del distrito de Avondale, 9 del distrito de Midway y 7 de Kingston. Cada página contenía 200 firmas; en los datos que aparecen a continuación se muestra el número de firmas invalidadas en cada página.

Utilice la información para evaluar las dos propuestas de Sara. Calcule la probabilidad de rechazar una página de acuerdo con cada enfoque. ¿Obtendría aproximadamente los mismos resultados si analizara cada firma? Proponga su propio plan y explique por qué podría ser mejor o peor que los dos planes propuestos por Sara.

Avondale	Midway	Kingston
9	19	38
14	22	39
11	23	41
8	14	39
14	22	41
6	17	39
10	15	39
13	20	
8	18	
8		
9		
12		
7		
13		

C. Geoff "aplica" su educación

Geoff Brown es gerente de una pequeña empresa de telemarketing y evalúa la tasa de ventas de sus empleados con experiencia para establecer niveles mínimos con el fin de hacer nuevas contrataciones. Durante las últimas semanas registró el número de llamadas exitosas por hora del personal. Estos datos, que se presentan a continuación, incluyen estadísticas resumidas que formuló con ayuda de un software de estadística. Geoff estudió en la universidad de la comunidad y ha oído sobre los distintos tipos de distribuciones de probabilidad (binomial, normal, hipergeométrica, de Poisson, etc.). ¿Puede dar algunos consejos a Geoff sobre el tipo de distribución que debe emplear para adaptarse a estos datos de la mejor manera posible y decidir cuándo aceptar a un empleado que está a prueba, una vez que alcanza el mayor grado de productividad? Es importante porque implica un incremento salarial para el empleado y, en el pasado, algunos trabajadores a prueba abandonaron el empleo debido a que se desalentaron porque no cumplieron con los requisitos.

Las llamadas de ventas exitosas por hora durante la semana del 14 de agosto son las siguientes:

4	2	3	1	4	5	5	2	3	2	2	4	5	2	5	3	3	0
1	3	2	8	4	5	2	2	4	1	5	5	4	5	1	2	4	

Estadística descriptiva:

N	MEDIA	MEDIANA	MDIATR	DESSTD	MEDIASE
35	3.229	3.000	3.194	1.682	0.284
MÍN	MÁX	Q1	Q3		
0.0	8.000	2.000	5.000		

¿Qué distribución piensa que debe utilizar Geoff para su análisis?

D. Tarjeta de crédito del banco CNP

Por lo general, antes que un banco emita una tarjeta de crédito clasifica (o califica) al cliente en función de la probabilidad de que resulte rentable. La siguiente se considera una tabla habitual de calificaciones:

Edad	**Menos de 25** (12 pts.)	**25-29** (5 pts.)	**30-34** (0 pts.)	**35+** (18 pts.)
Tiempo viviendo en la misma dirección	**<1 año** (9 pts.)	**1-2 años** (0 pts.)	**3-4 años** (13 pts.)	**5+ años** (20 pts.)
Antigüedad de su automóvil	**Ninguno** (18 pts.)	**0-1 año** (12 pts.)	**2-4 años** (13 pts.)	**5+ años** (3 pts.)
Pago mensual de automóvil	**Ninguno** (15 pts.)	**\$1-\$99** (6 pts.)	**\$100-\$299** (4 pts.)	**\$300+** (0 pts.)
Costo de vivienda	**\$1-\$199** (0 pts.)	**\$200-\$399** (10 pts.)	**Propia** (12 pts.)	**Vive con parientes** (24 pts.)
Cuenta de cheques o ahorros	**Ambas** (15 pts.)	**Solo cheques** (3 pts.)	**Solo ahorros** (2 pts.)	**Ninguna** (0 pts.)

La calificación es la suma de los puntos de los seis rubros. Por ejemplo, Sushi Brown tiene menos de 25 años (12 puntos); ha vivido en el mismo domicilio durante dos años (0 puntos); desde hace cuatro años es dueña de un automóvil (13 puntos), por el que realiza pagos de 75 dólares (6 puntos); realiza gastos domésticos de 200 dólares (10 puntos) y posee una cuenta de cheques (3 puntos). La calificación que obtendría sería de 44.

Después, con una segunda tabla, se convierten las calificaciones en probabilidades de rentabilidad del cliente. A continuación se muestra una tabla de esta clase.

Puntuación	30	40	50	60	70	80	90
Probabilidad	0.70	0.78	0.85	0.90	0.94	0.95	0.96

La puntuación de Sushi (44) se traduciría en una probabilidad de rentabilidad aproximada de 0.81. En otras palabras, 81% de los clientes como Sushi generarían dinero a las operaciones con tarjeta del banco.

A continuación se muestran los resultados de las entrevistas con los tres posibles clientes.

Nombre	David Born	Edward Brendan	Ann McLaughlin
Edad	42	23	33
Tiempo de vivir en el mismo domicilio	9	2	5
Antigüedad con el auto	2	3	7
Pago mensual del auto	$140	$99	$175
Costo de vivienda	$300	$200	Propietario
Cuenta de cheques o ahorros	Ambas	Solo de cheques	Ninguna

1. Califique a cada uno de estos clientes y calcule la probabilidad de que resulten rentables.
2. ¿Cuál es la probabilidad de que los tres resulten rentables?
3. ¿Cuál es la probabilidad de que ninguno sea rentable?
4. Determine la distribución de probabilidad total del número de clientes rentables entre este grupo de tres clientes.
5. Redacte un breve resumen de sus hallazgos.

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. ¿Bajo qué condiciones una probabilidad sería mayor a 1 o 100%? **1.** ________
2. Un ________ es la observación de alguna actividad o el acto de tomar algún tipo de medida. **2.** ________
3. Un ________ es la recolección de uno o más resultados de un experimento. **3.** ________
4. Una probabilidad ________ implica que dos o más eventos ocurrirán al mismo tiempo. **4.** ________
5. En una (5a) ________, el orden en que se cuentan los eventos es importante, pero en una (5b) ________ no lo es. **5. a.** ________ **5. b.** ________
6. En una distribución de probabilidad discreta, la suma de los posibles resultados es igual a ________. **6.** ________
7. ¿Cuál de los siguientes *NO* es un requisito para la distribución binomial? Probabilidad constante de éxito, tres o más resultados, el resultado de los conteos. **7.** ________
8. ¿Cuántas distribuciones normales existen? Elija una: 1, 10, 30, 1 000, o infinitas. **8.** ________
9. ¿Cuántas distribuciones estándar existen? Elija una: 1, 10, 30, 1 000, o infinitas. **9.** ________
10. ¿Cuál es la probabilidad de encontrar un valor z entre 0 y -0.76? **10.** ________
11. ¿Cuál es la probabilidad de encontrar un valor z mayor a 1.67? **11.** ________
12. Dos eventos son ________ si la ocurrencia de uno de ellos no afecta la del otro. **12.** ________
13. Dos eventos son ________ si por virtud de que ocurra uno, el otro no puede suceder. **13.** ________
14. ¿Cuál de los siguientes conceptos es falso con respecto a la distribución de probabilidad normal? Asintótico, familia de distribuciones, solo dos resultados, 50% de las observaciones son mayores que la media. **14.** ________
15. ¿Cuál de los siguientes conceptos describe mejor la forma de una distribución de probabilidad normal? Forma de campana, uniforme, forma de V, no hay forma constante. **15.** ________

Parte 2: Problemas

1. El contador Fred Friendly tiene que preparar 20 declaraciones de impuestos antes de la fecha límite del 15 de abril. Ya es de noche, así que decide hacer dos más antes de irse a casa. En su paquete de cuentas, 12 son personales, 5 comerciales y 3 pertenecen a organizaciones de caridad. Selecciona dos al azar. Determine la probabilidad de que:
   a. Ambas sean comerciales.
   b. Al menos una sea comercial.
2. El IRS reporta que 15% de las declaraciones donde el ingreso bruto ajustado asciende a más de un millón de dólares estará sujeto a una auditoría por computadora. Durante el ejercicio de 2013, el contador Fred Friendly completó 16 declaraciones donde el ingreso bruto ajustado superaba esa cantidad.
   a. ¿Cuál es la probabilidad de que exactamente una de estas declaraciones sea auditada?
   b. ¿Cuál es la probabilidad de al menos una sea auditada?
3. Fred trabaja en un despacho fiscal junto con otros cinco contadores. Hay cinco lugares de estacionamiento a un lado de la oficina. ¿En cuántas formas diferentes pueden disponerse los autos de los contadores en los cinco lugares? Asuma que todos usan su auto para ir a trabajar.
4. Fred decidió estudiar el número de exenciones reclamadas en las declaraciones personales de impuestos que preparó en 2007. Los datos se resumen en la tabla de la derecha:
   a. ¿Cuál es el número medio de exenciones por declaración?
   b. ¿Cuál es la varianza del número de exenciones por declaración?

Exenciones	Porcentaje
1	20
2	50
3	20
4	10

5. En un memorándum a todos los involucrados en la preparación de las declaraciones de impuestos, el IRS indicó que la cantidad media de reembolsos fue 1 600 dólares, con una desviación estándar de 850 dólares. Asuma que la distribución de las cantidades devueltas sigue una distribución normal.

a. ¿Qué porcentaje de las devoluciones estuvo entre 1 600 y 2 000 dólares?
b. ¿Qué porcentaje de las devoluciones estuvo entre 900 y 2 000 dólares?
c. De acuerdo con la información anterior, ¿qué porcentaje de las devoluciones fue inferior a cero dólares? Es decir, el contribuyente aún le debe al IRS.

6. Durante el ejercicio de 2013, Fred Friendly completó un total de 80 declaraciones. Desarrolló la siguiente tabla que resume la relación entre el número de dependientes económicos y el hecho de que el cliente recibiera o no una devolución.

	Dependientes			
Devolución	**1**	**2**	**3 o más**	**Total**
Sí	20	20	10	50
No	10	20	0	30
Total	30	40	10	80

a. ¿Qué nombre se le da a esta tabla?
b. ¿Cuál es la probabilidad de seleccionar un cliente que recibió una devolución?
c. ¿Cuál es la probabilidad de seleccionar un cliente que recibió una devolución o tenía un dependiente?
d. Dado que el cliente recibió una devolución, ¿cuál es la probabilidad de que tuviera un dependiente?
e. ¿Cuál es la probabilidad de seleccionar un cliente que *no* recibió devolución y tenía un dependiente?

7. El IRS permite a los contribuyentes elegir que la dependencia calcule la cantidad de la devolución de sus impuestos. Durante una época muy ocupada, el número de declaraciones que recibió el Centro de Servicio Springfield, que solicitó este servicio, sigue una distribución de Poisson con una media de tres por día. Determine la probabilidad de que en un día en particular:
a. No haya solicitudes.
b. Aparezcan exactamente tres solicitudes.
c. Se efectúen cinco o más solicitudes.
d. No haya solicitudes en dos días consecutivos.

8 Métodos de muestreo y teorema central del límite

EL INFORME ANUAL DE NIKE indica que el estadounidense promedio compra 6.5 pares de zapatos deportivos al año. Suponga que la desviación estándar de la población es de 2.1 y que se analizará una muestra de 81 clientes el próximo año. ¿Cuál es el error estándar de la media en este experimento? (vea el ejercicio 45 y el OA8-4).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA8-1 Explicar por qué se muestrean las poblaciones, describir cuatro métodos para seleccionar una muestra.

OA8-2 Definir un error de muestreo.

OA8-3 Definir la construcción de una distribución muestral de la media de la muestra.

OA8-4 Enunciar el teorema central del límite y definir el error estándar de la distribución muestral de la media.

OA8-5 Aplicar el teorema central del límite para calcular probabilidades.

Introducción

En los capítulos 2 a 4 se hizo hincapié en las técnicas para describir datos. Con el fin de ilustrar dichas técnicas, se organizaron las ganancias obtenidas por los 180 vehículos que el mes anterior vendió Applewood Auto Group en una distribución de frecuencias para calcular las diversas medidas de ubicación y dispersión. Dichas medidas, como la media y la desviación estándar, describen el precio de venta habitual y la dispersión de las ganancias. En esos capítulos se destacó la descripción de la condición de los datos; es decir, se describió algo que ya había sucedido.

En el capítulo 5 se comenzó a establecer el fundamento de la inferencia estadística con el estudio de la probabilidad. Recuerde que, en la inferencia estadística, el objetivo es determinar algo sobre una *población* a partir de solo una *muestra*. La población es todo el grupo de individuos u objetos en estudio, y la muestra es una parte o subconjunto de dicha población. En el capítulo 6 se ampliaron los conceptos de probabilidad al describir tres distribuciones de probabilidad discreta: binomial, hipergeométrica y de Poisson. En el capítulo 7 se describieron tres distribuciones de probabilidad continua: la uniforme, la normal y la exponencial. Las distribuciones de probabilidad abarcan todos los resultados viables de un experimento, así como la probabilidad asociada con cada resultado. Mediante las distribuciones de probabilidad se evaluó la posibilidad de que algo ocurra en el futuro.

En este capítulo comienza el estudio del muestreo, que es el proceso de selección de elementos de una población para hacer juicios o inferencias acerca de esta. Este capítulo se inicia con el análisis de los métodos para seleccionar una muestra de una población. Después, se señala cómo construir una distribución de la media de la muestra para entender la forma en que las medias muestrales tienden a acumularse en torno a la media de la población. Por último, se demuestra que, para cualquier población, la forma de la distribución de muestreo tiende a seguir la distribución de probabilidad normal.

ESTADÍSTICA EN ACCIÓN

Con la importancia del papel que desempeña la estadística inferencial en todas las ramas de la ciencia, es ya una necesidad disponer de fuentes amplias de números aleatorios. En 1927 se publicó el primer libro de números aleatorios, con 41 600 dígitos, generados por L. Tippett. En 1938, R. A. Fisher y E. Yates publicaron 15 000 dígitos aleatorios, generados con dos mazos de barajas. En 1955, RAND Corporation publicó un millón de dígitos aleatorios, generados por pulsos de frecuencia aleatorios de una ruleta electrónica. En 1970, las aplicaciones del muestreo requerían miles de millones de números aleatorios. Desde entonces se han creado métodos para generar, con ayuda de computadoras, dígitos "casi" aleatorios, por lo que se les llama *pseudoaleatorios*. Aún es motivo de debate la pregunta acerca de si un programa de computadora sirve para generar números aleatorios que de verdad lo sean.

Métodos de muestreo

OA8-1 Explicar por qué se muestrean las poblaciones, describir cuatro métodos para seleccionar una muestra.

En el capítulo 1 se mencionó que el propósito de la estadística inferencial consiste en determinar algo sobre una población a partir de una muestra. Una muestra es una porción o parte de la población de interés. En muchos casos, el muestreo resulta más accesible que el estudio de toda la población. En esta sección se explican las razones principales para muestrear y, enseguida, diversos métodos para elegir una muestra.

Razones para muestrear

Cuando se estudian las características de una población, existen diversas razones prácticas para preferir algunas partes (o muestras) de esta para observar y medir. He aquí algunas razones para muestrear:

1. **Establecer contacto con toda la población requiere mucho tiempo.** Un candidato para un puesto federal quizá desee determinar las posibilidades que tiene de resultar elegido. Una encuesta de muestreo en la que se utiliza el personal y las entrevistas de campo convencionales de una empresa especializada en encuestas tardaría uno o dos días. Con el mismo personal y los mismos entrevistadores, y laborando siete días a la semana, se requerirían 200 años para ponerse en contacto con toda la población en edad para votar. Aunque fuera posible reunir a un numeroso equipo de encuestadores, quizá no valdría la pena entrar en contacto con todos los votantes.
2. **El costo de estudiar todos los elementos de una población resulta prohibitivo**. Por lo general, las organizaciones que realizan encuestas de opinión pública y pruebas entre consumidores, como Harris International, CBS News Polls y Zogby International, entran en contacto con menos de 2 000 de las casi 60 millones de familias en Estados Unidos. Una organización que entrevista a consumidores en panel cobra cerca de 40 000 dólares por enviar muestras por correo y tabular las respuestas con el fin de probar un producto (como un cereal para el desayuno, alimento para gato o algún perfume). Esa prueba del producto con 60 millones de familias sería demasiado costosa para valer la pena.
3. **Es imposible verificar de manera física todos los elementos de la población.** Algunas poblaciones son infinitas. Es imposible verificar toda el agua del lago Erie en lo que se refiere a niveles de bacterias, así que se eligen muestras en diversos lugares de este. Las poblaciones

de peces, aves, serpientes o mosquitos son grandes, y se desplazan, nacen y mueren de manera continua. En lugar de intentar contar todos los patos que hay en Canadá o todos los peces del lago Pontchartrain, se hacen aproximaciones mediante diversas técnicas: se cuentan todos los patos que hay en un estanque, capturados al azar, se revisan las cestas de los cazadores o se colocan redes en lugares predeterminados en el lago.

4. **Algunas pruebas son de naturaleza destructiva.** Si los catadores de vino de Sutter Home Winery, California, bebieran todo el vino para evaluar la vendimia, acabarían con la cosecha y no quedaría nada disponible para la venta. Las placas de acero, cables y productos similares, en el área de producción industrial, deben contar con una resistencia mínima a la tensión. Para cerciorarse de que el producto satisface la norma mínima, el departamento de control de calidad elige una muestra de la producción. Cada pieza se somete a tensión hasta que se rompe y se registra el punto de ruptura (medido en libras por pulgada cuadrada). Es obvio que si se sometieran todos los cables o todas las placas a pruebas de resistencia a la tensión no habría productos disponibles para vender o utilizar. Por la misma razón, solo unas cuantas semillas se someten a pruebas de germinación en Burpee Seeds, Inc., antes de la temporada de siembra.
5. **Los resultados de la muestra son adecuados.** Aunque se contara con recursos suficientes, es difícil que la precisión de una muestra de 100% —toda la población— resulte esencial en la mayoría de los casos. Por ejemplo, el gobierno estadounidense utiliza una muestra de tiendas de comestibles distribuidas en ese país para determinar el índice mensual de precios de los alimentos, incluyendo los del pan, frijol, leche y otros productos de primera necesidad. Resulta poco probable que la inclusión de todas las tiendas de comestibles de Estados Unidos influya significativamente en el índice, pues los precios de los productos de primera necesidad no varían más de unos cuantos centavos de una cadena de tiendas a otra.

Muestreo aleatorio simple

Este es el tipo de muestreo más común.

MUESTREO ALEATORIO SIMPLE Muestra seleccionada de manera que cada elemento o individuo de la población tenga las mismas posibilidades de que se le incluya.

Para ejemplificar el muestreo aleatorio simple y la selección, suponga que una población de interés son los 750 jugadores de las Ligas Mayores de Béisbol en activo de los 30 equipos al terminar la temporada 2012. El presidente del sindicato de jugadores desea formar un comité de 10 jugadores para estudiar el tema de las conmociones cerebrales. Una forma de garantizar que cada jugador de la población tenga la misma oportunidad de ser elegido para formar parte del Comité de Conmociones Cerebrales es escribir cada uno de los 750 nombres en un pedazo de papel y colocar todos los papeles en una bolsa. Después de mezclar los papeles, realizar la primera selección sacando uno de ellos de la caja, identificando así al primer jugador. Ese pedazo de papel no se devuelve a la caja; por tanto, la probabilidad de cada selección aumenta. Sin embargo, las diferencias son muy pequeñas: la probabilidad de cada selección es aproximadamente 0.0013, redondeada a cuatro lugares decimales. Este proceso se repite nueve veces más para formar el comité.

Por supuesto, el proceso de escribir todos los nombres de los jugadores en un pedazo de papel se lleva mucho tiempo. Un método más conveniente de seleccionar una muestra aleatoria consiste en utilizar una **tabla de números aleatorios** como la del apéndice B.4. En este caso, el presidente del sindicato prepararía una lista de los 750 jugadores y le asignaría un número del 1 al 750 en un programa de computadora. Utilizando una tabla de números aleatorios, se elegiría al azar un punto de partida en esta y se seleccionarían 10 números de tres dígitos entre el 001 y el 750. También se puede usar una computadora para generar números aleatorios que correspondan a los 10 jugadores seleccionados para formar el comité. Como su nombre lo indica, la probabilidad de seleccionar cualquier número entre el 001 y el 750 es la misma. Así, la probabilidad de seleccionar al jugador 131 es la misma que seleccionar al jugador 722 o 382. Cuando se emplean números aleatorios para hacer selecciones, se elimina cualquier sesgo del proceso.

En el siguiente ejemplo se muestra cómo seleccionar números al azar utilizando una fracción de la tabla de números aleatorios que aparece enseguida. Primero, se elige un punto de partida en la tabla. Una forma de hacerlo es cerrar los ojos y señalar un número de la tabla; cualquiera servirá. Otra forma es elegir de manera fortuita una columna y una fila. Ahora suponga que el reloj marca las 3:04. Utilizando la hora, tres de la tarde, elija la tercera columna y enseguida, usando los minutos desplácese hacia abajo hasta la cuarta fila de números. El número es 03759. Como solo hay 750 jugadores, se utilizan los tres primeros dígitos de un número aleatorio de cinco dígitos. Por lo tanto, 037 es el número del primer jugador que se convertirá en miembro de la muestra. Para continuar seleccionando jugadores, se puede desplazar en cualquier dirección. Suponga que se mueve a la derecha. Los primeros tres dígitos del número a la derecha de 03759 son 447, el número del segundo jugador seleccionado para integrar el comité. El próximo número de tres dígitos a la derecha es 961. Omita 961, así como el siguiente número, 784, porque solo hay 750 jugadores. El tercer jugador seleccionado es el número 189. Continúe este proceso hasta tener 10 jugadores.

50525	57454	28455	68226	34656	38884	39018
72507	53380	53827	42486	54465	71819	91199
34986	74297	00144	38676	89967	98869	39744
68851	27305	03759	44723	96108	78489	18910
06738	62879	03910	17350	49169	03850	18910
11448	10734	05837	24397	10420	16712	94496
		Punto de partida	Segundo jugador			Tercer jugador

ESTADÍSTICA EN ACCIÓN

¿Es discriminación sacar ventaja del físico? Antes de contestar, considere un artículo reciente que apareció en *Personnel Journal*. Sus hallazgos indican que los hombres y mujeres atractivos ganan alrededor de 5% más que los que tienen una apariencia promedio, quienes, a su vez, ganan 5% más que sus compañeros poco agraciados. Esta preferencia afecta tanto a hombres como a mujeres; en gran variedad de ocupaciones, desde la construcción hasta la reparación de automóviles y los empleos de telemarketing, empleos para los que, según se cree, la apariencia no es importante.

Los paquetes estadísticos, como Minitab, y los de hojas de cálculo, como Excel, incluyen una herramienta para seleccionar una muestra aleatoria simple. En el siguiente ejemplo se emplea Excel para elegir una muestra aleatoria de una lista de datos.

EJEMPLO

Jane y Joe Millar administran el Foxtrot Inn, una pensión donde dan alojamiento y desayuno, localizada en Tryon, Carolina del Norte. El negocio tiene ocho habitaciones. A continuación se muestra el número de las que se rentaron diariamente durante junio de 2013. Utilice Excel para seleccionar una muestra de cinco noches de junio.

Junio	Habitaciones en renta	Junio	Habitaciones en renta	Junio	Habitaciones en renta
1	0	11	3	21	3
2	2	12	4	22	2
3	3	13	4	23	3
4	2	14	4	24	6
5	3	15	7	25	0
6	4	16	0	26	4
7	2	17	5	27	1
8	3	18	3	28	1
9	4	19	6	29	3
10	7	20	2	30	3

SOLUCIÓN

Excel selecciona la muestra aleatoria y arroja los resultados. En la primera fecha que se muestreó había cuatro habitaciones rentadas. En la segunda fecha muestreada de junio, se rentaron siete habitaciones. La información aparece en la columna D de la hoja de cálculo. Los pasos se incluyen en la sección "Comandos de software", en el apéndice C. Excel lleva a cabo el muestreo *con* reemplazo (es posible que el mismo día aparezca más de una vez en una muestra).

	A	B	C	D
1	Day of June	Rentals		Sample
2	1	0		4
3	2	2		7
4	3	3		4
5	4	2		3
6	5	3		1
7	6	4		
8	7	2		
9	8	3		
10	9	4		
11	10	7		
12	11	3		
13	12	4		
14	13	4		
15	14	4		

AUTOEVALUACIÓN 8-1

En la siguiente lista se incluyen los estudiantes que se matricularon en un curso de introducción a la estadística administrativa. Se eligen al azar tres de ellos, a quienes se formulan varias preguntas relacionadas con el contenido del curso y el método de enseñanza.

(a) Se escriben a mano los números de 00 hasta 45 en papeletas y se colocan en un recipiente. Los tres números seleccionados son 31, 7 y 25. ¿Qué estudiantes se van a incluir en la muestra?
(b) Ahora utilice la tabla de dígitos aleatorios (apéndice B.4) para seleccionar su propia muestra.
(c) ¿Qué haría si localizara el número 59 en la tabla de números aleatorios?

CSPM 264 01 BUSINESS & ECONOMIC STAT
8:00 AM 9:40 AM MW ST 118 LIND D

RANDOM NUMBER	NAME	CLASS RANK	RANDOM NUMBER	NAME	CLASS RANK
00	ANDERSON, RAYMOND	SO	23	MEDLEY, CHERYL ANN	SO
01	ANGER, CHERYL RENEE	SO	24	MITCHELL, GREG R	FR
02	BALL, CLAIRE JEANETTE	FR	25	MOLTER, KRISTI MARIE	SO
03	BERRY, CHRISTOPHER G	FR	26	MULCAHY, STEPHEN ROBERT	SO
04	BOBAK, JAMES PATRICK	SO	27	NICHOLAS, ROBERT CHARLES	JR
05	BRIGHT, M. STARR	JR	28	NICKENS, VIRGINIA	SO
06	CHONTOS, PAUL JOSEPH	SO	29	PENNYWITT, SEAN PATRICK	SO
07	DETLEY, BRIAN HANS	JR	30	POTEAU, KRIS E	JR
08	DUDAS, VIOLA	SO	31	PRICE, MARY LYNETTE	SO
09	DULBS, RICHARD ZALFA	JR	32	RISTAS, JAMES	SR
10	EDINGER, SUSAN KEE	SR	33	SAGER, ANNE MARIE	SO
11	FINK, FRANK JAMES	SR	34	SMILLIE, HEATHER MICHELLE	SO
12	FRANCIS, JAMES P	JR	35	SNYDER, LEISHA KAY	SR
13	GAGHEN, PAMELA LYNN	JR	36	STAHL, MARIA TASHERY	SO
14	GOULD, ROBYN KAY	SO	37	ST. JOHN, AMY J	SO
15	GROSENBACHER, SCOTT ALAN	SO	38	STURDEVANT, RICHARD K	SO
16	HEETFIELD, DIANE MARIE	SO	39	SWETYE, LYNN MICHELE	SO
17	KABAT, JAMES DAVID	JR	40	WALASINSKI, MICHAEL	SO
18	KEMP, LISA ADRIANE	FR	41	WALKER, DIANE ELAINE	SO
19	KILLION, MICHELLE A	SO	42	WARNOCK, JENNIFER MARY	SO
20	KOPERSKI, MARY ELLEN	SO	43	WILLIAMS, WENDY A	SO
21	KOPP, BRIDGETTE ANN	SO	44	YAP, HOCK BAN	SO
22	LEHMANN, KRISTINA MARIE	JR	45	YODER, ARLAN JAY	JR

ESTADÍSTICA EN ACCIÓN

Los métodos de muestreo aleatorio y sin sesgos son muy importantes para realizar inferencias estadísticas válidas. En 1936 se efectuó un sondeo de opinión para predecir el resultado de la carrera presidencial entre Franklin Roosevelt y Alfred Landon. Se enviaron 10 millones de papeletas en forma de postales retornables gratuitas a domicilios tomados de directorios telefónicos y registros de automóviles. Se contestó una alta proporción de papeletas, con 59% en favor de Landon y 41%, de Roosevelt. El día de la elección, Roosevelt obtuvo 61% de los votos; Landon, 39%. Sin duda, a mediados de la década de 1930 la gente que tenía teléfono y automóvil no era representativa de los votantes estadounidenses.

Muestreo aleatorio sistemático

El procedimiento de muestreo aleatorio simple resulta complicado en algunos estudios. Por ejemplo, Stood's Grocery Market necesita muestrear a sus clientes para estudiar el lapso de tiempo que pasan en la tienda. El muestreo aleatorio simple no es efectivo. Prácticamente no hay una lista de clientes, así que es imposible asignarles números aleatorios. En su lugar, es posible aplicar el **muestreo aleatorio sistemático** para seleccionar una muestra representativa. Aplicando este método para Stood's Grocery Market, usted decide seleccionar 100 clientes durante cuatro días, de lunes a jueves, 25 al día, comenzando el muestreo a distintas horas: 8:00, 11:00, 16:00 y 19:00. Registra los cuatro horarios y los cuatro días en una hoja de papel y los pone en dos sombreros, uno para los horarios y otro para los días. Elige un papel de cada sombrero para garantizar que cada día tendrá asignado un horario aleatorio. Suponga que comienza el lunes a las 16:00. Después, selecciona un número aleatorio entre 1 y 10: 6. El proceso inicia el lunes a las 16:00, escogiendo al sexto cliente que entra en la tienda. Después, elige cada décimo (16o., 26o., 36o.) cliente hasta alcanzar la meta de 25 y, para cada uno, mide el tiempo que pasa en la tienda.

MUESTREO ALEATORIO SISTEMÁTICO Se selecciona un punto aleatorio de inicio y posteriormente se elige cada *k*-ésimo miembro de la población.

El muestreo aleatorio simple se utiliza para seleccionar los días, los horarios y el punto de partida; pero el procedimiento sistemático se emplea para seleccionar al cliente real.

Antes de aplicar el muestreo aleatorio sistemático observe con cuidado el orden físico de la población; cuando este se relacione con la característica de la población, no lo utilice porque la muestra puede tener un sesgo. Por ejemplo, si quiere auditar las facturas en un cajón de archivo que se acomodaron en orden ascendente con base en los montos, el muestreo aleatorio sistemático no garantiza una muestra aleatoria y sin sesgos; por tanto, aplique otros métodos de muestreo.

Muestreo aleatorio estratificado

Cuando una población se divide en grupos a partir de ciertas características, el **muestreo aleatorio estratificado** garantiza que cada grupo o **estrato** se encuentre representado en la muestra. Por ejemplo, los estudiantes universitarios se pueden agrupar en alumnos de tiempo completo o de medio tiempo, por sexo (masculino o femenino) o grado (primero, segundo, tercero o cuarto). Usualmente, los estratos se forman con base en los atributos o características compartidos entre los miembros. Se toma una muestra aleatoria de cada uno en un número proporcional al tamaño del estrato comparado con la población; tras definirlos se aplica el muestreo aleatorio simple en cada grupo para formar la muestra.

MUESTRA ALEATORIA ESTRATIFICADA Una población se divide en subgrupos, denominados *estratos,* y se selecciona al azar una muestra de cada uno.

Por ejemplo, puede estudiar los gastos en publicidad de las 352 empresas más grandes de Estados Unidos. El objetivo del estudio consiste en determinar si las empresas con altos rendimientos sobre el capital (una medida de rentabilidad) gastan en publicidad más dinero que las empresas con un registro de bajo rendimiento o déficit. Para asegurar que la muestra sea una representación imparcial de las 352 empresas, estas se deben agrupar de acuerdo con su rendimiento porcentual sobre el capital. En la tabla 8.1 se incluyen los estratos y las frecuencias relativas. Si aplicara el muestreo aleatorio simple, las empresas del tercero y cuarto estratos tendrían una probabilidad alta de ser seleccionadas (0.87), mientras que las empresas de los demás estratos tendrían muchas menos (0.13). Podría no seleccionar ninguna de las empresas que aparecen en el primer o quinto estratos *sencillamente por azar*; no obstante, el muestreo aleatorio estratificado garantiza que por lo menos una empresa de estos aparezca en la muestra.

TABLA 8.1 Número seleccionado de una muestra aleatoria estratificada proporcional

Estrato	Probabilidad (recuperación de capital)	Número de empresas	Frecuencia relativa	Número muestreado
1	30% y más	8	0.02	1*
2	20 hasta 30%	35	0.10	5*
3	10 hasta 20%	189	0.54	27
4	0 hasta 10%	115	0.33	16
5	Déficit	5	0.01	1
Total		352	1.00	50

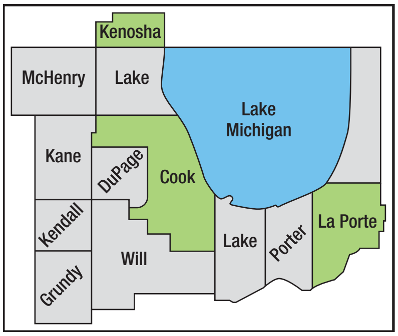
* 0.02 de 50 = 1, 0.10 de 50 = 5, etcétera.

Considere una selección de 50 compañías para llevar a cabo un estudio minucioso. Entonces, con base en la probabilidad seleccione de forma aleatoria una o (0.02 × 50) empresa del estrato 1; cinco (0.10 × 5), del estrato 2, etcétera. En este caso, el número de empresas en cada estrato es proporcional a la frecuencia relativa de este en la población. El muestreo estratificado ofrece la ventaja de que, en algunos casos, refleja con mayor fidelidad las características de la población que el muestreo aleatorio simple o el aleatorio sistemático.

Muestreo por conglomerados

Este es otro tipo común de muestreo; a menudo se emplea para reducir el costo de muestrear una población dispersa en cierta área geográfica.

MUESTREO POR CONGLOMERADOS La población se divide en conglomerados a partir de los límites naturales geográficos u otra clase. A continuación, estos se seleccionan al azar y se toma una muestra de forma aleatoria con elementos de cada grupo.



GRÁFICA 8.1 Condados de la gran zona metropolitana de Chicago, Illinois

Suponga que desea determinar la opinión de los residentes de la gran zona urbana de Chicago, Illinois, con referencia a las políticas federales y estatales de protección ambiental. Seleccionar una muestra aleatoria de residentes de la región y ponerse en contacto con cada persona requeriría mucho tiempo y resultaría muy costoso. Es mejor aplicar el muestreo por conglomerados y subdividir el estado en pequeñas unidades (o *unidades primarias*), tal vez por condados.

Hay 12 condados en la gran zona urbana de Chicago. Suponga que seleccionó al azar tres regiones: La Porte, Cook y Kenosha (vea la gráfica 8.1). Después, toma una muestra aleatoria de los residentes de cada uno de estos condados y los entrevista (esto se conoce como muestreo a través de una *unidad intermedia*). En este caso, la unidad intermedia es el condado (observe que se trata de una combinación de un muestreo por conglomerados y un muestreo aleatorio simple).

En el estudio de los métodos de muestreo de las secciones anteriores no se incluyen todos los métodos para el investigador. Si usted emprende un proyecto de investigación importante de marketing, finanzas, contabilidad u otras áreas, necesitará consultar libros dedicados exclusivamente a la teoría del muestreo y al diseño de muestras.

AUTOEVALUACIÓN 8-2

Consulte la autoevaluación 8.1 y la lista de alumnos de la sección “Muestreo aleatorio simple”. Suponga que en un muestreo aleatorio sistemático se debe elegir a cada noveno estudiante de la clase. Al principio se elige al azar al cuarto alumno de la lista; quien es el número 03. Recuerde que los números aleatorios comienzan con 00, entonces, ¿qué estudiantes se elegirán como miembros de la muestra?

EJERCICIOS

1. En la siguiente lista se registran las 24 tiendas de Marco’s Pizza en el condado Lucas; las cuales se identifican con números 00 hasta 23. También se indica si la tienda es propiedad de alguna corporación (C) o del administrador (A). Seleccione e inspeccione una muestra de cuatro establecimientos en relación con la conveniencia para el cliente, la seguridad, la higiene y otras características.

Número de identificación	Dirección	Tipo	Número de identificación	Dirección	Tipo
00	2607 Starr Av	C	12	2040 Ottawa River Rd	C
01	309 W Alexis Rd	C	13	2116 N Reynolds Rd	C
02	2652 W Central Av	C	14	3678 Rugby Dr	C
03	630 Dixie Hwy	A	15	1419 South Av	C
04	3510 Dorr St	C	16	1234 W Sylvania Av	C
05	5055 Glendale Av	C	17	4624 Woodville Rd	A
06	3382 Lagrange St	A	18	5155 S Main	a
07	2525 W Laskey Rd	C	19	106 E Airport Hwy	C
08	303 Louisiana Av	C	20	6725 W Central	A
09	149 Main St	C	21	4252 Monroe	C
10	835 S McCord Rd	A	22	2036 Woodville Rd	C
11	3501 Monroe St	A	23	1316 Michigan Av	A

   a. Los números aleatorios seleccionados son 08, 18, 11, 02, 41 y 54, ¿qué tiendas se eligieron?
   b. Utilice una tabla de números aleatorios para seleccionar su propia muestra de establecimientos.
   c. Una muestra consta de cada séptimo establecimiento, y el número 03 se selecciona como punto de partida, ¿qué establecimientos se incluirán en la muestra?
   d. Una muestra consta de tres establecimientos, de los cuales dos son propiedad corporativa y uno del administrador. Seleccione una muestra adecuada.

2. En la lista que aparece en la página siguiente se registran los 29 hospitales que se localizan en las regiones de Cincinnati (Ohio) y la región norte de Kentucky; los cuales se identifican con los números 00 hasta 28. También se menciona si se trata de un hospital general médico y quirúrgico (M/Q), o de especialidades (E). Calcule el promedio de enfermeras que trabajan medio tiempo en los hospitales del área.
   a. Seleccione una muestra aleatoria de siete hospitales. Los números aleatorios son: 09, 16, 00, 49, 54, 12 y 04, ¿qué hospitales se incluirán en la muestra?
   b. Utilice una tabla de números aleatorios para formar su propia muestra de cinco hospitales.

Número de identificación	Nombre	Dirección	Tipo
00	Bethesda North	10500 Montgomery Cincinnati, Ohio 45242	M/Q
01	Ft. Hamilton–Hughes	630 Eaton Avenue Hamilton, Ohio 45013	M/Q
02	Jewish Hospital–Kenwood	4700 East Galbraith Rd. Cincinnati, Ohio 45236	M/Q
03	Mercy Hospital–Fairfield	3000 Mack Road Fairfield, Ohio 45014	M/Q
04	Mercy Hospital–Hamilton	100 Riverfront Plaza Hamilton, Ohio 45011	M/Q
05	Middletown Regional	105 McKnight Drive Middletown, Ohio 45044	M/Q
06	Clermont Mercy Hospital	3000 Hospital Drive Batavia, Ohio 45103	M/Q
07	Mercy Hospital–Anderson	7500 State Road Cincinnati, Ohio 45255	M/Q
08	Bethesda Oak Hospital	619 Oak Street Cincinnati, Ohio 45206	M/Q
09	Children's Hospital Medical Center	3333 Burnet Avenue Cincinnati, Ohio 45229	M/Q
10	Christ Hospital	2139 Auburn Avenue Cincinnati, Ohio 45219	M/Q
11	Deaconess Hospital	311 Straight Street Cincinnati, Ohio 45219	M/Q
12	Good Samaritan Hospital	375 Dixmyth Avenue Cincinnati, Ohio 45220	M/Q
13	Jewish Hospital	3200 Burnet Avenue Cincinnati, Ohio 45229	M/Q
14	University Hospital	234 Goodman Street Cincinnati, Ohio 45267	M/Q
15	Providence Hospital	2446 Kipling Avenue Cincinnati, Ohio 45239	M/Q
16	St. Francis–St. George Hospital	3131 Queen City Avenue Cincinnati, Ohio 45238	M/Q
17	St. Elizabeth Medical Center, North Unit	401 E. 20th Street Covington, Kentucky 41014	M/Q
18	St. Elizabeth Medical Center, South Unit	One Medical Village Edgewood, Kentucky 41017	M/Q
19	St. Luke's Hospital West	7380 Turfway Drive Florence, Kentucky 41075	M/Q
20	St. Luke's Hospital East	85 North Grand Avenue Ft. Thomas, Kentucky 41042	M/Q
21	Care Unit Hospital	3156 Glenmore Avenue Cincinnati, Ohio 45211	E
22	Emerson Behavioral Science	2446 Kipling Avenue Cincinnati, Ohio 45239	E
23	Pauline Warfield Lewis Center for Psychiatric Treat.	1101 Summit Road Cincinnati, Ohio 45237	E
24	Children's Psychiatric No. Kentucky	502 Farrell Drive Covington, Kentucky 41011	E
25	Drake Center Rehab—Long Term	151 W. Galbraith Road Cincinnati, Ohio 45216	E
26	No. Kentucky Rehab Hospital—Short Term	201 Medical Village Edgewood, Kentucky	E
27	Shriners Burns Institute	3229 Burnet Avenue Cincinnati, Ohio 45229	E
28	VA Medical Center	3200 Vine Cincinnati, Ohio 45220	E

   c. Una muestra consta de cada quinto establecimiento, y el número 02 se selecciona como punto de partida, ¿qué hospitales se incluirán en la muestra?
   d. Una muestra consta de cuatro hospitales médicos y quirúrgicos, y uno de especialidades. Seleccione una muestra adecuada.

3. Abajo se muestra una lista de los 35 miembros de la Metro Toledo Automobile Dealers Association. Calcule el ingreso medio de los departamentos de servicios de los distribuidores. Los miembros se identifican con números 00 hasta 34.
   a. Seleccione una muestra aleatoria de doce distribuidores. Los números aleatorios son: 05, 20, 59, 21, 31, 28, 49, 38, 66, 08, 29 y 02, ¿qué distribuidores se incluirán en la muestra?
   b. Utilice una tabla de números aleatorios para seleccionar su propia muestra de cinco distribuidores.

Número de identificación	Distribuidor
00	Dave White Acura
01	Autofair Nissan
02	Autofair Toyota-Suzuki
03	George Ball's Buick GMC Truck
04	Yark Automotive Group
05	Bob Schmidt Chevrolet
06	Bowling Green Lincoln Mercury Jeep Eagle
07	Brondes Ford
08	Brown Honda
09	Brown Mazda
10	Charlie's Dodge
11	Thayer Chevrolet/Toyota
12	Spurgeon Chevrolet Motor Sales, Inc.
13	Dunn Chevrolet
14	Don Scott Chevrolet
15	Dave White Chevrolet Co.
16	Dick Wilson Infinity
17	Doyle Buick
18	Franklin Park Lincoln Mercury
19	Genoa Motors
20	Great Lakes Ford Nissan
21	Grogan Towne Chrysler
22	Hatfield Motor Sales
23	Kistler Ford, Inc.
24	Lexus of Toledo
25	Mathews Ford Oregon, Inc.
26	Northtowne Chevrolet
27	Quality Ford Sales, Inc.
28	Rouen Chrysler Jeep Eagle
29	Saturn of Toledo
30	Ed Schmidt Jeep Eagle
31	Southside Lincoln Mercury
32	Valiton Chrysler
33	Vin Divers
34	Whitman Ford

c. Una muestra consta de cada séptimo distribuidor, y el número 04 se selecciona como punto de partida, ¿qué distribuidores se incluirán en la muestra?

4. Enseguida se enumeran los 27 agentes de seguros de Nationwide Insurance en el área metropolitana de Toledo, Ohio. Los agentes se identifican con los números 00 hasta 26. Calcule el promedio de años que han laborado en Nationwide.

Número de identificación	Agente	Número de identificación	Agente	Número de identificación	Agente
00	**Bly Scott** 3332 W Laskey Rd	10	**Heini Bernie** 7110 W Centra	19	**Riker Craig** 2621 N Reynolds Rd
01	**Coyle Mike** 5432 W Central Av	11	**Hinckley Dave** 14 N Holland Sylvania Rd	20	**Schwab Dave** 572 W Dussel Dr
02	**Denker Brett** 7445 Airport Hwy			21	**Seibert John H** 201 S Main
03	**Denker Rollie** 7445 Airport Hwy	12	**Joehlin Bob** 3358 Navarre Av	22	**Smithers Bob** 229 Superior St
04	**Farley Ron** 1837 W Alexis Rd	13	**Keisser David** 3030 W Sylvania Av	23	**Smithers Jerry** 229 Superior St
05	**George Mark** 7247 W Central Av	14	**Keisser Keith** 5902 Sylvania Av	24	**Wright Steve** 105 S Third St
06	**Gibellato Carlo** 6616 Monroe St	15	**Lawrence Grant** 342 W Dussel Dr	25	**Wood Tom** 112 Louisiana Av
07	**Glemser Cathy** 5602 Woodville Rd	16	**Miller Ken** 2427 Woodville Rd	26	**Yoder Scott** 6 Willoughby Av
08	**Green Mike** 4149 Holland Sylvania Rd	17	**O'Donnell Jim** 7247 W Central Av		
09	**Harris Ev** 2026 Albon Rd	18	**Priest Harvey** 5113 N Summit St		

a. Seleccione una muestra aleatoria de nueve agentes. Los números aleatorios son: 02, 59, 51, 25, 14, 29, 77, 69 y 18, ¿qué agentes se incluirán en la muestra?

b. Utilice una tabla de números aleatorios para seleccionar su propia muestra de cuatro agentes.

c. Una muestra consta de cada séptimo distribuidor, y el número 04 se selecciona como punto de partida, ¿qué agentes se incluirán en la muestra?

OA8-2
Definir un error de muestreo.

"Error" de muestreo

En la sección anterior se estudiaron métodos de muestreo útiles para seleccionar una muestra que constituya una representación imparcial, o sin sesgos, de la población. Es importante señalar que, en cada método, la selección de cualquier posible muestra de determinado tamaño de una población tiene una posibilidad conocida que constituye otra forma de describir un método de muestreo sin sesgo.

Las muestras se emplean para determinar características de la población. Por ejemplo, con la media de una muestra se calcula la media de la población; no obstante, como la muestra forma parte o es una porción representativa de la población, es poco probable que su media sea *exactamente igual* a la de la población. Asimismo, es poco factible que la desviación estándar de la muestra sea *exactamente igual* a la de la población; por lo tanto, se puede esperar una diferencia entre un *estadístico de la muestra* y el *parámetro de la población* correspondiente; la cual recibe el nombre de **error de muestreo.**

ERROR DE MUESTREO Diferencia entre el estadístico de una muestra y el parámetro de la población correspondiente.

En el siguiente ejemplo se aclara el concepto de error de muestreo.

EJEMPLO

Revise el ejemplo anterior de la sección "Muestreo aleatorio simple", en el que se estudió el número de habitaciones rentadas en Foxtrot Inn, en Tryon, Carolina del Norte. La población se refiere al número de habitaciones rentadas durante cada uno de los 30 días de junio de 2013. Determine la media de la población. Utilice Excel u otro software de estadística para seleccionar tres muestras aleatorias de cinco días. Calcule la media de cada muestra y compárela con la media poblacional. ¿Cuál es el error de muestreo en cada caso?

SOLUCIÓN

Durante el mes se rentaron un total de 94 habitaciones. Por lo tanto, la media de las unidades que se rentaron por noche es de 3.13. Esta es la media de la población cuyo valor se designa con la letra griega μ.

$$\mu = \frac{\Sigma x}{N} = \frac{0 + 2 + 3 + \cdots + 3}{30} = \frac{94}{30} = 3.13$$

La primera muestra aleatoria de cinco noches dio como resultado el siguiente número de habitaciones rentadas: 4, 7, 4, 3 y 1. La media de esta muestra de cinco noches es de 3.8 habitaciones, que se representa como $\bar{x}_1$. La barra sobre la x recuerda que se trata de una media muestral, y el subíndice 1 indica que se trata de la media de la primera muestra.

$$\bar{x}_1 = \frac{\Sigma x}{n} = \frac{4 + 7 + 4 + 3 + 1}{5} = \frac{19}{5} = 3.80$$

El error de muestreo de la primera muestra es la diferencia entre la media poblacional (3.13) y la media muestral (3.80). De ahí que el error muestral sea $(\bar{x}_1 = \mu) = 3.80 - 3.13 = 0.67$. La segunda muestra aleatoria de cinco días de la población de 30 días de junio arrojó el siguiente número de habitaciones rentadas: 3, 3, 2, 3 y 6. La media de estos cinco valores es de 3.4, que se calcula de esta manera:

$$\bar{x}_2 = \frac{\Sigma x}{n} = \frac{3 + 3 + 2 + 3 + 6}{5} = 3.40$$

El error de muestreo es $(\bar{x}_2 = \mu) = 3.4 - 3.13 = 0.27$. En la tercera muestra aleatoria, la media fue de 1.80, y el error de muestreo fue de -1.33.

Cada una de estas diferencias, 0.67, 0.27 y -1.33, representa el error de muestreo cometido al calcular la media de la población. A veces estos errores son valores positivos, lo cual indica que la media muestral sobreexcedió la media poblacional; otras veces son negativos, lo cual indica que la media muestral es inferior a la media poblacional.

	A	B	C	D	E	F	G
1	Day of June	Rentals			Sample 1	Sample 2	Sample 3
2	1	0			4	3	0
3	2	2			7	3	0
4	3	3			4	2	3
5	4	2			3	3	3
6	5	3			1	6	3
7	6	4		Totals	19	17	9
8	7	2		Sample means	3.80	3.40	1.80
9	8	3					
10	9	4					
11	10	7					
12	11	3					
13	12	4					
14	13	4					
15	14	4					

En este caso, con una población de 30 valores y muestras de cinco, existe una gran cantidad de muestras posibles (exactamente 142 506). Para calcular este valor se aplica la fórmula de las combinaciones [5.10]. Cada una de las muestras cuenta con las mismas posibilidades de que se le seleccione y puede tener una media muestral diferente; es decir, un error de muestreo distinto. El valor del error de muestreo se basa en el valor particular de las 142 506 muestras posibles seleccionadas; por consiguiente, los errores de muestreo son aleatorios y se presentan al azar. Si se determinara la suma de estos errores en una gran cantidad de muestras, el resultado se aproximaría mucho a cero porque la media de la muestra constituye un *estimador sin sesgo* de la media de la población.

Distribución muestral de la media

OA8-3
Definir la construcción de una distribución muestral de la media de la muestra.

En la sección anterior se definió el error de muestreo y se presentaron los resultados de comparar un estadístico para una muestra (como la media de la muestra) con la media de la población; en otras palabras, cuando se usa la media muestral para estudiar la media de la población, ¿cómo se determina la exactitud de la estimación? Determine cómo:

- Un supervisor de calidad decide si una máquina está llenando botellas de 20 onzas con esa cantidad de refresco de cola basándose solamente en una muestra de 10 botellas llenas.
- CNN/*USA Today* o ABC News-*Washington Post* hacen pronósticos precisos sobre los años promedio de estudio de los votantes en una elección presidencial con base en una muestra de 1 200 electores registrados de una población de casi 90 millones.

Para responder estas preguntas, primero hay que precisar el concepto de *distribución muestral de la media.*

Las medias muestrales del ejemplo anterior varían de una muestra a la siguiente. La media de la primera muestra de 5 días fue de 3.80 habitaciones, y la de la segunda fue de 3.40 habitaciones. La media poblacional fue de 3.13 habitaciones. Si se organizan las medias de todas las muestras posibles de 5 días en una distribución de probabilidad, el resultado recibe el nombre de **distribución muestral de la media.**

DISTRIBUCIÓN MUESTRAL DE LA MEDIA Distribución de probabilidad de todas las posibles medias de las muestras de un determinado tamaño muestral de la población.

En el siguiente ejemplo se ilustra la construcción de una distribución muestral de la media. Se utiliza intencionalmente una población pequeña para resaltar la relación entre la media de la población y las diversas medias muestrales.

EJEMPLO

Tartus Industries cuenta con siete empleados de producción (a quienes se les considera la población). En la tabla 8.2 se incluyen los ingresos por hora de cada uno.

TABLA 8.2 Ingresos por hora de los empleados de producción en Tartus Industries

Empleado	Ingresos por hora	Empleado	Ingresos por hora
Joe	$7	Jan	$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

1. ¿Cuál es la media de la población?
2. ¿Cuál es la distribución muestral de la media de muestras de tamaño 2?
3. ¿Cuál es la media de la distribución muestral?
4. ¿Qué observaciones es posible hacer sobre la población y la distribución muestral?

SOLUCIÓN

He aquí las respuestas.

1. La media de la población es de 7.71 dólares, que se determina de la siguiente manera:

$$\mu = \frac{\Sigma x}{N} = \frac{\$7 + \$7 + \$8 + \$8 + \$7 + \$8 + \$9}{7} = \$7.71$$

   Identificamos la media de la población por medio de la letra griega μ. Recuerde que en capítulos anteriores se mencionó que las letras griegas representan parámetros poblacionales.

2. Para obtener la distribución muestral de la media se seleccionaron, sin reemplazos de la población, todas las muestras posibles de tamaño 2 y se calcularon las medias de cada una. Hay 21 muestras posibles, que se calcularon con la fórmula [5.10].

$$_NC_n = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

   donde $N = 7$ es el número de elementos de la población, y $n = 2$, el número de elementos de la muestra.

TABLA 8.3 Medias muestrales de todas las muestras posibles de dos empleados

Muestra	Empleados	Ingresos por hora	Suma	Media	Muestra	Empleados	Ingresos por hora	Suma	Media
1	Joe, Sam	\$7 \$7	\$14	\$7.00	12	Sue, Bob	\$8 \$8	\$16	\$8.00
2	Joe, Sue	7 8	15	7.50	13	Sue, Jan	8 7	15	7.50
3	Joe, Bob	7 8	15	7.50	14	Sue, Art	8 8	16	8.00
4	Joe, Jan	7 7	14	7.00	15	Sue, Ted	8 9	17	8.50
5	Joe, Art	7 8	15	7.50	16	Bob, Jan	8 7	15	7.50
6	Joe, Ted	7 9	16	8.00	17	Bob, Art	8 8	16	8.00
7	Sam, Sue	7 8	15	7.50	18	Bob, Ted	8 9	17	8.50
8	Sam, Bob	7 8	15	7.50	19	Jan, Art	7 8	15	7.50
9	Sam, Jan	7 7	14	7.00	20	Jan, Ted	7 9	16	8.00
10	Sam, Art	7 8	15	7.50	21	Art, Ted	8 9	17	8.50
11	Sam, Ted	7 9	16	8.00					

En la tabla 8.3 se ilustran las 21 medias muestrales de todas las muestras posibles de tamaño 2 que pueden tomarse de la población; estas se utilizan para construir la distribución de probabilidad (distribución muestral de la media) que se resume en la tabla 8.4.

TABLA 8.4 Distribución muestral de la media con $n = 2$

Media muestral	Número de medias	Probabilidad
\$7.00	3	0.1429
7.50	9	0.4285
8.00	6	0.2857
8.50	3	0.1429
	21	1.0000

3. Usando los datos de la tabla 8.3, la media de la distribución muestral de la media se obtiene al sumar las medias muestrales y dividir el resultado entre el número de muestras. La media de todas las medias muestrales se representa mediante $\mu_{\bar{x}}$. La μ recuerda que se trata de un valor poblacional, pues se tomaron en cuenta todas las muestras posibles de dos empleados de la población de ocho. El subíndice $\bar{x}$ indica que se trata de la distribución muestral de la media.

$$\mu_{\bar{x}} = \frac{\text{Suma de todas las medias muestrales}}{\text{Total de muestras}} = \frac{\$7.00 + \$7.50 + \$7.50 + \cdots + \$8.00 + \$8.50}{21}$$

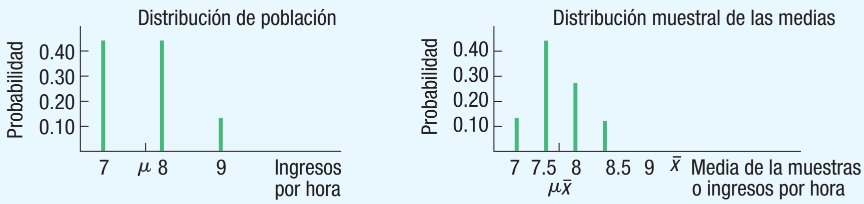
$$= \frac{\$162}{21} = \$7.71$$

4. Consulte la gráfica 8.2, donde se muestra la distribución poblacional basada en los datos de la tabla 8.2 y la distribución muestral de la media basada en los de la tabla 8.4; considere lo siguiente:

a. La media de la distribución muestral de la media (\$7.71) es igual a la media de la población: $\mu = \mu_{\bar{x}}$.

b. La dispersión de la distribución muestral de las medias (varía de \$7.00 hasta \$8.50) es menor que la dispersión de los valores de población (van de \$7.00 hasta \$9.00). Observe que, conforme se incrementa el tamaño de la muestra, se reduce la dispersión de la distribución muestral de las medias.

c. La forma de la distribución muestral de la media y la forma de la distribución de frecuencias de los valores de población son diferentes. La primera tiende a adoptar más forma de campana y a aproximarse a la distribución de probabilidad normal.



GRÁFICA 8.2 Distribución de los valores de población y distribución muestral de las medias

En resumen, se toman todas las posibles muestras aleatorias de una población y se calcula el estadístico muestral (la media de los ingresos percibidos) de cada una. Este ejemplo ilustra las importantes relaciones entre la distribución poblacional y la distribución muestral de la media:

1. La media de las medias de las muestras es exactamente igual a la media de la población.
2. La dispersión de la distribución muestral de la media es más estrecha que la distribución poblacional.
3. La distribución muestral de la media suele tener forma de campana y se aproxima a la distribución de probabilidad normal.

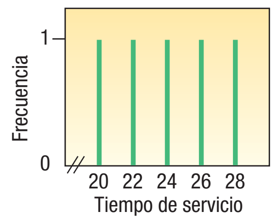
Dada una distribución de probabilidad normal (forma de campana), se aplican los conceptos del capítulo 7 para determinar la probabilidad de seleccionar una muestra con una media muestral específica. En la siguiente sección se resalta la importancia del tamaño de una muestra en relación con la distribución muestral de la media.

AUTOEVALUACIÓN 8-3

Los años de servicio de los ejecutivos que laboran en Standard Chemicals son los que aparecen a la derecha.

(a) De acuerdo con la fórmula de las combinaciones, ¿cuántas muestras de tamaño 2 son posibles?
(b) Elabore una lista de todas las muestras posibles de dos ejecutivos de la población y calcule las medias.
(c) Organice las medias en una distribución muestral.
(d) Compare la media poblacional y la media de las medias de las muestras.
(e) Compare la dispersión en la población con la dispersión de la distribución muestral de la media.
(f) A la derecha se muestra una gráfica con los valores de la población, ¿tienen estos una distribución normal (en forma de campana)?
(g) ¿Tiende la distribución muestral de la media que se calculó en el inciso (c) a adoptar forma de campana?

Nombre	Años
Señor Snow	20
Señora Tolson	22
Señor Kraft	26
Señora Irwin	24
Señor Jones	28



EJERCICIOS

5. Una población consta de los siguientes cuatro valores: 12, 12, 14 y 16.
   a. Enumere todas las muestras de tamaño 2 y calcule la media de cada muestra.
   b. Calcule la media de la distribución muestral de la media y la media de la población; compare ambos valores.
   c. Compare la dispersión en la población con la de las medias de las muestras.
6. Una población consta de los siguientes cinco valores: 2, 2, 4, 4 y 8.
   a. Enumere todas las muestras de tamaño 2 y calcule la media de cada muestra.
   b. Calcule la media de la distribución muestral de las medias y la media de la población; compare ambos valores.
   c. Compare la dispersión en la población con la de las medias de las muestras.
7. Una población consta de los siguientes cinco valores: 12, 12, 14, 15 y 20.
   a. Enumere todas las muestras de tamaño 3 y calcule la media de cada muestra.
   b. Calcule la media de la distribución muestral de las medias y la media de la población; compare ambos valores.
   c. Compare la dispersión de la población con la de las medias de las muestras.
8. Una población consta de los siguientes cinco valores: 0, 0, 1, 3 y 6.
   a. Enumere todas las muestras de tamaño 3 y calcule la media de cada muestra.
   b. Calcule la media de la distribución muestral de las medias y la media de la población; compare ambos valores.
   c. Compare la dispersión de la población con la de las medias de las muestras.
9. El despacho de abogados Tybo and Associates consta de seis socios. En la siguiente tabla se incluye el número de casos que en realidad atendió cada socio en los tribunales durante el mes previo.
   a. ¿Cuántas muestras de tamaño 3 son posibles?
   b. Enumere todas las muestras posibles de tamaño 3 y calcule el número medio de casos en cada muestra.

Socio	Número de casos
Ruud	3
Wu	6
Sass	3
Flores	3
Wilhelms	0
Schueller	1

c. Compare la media de la distribución muestral de las medias con la de la media poblacional.
d. En una gráfica similar a la 8.2, compare la dispersión en la población con la de las medias muestrales.

10. Mid-Motors Ford tiene cinco vendedores. Los cinco representantes de ventas y el número de automóviles que se vendieron la semana pasada aparecen a la derecha:
a. ¿Cuántas muestras de tamaño 2 son posibles?
b. Enumere todas las muestras posibles de tamaño 2 y calcule la media en cada muestra.
c. Compare la media de la distribución muestral de la media con la de la media poblacional.
d. En una gráfica similar a la 8.2, compare la dispersión de la población con la de la media de la muestra.

Representantes de ventas	Autos vendidos
Peter Hankish	8
Connie Stallter	6
Juan Lopez	4
Ted Barnes	10
Peggy Chu	6

Teorema central del límite

OA8-4 Enunciar el teorema central del límite y definir el error estándar de la distribución muestral de la media.

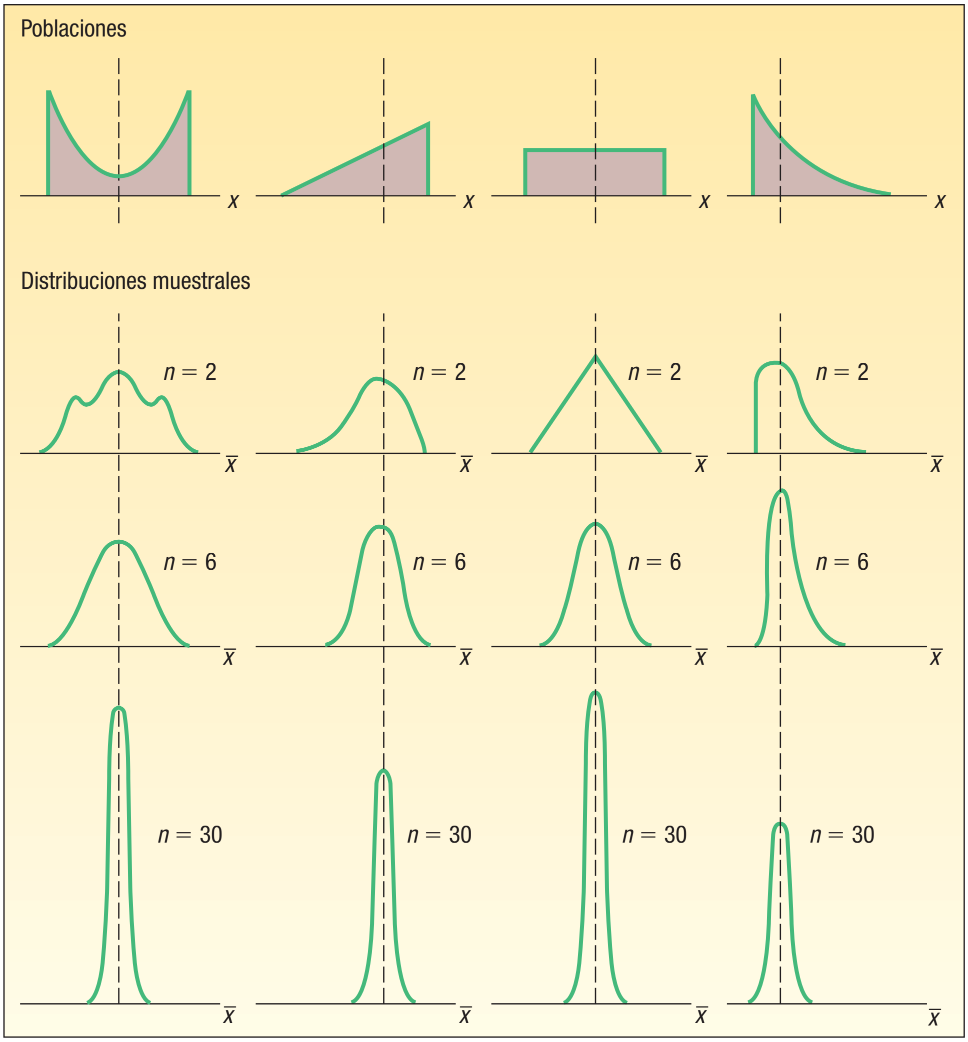
En esta sección se estudia el **teorema central del límite.** Su aplicación a la distribución muestral de medias se introdujo en la sección anterior; esta permite utilizar la distribución de probabilidad normal para crear intervalos de confianza de la media poblacional (vea el capítulo 9) y llevar a cabo pruebas de hipótesis (vea el capítulo 10). El teorema central del límite hace hincapié en que, en el caso de muestras aleatorias grandes, la forma de la distribución muestral de la media se aproxima a la distribución de probabilidad normal. La aproximación es más exacta en el caso de muestras grandes que en el de muestras pequeñas; lo cual es una de las conclusiones más útiles de la estadística porque permite razonar sobre la distribución de las medias muestrales sin ninguna información acerca de la forma de la distribución de la población de la que se toma la muestra. En otras palabras, el teorema central del límite se cumple en el caso de todas las distribuciones.

He aquí el enunciado formal del teorema central del límite.

TEOREMA CENTRAL DEL LÍMITE Si todas las muestras de un tamaño en particular se seleccionan de cualquier población, la distribución muestral de la media se aproxima a una distribución normal; esta mejora con muestras más grandes.

Si la población obedece a una distribución normal, entonces, en el caso de cualquier tamaño de muestra, la distribución muestral de las medias también será de naturaleza normal. Si la distribución poblacional es simétrica (pero no normal), la forma normal de la distribución muestral de las medias se presenta con muestras tan pequeñas como 10. Por otra parte, si se comienza con una distribución sesgada o con colas anchas, quizá se requieran muestras de 30 o más para registrar la característica de normalidad. Este concepto se resume en la gráfica 8.3 para diversas formas de población; observe la convergencia hacia una distribución normal sin que importe la forma de la distribución de la población. La mayoría de los especialistas en estadística consideran que una muestra de 30 o mayor es lo bastante grande para aplicar el teorema central del límite.

La idea de que la distribución muestral de las medias de una población que no es normal converge hacia la normalidad se ilustra en las gráficas 8.4, 8.5 y 8.6. En breve se analizará este ejemplo con más detalles; mientras tanto, en la gráfica 8.4 se muestra una distribución de probabilidad discreta con sesgo positivo. Hay varias muestras posibles de tamaño 5 que pueden seleccionarse de la población de esta gráfica; suponga que selecciona al azar 25 muestras de tamaño 5 cada una y calcula la media de cada muestra, los resultados se indican en la gráfica 8.5. Considere que la forma de la distribución muestral de las medias cambió la forma de la población original aunque solo seleccionó 25 de las diversas muestras posibles. En otras palabras, eligió 25 muestras al azar de tamaño 5 de una población positivamente sesgada, y encontró que la distribución muestral de las medias cambió en lo que se refiere a la forma de la población. A medida que toma muestras más grandes, es decir, $n = 20$ en lugar de $n = 5$, la distribución muestral de las medias se aproximará a la distribución normal. En la gráfica 8.6 se muestran los resultados de 25 muestras aleatorias de 20 observaciones cada una tomadas de la misma población. Observe la clara tendencia hacia la distribución de probabilidad normal; esta es la esencia del teorema central del límite. En el siguiente ejemplo se pone de relieve dicha condición.



GRÁFICA 8.3 Resultados del teorema central del límite para diversas poblaciones

EJEMPLO

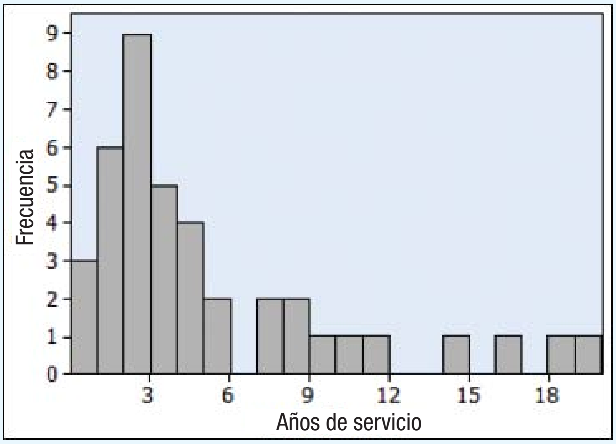
Ed Spence fundó su negocio de engranes hace 20 años; este creció a lo largo del tiempo y ahora cuenta con 40 empleados. Spence Sprockets, Inc., encara algunas decisiones importantes relacionadas con la atención médica de su personal. Antes de tomar una decisión definitiva sobre el programa de atención médica que va a comprar, Ed decide formar un comité de cinco trabajadores y pedirle que estudie el tema del cuidado de la salud y haga alguna recomendación sobre el plan que mejor convenga al personal. Ed cree que el punto de vista de los empleados más recientes en relación con el cuidado de la salud difiere de quienes tienen más experiencia. Si Ed selecciona al azar este comité, ¿qué puede esperar en términos del promedio de años que sus miembros llevan con Spence Sprockets? ¿Cuál es la forma de la distribución de los años de experiencia de todos el personal (la población) en comparación con la forma de la distribución muestral de la media? Los años de servicio (redondeados al año más cercano) de los 40 trabajadores que actualmente están en nómina en Spence Sprockers, Inc., son los siguientes:

11	4	18	2	1	2	0	2	2	4
3	4	1	2	2	3	3	19	8	3
7	1	0	2	7	0	4	5	1	14
16	8	9	1	1	2	5	10	2	3

SOLUCIÓN

En la gráfica 8.4 se muestra la distribución de frecuencias de los años de servicio de la población de los 40 empleados. ¿Por qué la distribución tiene un sesgo positivo? Como el negocio ha crecido en años recientes, la distribución indica que 29 de los 40 empleados han estado en la compañía durante menos de seis años. También hay 11 empleados que han trabajado en Spence Sprockers por más de seis años. En particular, cuatro de ellos han laborado en la compañía 12 años o más (cuente las frecuencias por arriba de 12). Así, existe una larga cola en la distribución de los años de servicio a la izquierda, esto es, la distribución tiene un sesgo positivo.

Sin embargo, como el negocio creció, el número de empleados se incrementó en los últimos cinco años. De los 40 empleados, 18 han laborado en la compañía dos años o menos.



GRÁFICA 8.4 Años de servicio de los empleados en Spence Sprockets, Inc.

Considere el primer problema de Ed Spence; a él le gustaría formar un comité de cinco empleados para que estudien la cuestión del cuidado de la salud y sugieran la cobertura de gastos médicos más adecuada para la mayoría de ellos. ¿Cómo elegiría al comité? Si lo selecciona al azar, ¿qué puede esperar respecto del tiempo medio de servicio de quienes lo integren?

Para comenzar, Ed registra en papeletas el tiempo de servicio de cada empleado y las coloca en una gorra de béisbol. Después las revuelve y selecciona cinco al azar. Los tiempos de servicio de estos empleados son: 1, 9, 0, 19 y 14 años. Por lo tanto, el tiempo medio de servicio de esta muestra es de 8.60 años. ¿Cómo se compara este resultado con la media de la población? En este momento, Ed no conoce la media de la población, aunque el número de empleados de la población es de solo 40, así que decide calcular la media del tiempo de servicio de *todos* sus empleados; la cual es de 4.8 años y se determina al sumar los tiempos de servicio de *todos* los empleados y dividir el total entre 40.

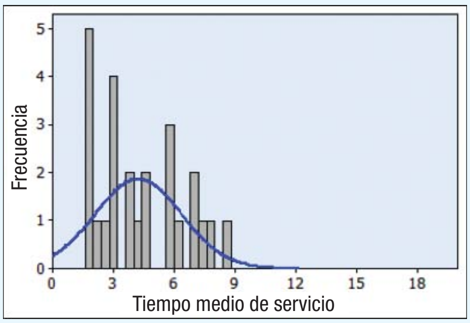
$$\mu = \frac{11 + 4 + 18 + \cdots + 2 + 3}{40} = 4.80$$

La diferencia entre la media de la muestra ($\bar{x}$) y la media de la población (μ) se llama **error de muestreo.** En otras palabras, la diferencia de 3.80 años entre la media poblacional de 4.80 y la media muestral de 8.60 es el error de muestreo. Este se debe al azar; por consiguiente, si Ed selecciona a estos cinco empleados para formar el comité, el tiempo medio de servicio de estos sería mayor que el de la media de la población.

¿Qué sucedería si Ed colocara de nuevo los papeles en la gorra y tomara otra muestra? ¿Esperaría que la media de esta segunda muestra fuera exactamente la misma que la anterior? Suponga que selecciona otra muestra de cinco empleados y encuentra que los tiempos de servicio de esta son de 7, 4, 4, 1 y 3. La media muestral es de 3.80 años. El resultado de seleccionar 25 muestras de cinco empleados cada una se registra en la tabla 8.5 y en la gráfica 8.5. En realidad hay 658 008 muestras de tamaño 5 que se pueden tomar de la población de 40 empleados, las cuales se determinan con la fórmula de las combinaciones [5.9] con 40 objetos tomados de 5 en 5. Observe la diferencia de forma de las distribuciones poblacional y muestral de medias; la población de tiempos de servicio de los empleados (gráfica 8.4) tiene un sesgo positivo, y la distribución de estas 25 medias muestrales no refleja el mismo sesgo positivo. También existe una diferencia en el rango de las medias muestrales en comparación con el rango de la población; esta varía de 0 a 19 años, mientras que las medias muestrales varían de 1.6 a 8.6 años.

TABLA 8.5 Veinticinco muestras aleatorias de cinco empleados

	Datos muestra						
Muestra	**Obs 1**	**Obs 2**	**Obs 3**	**Obs 4**	**Obs 5**	**Suma**	**Media**
A	1	9	0	19	14	43	8.6
B	7	4	4	1	3	19	3.8
C	8	19	8	2	1	38	7.6
D	4	18	2	0	11	35	7.0
E	4	2	4	7	18	35	7.0
F	1	2	0	3	2	8	1.6
G	2	3	2	0	2	9	1.8
H	11	2	9	2	4	28	5.6
I	9	0	4	2	7	22	4.4
J	1	1	1	11	1	15	3.0
K	2	0	0	10	2	14	2.8
L	0	2	3	2	16	23	4.6
M	2	3	1	1	1	8	1.6
N	3	7	3	4	3	20	4.0
O	1	2	3	1	4	11	2.2
P	19	0	1	3	8	31	6.2
Q	5	1	7	14	9	36	7.2
R	5	4	2	3	4	18	3.6
S	14	5	2	2	5	28	5.6
T	2	1	1	4	7	15	3.0
U	3	7	1	2	1	14	2.8
V	0	1	5	1	2	9	1.8
W	0	3	19	4	2	28	5.6
X	4	2	3	4	0	13	2.6
Y	1	1	2	3	2	9	1.8



GRÁFICA 8.5 Histograma de tiempos de servicio medio de 25 muestras de cinco empleados

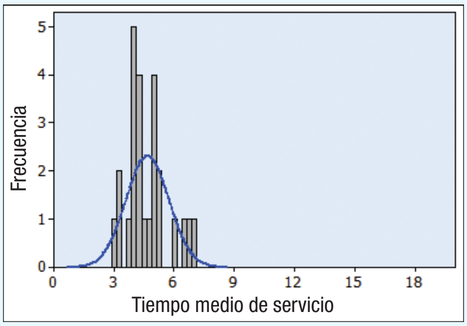
Cambiemos ahora el ejemplo aumentando el tamaño de la muestra de cinco empleados a 20. En la tabla 8.6 se indican los resultados de seleccionar 25 muestras de 20 empleados cada una y el cálculo de las medias muestrales. Estas medias muestrales se representan en la gráfica 8.6; compare la forma de esta distribución con la población (gráfica 8.4) y con la distribución muestral de medias cuando la muestra es de $n = 5$ (gráfica 8.5). Observe dos características importantes:

1. La forma de la distribución muestral de las medias es diferente a la de la población. La distribución de empleados que se muestra en la gráfica 8.4 tiene un sesgo positivo; no obstante, conforme se seleccionan muestras aleatorias de la población, cambia la forma de la distribución muestral de las medias. A medida que incrementa el tamaño de la muestra, la distribución muestral de las medias se aproxima a la distribución de probabilidad normal; este hecho se ilustra con el teorema central del límite.

TABLA 8.6 Muestras aleatorias y medias muestrales de 25 muestras de 20 empleados

	Datos muestra							
Muestra	**Obs 1**	**Obs 2**	**Obs 3**	**–**	**Obs 19**	**Obs 20**	**Suma**	**Media**
A	3	8	3	–	4	16	79	3.95
B	2	3	8	–	3	1	65	3.25
C	14	5	0	–	19	8	119	5.95
D	9	2	1	–	1	3	87	4.35
E	18	1	2	–	3	14	107	5.35
F	10	4	4	–	2	1	80	4.00
G	5	7	11	–	2	4	131	6.55
H	3	0	2	–	16	5	85	4.25
I	0	0	18	–	2	3	80	4.00
J	2	7	2	–	3	2	81	4.05
K	7	4	5	–	1	2	84	4.20
L	0	3	10	–	0	4	81	4.05
M	4	1	2	–	1	2	88	4.40
N	3	16	1	–	11	1	95	4.75
O	2	19	2	–	2	2	102	5.10
P	2	18	16	–	4	3	100	5.00
Q	3	2	3	–	3	1	102	5.10
R	2	3	1	–	0	2	73	3.65
S	2	14	19	–	0	7	142	7.10
T	0	1	3	–	2	0	61	3.05
U	1	0	1	–	9	3	65	3.25
V	1	9	4	–	2	11	137	6.85
W	8	1	9	–	8	7	107	5.35
X	4	2	0	–	2	5	86	4.30
Y	1	2	1	–	1	18	101	5.05

2. Hay menos dispersión en la distribución muestral de las medias que en la distribución de la población. En la población, los periodos de servicio variaron de 0 a 19 años. Cuando se seleccionaron muestras de tamaño 5, las medias de las muestras variaron de 1.6 a 8.6 años, y cuando seleccionaron muestras de 20, estas variaron de 3.05 a 7.10 años.



GRÁFICA 8.6 Histograma del tiempo medio de servicio de 25 muestras de 20 empleados

También es posible comparar la media de las medias de la muestra con la media de la población. La media de las 25 muestras de los 20 empleados de la tabla 8.6 es de 4.676 años.

$$\mu_{\bar{x}} = \frac{3.95 + 3.25 + \cdots + 4.30 + 5.05}{25} = 4.676$$

Se emplea el símbolo $\mu_{\bar{x}}$ para identificar la media de la distribución muestral de las medias. El subíndice recuerda que la distribución se refiere a la media muestral (se lee "mu subíndice X barra"); observe que la media de las medias muestrales (4.676 años) se encuentra muy próxima a la media de la población (4.80).

¿Qué concluye de este ejemplo? El teorema central del límite indica que, sin importar la forma de la distribución de la población, la distribución muestral de la media se aproximará a la distribución de probabilidad normal; cuanto mayor sea el número de observaciones en cada muestra, más evidente será la convergencia. El ejemplo de Spence Sprockets, Inc., demuestra el mecanismo del teorema central del límite. Comenzó con una población con sesgo positivo (gráfica 8.4). Después seleccionó 25 muestras aleatorias de cinco observaciones; calculó la media de cada muestra y, por último, organizó las 25 medias de muestra en una gráfica (8.5), registrando un cambio en la forma de la distribución muestral de las medias respecto de la de la población. El desplazamiento va de una distribución con sesgo positivo a una que tiene la forma de la distribución de probabilidad normal.

Para aclarar más los efectos del teorema central del límite, se incrementa el número de observaciones en cada muestra de 5 a 20; de estas, se seleccionan 25 muestras y se calcula la media de cada una; por último, estas medias muestrales se organizan en una gráfica (8.6). La forma del histograma de la gráfica 8.6 se desplaza claramente hacia la distribución de probabilidad normal.

En la gráfica 6.3 del capítulo 6 se muestran diversas distribuciones binomiales con una proporción de "éxitos" de 0.10, lo cual es otra demostración del teorema central del límite. Observe que, conforme n se incrementa de 7 hasta 12 y de 20 hasta 40, el perfil de las distribuciones de probabilidad se desplaza para acercarse cada vez más a una distribución de probabilidad normal. En la gráfica 8.6 también se muestra la convergencia hacia la normalidad conforme n se incrementa; esto confirma de nuevo el hecho de que, a medida que se incluyen más observaciones de la muestra de cualquier distribución poblacional, la forma de la distribución muestral de las medias se aproxima cada vez más a la distribución normal.

El teorema central del límite mismo (relea la definición que se proporcionó antes) no dice nada sobre la dispersión de la distribución muestral de medias ni sobre la comparación entre la media de la distribución muestral de medias y la media de la población; sin embargo, en el ejemplo de Spence Sprockets hay menor dispersión en la distribución de la media muestral que en la distribución de la población, lo cual indica la diferencia entre los rangos de la población y los de las medias muestrales. Advierta que la media de las medias de las muestras se encuentra cerca de la de la población; es posible demostrar que la media de la distribución muestral es la media poblacional (es decir, que $\mu_{\bar{x}} = \mu$), y si la desviación estándar de la población es σ, la de las medias muestrales es $\sigma/\sqrt{n}$, en la que n es el número de observaciones de cada muestra. Entonces, $\sigma/\sqrt{n}$ es el **error estándar de la media.** En realidad, el nombre completo es *desviación estándar de la distribución muestral de la media.*

ERROR ESTÁNDAR DE LA MEDIA	$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$	**[8.1]**

En esta sección llegamos a otras conclusiones importantes.

1. La media de la distribución muestral de medias será *exactamente* igual a la media poblacional si se seleccionan todas las muestras posibles del mismo tamaño de cualquier población; es decir,

$$\mu = \mu_{\bar{x}}$$

Aunque no se seleccionen todas las muestras, es de esperar que la media de la distribución muestral de medias se aproxime a la media poblacional.
2. Hay menos dispersión en la distribución muestral de las medias que en la población. Si la desviación estándar de la población es σ, la de la distribución muestral de medias es $\sigma/\sqrt{n}$. Observe que, cuando se incrementa el tamaño de la muestra, disminuye el error estándar de la media.

AUTOEVALUACIÓN 8-4

Repase los datos de Spence Sprockets, Inc. que aparecen en la gráfica 8.4 y seleccione al azar 10 muestras de 5 empleados cada una. Utilice los métodos descritos en el capítulo y la tabla de números aleatorios (apéndice B.6) para determinar los empleados que se incluirán en la muestra. Calcule la media de cada muestra y trace un esquema de las medias muestrales en una gráfica similar a la 8.4. ¿Cuál es la media de las 10 medias muestrales?

EJERCICIOS

11. El apéndice B.4 es una tabla de números aleatorios uniformemente distribuidos. De ahí que cada dígito tenga la misma probabilidad de presentarse.

a. Trace una gráfica que muestre la distribución de la población. ¿Cuál es la media de la población?

b. A continuación se registran los 10 primeros renglones de cinco dígitos del apéndice B.4; suponga que se trata de 10 muestras aleatorias de cinco valores cada una. Determine la media de cada muestra y trace una gráfica similar a la 8.4; compare la media de la distribución muestral de las medias con la media poblacional.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

0	2	7	1	1
9	4	8	7	3
5	4	9	2	1
7	7	6	4	0
6	1	5	4	5
1	7	1	4	7
1	3	7	4	8
8	7	4	5	5
0	8	9	9	9
7	8	8	0	4

12. Scrapper Elevator Company tiene 20 representantes de ventas que distribuyen su producto en Estados Unidos y Canadá. La cantidad de unidades que el mes anterior vendió cada representante se incluye a continuación. Suponga que estas cifras representan los valores la población.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2	3	2	3	3	4	2	4	3	2	2	7	3	4	5	3	3	3	3	5

a. Trace una gráfica que muestre la distribución de la población.

b. Calcule la media de la población.

c. Seleccione cinco muestras aleatorias de tamaño 5 cada una y calcule la media de cada muestra. Utilice los métodos descritos en el capítulo y en el apéndice B.4 para determinar los elementos que deben incluirse en la muestra.

d. Compare la media de la distribución muestral de medias con la media poblacional. ¿Esperaría que ambos valores fueran aproximadamente iguales?

e. Trace un histograma de las medias muestrales. ¿Observa alguna diferencia en la forma de la distribución muestral de las medias en comparación con la forma de la distribución de la población?

13. Considere que todas las monedas (un centavo de dólar, 25 centavos de dólar, etc.) que tenga en el bolsillo o monedero constituyen una población. Elabore una tabla de frecuencias, comience por el año en curso y cuente de manera regresiva, para registrar la antigüedad (en años) de las monedas. Por ejemplo, si el año en curso fuera 2013, una moneda que tiene impreso el año 2011 tendría dos años de antigüedad.

a. Trace un histograma u otro tipo de gráfica que muestre la distribución de la población.

b. Seleccione de manera aleatoria cinco monedas y registre su antigüedad media; repita el proceso 20 veces. Ahora trace un histograma u otro tipo de gráfica que muestre la distribución muestral de las medias.

c. Compare la forma de ambos histogramas.

14. Considere los dígitos de los números telefónicos de una página seleccionada al azar del directorio telefónico local como una población. Elabore una tabla de frecuencias con el último dígito de 30 números telefónicos seleccionados al azar. Por ejemplo, si el número telefónico es 55-55-97-04, registre un 4.

a. Trace un histograma u otro tipo de gráfica que muestre la distribución de la población. Con la distribución uniforme, calcule la media y la desviación estándar de la población.

b. Registre, asimismo, la media de la muestra de los últimos cuatro dígitos (97-04 resultaría una media de 5). Ahora elabore un histograma u otro tipo de gráfica que muestre la distribución muestral de las medias.

c. Compare la forma de ambos histogramas.

Uso de la distribución muestral de la media

OA8-5 Aplicar el teorema central del límite para calcular probabilidades.

El análisis anterior reviste importancia, pues la mayoría de las decisiones que se toman en los negocios tienen como fundamento los resultados de un muestreo. He aquí algunos ejemplos.

1. Arm and Hammer Company desea comprobar que su detergente para lavandería realmente contiene 100 onzas líquidas, como indica la etiqueta. Los registros históricos de los procesos de llenado indican que la cantidad media por recipiente es de 100 onzas líquidas y que la desviación estándar es de dos onzas líquidas. A las diez de la mañana el técnico de calidad verifica 40 recipientes y encuentra que la cantidad media por recipiente es de 99.8 onzas líquidas. ¿Debe el técnico interrumpir el proceso de llenado?
2. A.C. Nielsen Company proporciona información a las empresas que se anuncian en televisión. Las investigaciones indican que, en promedio, los adultos estadounidenses ven televisión 6.0 horas al día y la desviación estándar es de 1.5 horas. En el caso de una muestra de 50 adultos que viven en el área más poblada de Boston, ¿sería razonable seleccionar al azar una muestra y encontrar que sus individuos ven un promedio de 6.5 horas al día?
3. Haughton Elevator Company pretende formular especificaciones relacionadas con el número de personas que pueden desplazarse en un elevador nuevo de gran capacidad. Suponga que el peso medio de un adulto es de 160 libras, y que la desviación estándar es de 15 libras. Ahora bien, los pesos no siguen una distribución de probabilidad normal porque tienen un sesgo positivo. ¿Cuál es la probabilidad de que, en una muestra de 30 adultos, el peso medio sea de 170 libras o más?

Es posible responder las preguntas en cada una de estas situaciones utilizando las ideas que se analizaron en la sección anterior. En cada caso hay una población con información acerca de su media y su desviación estándar. Usando esta información y el tamaño de la muestra se determina la distribución de las medias muestrales y se calcula la probabilidad de que una media muestral caiga dentro de cierto rango. La distribución de muestreo sigue a la distribución de probabilidad normal con dos condiciones:

1. Cuando se sabe que las muestras se toman de poblaciones regidas por la distribución normal; en este caso, el tamaño de la muestra no constituye un factor.
2. El tamaño de la muestra es importante cuando se desconoce la forma de la distribución de la población o se sabe que no es normal. En general, la distribución muestral estaría normalmente distribuida a medida que el tamaño de la muestra se aproxima al infinito; en la práctica, una distribución de muestreo estaría cerca de una distribución normal con muestras de al menos 30 observaciones.

Se aplica la fórmula [7.5] del capítulo anterior para convertir cualquier distribución normal en una distribución normal estándar y calcular los valores *z*. Se puede usar la tabla de la distribución normal estándar, del apéndice B.3, para determinar la probabilidad de seleccionar un valor *z* que caerá dentro de un rango específico; la fórmula que se emplea para establecerlo es:

$$z = \frac{x - \mu}{\sigma}$$

En esta fórmula, *x* es el valor de la variable aleatoria; μ es la media de la población y σ es la desviación estándar de la población.

Sin embargo, la mayor parte de las decisiones de negocios se refiere a una muestra, no a una sola observación. Así, lo importante es la distribución de $\overline{X}$ (la media muestral), en lugar de *X* (el valor de una observación). Este es el primer cambio que se aplica a la fórmula [7.5]. El segundo consiste en emplear el error estándar de la media de *n* observaciones en lugar de la desviación estándar de la población. Es decir, se usa $\sigma/\sqrt{n}$ en el denominador en vez de σ. Por consiguiente, para determinar la probabilidad de una media muestral con rango específico primero aplique la fórmula para determinar el valor *z* correspondiente. Después, consulte el apéndice B.3 o un software estadístico para determinar la probabilidad.

CÁLCULO DEL VALOR *z* DE $\overline{x}$ CUANDO SE CONOCE LA DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \quad \textbf{[8.2]}$$

En el siguiente ejemplo se muestra la aplicación.

EJEMPLO

El departamento de control de calidad de Cola, Inc., conserva registros sobre la cantidad de bebida de cola en su botella gigante; la cantidad real en cada botella es de primordial importancia, pero varía en una mínima cantidad entre estas. La empresa no desea llenar botellas con menos líquido del debido, pues tendría problemas en lo que se refiere a la confiabilidad de la marca. Por otra parte, no puede colocar producto de más en las botellas porque lo regalaría, lo cual reduciría sus utilidades. Los registros del departamento de control de calidad indican que la cantidad de bebida de cola tiene una distribución de probabilidad normal, la media por botella es de 31.2 onzas y la desviación estándar de la población es de 0.4 onzas. Hoy, a las 8:00 horas, el técnico de control de calidad seleccionó al azar 16 botellas de la línea de llenado; la cantidad media de bebida en las botellas es de 31.38 onzas. ¿Es un resultado poco probable? ¿Es probable que el proceso permita colocar demasiada bebida en las botellas? En otras palabras, ¿es poco común el error de muestreo de 0.18 onzas?

SOLUCIÓN

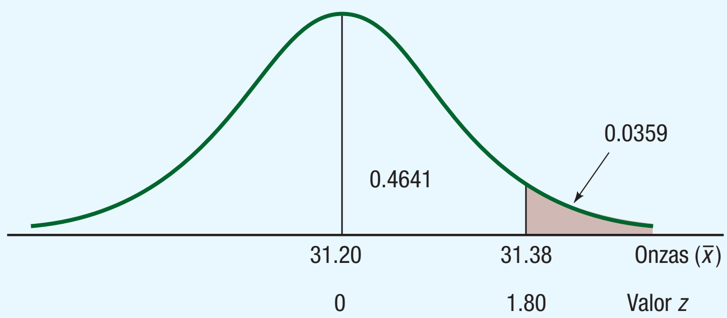
Utilice los resultados de la sección anterior para determinar la probabilidad de seleccionar una muestra de 16 (n) botellas de una población normal con una media de 31.2 (μ) onzas y una desviación estándar de la población de 0.4 (σ) onzas, y encontrar que la media muestral es de 31.38 ($\bar{x}$) o superior. Mediante la fórmula [8.2] se determina el valor de z.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{31.38 - 31.20}{0.4/\sqrt{16}} = 1.80$$

El numerador de esta ecuación, $\bar{x} = \mu = 31.38 - 31.20 = 0.18$, es el error muestral. El denominador, $\sigma/\sqrt{n} = 0.4/\sqrt{16} = 0.1$, es el error estándar de la distribución muestral de la media; así, los valores z expresan el error muestral en unidades estándar; en otras palabras, el error estándar.

Después, calcule la probabilidad de un valor z mayor que 1.80. En el apéndice B.3 localice la probabilidad correspondiente a un valor z de 1.80; este valor es de 0.4641. La probabilidad de un valor z mayor que 1.80 es de 0.0359, el cual se calcula con la resta $0.5000 - 0.4641$.

En conclusión, no es probable —menos de 4% de probabilidad— seleccionar una muestra de 16 observaciones de una población normal con una media de 31.2 onzas y una desviación estándar poblacional de 0.4 onzas, ni determinar que la media de la muestra es igual o mayor que 31.38 onzas; por tanto, en el proceso se vierte demasiada bebida en las botellas. El técnico de control de calidad debe entrevistarse con el supervisor de producción para sugerir la reducción de la cantidad de líquido en cada botella. La información se resume en la gráfica 8.7.



GRÁFICA 8.7 Distribución muestral de la cantidad media de bebida de cola en una botella gigante

AUTOEVALUACIÓN 8-5

Consulte la información relativa a Cola, Inc., y suponga que el técnico de control de calidad seleccionó una muestra de 16 botellas gigantes con un promedio de 31.08 onzas. ¿Qué concluye sobre el proceso de llenado?

EJERCICIOS

15. Una población normal tiene una media de 60 y una desviación estándar de 12; seleccione una muestra aleatoria de 9 y calcule la probabilidad de que la media muestral:
a. Sea mayor que 63.
b. Sea menor que 56.
c. Se encuentre entre 56 y 63.

16. Una población normal tiene una media de 75 y una desviación estándar de 5; seleccione una muestra de 40 y calcule la probabilidad de que la media muestral:
a. Sea menor que 74.
b. Se encuentre entre 74 y 76.
c. Se encuentre entre 76 y 77.
d. Sea mayor que 77.

17. En el sur de California, la renta de un departamento con una recámara tiene una distribución normal, una media de 2 200 dólares mensuales y una desviación estándar de 250 dólares. La distribución del costo mensual no se rige por la distribución normal; de hecho, tiene un sesgo positivo. ¿Cuál es la probabilidad de seleccionar una muestra de 50 departamentos de una recámara y hallar que la media mínima es de 1 950 dólares mensuales?

18. De acuerdo con un estudio del Internal Revenue Service, los contribuyentes tardan 330 minutos en promedio en preparar, copiar y archivar en un medio electrónico la forma fiscal 1040. Esta es una distribución normal de tiempos, y la desviación estándar es de 80 minutos. Un organismo de control selecciona una muestra aleatoria de 40 personas.
a. ¿Cuál es el error estándar de la media de este ejemplo?
b. ¿Cuál es la probabilidad de que la media de la muestra sea mayor que 320 minutos?
c. ¿Cuál es la probabilidad de que la media de la muestra se encuentre entre 320 y 350 minutos?
d. ¿Cuál es la probabilidad de que la media de la muestra sea superior que 350 minutos?

RESUMEN DEL CAPÍTULO

I. Hay muchas razones para realizar el muestreo de una población.
A. Los resultados de una muestra permiten calcular adecuadamente el valor del parámetro poblacional, con lo cual se ahorra tiempo y dinero.
B. Entrar en contacto con todos los miembros de la población consume demasiado tiempo.
C. Resulta imposible verificar y localizar a todos los miembros de la población.
D. El costo de estudiar a todos los elementos de la población resulta prohibitivo.
E. En una prueba con frecuencia se destruye el elemento de la muestra y no se puede regresar a la población.

II. En una muestra sin sesgo, todos los miembros de la población tienen la posibilidad de ser seleccionados para la muestra. Existen diversos métodos de muestreo de probabilidad.
A. En una muestra aleatoria simple, todos los miembros de la población tienen la misma posibilidad de ser seleccionados para la muestra.
B. En una muestra sistemática, se selecciona un punto de partida aleatorio y después se selecciona cada *k*-ésimo elemento subsiguiente de la población para formar la muestra.
C. En una muestra estratificada, la población se divide en varios grupos, a los que se denomina *estratos,* y enseguida se selecciona una muestra aleatoria de cada uno.
D. En el muestreo por conglomerados, la población se divide en unidades primarias y después se toman las muestras de las unidades primarias.

III. El error de muestreo es la diferencia entre un parámetro poblacional y un estadístico de la muestra.

IV. La distribución muestral de la media es una distribución de probabilidad de todas las posibles medias muestrales del mismo tamaño de muestra.
A. Para un tamaño de muestra dado, la media de todas las posibles medias muestrales tomadas de una población es igual a la media de la población.
B. Existe una menor variación en la distribución de las medias muestrales que en la distribución de la población.
C. El error estándar de la media mide la variación de la distribución muestral de las medias. El error estándar se calcula de la siguiente manera:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \textbf{[8.1]}$$

D. Si la población se rige por una distribución normal, la distribución muestral de la media también se regirá por la distribución normal con muestras de cualquier tamaño. Si la población no está normalmente distribuida, la distribución del muestreo de la media muestral se aproximará a una distribución normal cuando el tamaño de la muestra sea de al menos 30. Asuma que conoce la desviación estándar de la población; para determinar la probabilidad de que una media muestral caiga dentro de determinada región, se aplica la fórmula:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \textbf{[8.2]}$$

CLAVE DE PRONUNCIACIÓN

Símbolo	**Significado**	**Pronunciación**
$\mu_{\bar{x}}$	Media de la distribución muestral de la media	*mu subíndice x barra*
$\sigma_{\bar{x}}$	Error estándar de la población de la media muestral	*sigma subíndice x barra*

EJERCICIOS DEL CAPÍTULO

19. Las tiendas de venta al menudeo en el centro comercial de North Towne Square numeradas desde 00 hasta 24 son las siguientes:

00	Elder-Beerman	09	Lion Store	18	County Seat
01	Sears	10	Bootleggers	19	Kid Mart
02	Deb Shop	11	Formal Man	20	Lerner
03	Frederick's of Hollywood	12	Leather Ltd.	21	Coach House Gifts
04	Petries	13	B Dalton Bookseller	22	Spencer Gifts
05	Easy Dreams	14	Pat's Hallmark	23	CPI Photo Finish
06	Summit Stationers	15	Things Remembered	24	Regis Hairstylists
07	E. B. Brown Opticians	16	Pearle Vision Express		
08	Kay-Bee Toy & Hobby	17	Dollar Tree		

a. Si selecciona los números aleatorios 11, 65, 86, 62, 06, 10, 12, 77 y 04, ¿con qué tiendas es necesario ponerse en contacto para realizar una encuesta?
b. Utilice el apéndice B.6 para seleccionar una muestra aleatoria de cuatro tiendas.
c. Debe aplicar un procedimiento de muestreo sistemático; es necesario ponerse en contacto con la primera tienda y después con cada tercer establecimiento. ¿Con qué tiendas entrará en contacto?

20. Medical Mutual Insurance investiga el costo de una visita de rutina a consultorios de médicos familiares en el área de Rochester, Nueva York. La siguiente lista constituye una muestra de 39 médicos familiares de la región. Es preciso seleccionar a los médicos de forma aleatoria y establecer comunicación con ellos para conocer el monto de sus honorarios. Los 39 médicos se codificaron desde 00 hasta 38; también se indica si el médico trabaja solo en su consultorio (S), participa con un socio (P) o tiene un consultorio en grupo (G).

Número	Médico	Tipo de consultorio	Número	Médico	Tipo de consultorio
00	R. E. Scherbarth, M.D.	S	11	Wendy Martin, M.D.	S
01	Crystal R. Goveia, M.D.	P	12	Denny Mauricio, M.D.	P
02	Mark D. Hillard, M.D.	P	13	Hasmukh Parmar, M.D.	P
03	Jeanine S. Huttner, M.D.	P	14	Ricardo Pena, M.D.	P
04	Francis Aona, M.D.	P	15	David Reames, M.D.	P
05	Janet Arrowsmith, M.D.	P	16	Ronald Reynolds, M.D.	G
06	David DeFrance, M.D.	S	17	Mark Steinmetz, M.D.	G
07	Judith Furlong, M.D.	S	18	Geza Torok, M.D.	S
08	Leslie Jackson, M.D.	G	19	Mark Young, M.D.	P
09	Paul Langenkamp, M.D.	S	20	Gregory Yost, M.D.	P
10	Philip Lepkowski, M.D.	S	21	J. Christian Zona, M.D.	P

(*continúa*)

(*continuación*)

Número	Médico	Tipo de consultorio	Número	Médico	Tipo de consultorio
22	Larry Johnson, M.D.	P	31	Jeanne Fiorito, M.D.	P
23	Sanford Kimmel, M.D.	P	32	Michael Fitzpatrick, M.D.	P
24	Harry Mayhew, M.D.	S	33	Charles Holt, D.O.	P
25	Leroy Rodgers, M.D.	S	34	Richard Koby, M.D.	P
26	Thomas Tafelski, M.D.	S	35	John Meier, M.D.	P
27	Mark Zilkoski, M.D.	G	36	Douglas Smucker, M.D.	S
28	Ken Bertka, M.D.	G	37	David Weldy, M.D.	P
29	Mark DeMichiei, M.D.	G	38	Cheryl Zaborowski, M.D.	P
30	John Eggert, M.D.	P			

a. Los números aleatorios que se obtuvieron del apéndice B.6 son 31, 94, 43, 36, 03, 24, 17 y 09, ¿con qué médicos se debe establecer comunicación?
b. Seleccione una muestra aleatoria con los números aleatorios del apéndice B.6.
c. La muestra debe incluir a cada quinto médico. El número 04 se selecciona como punto de partida, ¿con qué médicos se debe establecer contacto?
d. Una muestra debe constar de dos médicos que trabajan solos (S), dos que participan con un socio (P) y uno con consultorio en grupo (G). Seleccione la muestra correspondiente. Explique su procedimiento.

21. Una población consiste en los siguientes tres valores: 1, 2 y 3.
a. Muestreando con reemplazos, enumere todas las muestras posibles de tamaño 2 y calcule la media de cada muestra.
b. Encuentre las medias de la distribución de la media muestral y la media poblacional; compare ambos valores.
c. Compare la dispersión de la población con la de la media muestral.
d. Describa las formas de ambas distribuciones.

22. En el departamento de educación de la UR University, los registros de los estudiantes sugieren que la población estudiantil pasa un promedio de 5.5 horas a la semana practicando deportes organizados. La desviación estándar de la población es 2.2 horas a la semana. Basándose en una muestra de 121 estudiantes, Healthy Lifestyles Incorporated (HLI) desea aplicar el teorema central del límite para realizar varias estimaciones.
a. Calcule el error estándar de la media muestral.
b. ¿Cuál es la probabilidad de que HLI encuentre una media muestral entre 5.0 y 6.0 horas?
c. Calcule la probabilidad de que la media muestral esté entre 5.3 y 5.7 horas.
d. ¿Qué tan extraño sería obtener una media muestral mayor a 6.5 horas?

23. El fabricante de eComputers, que manufactura una computadora económica, concluyó el diseño de un nuevo modelo portátil. A los altos ejecutivos de eComputers les gustaría obtener ayuda para poner precio al nuevo dispositivo. Se solicitaron los servicios de empresas de investigación de mercados y se le pidió preparar una estrategia de precios. Marketing-Gets-Results probó la nueva computadora portátil de eComputers con 50 consumidores elegidos al azar, quienes indicaron que tenían planes de adquirir la computadora el año entrante. Una segunda empresa de investigación de mercados, llamada Marketing-Reaps-Profits, probó en el mercado el mismo artículo con 200 propietarios de computadoras portátiles. ¿Cuál de las pruebas de las empresas de investigación de mercados resulta más útil? Explique las razones.

24. Responda las siguientes preguntas en uno o dos enunciados bien construidos.
a. ¿Qué sucede con el error estándar de la media si aumenta el tamaño de la muestra?
b. ¿Qué sucede con la distribución muestral de la media si aumenta el tamaño de la muestra?
c. Cuando se utiliza la distribución de la media muestral para aproximar la media poblacional, ¿cuál es el beneficio de utilizar tamaños muestrales más grandes?

25. Hay 25 moteles en Goshen, Indiana; el número de habitaciones en cada uno es el siguiente:

90	72	75	60	75	72	84	72	88	74	105	115	68	74	80	64	104	82	48	58	60	80	48	58	100

a. Utilice la tabla de números aleatorios (apéndice B.6) para seleccionar una muestra aleatoria de cinco moteles de esta población.
b. Obtenga una muestra sistemática seleccionando un punto de partida aleatorio entre los primeros cinco moteles y después elija cada quinto motel.
c. Suponga que los últimos cinco moteles son *de tarifas rebajadas.* Describa la forma en que seleccionaría una muestra aleatoria de tres moteles normales y dos de tarifas rebajadas.

26. Como parte de su programa de servicio al cliente, United Airlines seleccionó de forma aleatoria a 10 pasajeros de un vuelo que parte de Chicago a Tampa a las 9:00 horas. A cada pasajero de la muestra se le hará una entrevista a fondo en relación con las instalaciones, servicios, alimentos, etcétera, en los aeropuertos. Para identificar la muestra, a cada pasajero se le proporcionó un número al abordar la nave. El primer número fue 001 y el último, 250.

a. Seleccione al azar 10 números con ayuda del apéndice B.4.

b. La muestra de 10 pudo seleccionarse con una muestra sistemática. Elija el primer número con ayuda del apéndice B.4 y, después, mencione los números de las personas a las que entrevistará.

c. Evalúe ambos métodos; señale las ventajas y posibles desventajas.

d. ¿De qué otra forma se puede seleccionar una muestra aleatoria de los 250 pasajeros?

27. Suponga que el profesor de estadística le aplicó seis exámenes durante el semestre. Usted obtuvo las siguientes calificaciones (porcentaje corregido): 79, 64, 84, 82, 92 y 77. En lugar de promediar las seis calificaciones, el profesor le indicó que escogería dos al azar y calcularía el porcentaje final con base en estos datos.

a. ¿Cuántas muestras de dos calificaciones se pueden tomar?

b. Enumere todas las muestras posibles de tamaño 2 y calcule la media de cada una.

c. Calcule la media de la distribución muestral y compárela con la media de la población.

d. Si usted fuera estudiante, ¿le gustaría este sistema? ¿Sería diferente el resultado si se eliminara la calificación más baja? Redacte un breve informe.

28. En la oficina del First National Bank, ubicada en el centro de la ciudad, hay cinco cajeros automáticos. La semana pasada cada uno de los cajeros incurrió en el siguiente número de errores: 2, 3, 5, 3 y 5.

a. ¿Cuántas muestras de dos cajeros se pueden seleccionar?

b. Escriba todas las muestras posibles de tamaño 2 y calcule la media de cada una.

c. Calcule la media de la distribución muestral y compárela con la media de la población.

29. El departamento de control de calidad tiene cinco empleados técnicos en el turno matutino. A continuación se indica el número de veces que cada técnico indicó al supervisor de producción que interrumpiera el proceso durante la última semana.

Técnico	Interrupciones	Técnico	Interrupciones
Taylor	4	Rousche	3
Hurley	3	Huang	2
Gupta	5		

a. ¿Cuántas muestras de dos técnicos se forman con esta población?

b. Enumere todas las muestras de dos observaciones que se pueden tomar y calcule la media de cada muestra.

c. Compare la media de la distribución muestral con la media de la población.

d. Compare la forma de la distribución de la población con la forma de la distribución muestral de la media.

30. The Appliance Center cuenta con seis representantes de ventas en su sucursal del norte de Jacksonville. A continuación se indica el número de refrigeradores que vendió cada uno durante el último mes.

Vendedor	Refrigeradores vendidos	Vendedor	Refrigeradores vendidos
Zina Craft	54	Jan Niles	48
Woon Junge	50	Molly Camp	50
Ernie DeBrul	52	Rachel Myak	52

a. ¿Cuántas muestras de tamaño 2 se pueden tomar?

b. Seleccione todas las muestras posibles de tamaño 2 y calcule la cantidad media de refrigeradores vendidos.

c. Organice las medias de las muestras en una distribución de frecuencias.

d. ¿Cuál es la media de la población? ¿Cuál es la media de las medias de la muestra?

e. ¿Cuál es la forma de la distribución de la población?

f. ¿Cuál es la forma de la distribución muestral de la media?

31. Power +, Inc., produce baterías AA que se usan en autos de control remoto. La vida media de estas sigue una distribución normal de probabilidades con una media de 35.0 horas y una desviación estándar de 5.5 horas. Como parte de su programa de control de calidad, Power +, Inc., prueba muestras de 25 baterías.

a. ¿Qué se puede decir sobre la forma de la distribución muestral de la media?

b. ¿Cuál es el error estándar de la distribución muestral de la media?

c. ¿Qué proporción de las muestras tendrá una media de vida útil superior a 36 horas?
d. ¿Qué proporción de la muestra tendrá una media de vida útil mayor que 34.5 horas?
e. ¿Qué proporción de la muestra tendrá una media de vida útil entre 34.5 y 36 horas?

32. CRA CDs, Inc., desea que las extensiones medias de los "cortes" de un CD sean de 135 segundos (2:15 minutos). Esto permitirá a los disc jockeys contar con tiempo de sobra para incluir comerciales entre cada segmento de 10 minutos. Suponga que la distribución de la extensión de los cortes sigue una distribución normal con una desviación estándar de la población de ocho segundos, y que selecciona una muestra de 16 cortes de varios CD vendidos por CRA CDs, Inc.
a. ¿Qué puede decir sobre la forma de la distribución muestral de la media?
b. ¿Cuál es el error estándar de la media?
c. ¿Qué porcentaje de las medias muestrales será superior a 140 segundos?
d. ¿Qué porcentaje de las medias muestrales será superior a 128 segundos?
e. ¿Qué porcentaje de las medias muestrales será superior a 128 segundos e inferior a 140?

33. Estudios recientes indican que la mujer común de 50 años de edad gasta 350 dólares anuales en productos de cuidado personal. La distribución de las sumas que se gastan se rige por una distribución normal con una desviación estándar de 45 dólares anuales. Se selecciona una muestra aleatoria de 40 mujeres. La cantidad media que gasta dicha muestra es de 335 dólares. ¿Cuál es la probabilidad de hallar una media muestral igual o superior a la de la población indicada?

34. Información en poder del American Institute of Insurance indica que la cantidad media de seguros de vida por familia en Estados Unidos asciende a 110 000 dólares. Esta es una distribución normal y su desviación estándar es de 40 000 dólares.
a. Si selecciona una muestra aleatoria de 50 familias, ¿cuál es el error estándar de la media?
b. ¿Cuál es la forma de la distribución muestral de la media?
c. ¿Cuál es la probabilidad de seleccionar una muestra con una media mínima de 112 000 dólares?
d. ¿Cuál es la probabilidad de seleccionar una muestra con una media superior a 100 000 dólares?
e. Determine la probabilidad de seleccionar una muestra con una media superior a 100 000 dólares e inferior a 112 000.

35. La edad media a la que los hombres se casan por primera vez en Estados Unidos se rige por la distribución normal, con una media de 24.8 años. La desviación estándar de la distribución es de 2.5 años. En el caso de una muestra aleatoria de 60 hombres, ¿cuál es la probabilidad de que la edad a la que se casaran por primera vez sea menor de 25.1 años?

36. Un estudio reciente que llevó a cabo la Greater Los Angeles Taxi Drivers Association mostró que la tarifa media por servicio de Hermosa Beach al Aeropuerto Internacional de Los Ángeles es de 18.00 dólares, y la desviación estándar, de 3.50 dólares. Seleccione una muestra de 15 tarifas.
a. ¿Cuál es la probabilidad de que la media de la muestra se encuentre entre 17.00 y 20.00 dólares?
b. ¿Qué debe suponer para llevar a cabo el cálculo anterior?

37. Crosset Trucking Company afirma que el peso medio de sus camiones cuando se encuentran completamente cargados es de 6 000 libras, y la desviación estándar, de 150 libras. Considerando que la población se rige por la distribución normal, se seleccionan al azar 40 camiones y se pesan. ¿Dentro de qué límites se presentará 95% de las medias de la muestra?

38. La cantidad media de abarrotes que compra cada cliente en Churchill Grocery Store es de 23.50 dólares, con una desviación estándar de 5.00 dólares. Suponga que la distribución de cantidades compradas sigue la distribución normal. Considere una muestra de 50 clientes y conteste las siguientes preguntas.
a. ¿Cuál es la probabilidad de que la media de la muestra sea de de al menos 25.00 dólares?
b. ¿Cuál es la probabilidad de que la media de la muestra sea superior a 22.50 dólares e inferior a 25.00?
c. ¿Dentro de qué límites se presentará 90% de las medias muestrales?

39. El desempeño medio en una prueba de condición física a estudiantes atletas de la División I es de 947, con una desviación estándar de 205. Si selecciona una muestra aleatoria de 60 estudiantes, ¿cuál es la probabilidad de que la media se encuentre por debajo de 900?

40. Suponga que lanza un dado dos veces.
a. ¿Cuántas muestras se pueden seleccionar?
b. Enumere cada una de las muestras posibles y calcule la media.
c. En un esquema similar al de la gráfica 8.2, compare la distribución muestral de la media con la distribución de la población.
d. Calcule la media y la desviación estándar de cada distribución y compárelas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

41. En la tabla de la página siguiente se muestra una lista de los 50 estados asignados con los números 0 hasta 49.
a. Usted pretende seleccionar una muestra de 12 elementos de la lista. Los números aleatorios seleccionados son 45, 15, 81, 09, 39, 43, 90, 26, 06, 45, 01 y 42. ¿Qué estados se incluyen en la muestra?

Número	Estado	Número	Estado
0	Alabama	25	Montana
1	Alaska	26	Nebraska
2	Arizona	27	Nevada
3	Arkansas	28	New Hampshire
4	California	29	New Jersey
5	Colorado	30	New Mexico
6	Connecticut	31	New York
7	Delaware	32	North Carolina
8	Florida	33	North Dakota
9	Georgia	34	Ohio
10	Hawaii	35	Oklahoma
11	Idaho	36	Oregon
12	Illinois	37	Pennsylvania
13	Indiana	38	Rhode Island
14	Iowa	39	South Carolina
15	Kansas	40	South Dakota
16	Kentucky	41	Tennessee
17	Louisiana	42	Texas
18	Maine	43	Utah
19	Maryland	44	Vermont
20	Massachusetts	45	Virginia
21	Michigan	46	Washington
22	Minnesota	47	West Virginia
23	Mississippi	48	Wisconsin
24	Missouri	49	Wyoming

b. Usted desea utilizar una muestra sistemática de cada sexto elemento y elige el dígito 02 como punto de partida. ¿Qué estados incluirá?

42. Human Resource Consulting (HRC) lleva a cabo un sondeo con una muestra de 60 empresas de Twin Cities, Minnesota, con el fin de estudiar los costos del cuidado de la salud del cliente. Uno de los elementos que se estudia es el deducible anual que deben pagar los empleados. El Minnesota Department of Labor informa que la media de esta distribución es de 502 dólares, con una desviación estándar de 100 dólares.

a. Calcule el error estándar de la media muestral de HRC.
b. ¿Cuál es la probabilidad de que HRC encuentre una media muestral entre 477 y 527 dólares?
c. Calcule la probabilidad de que la media muestral oscile entre 492 y 512 dólares.
d. ¿Cuál es la probabilidad de que la media muestral sea superior a 550 dólares?

43. La década pasada, el número medio de ataques de hackers a miembros de la Information Systems Security Association fue de 510 por año, con una desviación estándar de 14.28 ataques; el número de ataques por año está normalmente distribuido. Suponga que nada cambia en este ambiente.

a. ¿Cuál es la probabilidad de que este grupo sufra un promedio superior a 600 ataques los próximos 10 años?
b. Calcule la probabilidad de que experimenten un promedio de entre 500 y 600 ataques durante los próximos 10 años.
c. ¿Cuál es la probabilidad de que experimenten un promedio inferior a 500 ataques durante los próximos 10 años?

44. Un economista utiliza el precio de un galón de leche como medida para la inflación, y descubre que el precio promedio es de 3.50 dólares por galón, con una desviación estándar de la población de 0.33 dólares. Usted decide muestrear 40 tiendas de abarrotes, recabar los precios del galón de leche y calcular el precio medio de la muestra.

a. ¿Cuál es el error estándar de la media de este experimento?
b. ¿Cuál es la probabilidad de que la media de la muestra oscile entre 3.46 y 3.54 dólares?
c. ¿Cuál es la probabilidad de que la diferencia entre la media muestral y la media poblacional sea inferior a un centavo de dólar?
d. ¿Cuál es la probabilidad de que la media de la muestra sea superior a 3.60 dólares?

45. El informe anual de Nike indica que el estadounidense promedio compra 6.5 pares de zapatos deportivos al año. Suponga que la desviación estándar de la población es de 2.1 y que se analizará una muestra de 81 clientes el próximo año.

a. ¿Cuál es el error estándar de la media en este experimento?
b. ¿Cuál es la probabilidad de que la media de la muestra se encuentre entre 6.0 y 7.0 pares de zapatos deportivos?
c. ¿Cuál es la probabilidad de que la diferencia entre la media muestral y la media poblacional sea inferior a 0.25 pares?
d. ¿Cuál es la probabilidad de que la media muestral sea superior a 7.0 pares?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

46. Consulte los datos sobre Real Estate, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.
Utilice un software estadístico para calcular la media y la desviación estándar de la distribución de los precios de venta de las casas. Suponga que esta es la población; calcule la media y la desviación estándar de la muestra. Determine la probabilidad de encontrar una media de la muestra de este tamaño o más grande de la población.

47. Consulte los datos sobre béisbol 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. En la última década, la asistencia media por equipo siguió una distribución normal, con una media de 2.25 millones por equipo y una desviación estándar de 0.70 millones. Utilice un software estadístico para calcular la asistencia media por equipo durante la temporada 2012. Determine la probabilidad de una media muestral de este tamaño o mayor de la población.

48. Consulte los datos de los autobuses del Distrito Escolar Buena. La información que proporciona el fabricante de autobuses escolares sugiere que el costo medio de mantenimiento mensual es de 455 dólares por unidad. Utilice un software estadístico para encontrar la media y la desviación estándar de los autobuses de Buena; ¿esta información concuerda con la reportada por el fabricante? Específicamente, ¿cuál es la probabilidad de que la media muestral sea menor que la de Buena, dados los datos del fabricante?

Estimación e intervalos de confianza

9

LA AMERICAN RESTAURANT ASSOCIATION recopiló información sobre las veces que los matrimonios jóvenes comen fuera de casa cada semana. Una encuesta de 60 parejas demostró que la cantidad media de comidas fuera de casa era de 2.76 por semana, con una desviación estándar de 0.75; construya el intervalo de confianza de 99% para la media de la población (vea el ejercicio 36 y el objetivo de aprendizaje OA9-2).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA9-1 Calcular e interpretar un estimador puntual de la media poblacional.

OA9-2 Calcular e interpretar un intervalo de confianza para una media poblacional.

OA9-3 Calcular e interpretar un intervalo de confianza para una proporción de la población.

OA9-4 Calcular el tamaño de la muestra necesario para estimar una proporción de la población o una media poblacional.

OA9-5 Ajustar un intervalo de confianza para poblaciones finitas.

ESTADÍSTICA EN ACCIÓN

En un lugar visible de la ventanilla de todos los automóviles nuevos de Estados Unidos aparece una calcomanía con un cálculo aproximado del ahorro de gasolina, según lo requiere la Environmental Protection Agency (EPA). Con frecuencia, el ahorro de gasolina constituye un factor importante para que el consumidor elija un automóvil nuevo debido a los costos del combustible o a cuestiones ambientales. Por ejemplo, los cálculos aproximados del rendimiento de combustible de un BMW 3281 Sedán 2010 (automático, de 6 cilindros) son de 18 millas por galón (mpg) en la ciudad y de 28 mpg en carretera. La EPA reconoce que el verdadero ahorro de gasolina puede diferir de los cálculos aproximados: "Ninguna prueba puede simular todas las combinaciones de condiciones y climas posibles, del comportamiento del conductor y hábitos en el cuidado del automóvil. El rendimiento real depende de cómo, cuándo y dónde se maneje el vehículo. La EPA descubrió que las mpg que obtiene la mayoría de los conductores difieren de los cálculos aproximados por muy poco". De hecho, la calcomanía del parabrisas también incluye una estimación del intervalo relativo al ahorro de combustible; por ejemplo, 14 a 22 mpg en ciudad y de 23 a 33 mpg en carretera.

Introducción

En el capítulo anterior se inició el estudio de la estadística inferencial y se presentaron las razones para el muestreo y los métodos para hacerlo. Las razones del muestreo son las siguientes:

- Contactar a toda la población consume demasiado tiempo.
- El costo de estudiar todos los elementos de la población es muy alto.
- Por lo general, los resultados de la muestra resultan adecuados.
- Algunas pruebas resultan destructivas.
- Es imposible revisar todos los elementos.

Existen varios métodos de muestreo; el aleatorio simple es el que más se utiliza. En este, cada miembro de la población posee las mismas posibilidades de ser seleccionado como parte de la muestra. Otros métodos de muestreo son el sistemático, el estratificado y el muestreo por conglomerados.

En el capítulo 8 se asume que se conoce la información relacionada con la media, la desviación estándar o la forma de la población. Dicha información no se encuentra disponible en la mayoría de las situaciones de negocios. En realidad, el propósito del muestreo es calcular de forma aproximada algunos de estos valores. Por ejemplo, se selecciona una muestra de una población y se utiliza la media de esta para aproximar la media de la población.

En este capítulo se estudian diversos aspectos importantes del muestreo. El primer paso es el estudio del **estimador puntual**, el cual consiste en un solo valor (punto) deducido de una muestra para estimar el de una población. Por ejemplo, suponga que se elige una muestra de 50 ejecutivos de nivel medio y a cada uno les pregunta cuántas horas laboró la semana pasada. Se calcula la media de esta muestra de 50 y se utiliza el valor de la media muestral como estimador puntual de la media poblacional desconocida; sin embargo, un estimador puntual es un solo dato. Un enfoque que arroja más información consiste en presentar un intervalo de valores del cual se espera estimar el parámetro poblacional. Dicho intervalo recibe el nombre de **intervalo de confianza.** En los negocios, a menudo es necesario determinar el tamaño de una muestra. ¿A cuántos electores debe contactar una compañía dedicada a realizar encuestas con el fin de predecir los resultados de las elecciones? ¿Cuántos productos se necesitan analizar para garantizar el nivel de calidad? En este capítulo también se explica una estrategia para determinar el tamaño adecuado de la muestra.

OA9-1 Calcular e interpretar un estimador puntual de la media poblacional.

Estimadores puntuales e intervalos de confianza de una media

Un estimador puntual es un estadístico único para calcular un parámetro poblacional. Suponga que Best Buy, Inc., desea estimar la edad media de los compradores de televisores led de alta definición; selecciona una muestra aleatoria de 50 clientes recientes, determina la edad de cada uno y calcula la edad media de estos. La media de esta muestra es un estimador puntual de la media de la población.

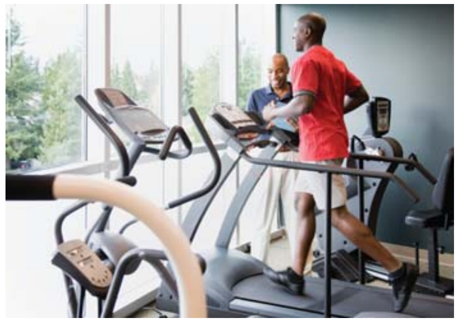
ESTIMADOR PUNTUAL Estadístico calculado a partir de información de la muestra para estimar el parámetro poblacional.

En los siguientes ejemplos se ilustran los estimadores puntuales de medias poblacionales.

1. El turismo constituye una fuente importante de ingresos para muchos países caribeños, como Barbados. Suponga que la Oficina de Turismo de Barbados desea un cálculo aproximado de la cantidad media que gastan los turistas que visitan el país. No resultaría viable ponerse en contacto con cada turista; por consiguiente, se seleccionan 500 turistas al azar en el momento en que salen del país y se les preguntan detalles de los gastos que realizaron durante su visita a la isla. La cantidad media que gastó la muestra de 500 turistas constituye un cálculo aproximado del parámetro poblacional desconocido. Es decir, `la media muestral es el estimador puntual de la media poblacional.
2. Litchfield Home Builders, Inc., construye casas en la zona sureste de Estados Unidos. Una de las principales preocupaciones de los compradores es la fecha en que concluirán las obras.

Hace poco Litchfield comunicó a sus clientes: “Su casa quedará terminada en 45 días a partir de la fecha de instalación de los muros”. El departamento de atención a clientes de Litchfield desea comparar este ofrecimiento con experiencias recientes. Una muestra de 50 casas terminadas este año reveló que el número medio de días de trabajo (o hábiles) a partir del inicio de la construcción de los muros a la terminación de la casa fue de 46.7. ¿Es razonable concluir que la media poblacional aún es de 45 días y que la diferencia entre la media muestral (46.7 días) y la media de población propuesta es un error de muestreo? En otras palabras, ¿la media muestral difiere en forma significativa de la media poblacional?

3. Estudios médicos recientes indican que el ejercicio constituye una parte importante de la salud general de una persona. El director de recursos humanos de OCF, fabricante importante de vidrio, desea calcular la cantidad de horas semanales que los empleados dedican al ejercicio. Una muestra de 70 empleados revela que la cantidad media de horas de ejercicio durante la semana pasada fue de 3.3. Este valor es un estimador puntual de la media poblacional desconocida.



La media muestral, $\bar{x}$, no es el único estimador puntual de un parámetro poblacional. Por ejemplo, p (una proporción muestral) es un estimador puntual de π (la proporción poblacional); y s (la desviación estándar muestral) es un estimador puntual de σ (la desviación estándar poblacional).

Intervalos de confianza de una media poblacional

OA9-2 Calcular e interpretar un intervalo de confianza para una media poblacional.

Ahora bien, un estimador puntual solo cuenta parte de la historia. Aunque se espera que este se aproxime al parámetro poblacional, sería conveniente medir su verdadera proximidad. Un intervalo de confianza sirve para este propósito. Por ejemplo, se estima que el ingreso anual medio de los trabajadores de la construcción en el área de Nueva York a Nueva Jersey es de 85 000 dólares. Un intervalo de este valor aproximado puede oscilar entre 81 000 y 89 000 dólares. Es preciso generar un enunciado probabilístico para describir cuánto es posible confiar en que el parámetro poblacional se encuentre en el intervalo; por ejemplo, se cuenta con 90% de confianza de que el ingreso anual medio de los trabajadores de la construcción en el área de Nueva York a Nueva Jersey se encuentra entre 81 000 y 89 000 dólares.

INTERVALO DE CONFIANZA Conjunto de valores que se forma a partir de una muestra de forma que exista la probabilidad de que el parámetro poblacional ocurra dentro de dicho conjunto con una probabilidad específica. La probabilidad específica se llama *nivel de confianza*.

Para calcular el intervalo de confianza para una media poblacional se consideran dos situaciones:

- Los datos de la muestra se utilizan para calcular μ con $\bar{x}$, mientras que la desviación estándar de la población (σ) es conocida.
- Los datos de la muestra se utilizan para calcular μ con $\bar{x}$, mientras que la desviación estándar de la población (σ) es desconocida.

Existen diferencias importantes en las suposiciones entre ambas situaciones. Considere primero el caso donde se conoce σ.

Desviación estándar de la población conocida (σ)

Un intervalo de confianza se calcula con el empleo de dos estadísticos: la media muestral, $\bar{x}$, y la desviación estándar. De los capítulos anteriores usted sabe que la desviación estándar es un estadístico importante porque mide la dispersión, o variación, de una población o una distribución muestral. Cuando se calcula un intervalo de confianza, se utiliza la desviación estándar para estimar el rango del intervalo de confianza.

Para demostrar la idea del intervalo de confianza, es preciso comenzar con una suposición simple: se conoce el valor de la desviación estándar de la población, σ. En general, se tiene la desviación estándar de la población en situaciones con una larga historia de recolección de datos; por ejemplos, Sobra la *s* el monitoreo de los procesos de llenado de botellas de refresco o de cajas de cereal, y los resultados de la prueba de razonamiento SAT (para admisión a la universidad). Conocer σ permite simplificar el desarrollo del intervalo de confianza porque es posible utilizar la distribución normal estándar que se estudió en el capítulo 8.

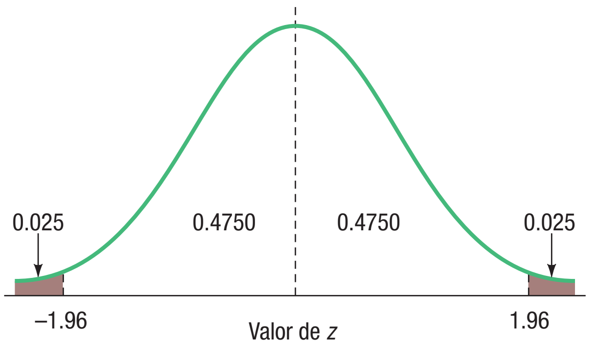
Recuerde que la distribución muestral de la media es la distribución de todas las medias muestrales, $\bar{x}$, con tamaño de la muestra, *n*, de una población, y que se conoce la desviación estándar de la población, σ. A partir de esta información, y del teorema central del límite, se sabe que la distribución muestral siguen una distribución de probabilidad normal con una media μ y una desviación estándar $\sigma/\sqrt{n}$. Recuerde que este valor recibe el nombre de error estándar.

Los resultados del teorema central del límite permiten afirmar lo siguiente con respecto a los intervalos de confianza utilizando el estadístico-*z*:

1. De todos los intervalos de confianza calculados a partir de muestras aleatorias seleccionadas de una población, 95% contendrá la media poblacional. Estos intervalos se calculan utilizando un estadístico-*z* igual a 1.96.
2. De todos los intervalos de confianza calculados a partir de muestras aleatorias seleccionadas de una población, 90% contendrá la media poblacional. Estos intervalos de confianza se calculan utilizando un estadístico-*z* igual a 1.65.

Los intervalos de confianza calculados de esta manera proporcionan ejemplos de los *niveles de confianza* y reciben el nombre de **intervalo de confianza de 95%** e **intervalo de confianza de 90%.** Por lo tanto, 95% y 90% son los niveles de confianza y se refieren al porcentaje de intervalos similarmente construidos que incluirían el parámetro a calcular, en este caso, μ, la media poblacional.

¿Cómo se obtienen 1.96 y 1.65? Primero, busque el valor *z* para el intervalo de confianza 95%. El siguiente diagrama y la tabla 9.1 (página siguiente) son de utilidad; en dicha tabla se muestra una reproducción del apéndice B.3, la tabla de valores estándar normales. Sin embargo, se eliminan varias filas y columnas para ubicar mejor las que interesan.



1. Primero, se divide el nivel de confianza a la mitad, así que 0.9500/2 = 0.4750.
2. Después se ubica 0.4750 en el cuerpo de la tabla 9.1; observe que esta cantidad se encuentra en la intersección de una fila y una columna.
3. Localice el valor de la fila correspondiente en el margen izquierdo, que es 1.9, y el de la columna en el margen superior, que es 0.06. Sumándolos se obtiene un valor *z* de 1.96.
4. Así, la probabilidad de encontrar un valor *z* entre 0 y 1.96 es 0.4750.
5. De la misma forma, y dado que la distribución normal es simétrica, la probabilidad de encontrar un valor *z* entre −1.96 y 0 también es 0.4750.
6. Cuando se suman ambas cantidades, la probabilidad de que un valor *z* esté entre −1.96 y 1.96 es 0.9500.

Para el nivel de confianza 90% se siguen los mismos pasos. Primero, la mitad del intervalo de confianza deseado es 0.4500; esta cantidad no se revela de manera exacta en la tabla 9.1; sin embargo, está entre dos valores (0.4495 y 0.4505); entonces, tal como en el paso tres, se localzan en la tabla. El primero (0.4495) corresponde a un valor *z* de 1.64, y el segundo (0.4505), a uno de 1.65. Para ser conservadores, seleccione el mayor de estos (1.65); el nivel exacto de confianza es 90.1%, o 2(0.4505). Enseguida, la probabilidad de hallar un valor *z* entre −1.65 y 0 es 0.4505, y la probabilidad de que este se ubique entre −1.65 y 1.65 es 0.9010.

¿Cómo se determina el intervalo de confianza de 95%? La amplitud del intervalo se determina por medio de dos factores: 1) el nivel de confianza, como se describe en la sección anterior y 2) el

TABLA 9.1 Tabla de valores estándar normales para valores seleccionados

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884

tamaño del error estándar de la media. Para encontrar el error estándar de la media, recuerde del capítulo anterior (vea la fórmula [8.1]) que el error estándar de la media indica la variación de la distribución de las medias muestrales. Se trata, en realidad, de la desviación estándar de la distribución muestral de medias. La fórmula se repite enseguida:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

donde:

- $\sigma_{\bar{x}}$ símbolo es el símbolo del error estándar de la media; se utiliza la letra griega porque se trata de un valor poblacional, y el subíndice x recuerda que se refiere a la distribución de medias muestrales;
- σ es la desviación estándar poblacional;
- n es el número de observaciones en la muestra.

Dos valores influyen en el tamaño del error estándar. El primero es la desviación estándar de la población: mientras mayor sea la desviación estándar de la población, σ, mayor será $\sigma/\sqrt{n}$. Si la población es homogénea, de modo que genere una desviación estándar poblacional pequeña, el error estándar también será pequeño. Sin embargo, la cantidad de valores de la muestra también afecta al error estándar: una muestra grande generará un error estándar pequeño en la estimación, lo que indicará que hay menos variabilidad en las medias muestrales.

Los cálculos en el caso de un intervalo de confianza de 95% se resumen con la siguiente fórmula:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

De manera similar, un intervalo de confianza de 90.1% se calcula de la siguiente manera:

$$\bar{x} \pm 1.65 \frac{\sigma}{\sqrt{n}}$$

Considere que 1.96 y 1.65 son valores z correspondientes a los intervalos de confianza 95% y 90.1%, respectivamente; sin embargo, no son exclusivos. Es posible seleccionar cualquier nivel de confianza entre 0% y 100% y encontrar el valor correspondiente a z. En general, un intervalo de confianza de la media poblacional, cuando se conoce la desviación estándar poblacional, se calcula de la siguiente manera:

INTERVALO DE CONFIANZA DE LA MEDIA POBLACIONAL CON UNA σ CONOCIDA

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \qquad \textbf{[9.1]}$$

Para explicar estos conceptos, considere el siguiente ejemplo. Del Monte Foods, Inc., distribuye duraznos en trozo en latas de 4.5 onzas. Para garantizar que cada lata contenga por lo menos la cantidad que se requiere, Del Monte establece que el proceso de llenado debe verter 4.51 onzas de duraznos y almíbar en cada lata. Así, 4.51 es la media poblacional. Por supuesto, no todas las latas contendrán exactamente 4.51 onzas de duraznos y almíbar, algunas latas tendrán más y otras, menos. A partir de datos históricos, Del Monte sabe que 0.04 onzas es la desviación estándar del pro-

ceso de llenado y que la cantidad, en onzas, sigue una distribución de probabilidad normal. El técnico en control de calidad selecciona una muestra de 64 latas de cada turno, mide la cantidad en cada una, calcula la cantidad media de llenado y desarrolla un intervalo de confianza de 95% para la media poblacional. Utilizando el intervalo de confianza, ¿el proceso está llenando las latas con la cantidad deseada? La última muestra de 64 latas tuvo una media muestral de 4.507 onzas. Basándose en esta información, el intervalo de confianza de 95% es:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 4.507 \pm 1.96 \frac{0.04}{\sqrt{64}} = 4.507 \pm 0.0098$$

El intervalo de confianza de 95% estima que la media poblacional está entre 4.4972 y 4.5168 onzas de duraznos y almíbar. Recuerde que el proceso está programado para llenar cada lata con 4.51 onzas; como la cantidad deseada de llenado está en este intervalo, se concluye que el proceso de llenado logra los resultados esperados. En otras palabras, es razonable concluir que la media muestral de 4.507 pudo provenir de una distribución poblacional con una media de 4.51 onzas.

En este ejemplo se observa que la media poblacional de 4.51 onzas está en el intervalo de confianza; pero este no siempre es así. Si se seleccionan 100 muestras de 64 latas de la población, se calcula la media muestral y se desarrolla un intervalo de confianza basado en cada muestra, sería factible encontrar la media poblacional en aproximadamente 95 de los 100 intervalos; o, por el contrario, cerca de cinco de los intervalos no contendrán la media poblacional. Recuerde, del capítulo 8, que esto se llama error muestral. En el siguiente ejemplo se detallan los muestreos repetidos de una población.

EJEMPLO

La American Management Association estudia el ingreso medio de los gerentes de tiendas de la industria del menudeo. Una muestra aleatoria de 49 gerentes revela una media muestral de 45 420 dólares y una desviación estándar de 2 050 dólares. A la asociación le gustaría responder las siguientes preguntas:

1. ¿Cuál es la media de la población?
2. ¿Cuál es un rango de valores razonable para la media poblacional?
3. ¿Cómo se deben interpretar estos resultados?

SOLUCIÓN

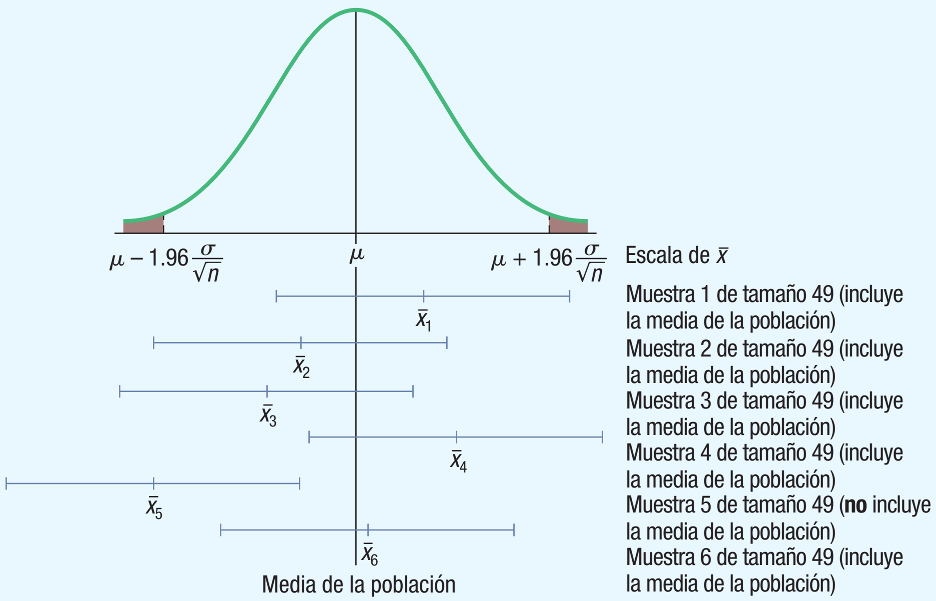
En general, las distribuciones de los salarios e ingresos tienen un sesgo positivo, pues unos cuantos individuos ganan considerablemente más que otros, lo cual sesga la distribución en dirección positiva. Por fortuna, el teorema central del límite estipula que, si se selecciona una muestra grande, la distribución de las medias muestrales tenderá a seguir la distribución normal; en este caso, una muestra de 49 gerentes es lo bastante grande para suponer que la distribución presentará dicha tendencia. A continuación se responden las preguntas planteadas en el ejemplo.

1. **¿Cuál es la media de la población?** En este caso se ignora, pero se sabe que la media de la muestra es de 45 420 dólares. De ahí que la mejor estimación del valor de población sea el estadístico de la muestra correspondiente; por consiguiente, la media de la muestra de 45 420 dólares constituye un *estimador puntual* de la media poblacional desconocida.
2. **¿Cuál es el rango de valores razonable para la media poblacional?** La asociación decide utilizar un nivel de confianza de 95%. Para determinar el intervalo de confianza correspondiente, se aplica la fórmula [9.1]:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} = \$45\,420 \pm 1.96 \frac{\$2\,050}{\sqrt{49}} = \$45\,420 \pm \$574$$

   Los límites del intervalo de confianza son 44 846 y 45 994 dólares. El grado o nivel de confianza es de 95%, y el intervalo de confianza abarca de 45 846 hasta 45 994 dólares. A $\pm$\$574 se le llama margen de error.
3. **¿Cómo se deben interpretar estos resultados?** Suponga que usted selecciona varias muestras de 256 gerentes, tal vez varios cientos. Para cada muestra, calcula la media y después construye un intervalo de confianza de 95%, como en la sección anterior. Puede esperar que alrededor de 95% de estos intervalos de confianza contenga la media de la *población.* Cerca de

5% de los intervalos no contendrán el ingreso anual medio poblacional, μ; no obstante, un intervalo de confianza particular contiene el parámetro poblacional o no lo contiene. En el siguiente diagrama se muestran los resultados de seleccionar muestras de la población de gerentes de la industria del menudeo: se calcula la media de cada una y, posteriormente, con la fórmula [9.1] se determina un intervalo de confianza de 95% de la media poblacional. Observe que no todos los intervalos incluyen la media poblacional; los dos puntos extremos de la quinta muestra son inferiores a la media poblacional. Esto se debe al error de muestreo, que constituye el riesgo que se asume cuando se selecciona el nivel de confianza.



Simulación por computadora

Con ayuda de una computadora es posible crear fácilmente muestras aleatorias del tamaño deseado, *n*, de una población y calcular la media para cada muestra de *n* valores, con sus correspondientes numéricos. A partir de la media muestral, la desviación estándar de la población y el nivel de confianza, se determina el intervalo de confianza de cada muestra. Después, usando todas las muestras y los intervalos de confianza, se encuentra la frecuencia con la que la media poblacional está incluida en los intervalos de confianza. En el siguiente ejemplo se muestra precisamente eso.

EJEMPLO

Tras varios años en el negocio de renta de automóviles, Town Bank sabe que la distancia media recorrida en un contrato de cuatro años es de 50 000 millas, y la desviación estándar, de 5 000. Estos son los valores poblacionales. Suponga que Town Bank quiere experimentar con la idea de hacer un muestreo para estimar la media poblacional de 50 000 millas; entonces decide elegir un tamaño de muestra de 30 observaciones y un intervalo de confianza de 95% para estimar la media poblacional. Con base en el experimento, encuentre la proporción de intervalos de confianza de 95% que incluirán la media poblacional de 50 000. Se espera que cerca de 95%, o 57 de 60 intervalos, incluyan la media poblacional. Para facilitar los cálculos, trabaje en miles de millas, en lugar de unidades de milla.

SOLUCIÓN

Mediante un software estadístico se generan 60 muestras aleatorias de 30 y se calculan las medias muestrales de cada una. Después, utilizando la *n* de 30 y un error estándar de 0.913 ($\sigma/\sqrt{n} = 5/\sqrt{30}$) se calcula un intervalo de confianza para cada muestra. A continuación se muestran los resultados del experimento.

	Observaciones muestrales													Media	Límites de confianza 95%	
Muestra	**1**	**2**	**3**	**4**	**5**	**–**	**–**	**–**	**26**	**27**	**28**	**29**	**30**	**muestral**	**Límite inferior**	**Límite superior**
1	56	47	47	48	58	–	–	–	55	62	48	61	57	51.6	49.811	53.389
2	55	51	52	40	53	–	–	–	47	54	55	55	45	50.77	48.981	52.559
3	42	46	48	46	41	–	–	–	50	52	50	47	45	48.63	46.841	50.419
4	52	49	55	47	49	–	–	–	46	56	49	43	50	49.9	48.111	51.689
5	48	50	53	48	45	–	–	–	46	51	61	49	47	49.03	47.241	50.819
6	49	44	47	46	48	–	–	–	51	44	51	52	43	47.73	45.941	49.519
7	50	53	39	50	46	–	–	–	55	47	43	50	57	50.2	48.411	51.989
8	47	51	49	58	44	–	–	–	49	57	54	48	48	51.17	49.381	52.959
9	51	44	47	56	45	–	–	–	45	51	49	49	52	50.33	48.541	52.119
10	45	44	52	52	56	–	–	–	52	51	52	50	48	50	48.211	51.789
11	43	52	54	46	54	–	–	–	43	46	49	52	52	51.2	49.411	52.989
12	57	53	48	42	55	–	–	–	49	44	46	46	48	49.8	48.011	51.589
13	53	39	47	51	53	–	–	–	42	44	44	55	58	49.6	47.811	51.389
14	56	55	45	43	57	–	–	–	48	51	52	55	47	49.03	47.241	50.819
15	49	50	39	45	44	–	–	–	49	43	44	51	51	49.37	47.581	51.159
16	46	44	55	53	55	–	–	–	44	53	53	43	44	50.13	48.341	51.919
17	64	52	55	55	43	–	–	–	58	46	52	58	55	52.47	50.681	54.259
18	57	51	60	40	53	–	–	–	50	51	53	46	52	50.1	48.311	51.889
19	50	49	51	57	45	–	–	–	53	52	40	45	52	49.6	47.811	51.389
20	45	46	53	57	49	–	–	–	49	43	43	53	48	49.47	47.681	51.259
21	52	45	51	52	45	–	–	–	43	49	49	58	53	50.43	48.641	52.219
22	48	48	52	49	40	–	–	–	50	47	54	51	45	47.53	45.741	49.319
23	48	50	50	53	44	–	–	–	48	57	52	44	39	49.1	47.311	50.889
24	51	51	40	54	52	–	–	–	54	45	50	57	48	50.13	48.341	51.919
25	48	63	41	52	41	–	–	–	48	50	48	44	53	49.33	47.541	51.119
26	47	45	48	59	49	–	–	–	44	47	49	55	42	49.63	47.841	51.419
27	52	45	60	51	52	–	–	–	52	50	54	46	52	49.4	47.611	51.189
28	46	48	46	57	51	–	–	–	51	50	51	41	52	49.33	47.541	51.119
29	46	48	45	42	48	–	–	–	49	43	59	46	50	48.27	46.481	50.059
30	55	48	47	48	48	–	–	–	47	59	54	51	42	50.53	48.741	52.319
31	58	49	56	46	46	–	–	–	44	51	47	51	46	50.77	48.981	52.559
32	53	54	52	58	55	–	–	–	53	52	45	44	51	50	48.211	51.789
33	50	57	56	51	51	–	–	–	58	47	50	56	46	49.7	47.911	51.489
34	61	48	49	53	54	–	–	–	46	46	56	45	54	50.03	48.241	51.819
35	43	42	43	46	49	–	–	–	49	49	56	51	45	49.43	47.641	51.219
36	39	48	48	51	44	–	–	–	54	52	47	50	52	50.07	48.281	51.859
37	48	43	57	42	54	–	–	–	52	50	59	50	52	50.17	48.381	51.959
38	55	43	49	57	45	–	–	–	41	51	51	52	52	49.5	47.711	51.289
39	47	49	58	54	54	–	–	–	50	56	51	56	58	50.37	48.581	52.159
40	47	56	41	50	54	–	–	–	46	56	61	61	45	51.6	49.811	53.389
41	48	47	42	47	62	–	–	–	44	47	49	55	43	49.43	47.641	51.219
42	46	49	43	36	52	–	–	–	45	51	46	51	43	47.67	45.881	49.459
43	44	48	49	48	51	–	–	–	47	52	51	48	49	49.63	47.841	51.419
44	45	52	54	54	49	–	–	–	49	45	53	50	52	49.07	47.281	50.859
45	54	46	54	45	48	–	–	–	55	38	56	50	62	49.53	47.741	51.319
46	48	50	49	52	51	–	–	–	53	57	58	46	50	49.9	48.111	51.689
47	54	55	46	55	50	–	–	–	56	54	50	55	51	50.5	48.711	52.289
48	45	47	47	63	44	–	–	–	45	53	42	53	50	50.1	48.311	51.889
49	47	47	48	54	56	–	–	–	50	48	54	49	51	49.93	48.141	51.719
50	45	61	51	45	54	–	–	–	55	52	47	45	53	51.03	49.241	52.819
51	49	62	43	49	48	–	–	–	49	58	42	58	52	51.07	49.281	52.859
52	54	52	62	43	54	–	–	–	51	57	49	58	55	50.17	48.381	51.959
53	46	50	59	56	46	–	–	–	50	51	52	54	53	50.47	48.681	52.259
54	52	50	48	48	58	–	–	–	58	52	43	61	54	51.77	49.981	53.559

(*continúa*)

(*continuación*)

	Observaciones muestrales													Media	Límites de confianza 95%	
Muestra	**1**	**2**	**3**	**4**	**5**	**–**	**–**	**–**	**26**	**27**	**28**	**29**	**30**	**muestral**	**Límite inferior**	**Límite superior**
55	45	44	46	56	46	–	–	–	43	45	63	48	56	49.37	47.581	51.159
56	60	50	56	51	43	–	–	–	45	43	49	59	54	50.37	48.581	52.159
57	59	56	43	47	52	–	–	–	49	54	50	50	57	49.53	47.741	51.319
58	52	55	48	51	40	–	–	–	53	51	51	52	47	49.77	47.981	51.559
59	53	50	44	53	52	–	–	–	47	50	55	46	51	50.07	48.281	51.859
60	55	54	50	52	43	–	–	–	57	50	48	47	53	52.07	50.281	53.859

Para explicar, en la primera fila, el software estadístico calculó 30 observaciones aleatorias basadas en una distribución con una media de 50 y una desviación estándar de 5. Por razones de espacio, solo se muestran las que van de 1 hasta 5 y de 26 hasta 30. La media de la primera muestra se calcula y se registra como 51.6. En las columnas siguientes se muestran los límites inferior y superior del intervalo de confianza de 95% para la primera muestra. A continuación se muestra el cálculo del intervalo de confianza para la primera muestra:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = 51.6 \pm 1.96 \frac{5}{\sqrt{30}} = 51.6 \pm 1.789$$

Este cálculo se repite para todas las demás muestras. Los resultados del experimento indican que 56, o 93.33%, de los 60 intervalos de confianza de 95% incluyen la media poblacional de 50. Cuatro, o 6.67%, no incluirán la media poblacional. Ese 6.67% está más cercano al estimado de que 5% de los intervalos no incluirían la media poblacional. Los intervalos particulares son las muestras 6, 17, 22 y 42, y están sombreados. Este es otro ejemplo de error muestral, o la posibilidad de que una muestra aleatoria en particular puede no ser una buena representación de la población, y que el intervalo de confianza basado en la muestra no incluye el valor del parámetro de la población.

AUTOEVALUACIÓN 9-1

Bun-and-Run es una franquicia de comida rápida de la zona noreste, la cual se especializa en hamburguesas de media onza, y sándwiches de pescado y pollo; también ofrece refrescos y papas a la francesa. El departamento de planeación de la firma informa que la distribución de ventas diarias de los restaurantes tiende a seguir una distribución normal y que la desviación estándar de la distribución de ventas diarias es de 3 000 dólares. Una muestra de 40 mostró que las ventas medias diarias suman 20 000 dólares.

(a) ¿Cuál es la media de la población?
(b) ¿Cuál es el mejor estimador de la media de la población? ¿Qué nombre recibe este valor?
(c) Construya un intervalo de confianza de 95% de la media poblacional.
(d) Interprete el intervalo de confianza.

EJERCICIOS

1. Se toma una muestra de 49 observaciones de una población normal con una desviación estándar de 10, la media de la muestra es de 55. Determine el intervalo de confianza de 99% de la media poblacional.

2. Se toma una muestra de 81 observaciones de una población normal con una desviación estándar de 5.0, la media de la muestra es de 40. Determine el intervalo de confianza de 95% de la media poblacional.

3. Se selecciona una muestra de 250 observaciones de una población normal en la cual la desviación estándar poblacional es de 25 y la media de la muestra es de 20.
   a. Determine el error estándar de la media.
   b. Explique por qué se debe utilizar la fórmula [9.1] para determinar el intervalo de confianza de 95%.
   c. Determine el intervalo de confianza de 95% de la media de la población.

4. Suponga que desea un nivel de confianza de 85%. ¿Qué valor utilizaría como z en la fórmula [9.1]?

5. Una empresa de investigación llevó a cabo una encuesta para determinar la cantidad media que los fumadores gastan en cigarrillos durante una semana. La empresa descubrió que la distribución de dichas cantidades tendía a seguir una distribución normal, con una desviación estándar de 5.00 dólares. Una muestra de 49 fumadores reveló que $\bar{x} = \$20$.
   a. ¿Cuál es el estimador puntual de la media de la población? ¿Qué indica esto?
   b. Con el nivel de confianza de 95%, determine el intervalo de confianza de μ. ¿Qué significa esto?

6. Repase el ejercicio anterior. Suponga que se tomó una muestra de 64 fumadores (en lugar de 49) y que la media muestral es la misma.
   a. ¿Cuál es el estimador del intervalo de confianza de 95% de μ?
   b. ¿Por qué este intervalo de confianza es más reducido que el que se determinó en el ejercicio anterior?
7. Bob Nale es propietario de Nale's Texaco GasTown; a él le gustaría estimar la cantidad de galones de gasolina que vendió. Suponga que la cantidad de galones vendidos tiende a seguir una distribución normal, con una desviación estándar de 2.30 galones. De acuerdo con sus registros, selecciona una muestra aleatoria de 60 ventas y descubre que la cantidad media de galones vendidos es de 8.60.
   a. ¿Cuál es el estimador puntual de la media poblacional?
   b. Establezca un intervalo de confianza de 99% para la media poblacional.
   c. Interprete el significado del punto anterior.
8. La doctora Patton es profesora de inglés. Hace poco contó el número de faltas de ortografía que cometió un grupo de estudiantes en sus ensayos. Observó que la distribución de las faltas de ortografía por ensayo se regía por la distribución normal con una desviación estándar de 2.44 palabras por ensayo. En su clase de 40 alumnos de las 10:00 horas, el número medio de palabras con faltas de ortografía fue de 6.05; construya un intervalo de confianza de 95% del número medio de palabras con faltas de ortografía en la población de ensayos.

Desviación estándar poblacional σ desconocida

En la sección anterior se supuso que se conocía la desviación estándar de la población. En el caso de las latas de duraznos de 4.5 onzas de Del Monte, quizá había una gran cantidad de mediciones del proceso de llenado. Por consiguiente, resulta razonable suponer que se dispone de la desviación estándar de la población; sin embargo, en la mayoría de los casos de muestreo no se conoce la desviación estándar de la población (σ). He aquí algunos ejemplos en los que se pretende estimar las medias poblacionales y no se conocen las desviaciones estándares. Suponga que cada uno de los siguientes estudios se relaciona con estudiantes de la West Virginia University.

- El decano de la Facultad de Administración desea estimar la cantidad media de horas de estudiantes de tiempo completo con trabajos remunerativos cada semana. Selecciona una muestra de 30 estudiantes, se pone en contacto con cada uno de ellos y les pregunta cuántas horas laboraron la semana pasada. De acuerdo con la información de la muestra, puede calcular la media muestral, pero no es probable que conozca o pueda determinar la desviación estándar *poblacional* (σ) que se requiere en la fórmula [9.1]. Puede calcular la desviación estándar de la muestra y utilizarla como estimador, pero quizá no conozca la desviación estándar de la población.
- La docente a cargo del asesoramiento de los estudiantes desea estimar la distancia que el estudiante común viaja cada día de su casa a la escuela. Ella selecciona una muestra de 40 estudiantes, los contacta y determina la distancia que recorre cada uno, de su casa al centro universitario. De acuerdo con los datos de la muestra, calcula la distancia media de viaje, es decir, $\bar{x}$. No es probable que conozca la desviación estándar de la población, lo cual, nuevamente, torna obsoleta la fórmula [9.1].
- El director de créditos estudiantiles desea conocer el monto medio de créditos estudiantiles en el momento de la graduación. Él selecciona una muestra de 20 estudiantes graduados y se pone en contacto con cada uno para obtener la información; con base en esta, estima la cantidad media. Sin embargo, para establecer un intervalo de confianza con la fórmula [9.1] es necesaria la desviación estándar de la población, y no es probable que esta información se encuentre disponible.

Por fortuna, se utiliza la desviación estándar de la muestra para estimar la desviación estándar poblacional. Es decir, se utiliza s (la desviación estándar de la muestra) para estimar σ (la desviación estándar de la población); no obstante, al hacerlo no es posible utilizar la fórmula [9.1]. Al desconocer σ, no se puede utilizar la distribución z; sin embargo, hay una solución: utilizar la desviación estándar de la media y sustituir la distribución z con la distribución t.

La distribución t es una distribución de probabilidad continua, con muchas características similares a las de la distribución z. William Gosset, experto cervecero, fue el primero en estudiar la distribución t.

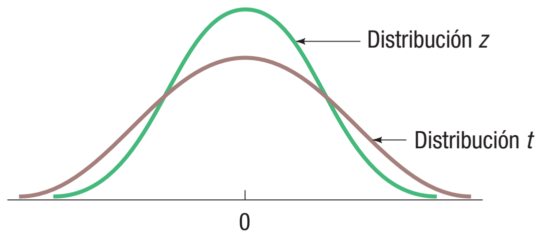
Estaba especialmente interesado en el comportamiento exacto de la distribución del siguiente estadístico:

ESTADÍSTICA EN ACCIÓN

William Gosset nació en Inglaterra en 1876 y murió allí en 1937. Trabajó muchos años en Arthur Guinness, Sons and Company. En realidad, en sus últimos años estuvo a cargo de Guiness Brewery en Londres; esta empresa prefería que sus empleados utilizaran seudónimos cuando publicaban trabajos, de modo que, en 1908, cuando Gosset escribió "The Probable Error of a Mean", utilizó el nombre de *Student*. En este artículo describió por primera vez las propiedades de la distribución t.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Aquí, s es un estimador de σ. Le preocupaba en particular la discrepancia entre s y σ cuando s se calculaba a partir de una muestra muy pequeña. La distribución t y la distribución normal estándar se muestran en la gráfica 9.1; observe que la distribución t es más plana y que se extiende más que la distribución normal estándar. Esto se debe a que la desviación estándar de la distribución t es más larga que la distribución normal estándar.

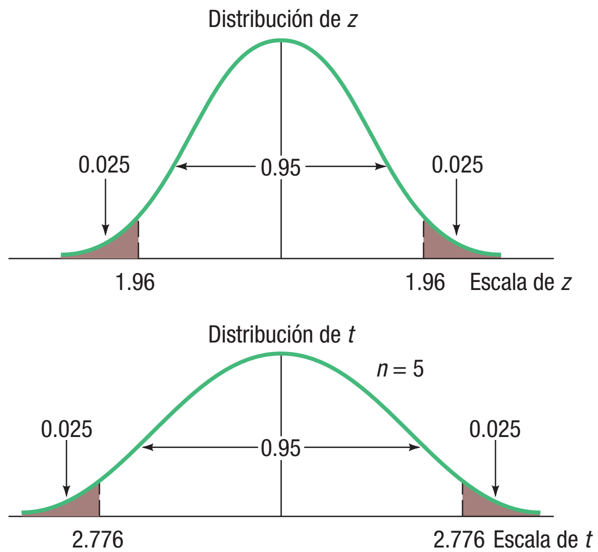


GRÁFICA 9.1 Distribución normal estándar y distribución *t* de Student

Las siguientes características de la distribución t se basan en el supuesto de que la población de interés es de naturaleza normal, o casi normal.

- Como en el caso de la distribución *z,* es continua.
- Como en el caso de la distribución *z,* tiene forma de campana y es simétrica.
- No existe una distribución *t,* sino una familia de distribuciones *t.* Todas estas tienen una media de 0, y sus desviaciones estándares difieren de acuerdo con el tamaño de la muestra, *n.* Existe una distribución *t* para un tamaño de muestra de 20, otro para un tamaño de muestra de 22, etc. La desviación estándar de una distribución *t* con cinco observaciones es mayor que en el caso de una distribución *t* con 20.
- La distribución *t* se extiende más y es más plana por el centro que la distribución normal estándar (vea la gráfica 9.1); sin embargo, conforme se incrementa el tamaño de la muestra, la distribución *t* se aproxima a la distribución normal estándar porque los errores que se cometen al utilizar *s* para estimar σ disminuyen con muestras mayores.

Como la distribución t de Student posee mayor dispersión que la distribución z, t en un nivel de confianza dado tiene una magnitud mayor que el valor z correspondiente. En la gráfica 9.2 se muestran los valores de z para un nivel de confianza de 95% y de t para el mismo nivel de confianza cuando el tamaño de la muestra es de $n = 5$. En breve se explicará cómo se obtuvo el valor real de t; por el momento, observe que, con el mismo nivel de confianza, la distribución t es más plana o más amplia que la distribución normal estándar.



GRÁFICA 9.2 Valores de *z* y *t* para el nivel de confianza de 95%

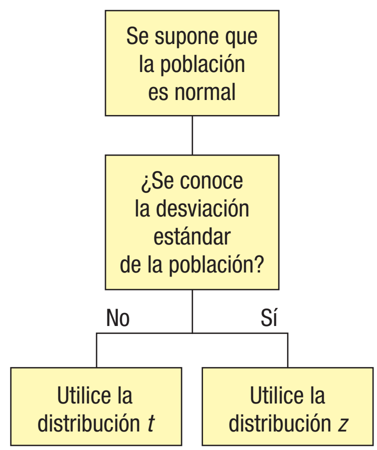
Para crear un intervalo de confianza de la media poblacional con la distribución t, se ajusta la fórmula [9.1] de la siguiente manera.

INTERVALO DE CONFIANZA DE LA MEDIA POBLACIONAL CON UNA σ DESCONOCIDA	$\bar{x} \pm t \dfrac{s}{\sqrt{n}}$	**[9.2]**

Para crear un intervalo de confianza de la media poblacional con una desviación estándar desconocida:

1. Suponga que la población muestreada es normal o aproximadamente normal. De acuerdo con el teorema central del límite, se sabe que este supuesto puede ser cuestionable en el caso de muestras pequeñas, y es más válida en el de muestras mayores.
2. Estime la desviación estándar de la población (σ) con la desviación estándar de la muestra (s).
3. Utilice la distribución t en lugar de la distribución z.

Es preciso hacer una aclaración en este momento: la decisión de utilizar t o z se basa en el hecho de que se conozca σ (la desviación estándar poblacional). Si se conoce, se utiliza z; en caso contrario, se debe utilizar t. En la gráfica 9.3 se resume el proceso de toma de decisión.



GRÁFICA 9.3 Cómo determinar cuándo usar la distribución *z* o la distribución *t*

En el siguiente ejemplo se ilustra un intervalo de confianza de una media poblacional cuando no se conoce la desviación estándar de la población y sirve para determinar el valor apropiado de t en una tabla.

EJEMPLO

Un fabricante de llantas desea investigar la durabilidad de sus productos. Una muestra de 10 llantas que recorrieron 50 000 millas reveló una media muestral de 0.32 pulgadas de cuerda restante con una desviación estándar de 0.09 pulgadas; construya un intervalo de confianza de 95% para la media poblacional. ¿Sería razonable que el fabricante concluyera que después de 50 000 millas el grosor medio poblacional de cuerda restante es de 0.30 pulgadas?

SOLUCIÓN

Para comenzar, suponga que la distribución de la población es normal. En este caso no hay muchas evidencias, pero tal vez la hipótesis es razonable. No se conoce la desviación estándar de la población, pero sí la desviación estándar de la muestra (0.09 pulgadas). Se aplica la fórmula [9.2]:

$$\bar{x} \pm t \frac{s}{\sqrt{n}}$$

De acuerdo con la información dada, $x = 0.32$, $s = 0.09$ y $n = 10$. Para hallar el valor de t, utilice el apéndice B.5 (una parte del cual se reproduce en la tabla 9.2.

TABLA 9.2 Una parte de la distribución t

	Intervalos de confianza				
	80%	90%	95%	98%	99%
	Nivel de significancia de una prueba de una cola				
gl	0.10	0.05	0.025	0.010	0.005
	Nivel de significancia de una prueba de dos colas				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	$\boxed{2.262}$	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

El primer paso para localizar t consiste en desplazarse a lo largo de las columnas “Intervalos de confianza” hasta el nivel de confianza que se requiere. En este caso, se busca el nivel de confianza de 95%, así que vaya a la columna con el encabezamiento “95%”. La columna del margen izquierdo se identifica como “gl”, palabras que se refieren al número de grados de libertad, esto es, el número de observaciones incluidas en la muestra menos el número de muestras, el cual se escribe $n - 1$. En este caso es de $10 - 1 = 9$. ¿Por qué se decidió que había nueve grados de libertad? Cuando se utilizan estadísticas de la muestra, es necesario determinar el número de valores que se encuentran *libres para variar.*

Para ilustrarlo, suponga que la media de cuatro números es 5. Los cuatro números son 7, 4, 1 y 8. Las desviaciones respecto de la media de estos números deben sumar 0. Las desviaciones de $+2$, -1, -4 y $+3$ suman 0; si se conocen las desviaciones de $+2$, -1 y -4, el valor de $+3$ se fija (se restringe) con el fin de satisfacer la condición de que la suma de las desviaciones debe totalizar 0. Por consiguiente, un grado de libertad se pierde en un problema de muestreo que implique la desviación estándar de la muestra, pues se conoce un número (la media aritmética). En el caso de un nivel de confianza de 95% y nueve grados de libertad, seleccione la fila con esa cantidad de grados de libertad. El valor de t es 2.262.

Para determinar el intervalo de confianza se sustituyen los valores en la fórmula [9.2]:

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 0.32 \pm 2.262\frac{0.09}{\sqrt{10}} = 0.32 \pm 0.064$$

Los puntos extremos del intervalo de confianza son 0.256 y 0.384. ¿Cómo se interpreta este resultado? Si se repitiéra este estudio 200 veces, calculando el intervalo de confianza de 95% con cada media de la muestra y la desviación estándar, 190 intervalos incluirían la media poblacional; 10 intervalos no la incluirían; este es el efecto del error muestral. Otra interpretación es concluir que la media poblacional se encuentra en este intervalo. El fabricante puede estar seguro (95% seguro) de que la profundidad media de las cuerdas oscila entre 0.256 y 0.384 pulgadas. Como el valor de 0.30 se encuentra en este intervalo, es posible que la media de la población sea de 0.30 pulgadas.

He aquí otro ejemplo para explicar el uso de los intervalos de confianza. Suponga que un artículo publicado en el periódico local indica que el tiempo medio para vender una residencia de la zona es de 60 días. Usted selecciona una muestra aleatoria de 20 residencias que se vendieron en el último año y encuentra que el tiempo medio de venta es de 65 días. De acuerdo con los datos de la muestra, crea un intervalo de confianza de 95% de la media de la población y descubre que los puntos extremos son 62 y 68 días. ¿Cómo interpreta este resultado? Puede confiar de manera razonable en que la media poblacional se encuentre dentro de este intervalo. El valor propuesto para la media poblacional, es decir, 60 días, no se incluye en el intervalo; por tanto, no es probable que la media poblacional sea de 60 días. La evidencia indica que la afirmación del periódico local puede no ser correcta; en otras palabras, parece poco razonable obtener la muestra que usted tomó de una población que tenía un tiempo de venta medio de 60 días.

En el siguiente ejemplo se indican detalles adicionales para determinar e interpretar el intervalo de confianza. Se usó Minitab para realizar los cálculos.

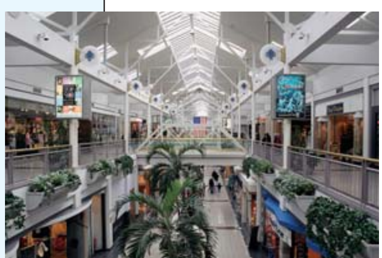
EJEMPLO

El gerente de Inlet Square Mall, cerca de Ft. Myers, Florida, desea estimar la cantidad media que gastan los clientes que visitan el centro comercial. Una muestra de 20 clientes revela las siguientes cantidades.

\$48.16	\$42.22	\$46.82	\$51.45	\$23.78	\$41.86	\$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

¿Cuál es la mejor estimación de la media poblacional? Determine un intervalo de confianza de 95% e interprete el resultado. ¿Concluiría de forma razonable que la media poblacional es de 50 o 60 dólares?

SOLUCIÓN



El gerente del centro comercial supone que la población de las cantidades gastadas sigue la distribución normal. En este caso es una suposición razonable; además, la técnica del intervalo de confianza resulta muy poderosa y tiende a consignar cualquier error del lado conservador si la población no es normal. No cabe suponer una condición normal cuando la población se encuentra pronunciadamente sesgada o la distribución tiene colas *gruesas*. En el capítulo 16 se exponen métodos para manejar este problema cuando no es posible suponer una condición normal. En este caso, resulta razonable suponer una condición normal.

No se conoce la desviación estándar de la población; de ahí que resulte adecuado utilizar la distribución t y la fórmula [9.2] para encontrar el intervalo de confianza. Se utiliza el software Minitab para hallar la media y la desviación estándar de esta muestra. Los resultados se indican en la página siguiente.

El gerente del centro comercial no conoce la media poblacional. La media muestral constituye la mejor aproximación de dicho valor. De acuerdo con el resultado de Minitab, la media es de 49.35

Worksheet 1 ***

	C1
	amount
1	48.16
2	42.22
3	46.82
4	51.45
5	23.78
6	41.86
7	54.86
8	37.92
9	52.64
10	48.59
11	50.82
12	46.94

Session

```
Descriptive Statistics: amount

Variable   N  N*   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3  Maximum
amount    20   0  49.35     2.02   9.01    23.78  44.62   50.00  54.31    61.83

One-Sample T: amount

Variable   N   Mean  StDev  SE Mean       95% CI
amount    20  49.35   9.01     2.02  (45.13, 53.57)
```

dólares, que constituye la mejor aproximación, la *estimación puntual,* de la media poblacional desconocida.

Se aplica la fórmula [9.2] para determinar el intervalo de confianza; el valor de t se localiza en el apéndice B.5. Hay $n - 1 = 20 - 1 = 19$ grados de libertad. Al desplazarse por el renglón con 19 grados de libertad a la columna del intervalo de confianza de 95%, el valor de esta intersección es de 2.093; estos se sustituyen en la fórmula [9.2] para encontrar el intervalo de confianza.

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = \$49.35 \pm 2.093\frac{\$9.01}{\sqrt{10}} = \$49.35 \pm \$4.22$$

Los puntos extremos del intervalo de confianza son $45.13 y $53.57. Resulta razonable concluir que la media poblacional se encuentra en dicho intervalo.

El gerente de Inlet Square se preguntaba si la media poblacional podría haber sido 50 o 60 dólares. Como el valor $50 se encuentra dentro del intervalo de confianza, resulta razonable que la media poblacional sea de 50 dólares. Dado que el valor $60 no se encuentra en el intervalo de confianza, se concluye que no es probable que la media poblacional sea de 60 dólares.

shopping

	A	B	C	D	E
1	Cantidad		Cantidad		
2	48.16				
3	42.22		Media	49.35	
4	46.82		Error estándar	2.02	
5	51.45		Mediana	50.00	
6	23.78		Moda	#N/A	
7	41.86		Desviación estándar	9.01	
8	54.86		Varianza de la muestra	81.22	
9	37.92		Curtosis	2.26	
10	52.64		Sesgo	-1.00	
11	48.59		Rango	38.05	
12	50.82		Mínimo	23.78	
13	46.94		Máximo	61.83	
14	61.83		Suma	986.96	
15	61.69		Cuenta	20.00	
16	49.17		Nivel de confianza (95.0%)	4.22	
17	61.46				
18	51.35				
19	52.68				
20	58.84				
21	43.88				

Los cálculos para construir un intervalo de confianza también se encuentran disponibles en Excel; la salida se indica a la izquierda. Observe que la media de la muestra ($49.35) y la desviación estándar de la muestra ($9.01) son las mismas que en los cálculos de Minitab. En la información de Excel, el último renglón de la salida también incluye el margen de error, que es la cantidad que se suma, y se resta de la media muestral para formar los puntos extremos del intervalo de confianza; este valor se determina a partir de la expresión

$$t\frac{s}{\sqrt{n}} = 2.093\frac{\$9.01}{\sqrt{10}} = \$4.22$$

Antes de resolver los ejercicios del intervalo de confianza, es preciso señalar una útil característica de la distribución t: esta permite utilizar la tabla t para encontrar con rapidez tanto los valores z como los t. Previamente en esta sección, se detallaron las características de la distribución t. El último punto fue que a medida que crece el tamaño de la muestra, la distribución t se aproxima a la distribución z; de hecho, cuando se alcanza una muestra infinitamente grande, la distribución t es exactamente igual a la distribución z.

Para explicar, en la tabla 9.3 se muestra una porción del apéndice B.5, omitiendo los grados de libertad entre 4 y 99. Para encontrar el valor z adecuado para el intervalo de confianza de 95%, comience ubicándose en la sección de intervalos de confianza y seleccionando la columna encabezada por "95%". Desplácese hacia abajo por esa columna hasta la última fila, etiquetada ∞ o grados infinitos de libertad. El valor reportado es 1.960, el mismo que se encontró utilizando la distribución normal estándar en el apéndice B.3. Esto confirma la convergencia de la distribución t con la distribución z.

¿Qué significa esto? Que en vez de buscar en el cuerpo de la tabla de valores *z*, se puede ir a la última fila de la tabla *t* y encontrar el valor adecuado para construir un intervalo de confianza. Un beneficio adicional es que los valores tienen tres lugares decimales. Así, utilizando esta tabla para un intervalo de confianza de 90%, baje por la columna "90%" y vea en 1.645, que es un valor *z* más preciso que puede usarse para dicho nivel de confianza. También hay otros con tres lugares decimales disponibles para los intervalos de confianza de 98% y 99%. Observe que se usará la tabla *t*, que se resume en la tabla 9.3, para encontrar los valores *z* con tres decimales para todos los ejercicios y problemas siguientes.

TABLA 9.3 Distribución de la *t* de Student

gl (grados de libertad)	Intervalo de confianza					
	80%	90%	95%	98%	99%	99.9%
	Nivel de significancia para una prueba de una cola, α					
	0.1	0.05	0.02	0.01	0.001	0.0005
	Nivel de significancia para una prueba de dos colas, α					
	0.2	0.1	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	1.290	1.660	1.984	2.364	2.626	3.390
120	1.289	1.658	1.980	2.358	2.617	3.373
140	1.288	1.656	1.977	2.353	2.611	3.361
160	1.287	1.654	1.975	2.350	2.607	3.352
180	1.286	1.653	1.973	2.347	2.603	3.345
200	1.286	1.653	1.972	2.345	2.601	3.340
∞	1.282	1.645	1.960	2.326	2.576	3.291

AUTOEVALUACIÓN 9-2

Dottie Kleman es la "Cookie Lady"; ella hornea y vende galletas en 50 lugares del área de Filadelfia. La señora Kleman está interesada en el ausentismo de sus trabajadoras; la siguiente información se refiere al número de días de ausencias de una muestra de 10 trabajadoras durante el último periodo de pago de dos semanas.

4	1	2	2	1	2	2	1	0	3

(a) Determine la media y la desviación estándar de la muestra.
(b) ¿Cuál es la media de la población? ¿Cuál es la mejor estimación de dicho valor?
(c) Construya un intervalo de confianza de 95% para la media poblacional.
(d) Explique la razón por la que se utiliza la distribución *t* como parte del intervalo de confianza.
(e) ¿Es razonable concluir que la trabajadora común no falta ningún día durante un periodo de pago?

EJERCICIOS

9. Utilice el apéndice B.5 para localizar el valor *t* en las siguientes condiciones.
  a. El tamaño de la muestra es de 12, y el nivel de confianza, de 95%.
  b. El tamaño de la muestra es de 20, y el nivel de confianza, de 90%.
  c. El tamaño de la muestra es de 8, y el nivel de confianza, de 99%.

10. Utilice el apéndice B.5 para localizar el valor de *t* en las siguientes condiciones.
  a. El tamaño de la muestra es de 15, y el nivel de confianza, de 95%.
  b. El tamaño de la muestra es de 24, y el nivel de confianza, de 98%.
  c. El tamaño de la muestra es de 12, y el nivel de confianza, de 90%.

11. El propietario de Britten's Egg Farm desea calcular la cantidad media de huevos que pone cada gallina; una muestra de 20 aves indica que ponen un promedio de 20 huevos al mes, con una desviación estándar de 2.00 huevos al mes.
  a. ¿Cuál es el valor de la media de la población? ¿Cuál es el mejor estimador de este?
  b. Explique por qué necesita utilizar la distribución *t*. ¿Qué suposiciones necesita hacer?

c. ¿Cuál es el valor de *t* en un intervalo de confianza de 95%?
d. Construya el intervalo de confianza de 95% de la media de población.
e. ¿Es razonable concluir que la media poblacional es de 21 huevos? ¿Y de 25 huevos?

12. La industria estadounidense de lácteos desea calcular el consumo medio de leche por año. Una muestra de 16 personas revela que el consumo medio anual es de 60 galones, con una desviación estándar de 20 galones. Asuma que la distribución de la población es normal.
a. ¿Cuál es el valor de la media poblacional? ¿Cuál es el mejor estimador de este?
b. Explique por qué necesita utilizar la distribución *t*. ¿Qué suposiciones necesita hacer?
c. ¿Cuál es el valor de *t* en un intervalo de confianza de 90%?
d. Construya el intervalo de confianza de 90% de la media de población.
e. ¿Es razonable concluir que la media poblacional es de 63 galones?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

13. Merrill Lynch Securities y Health Care Retirement, Inc., son dos grandes empresas ubicadas en el centro de Toledo, Ohio. Contemplan ofrecer de forma conjunta servicio de guardería para sus empleados y, como parte del estudio de viabilidad del proyecto, desean calcular el costo medio semanal por el cuidado de los niños. Una muestra de 10 empleados que recurren al servicio de guardería revela las siguientes cantidades gastadas la semana anterior.

$107	$92	$97	$95	$105	$101	$91	$99	$95	$104

Construya el intervalo de confianza de 90% de la media poblacional e interprete el resultado.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14. Greater Pittsburgh Area Chamber of Commerce desea calcular el tiempo medio que los trabajadores que laboran en el centro de la ciudad necesitan para llegar al trabajo. Una muestra de 15 trabajadores revela las siguientes cantidades de minutos de viaje.

29	38	38	33	38	21	45	34
40	37	37	42	30	29	35	

Construya el intervalo de confianza de 98% de la media poblacional e interprete el resultado.

OA9.3 Calcular e interpretar un intervalo de confianza para una proporción de la población.

Intervalo de confianza de una proporción

En el material expuesto hasta ahora en este capítulo se utiliza la escala de medición de razón; es decir, se emplean variables como ingresos, pesos, distancias y edades. Ahora se considerarán casos como los siguientes:

- El director de servicios profesionales de Southern Technical Institute informa que 80% de sus graduados entra en el mercado laboral en un puesto relacionado con su área de estudio.
- Un representante de ventas afirma que 45% de las ventas de Burger King se lleva a cabo en la ventana de servicio para automóviles.

- Un estudio de las casas del área de Chicago indicó que 85% de las construcciones nuevas cuenta con sistema de aire acondicionado central.
- Una encuesta reciente entre hombres casados de entre 35 y 50 años de edad encontró que 63% creía que ambos cónyuges deben aportar dinero.

Estos ejemplos ilustran la escala de medición nominal cuando el resultado se limita a dos valores; en tales casos, uno de estas observaciones se clasifica en uno o más grupos mutuamente excluyentes. Por ejemplo, un graduado de Southern Tech entra al mercado laboral en un puesto relacionado con su campo de estudio o no lo hace. Un consumidor de Burger King compra en la ventana de servicio para automóviles o no lo hace. Se puede hablar de los grupos en términos de proporciones.

PROPORCIÓN Fracción, razón o porcentaje que indica la parte de la muestra de la población que posee un rasgo de interés particular.

Como ejemplo de proporción, una encuesta reciente indicó que 92 de cada 100 entrevistados estaban de acuerdo con el horario de verano para ahorrar energía. La proporción de la muestra es de 92/100, 0.92 o 92%. Si *p* representa la proporción de la muestra, *x* es el número de éxitos y *n* es el número de elementos de la muestra, es posible determinar una proporción muestral de la siguiente manera:

PROPORCIÓN MUESTRAL $$p = \frac{x}{n}$$ **[9.3]**

La proporción de la población se define por medio de π; es decir, el porcentaje de éxitos en la población. Recuerde, del capítulo 6, que π es la proporción de *éxitos* en una distribución binomial, lo cual permite continuar la práctica de utilizar letras griegas para identificar parámetros de población, y letras latinas para identificar estadísticas muestrales.

Para crear el intervalo de confianza de una proporción, es necesario cumplir con los siguientes supuestos:

1. Las condiciones binomiales, estudiadas en el capítulo 6, han quedado satisfechas; en resumen, estas condiciones son:
   a. Los datos de la muestra son el número de éxitos en n ensayos.
   b. Solo hay dos posibles resultados (lo normal es referirse a uno de estos como "éxito" y al otro como "fracaso").
   c. La probabilidad de un éxito permanece igual entre un ensayo y el siguiente.
   d. Los ensayos son independientes; es decir, el resultado no influye en el resultado de otra.
2. Los valores $n\pi$ y $n(1 - \pi)$ deben ser mayores o iguales que cinco. Esta condición permite recurrir al teorema central del límite y emplear la distribución normal estándar, es decir, z, para completar un intervalo de confianza.

El desarrollo del estimador puntual de la proporción de la población y el intervalo de confianza de una proporción de población es similar a hacerlo para una media; para ilustrarlo considere lo siguiente: John Gail es candidato para representar al tercer distrito de Nebraska ante el Congreso. De una muestra aleatoria de 100 electores en el distrito, 60 indican que planean votar por él en las próximas elecciones. La proporción de la muestra es de 0.60, pero no se conoce la proporción poblacional; es decir, no se conoce qué proporción de electores de la *población* votará por Gail. El valor de la muestra, 0.60, es el mejor estimador del parámetro poblacional desconocido; así, p, que es de 0.60, constituye un estimador de π, que no se conoce.

Para crear el intervalo de confianza de una proporción de población se aplica la fórmula:

INTERVALO DE CONFIANZA DE LA PROPORCIÓN DE UNA POBLACIÓN $$p \pm z\sqrt{\frac{p(1-p)}{n}}$$ **[9.4]**

Un ejemplo ayudará a explicar los detalles para determinar un intervalo de confianza y el resultado.

ESTADÍSTICA EN ACCIÓN

Los resultados de muchas encuestas que aparecen en periódicos, revistas de noticias y televisión utilizan intervalos de confianza; por ejemplo, una encuesta reciente de 800 televidentes de Toledo, Ohio, reveló que 44% observaba las noticias de la noche en la estación local afiliada a CBS. El artículo también indicó que el margen de error fue de 3.4%; el cual es, en realidad, la cantidad que se suma y resta del estimador puntual para determinar los puntos extremos de un intervalo de confianza. De acuerdo con la fórmula [9.4] y el nivel de confianza de 95%:

$$z\sqrt{\frac{p(1-p)}{n}}$$

$$= 1.96\sqrt{\frac{0.44(1-0.44)}{800}}$$

$$= 0.034$$

EJEMPLO

El sindicato que representa a Bottle Blowers of America (BBA) considera la propuesta de fusión con Teamsters Union. De acuerdo con el reglamento del sindicato de BBA, por lo menos tres cuartas partes de los miembros del sindicato deben aprobar cualquier fusión. Una muestra aleatoria de 2 000 miembros actuales de BBA revela que 1 600 planean apoyar la propuesta. Determine el estimador de la proporción poblacional y el intervalo de confianza de 95% de la proporción poblacional. Si fundamenta su decisión en esta información de la muestra, ¿puede concluir que la proporción necesaria de miembros del BBA favorece la fusión? ¿Por qué?

SOLUCIÓN

Primero estime la proporción de la muestra de acuerdo con la fórmula [9.3]; la cual es de 0.80, y se calcula de la siguiente manera:

$$p = \frac{x}{n} = \frac{1\,600}{2\,000} = 0.80$$

Por consiguiente, se calcula que 80% de la población favorece la propuesta de fusión. Determine el intervalo de confianza de 95% con ayuda de la fórmula [9.4]; el valor z correspondiente al nivel de confianza de 95% es de 1.96.

$$p \pm z\sqrt{\frac{p(1-p)}{n}} = 0.80 \pm 1.96\sqrt{\frac{0.80(1-0.80)}{2\,000}} = 0.80 \pm 0.018$$

Los puntos extremos del intervalo de confianza son 0.782 y 0.818. El punto extremo más bajo es mayor que 0.75; así, es probable que se apruebe la propuesta de fusión, pues el estimador del intervalo incluye valores superiores a 75% de los miembros del sindicato.

He aquí un repaso de la interpretación del intervalo de confianza: si la encuesta se aplicó 100 veces con 100 muestras distintas, los intervalos de confianza construidos a partir de 95 de las muestras contendrán la verdadera proporción de la población; además, la interpretación de un intervalo de confianza resulta de mucha utilidad en la toma de decisiones y desempeña un papel muy importante, en especial la noche de las elecciones. Por ejemplo, Cliff Obermeyer se postula para representar ante el Congreso al 6o. distrito de Nueva Jersey. Suponga que se entrevista a los electores que acaban de votar y 275 indican que votaron por él. Considere que 500 electores es una muestra aleatoria de quienes votan en el 6o. distrito; esto significa que 55% de los electores de la muestra votó por Obermeyer. De acuerdo con la fórmula [9.3]:

$$p = \frac{x}{n} = \frac{275}{500} = 0.55$$

Ahora, para estar seguro de su triunfo, Obermeyer debe ganar *más de* 50% de los votos de la población de electores. En este momento se conoce un estimador puntual (0.55) de la población de electores que votarán por él, pero no se conoce el porcentaje total de la población. En estas circunstancias, la pregunta es: ¿es posible tomar una muestra de 500 electores de una población en la que 50% o menos apoye a Obermeyer para encontrar que 55% de la muestra lo favorece? En otras palabras, ¿el error de muestreo, que es $p - \pi = 0.55 - 0.50 = 0.05$, se debe al azar, o la población de electores que apoya a Obermeyer es superior a 0.50? Al establecer un intervalo de confianza de la proporción de la muestra y hallar que el punto inferior es mayor a 0.50, se concluye que la proporción de electores que apoya a Obermeyer es mayor que 0.50. Esto significa que, de hecho, puede resultar electo. ¿Qué pasa si 0.50 pertenece al intervalo? Entonces se concluye que no es seguro tener mayoría ni es posible asegurar que será electo. En este caso, si se utiliza el nivel de significancia de 95% y la fórmula [9.4], se tiene que:

$$p \pm z\sqrt{\frac{p(1-p)}{n}} = 0.55 \pm 1.96\sqrt{\frac{0.55(1-0.55)}{500}} = 0.55 \pm 0.44$$

Así, los puntos extremos del intervalo de confianza son: $0.55 - 0.044 = 0.506$ y $0.55 + 0.044 = 0.594$. El valor 0.50 no pertenece al intervalo; por lo tanto, se concluye que probablemente *más de 50%* de los electores apoya a Obermeyer, lo cual es suficiente para que sea elegido.

¿Alguna vez se utiliza este procedimiento? Sí, es exactamente el procedimiento que utilizan las cadenas de televisión, revistas de noticias y encuestas de salida en la noche de las elecciones.

AUTOEVALUACIÓN 9-3

Se llevó a cabo una encuesta de mercado para calcular la proporción de amas de casa que reconocerían el nombre de la marca de un limpiador a partir de la forma y color del envase. De las 1 400 amas de casa de la muestra, 420 identificaron la marca por su nombre.

(a) Estime el valor de la proporción de la población.
(b) Construya el intervalo de confianza de 99% de la proporción poblacional.
(c) Interprete sus conclusiones.

EJERCICIOS

15. El propietario de West End Kwick Fill Gas Station desea determinar la proporción de clientes que utilizan tarjeta de crédito o débito para pagar la gasolina en el área de las bombas. Entrevista a 100 clientes y descubre que 80 pagaron de esta manera.

a. Estime el valor de la proporción de la población.
b. Construya el intervalo de confianza de 95% de la proporción poblacional.
c. Interprete sus conclusiones.

16. María Wilson considera postularse para la alcaldía de la ciudad de Bear Gulch, Montana; pero antes de solicitar la postulación decide realizar una encuesta entre los electores de Bear Gulch. Una muestra de 400 electores revela que 300 la apoyarían en las elecciones de noviembre.
  a. Estime el valor de la proporción de la población.
  b. Calcule el error estándar de la proporción.
  c. Construya el intervalo de confianza de 99% de la proporción poblacional.
  d. Interprete sus resultados.

17. La televisora Fox TV considera reemplazar uno de sus programas de investigación ciminal, que se transmite durante las horas de mayor audiencia, por una nueva comedia orientada a la familia. Antes de tomar una decisión definitiva, los ejecutivos estudian una muestra de 400 telespectadores. Después de ver la comedia, 250 afirmaron que la verían y sugirieron reemplazar el programa de investigación criminal.
  a. Estime el valor de la proporción de la población.
  b. Construya el intervalo de confianza de 99% de la proporción poblacional.
  c. Interprete los resultados que obtuvo.

18. Schadek Silkscreen Printing, Inc., compra tazas de plástico para imprimir en ellas logotipos de eventos deportivos, graduaciones, cumpleaños u otras ocasiones importantes. Zack Schadek, el propietario, recibió un envío grande esta mañana. Para asegurarse de la calidad del envío, seleccionó una muestra aleatoria de 300 tazas y halló que 15 estaban defectuosas.
  a. ¿Cuál es la proporción aproximada de tazas defectuosas en la población?
  b. Construya el intervalo de confianza de 95% de la proporción de tazas defectuosas.
  c. Zack acordó con su proveedor que devolverá lotes con 10% o más de artículos defectuosos. ¿Debe devolver este lote? Explique su respuesta.

Elección del tamaño adecuado de una muestra

OA9-4
Calcular el tamaño de la muestra necesario para estimar una proporción de la población o una media poblacional.

Una variable importante cuando se trabaja con intervalos de confianza es el tamaño de la muestra; sin embargo, en la práctica, no es una variable, sino una decisión que se toma para que la estimación del parámetro de población sea bueno. Esta decisión se basa en tres variables:

1. El margen de error que tolerará el investigador.
2. El nivel de confianza deseado.
3. La variación o dispersión de la población que se estudia.

La primera variable es el *margen de error.* El máximo error admisible, designado *E,* es la magnitud que se suma y resta de la media muestral (o proporción muestral) para determinar los puntos extremos del intervalo de confianza. Por ejemplo, en un estudio de salarios se desea estimar el sueldo promedio de la población con un margen de error de más o menos 1 000 dólares. O en una encuesta de opinión se desea calcular la proporción de la población con un margen de error de más o menos 3.5%; esto es, la magnitud del error que se tolerará al estimar un parámetro poblacional. Quizás se pregunte por qué no elegir márgenes pequeños de error. Existe una compensación entre el margen de error y el tamaño de la muestra; un margen de error pequeño requiere una muestra más grande y más tiempo y dinero para recolectarla. Un margen de error más grande genera una muestra más pequeña y un intervalo de confianza más amplio.

La segunda elección es el *nivel de confianza.* Al trabajar con un intervalo de confianza, lógicamente se elegirán niveles de confianza relativamente altos, como de 95 y 99%, que son los más comunes. Para calcular el tamaño de la muestra se necesita un estadístico-z que corresponda al nivel de confianza elegido. El nivel de confianza de 95% corresponde al valor z de 1.96, y el nivel de confianza de 90%, a uno de 1.645 (se usa la tabla de valores t). Observe que las muestras más grandes (con su consecuente requerimiento de más tiempo y dinero para recolectarlas) corresponden a niveles de confianza más altos, y que se utiliza un estadístico z.

El tercer factor en la determinación del tamaño de una muestra es la *desviación estándar de la población*; si esta se encuentra muy dispersa, se requerirá una muestra grande. Por el contrario, si se encuentra concentrada (homogénea), el tamaño de muestra que se requiere será menor; no obstante, puede ser necesario utilizar un estimador de la desviación estándar de la población. He aquí algunas sugerencias para determinar dicho estimador.

1. **Realice un estudio piloto.** Este es el método más común. Suponga que desea un cálculo aproximado de la cantidad de horas que trabajan a la semana los estudiantes matriculados en la Facultad de Administración de la Universidad de Texas. Para probar la validez del cuestiona-

rio, se aplica a una pequeña muestra de estudiantes; a partir de esta pequeña muestra se calcula la desviación estándar de la cantidad de horas que trabajan y se utiliza este valor como la desviación estándar de la población.

2. **Utilice un estudio comparativo.** Aplique este enfoque cuando se encuentre disponible un estimador de la desviación estándar de otro estudio. Suponga que quiere calcular la cantidad de horas semanales que trabajan los recolectores de basura. La información de ciertas dependencias estatales o federales que normalmente estudian la fuerza de trabajo puede ser útil para obtener un cálculo aproximado de la desviación estándar.
3. **Emplee un enfoque basado en el intervalo.** Para aplicar este enfoque es preciso conocer o contar con un cálculo de los valores máximo y mínimo de la población (recuerde, del capítulo 3, en el que se explicó la regla empírica, que se podía esperar que casi todos estos se encontraran a más o menos tres desviaciones estándares de la media si la distribución seguía la distribución normal). Por consiguiente, la distancia entre los valores máximo y mínimo es de seis desviaciones estándar. Puede calcular la desviación estándar como un sexto del rango; por ejemplo, la directora de operaciones del University Bank desea un cálculo aproximado del número de cheques que expiden cada mes los estudiantes universitarios. Ella cree que la distribución del número de cheques es normal. La cantidad mínima de cheques expedidos cada mes es 2, y la máxima, 50. El rango de la cantidad de cheques que se expiden por mes es 48, que se determina al restar $50 - 2$. El estimador de la desviación estándar es entonces de ocho cheques mensuales: 48/6.

Tamaño de la muestra para calcular una media poblacional

Para calcular una media poblacional se expresa la interacción entre estos tres factores y el tamaño de la muestra mediante la fórmula siguiente; observe que esta es el margen de error que se utiliza para calcular los puntos extremos de los intervalos de confianza para estimar una media poblacional.

$$E = z\frac{\sigma}{\sqrt{n}}$$

Al despejar n en esta ecuación se obtiene el siguiente resultado:

TAMAÑO DE LA MUESTRA PARA ESTIMAR LA MEDIA DE LA POBLACIÓN

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad \textbf{[9.5]}$$

donde:

n es el tamaño de la muestra;
z es el valor normal estándar correspondiente al nivel de confianza deseado;
σ es la desviación estándar de la población;
E es el error máximo admisible.

El resultado de este cálculo no siempre es un número entero; en tal caso, se acostumbra redondear *cualquier* resultado fraccionario. Por ejemplo, 201.21 se redondea a 202.

EJEMPLO

Un estudiante de administración pública desea determinar la cantidad media que ganan al mes los miembros de los consejos ciudadanos de las grandes ciudades. El error al calcular la media debe ser inferior a 100 dólares, con un nivel de confianza de 95%. El estudiante encontró un informe del Departamento del Trabajo en el que la desviación estándar es de 1 000 dólares. ¿Cuál es el tamaño de la muestra que se requiere?

SOLUCIÓN

El error máximo admisible, E, es de 100 dólares. El valor z de un nivel de confianza de 95% es de 1.96, y el estimador de la desviación estándar, de 1 000 dólares. Al sustituir estos valores en la fórmula [9.5] se obtiene el tamaño de la muestra que se requiere:

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{(1.96)(\$1\ 000)}{\$100}\right)^2 = (19.6)^2 = 384.16$$

El valor calculado de 384.16 se redondea a 385. Se requiere una muestra de 385 para satisfacer las especificaciones; si el estudiante desea incrementar el nivel de confianza, por ejemplo, a 99%, se requerirá una muestra más grande. El valor z correspondiente al nivel de confianza de 99% es 2.576.

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{(2.576)(\$1\ 000)}{\$100}\right)^2 = (25.76)^2 = 663.58$$

Se recomienda una muestra de 664. Observe cuánto modificó el tamaño de la muestra el cambio en el nivel de confianza. Un incremento del nivel de confianza de 95% al de 99% dio como resultado un incremento de 279 valores o 72% [(664/385) $\times$ 100]; esto puede incrementar mucho el costo del estudio, en términos de tiempo y dinero. De ahí que deba considerarse con cuidado el nivel de confianza.

Tamaño de la muestra para calcular la proporción de una población

Para determinar el tamaño de la muestra en el caso de una proporción es necesario especificar estas mismas tres variables:

1. El margen de error.
2. El nivel de confianza deseado.
3. La variación o dispersión de la población a estudiar.

En el caso de la distribución binomial, el margen de error es:

$$E = z\sqrt{\frac{\pi(1-\pi)}{n}}$$

Si la ecuación se resuelve para despejar n se obtiene lo siguiente:

TAMAÑO DE LA MUESTRA DE LA PROPORCIÓN DE LA POBLACIÓN

$$n = \pi(1-\pi)\left(\frac{z}{E}\right)^2 \qquad \textbf{[9.6]}$$

donde:

n es el tamaño de la muestra;
z es el valor normal estándar correspondiente al nivel de confianza deseado;
π es la proporción de la población;
E es el máximo error tolerable.

Las elecciones del estadístico-z y el margen de error E son las mismas que para calcular la media poblacional; sin embargo, la desviación estándar de la población de una distribución normal está representada por $\pi(1-\pi)$. Para encontrar el valor de una proporción de la población, halle un estudio similar o conduzca un estudio piloto; si no se puede encontrar uno confiable, entonces use un valor de π de 0.50. Observe que $\pi(1-\pi)$ es mayor cuando se utiliza 0.50 y, por lo tanto, sin una buena estimación de la proporción de la población, se sobrestima el tamaño de la muestra. Esta diferencia no afecta al estimador de la proporción de la población.

EJEMPLO

En el estudio del ejemplo anterior también se calcula la proporción de ciudades que cuentan con recolectores de basura privados. El estudiante desea que el margen de error se encuentre a 0.10 de la proporción de la población; el nivel de confianza deseado es de 90%, y no se encuentra disponible ningún estimador de la proporción de la población. ¿Cuál es el tamaño de la muestra que se requiere?

SOLUCIÓN

El estimador de la proporción de la población se encuentra a 0.10, por lo que $E = 0.10$. El nivel de confianza deseado es de 0.90, que corresponde a un valor z de 1.645. Como no se encuentra disponible ningún estimador de la población, se utiliza 0.50; la cantidad de valores que se sugiere es

$$n = (0.5)(1 - 0.5)\left(\frac{1.645}{0.10}\right)^2 = 67.65$$

El investigador necesita una muestra aleatoria de 68 ciudades.

AUTOEVALUACIÓN 9-4

El secretario académico de la universidad desea calcular el promedio aritmético de las calificaciones de los estudiantes que se graduaron durante los últimos 10 años. Los promedios oscilan entre 2.0 y 4.0; el promedio se va a calcular a 0.05 más o menos de la media poblacional, y se calcula que la desviación estándar es de 0.279. Utilice el nivel de confianza de 99%. ¿Ayudaría al secretario a determinar cuántas boletas tiene que estudiar?

EJERCICIOS

19. Se calcula que una población tiene una desviación estándar de 10. Desea estimar la media de la población a menos de 2.00 unidades del error máximo admisible, con un nivel de confianza de 95%. ¿De qué tamaño debe ser la muestra?

20. Quiere estimar la media de la población a menos de 5.00, con un nivel de confianza de 99%. Se calcula que la desviación estándar es de 15. ¿De qué tamaño debe ser la muestra?

21. El estimador de la proporción poblacional debe estar a más o menos 0.05, con un nivel de confianza de 95%. El mejor estimador de la proporción poblacional es de 0.15. ¿De qué tamaño debe ser la muestra que se requiere?

22. El estimador de la proporción poblacional debe estar a más o menos de 0.10, con un nivel de confianza de 99%. El mejor estimador de la proporción poblacional es de 0.45. ¿De qué tamaño debe ser la muestra que se requiere?

23. Se planea llevar a cabo una encuesta para determinar el tiempo medio que ven televisión los ejecutivos corporativos. Una encuesta piloto indicó este es de 12 horas semanales, con una desviación estándar de 3.00 horas. Se desea que el estimador de la media de quienes ven televisión esté a menos de un cuarto de hora. Se utilizará el nivel de confianza de 95%. ¿A cuántos ejecutivos debe entrevistarse?

24. Un procesador de zanahorias corta las hojas, lava los vegetales y los inserta en un paquete. En una caja se guardan 20 paquetes para enviarse. Para controlar el peso de las cajas se revisaron unas cuantas; el peso medio fue de 20.4 libras, y la desviación estándar, de 0.5 libras. ¿Cuántas cajas debe tener la muestra para conseguir una confianza de 95% de que la media de la muestra no difiere de la media de la población por más de 0.2 libras?

25. Suponga que el presidente de Estados Unidos desea un cálculo de la proporción de la población que apoya su actual política relacionada con las revisiones del sistema de seguridad social; él quiere que el cálculo se encuentre a menos de 0.04 de la proporción real. Suponga un nivel de confianza de 95%. Los asesores políticos del presidente calculan que la proporción que apoya la actual política es de 0.60.
- **a.** ¿De qué tamaño debe ser la muestra que se requiere?
- **b.** ¿De qué tamaño debe ser una muestra si no hubiera disponible ningún estimador de la proporción que apoya la actual política?

26. Las encuestas anteriores revelan que 30% de los turistas que van a Las Vegas a jugar durante el fin de semana gasta más de 1 000 dólares cada uno. La gerencia desea actualizar este porcentaje.
- **a.** El nuevo estudio utilizará el nivel de confianza de 90%. El estimador estará a menos de 1% de la proporción de la población. ¿Cuál es el tamaño necesario de la muestra?
- **b.** La gerencia indicó que el tamaño de la muestra determinado es demasiado grande. ¿Qué se puede hacer para reducir la muestra? Con base en su sugerencia, vuelva a calcular el tamaño de la muestra.

OA9-5
Ajustar un intervalo de confianza para poblaciones finitas.

Factor de corrección de una población finita

Las poblaciones de las que se han tomado muestras hasta ahora han sido muy grandes o infinitas. ¿Qué sucedería si la población de la que se toma la muestra no fuera tan grande? Es necesario realizar algunos ajustes en la forma de calcular el error estándar de las medias muestrales y en el error estándar de sus proporciones.

Una población con un límite superior es *finita.* Por ejemplo, hay 21 179 estudiantes en la matrícula de la Eastern Illinois University; hay 40 empleados en Spence Sprockets; Chrysler ensambló 917 Jeeps Wrangler en la planta de Alexis Avenue el día de ayer; o había 65 pacientes programados para cirugía en St. Rose Memorial Hospital en Sarasota el día de ayer. Una población finita tal vez sea muy pequeña pues puede constar de todos los estudiantes registrados para este curso; también es posible que sea muy grande, como todas las personas de la tercera edad que viven en Florida.

En el caso de una población finita, en la que el número total de objetos o individuos es N y el número de objetos o individuos incluidos en la muestra es n, es necesario ajustar los errores muestrales en las fórmulas de los intervalos de confianza. En otras palabras, para determinar el intervalo de confianza de la media, se ajusta el error estándar de la media en las fórmulas [9.1] y [9.2]. Si quiere determinar el intervalo de confianza de una proporción, ajuste el error estándar de la proporción en la fórmula [9.4].

Este ajuste recibe el nombre de **factor de corrección de una población finita** que con frecuencia se abrevia *FCP*, el cual es:

$$FPC = \sqrt{\frac{N - n}{N - 1}}$$

¿Por qué es necesario aplicar un factor y cuál es el efecto de hacerlo? Por lógica, si la muestra es un porcentaje significativo de la población, el estimador es más preciso. Observe el efecto del término $(N - n)/(N - 1)$. Suponga que la población es de 1 000, y la muestra, de 100. Entonces esta razón es de $(1\,000 - 100)/(1\,000 - 1)$, o 900/999. Al extraer la raíz cuadrada se obtiene el factor de corrección 0.9492; si dicho factor se multiplica por el error estándar, este se *reduce* aproximadamente 5% $(1 - 0.9492 = 0.0508)$. Reducir la magnitud del error estándar da como resultado un intervalo menor de valores al calcular la media poblacional o la proporción poblacional. Si la muestra es de 200, el factor de corrección es de 0.8949, lo cual significa que el error estándar se redujo más de 10%. En la tabla 9.4 se muestran los efectos de diversos tamaños de muestras.

TABLA 9.4 Factor de corrección de una población finita de muestras seleccionadas cuando la población es de 1 000

Tamaño de la muestra	Fracción de la población	Factor de corrección
10	0.010	0.9955
25	0.025	0.9879
50	0.050	0.9752
100	0.100	0.9492
200	0.200	0.8949
500	0.500	0.7075

Así, si quisiera construir un intervalo de confianza de una media a partir de una población finita sin conocer la desviación estándar de la población, la fórmula [9.2] se ajusta de la siguiente manera:

$$\bar{x} \pm t\frac{s}{\sqrt{n}}\left(\sqrt{\frac{N - n}{N - 1}}\right)$$

Haría un ajuste similar en la fórmula [9.4], en caso de una proporción.

En el siguiente ejemplo se resumen los pasos para determinar un intervalo de confianza de una media.

EJEMPLO

Hay 250 familias en Scandia, Pennsylvania. Una muestra aleatoria de 40 de estas familias revela que la contribución anual media a la iglesia fue de 450 dólares, y la desviación estándar, de 75 dólares.

1. ¿Cuál es la media de la población? ¿Cuál es el mejor estimador de la media poblacional?
2. Construya el intervalo de confianza de 90% de la media de la población. ¿Cuáles son los puntos extremos del intervalo de confianza?
3. Utilizando el intervalo de confianza, explique por qué la media poblacional podría ser de 445 dólares. ¿Podría ser de 425 dólares? ¿Por qué?

SOLUCIÓN

Primero observe que la población es finita. Es decir, existe un límite para el número de familias que viven en Scandia, en este caso, 250.

1. No conoce la media poblacional, que es el valor que quiere calcular. El mejor estimador de la media poblacional es la media de la muestra, que es de 450 dólares.
2. La fórmula para determinar el intervalo de confianza para la media de la población es la siguiente:

$$\bar{x} \pm t \frac{s}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

En este caso, sabe que $\bar{x} = 450$, $s = 75$, $N = 250$ y $n = 40$. Como no conoce la desviación estándar de la población, utiliza la distribución *t.* Para hallar el valor apropiado de *t* recurra al apéndice B.5, recorra la parte superior del renglón hasta la columna con el encabezamiento de 90%; los grados de libertad son: $gl = n - 1 = 40 - 1 = 39$; así, vaya a la celda en la que el renglón de *gl* de 39 interseca la columna con el encabezamiento de 90%; el valor es de 1.685 y, al sustituirlos en la fórmula, obtiene:

$$\bar{x} \pm t \frac{s}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

$$= \$450 \pm 1.685 \frac{\$75}{\sqrt{40}}\left(\sqrt{\frac{250-40}{250-1}}\right) = \$450 \pm \$19.98\sqrt{0.8434} \pm \$450 \pm \$18.35$$

Los puntos extremos del intervalo de confianza son \$431.65 y \$468.35.
3. Es probable que la media poblacional sea superior a 431.65 dólares e inferior a 468.35. En otras palabras, ¿la media de la población puede ser de 445 dólares? Sí, pero no es probable que sea de 425 dólares. ¿Por qué? Porque el valor \$445 se encuentra dentro del intervalo de confianza y \$425 no pertenece al intervalo de confianza.

AUTOEVALUACIÓN 9-5

El mismo estudio relacionado con las contribuciones para la iglesia en Scandia reveló que 15 de las 40 familias tomadas de la muestra asiste regularmente a la iglesia; construya el intervalo de confianza de 95% de la población de familias que asiste con regularidad.

EJERCICIOS

27. Se seleccionan al azar 36 artículos de una población de 300; la media de la muestra es de 35, y la desviación estándar, de 5; construya el intervalo de confianza de 95% de la media poblacional.

28. Se seleccionan al azar 45 elementos de una población de 500; la media de la muestra es de 40, y la desviación estándar, de 9; construya el intervalo de confianza de 99% de la media poblacional.

29. La asistencia al juego de béisbol del equipo de Las Ligas Menores, Savannah Colts, la noche anterior fue de 400; una muestra aleatoria de 50 asistentes reveló que la cantidad media de refrescos consumidos por persona fue de 1.86, con una desviación estándar de 0.50; construya el intervalo de confianza de 99% de la cantidad media de refrescos consumidos por persona.

30. Hay 300 soldadores en Maine Shipyards Corporation; una muestra de 30 de ellos reveló que 18 se graduaron en un curso de soldadura certificado; construya el intervalo de confianza de 95% de la proporción de soldadores graduados en un curso de soldadura certificado.

RESUMEN DEL CAPÍTULO

I. Un estimador puntual es un solo valor (estadístico) para estimar el de la población (parámetro).

II. Un intervalo de confianza es un conjunto de valores entre los cuales se espera que ocurra el parámetro de la población.

A. Los factores que determinan la magnitud de un intervalo de confianza de una media son:

1. El número de observaciones en la muestra, *n.*
2. La variabilidad en la población, normalmente calculada por la desviación estándar de la muestra, *s.*
3. El nivel de confianza.

a. Para determinar los límites de confianza cuando se conoce la desviación estándar de la población se utiliza la distribución z; cuya fórmula es:

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} \quad \textbf{[9.1]}$$

b. Para determinar los límites de confianza cuando no se conoce la desviación estándar de la población se utiliza la distribución t; cuya fórmula es:

$$\bar{x} \pm t\frac{s}{\sqrt{n}} \quad \textbf{[9.2]}$$

III. Las principales características de la distribución t son:
A. Es una distribución continua.
B. Tiene forma de campana y es simétrica.
C. Es plana, o más amplia, que la distribución normal estándar.
D. Existe una familia de distribuciones t, según el número de grados de libertad.

IV. Una proporción es una razón, fracción o porcentaje que indica la parte de la muestra o población que posee una característica particular.
A. Una proporción muestral se determina al dividir x (el número de éxitos) entre n (el número de observaciones).
B. Se construyó un intervalo de confianza de una proporción muestral con la siguiente fórmula:

$$p \pm z\sqrt{\frac{p(1-p)}{n}} \quad \textbf{[9.4]}$$

V. Es posible determinar un tamaño apropiado de muestra para calcular tanto medias como proporciones.
A. Hay tres factores que determinan el tamaño de una muestra cuando desea calcular la media.
1. El margen de error máximo, E.
2. El nivel de confianza deseado.
3. La variación en la población.
4. La fórmula para determinar el tamaño muestral de la media es:

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad \textbf{[9.5]}$$

B. Hay tres factores que determinan el tamaño de una muestra cuando se desea calcular una proporción.
1. El margen de error, E.
2. El nivel de confianza deseado.
3. Un valor de π para calcular la variación en la población.
4. La fórmula para determinar el tamaño muestral de una proporción es:

$$n = \pi(1-\pi)\left(\frac{z}{E}\right)^2 \quad \textbf{[9.6]}$$

VI. En el caso de una población finita, el error estándar se ajusta con el factor fórmula $\sqrt{\frac{N-n}{N-1}}$.

EJERCICIOS DEL CAPÍTULO

31. Una muestra aleatoria de líderes de grupo, supervisores y personal similar de General Motors reveló que, en promedio, pasan 6.5 años en su trabajo antes de ascender. La desviación estándar de la muestra fue de 1.7 años; construya el intervalo de confianza de 95%.

32. A un inspector de carne del estado de Iowa se le encargó calcular el peso neto medio de los paquetes de carne molida con la etiqueta “Tres libras”. Por supuesto, se da cuenta de que los paquetes no pesan precisamente tres libras. Una muestra de 36 paquetes revela que el peso medio es de 3.01 libras, con una desviación estándar de 0.03 libras.
a. ¿Cuál es el estimador de la media poblacional?
b. Determine el intervalo de confianza de 95% de la media poblacional.

33. Como parte de su paquete promocional, la Cámara de Comercio de Milwaukee desea tener una estimación del costo medio mensual de un apartamento de una recámara. Una muestra aleatoria de 40 apartamentos disponibles para renta reveló que el costo medio mensual era de 323 dólares, y la desviación estándar de la muestra, de 25 dólares.

a. Construya un intervalo de confianza de 98% para la media poblacional.
b. ¿Sería razonable concluir que la media poblacional es de 350 dólares mensuales?

34. Una encuesta reciente a 50 ejecutivos despedidos reveló que tardaron 26 semanas en conseguir otro empleo. La desviación estándar de la muestra fue de 6.2 semanas; construya el intervalo de confianza de 95% de la media de población. ¿Es razonable que la media poblacional sea de 28 semanas? Justifique su respuesta.

35. Marty Rowatti recién asumió el puesto de director de la YMCA de South Jersey. Le gustaría contar con datos recientes sobre el tiempo que sus miembros actuales han pertenecido a la institución. Para investigarlo, suponga que selecciona una muestra aleatoria de 40 miembros actuales; el tiempo medio de membresía de quienes se encuentran en la muestra es de 8.32 años, y la desviación estándar, de 3.07 años.
a. ¿Cuál es la media de la población?
b. Construya un intervalo de confianza de 90% para la media poblacional.
c. La directora anterior, en el breve informe que preparó al retirarse, indicó que ahora el tiempo medio de membresía era de “casi 10 años”. ¿Confirma la información esta aseveración? Cite evidencias.

36. La American Restaurant Association recopiló información sobre las veces que los matrimonios jóvenes comen fuera de casa cada semana. Una encuesta de 60 parejas demostró que la cantidad media de comidas fuera de casa era de 2.76 por semana, con una desviación estándar de 0.75; construya el intervalo de confianza de 99% para la media de la población.

37. La National Collegiate Athletic Association (NCAA) informó que la cantidad media de horas semanales que los asistentes de los entrenadores de futbol invierten en entrenamiento y reclutamiento durante la temporada es de 70. Una muestra aleatoria de 50 asistentes indicó que la media de la muestra es de 68.6 horas, con una desviación estándar de 8.2 horas.
a. De acuerdo con los datos de la muestra, construya el intervalo de confianza de 99% de la media de la población.
b. ¿Incluye el intervalo de confianza el valor que sugiere la NCAA? Interprete este resultado.
c. Suponga que decidió cambiar el intervalo de confianza de 99% a 95%. Sin realizar cálculos, ¿aumentará el intervalo, se reducirá o permanecerá igual? ¿Qué valores de la fórmula cambiarán?

38. El Departamento de Recursos Humanos de Electronics, Inc., desea incluir un plan dental como parte del paquete de prestaciones. La pregunta que se plantea es: ¿cuánto invierte un empleado común y su familia en gastos dentales al año? Una muestra de 45 empleados revela que la cantidad media que se invirtió el año previo fue de 1 820 dólares, con una desviación estándar de 660 dólares.
a. Construya el intervalo de confianza de 95% de la media poblacional.
b. Al presidente de Electronics, Inc., se le proporcionó la información del punto anterior. Este indicó que podía pagar 1 700 dólares de gastos dentales por empleado. ¿Es posible que la media poblacional pudiera ser de 1 700 dólares? Justifique su respuesta.

39. Un estudiante llevó a cabo un estudio e informó que el intervalo de confianza de 95% de la media variaba entre 46 y 54; y aseguró que la media de la muestra era de 50; la desviación estándar, de 16, y que la muestra era de por lo menos 30 elementos; pero no recordó el número exacto. ¿Puede usted ayudarle?

40. Un estudio reciente llevado a cabo por la American Automobile Dealers Association reveló que la cantidad media de utilidades por automóvil vendido en una muestra de 20 concesionarias fue de 290 dólares, con una desviación estándar de 125 dólares; construya el intervalo de confianza de 95% de la media poblacional.

41. Un estudio de 25 graduados de universidades de cuatro años llevado a cabo por la American Banker´s Association reveló que la cantidad media que debía un estudiante por concepto de crédito estudiantil era de 14 381 dólares. La desviación estándar de la muestra fue de 1 892 dólares; construya el intervalo de confianza de 90% de la media poblacional. ¿Es razonable concluir que la media de la población en realidad es de 15 000 dólares? Indique por qué.

42. Un factor importante en la venta de propiedades residenciales es la cantidad de personas que le echan un vistazo a las casas. Una muestra de 15 casas vendidas recientemente en el área de Buffalo, Nueva York, reveló que el número medio de personas que ven las casas fue de 24, y la desviación estándar de la muestra, de 5 personas; construya el intervalo de confianza de 98% de la media poblacional.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

43. Warren County Telephone Company afirma en su informe anual que “el consumidor habitual gasta 60 dólares mensuales en el servicio local y de larga distancia”. Una muestra de 12 clientes reveló las cantidades que gastaron el mes anterior.

\$64	\$66	\$64	\$66	\$59	\$62	\$67	\$61	\$64	\$58	\$54	\$66

a. ¿Cuál es el estimador puntual de la media poblacional?
b. Construya el intervalo de confianza de 90% de la media poblacional.

c. ¿Es razonable la afirmación de la compañía de que el "consumidor habitual" gasta 60 dólares mensuales? Justifique su respuesta.

44. El fabricante de una nueva línea de impresoras de inyección de tinta desea incluir, como parte de su publicidad, el número de páginas que el usuario puede imprimir con un cartucho. Una muestra de 10 cartuchos reveló el siguiente número de páginas impresas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2 698	2 028	2 474	2 395	2 372	2 475	1 927	3 006	2 334	2 379

a. ¿Cuál es el estimador puntual de la media poblacional?
b. Construya el intervalo de confianza de 95% de la media poblacional.

45. La doctora Susan Benner es psicóloga industrial. En este momento estudia el estrés en los ejecutivos de las compañías de internet; por tanto, elaboró un cuestionario que cree que mide el estrés. Un resultado de 80 indica un nivel de estrés peligroso. Una muestra aleatoria de 15 ejecutivos reveló los siguientes niveles de estrés.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

94	78	83	90	78	99	97	90	97	90	93	94	100	75	84

a. Determine el nivel medio de estrés de esta muestra. ¿Cuál es el estimador puntual de la media poblacional?
b. Construya el intervalo de confianza de 95% de la media poblacional.
c. De acuerdo con la doctora, Benner, ¿es razonable concluir que el nivel medio de estés de los ejecutivos de internet es de 80? Explique.

46. Como requisito para obtener el empleo, los candidatos de Fashion Industries deben pasar por una prueba de drogas. De los últimos 220 solicitantes, 14 reprobaron; construya el nivel de confianza de 99% de la proporción de solicitantes que no pasan la prueba. ¿Es razonable concluir que más de 10% de los solicitantes no la superan?

47. Fashion Industries aplica pruebas aleatorias a sus empleados a lo largo del año; durante el ciclo anterior, de las 400 pruebas aleatorias aplicadas, 14 no pasaron. Desarrolle un intervalo de confianza de 99% para la proporción de sujetos que fallaron en la prueba. ¿Es razonable concluir que menos de 5% de los empleados no pasan la prueba aleatoria de drogas? Explique su respuesta.

48. Durante un debate nacional sobre cambios en el sistema de salud, un servicio de noticias por cable realizó una encuesta de opinión entre 500 pequeños propietarios de empresas. Se reveló que 65% de estos no aprueban los cambios; construya el intervalo de confianza de 95% de la proporción que se opone a dichos cambios en el sistema de salud. Comente los resultados.

49. En York County, Carolina del Sur, hay 20 000 votantes. Una muestra aleatoria de 500 electores de esa localidad reveló que 350 planean apoyar el regreso al senado de Louella Miller; construya el intervalo de confianza de 99% de la proporción de votantes en el condado que planea favorecer a Millar. A partir de la información de esta muestra, ¿es razonable concluir que la señora Miller recibirá una mayoría de votos?

50. En una encuesta para medir la popularidad del presidente, se pidió a una muestra aleatoria de 1 000 electores que marcara una de las siguientes afirmaciones:
1. El presidente hace un buen trabajo.
2. El presidente realiza un trabajo deficiente.
3. Prefiero no opinar.
Un total de 500 entrevistados eligió la primera afirmación e indicó que considera que el presidente realiza un buen trabajo.
a. Construya el intervalo de confianza de 95% de la proporción de entrevistados que piensan que el presidente hace un buen trabajo.
b. Con base en el intervalo del punto anterior, ¿es razonable concluir que la mayoría (más de la mitad) de la población considera que el presidente realiza un buen trabajo?

51. Edward Wilkin, jefe de la policía de River City, informa que hubo 500 infracciones de tránsito el mes anterior. Una muestra de 35 de estas infracciones mostró que la suma media de estas fue de 54 dólares, con una desviación estándar de 4.50 dólares; construya el intervalo de confianza de 95% de la suma media de una multa en River City.

52. El First National Bank de Wilson tiene 650 clientes con cuentas de cheques. Una encuesta reciente a 50 de ellos mostró que 26 tenían una tarjeta Visa con el banco; construya el intervalo de confianza de 99% de la proporción de clientes con cuenta de cheques que tienen una tarjeta Visa con el banco.

53. Se estima que 60% de los hogares en Estados Unidos contrata televisión por cable. A usted le gustaría verificar esta afirmación para su clase de comunicación masiva; si desea que su estimador se encuentre a menos de cinco puntos porcentuales con un nivel de confianza de 95%, ¿qué tamaño de muestra requiere?

54. Usted necesita calcular la cantidad media de días que viajan al año los vendedores. La media de un pequeño estudio piloto fue de 150 días, con una desviación estándar de 14 días. Si usted debe calcular la media poblacional a menos de dos días, ¿a cuántos vendedores debe incluir en la muestra? Utilice un intervalo de confianza de 90%.

55. Usted va a llevar a cabo el sondeo de una muestra para determinar el ingreso medio familiar en un área rural del centro de Florida. La pregunta es: ¿a cuántas familias se debe incluir en la muestra? En una muestra piloto de 10 familias, la desviación estándar fue de 500 dólares. El patrocinador de la encuesta desea que usted utilice un nivel de confianza de 95%. El estimador debe estar dentro de un margen de 100 dólares. ¿A cuántas familias debe entrevistar?

56. *Families USA,* revista mensual que trata temas relacionados con la salud y sus costos, encuestó a 20 de sus suscriptores. Encontró que las primas anuales de seguros de salud para una familia con cobertura de una empresa promediaron 10 979 dólares. La desviación estándar de la muestra fue de 1 000 dólares.
- **a.** Con base en la información de esta muestra, construya el intervalo de confianza de 90% de la prima anual media de la población.
- **b.** ¿De qué tamaño debe ser la muestra para que la media poblacional se encuentre dentro de un margen menor a 250 dólares, con 99% de confianza?

57. La presurización en la cabina del avión influye en la comodidad de los pasajeros; una presurización más alta permite un ambiente más cercano a lo normal y un vuelo más relajado. Un estudio que llevó a cabo un grupo de usuarios de aerolíneas registró la presión de aire correspondiente a 30 vuelos elegidos de forma aleatoria, y reveló una presión equivalente media de 8 000 pies, con una desviación estándar de 300 pies.
- **a.** Establezca un intervalo de confianza de 99% para la presión equivalente de la media poblacional.
- **b.** ¿De qué tamaño necesita ser la muestra para que la media de la población se encuentre dentro de un margen de 25 pies, con una confianza de 95%?

58. Una muestra aleatoria de 25 personas empleadas por las autoridades del estado de Florida estableció que ganaban un salario promedio (con prestaciones) de 65.00 dólares por hora; la desviación estándar es de 6.25 dólares por hora.
- **a.** ¿Cuál es la media de la población? ¿Cuál es el mejor estimador de la media poblacional?
- **b.** Construya el intervalo de confianza de 99% del salario medio de la población (con prestaciones) de estos empleados.
- **c.** ¿De qué tamaño debe ser la muestra para calcular la media de la población con un error admisible de un dólar, con una confianza de 95%?

59. Una alianza cinematográfica utilizó una muestra aleatoria de 50 ciudadanos estadounidenses para calcular que el estadounidense común vio videos y películas en DVD 78 horas el año previo. La desviación estándar de esta muestra fue de nueve horas.
- **a.** Construya el intervalo de confianza de 95% de la cantidad media poblacional de horas empleadas en ver videos y películas en DVD el año anterior.
- **b.** ¿De qué tamaño debe ser la muestra para que resulte 90% confiable de que la media de la muestra se encuentra dentro de un margen de 1.0 horas de la media de la población?

60. Dylan Jones lleva registros meticulosos de la eficiencia en el gasto de combustible de su nuevo auto. Después de las primeras nueve veces que llenó el tanque, encontró que el rendimiento medio era de 23.4 millas por galón (mpg), con una desviación estándar muestral de 0.9 mpg.
- **a.** Calcule el intervalo de confianza del 95% para su rendimiento.
- **b.** ¿Cuántas veces debe llenar el tanque de gasolina para obtener un margen de error por debajo de 0.1 mpg?

61. Una encuesta a 36 propietarios de iPhone seleccionados al azar mostró que el precio de compra tiene una media de 416 dólares, con una desviación estándar de 180 dólares.
- **a.** Calcule el error estándar de la media muestral.
- **b.** Calcule el intervalo de confianza de 95% de la media.
- **c.** ¿De qué tamaño debe ser la muestra para estimar la media poblacional dentro de 10 dólares?

62. Usted planea llevar a cabo una encuesta para hallar la proporción de fuerza laboral con dos o más trabajos. Decide basarse en un nivel de confianza de 95%, y establece que la proporción estimada debe encontrarse en un margen de menos de 2% de la proporción poblacional. Una encuesta piloto revela que 10% de los 50 entrevistados tenían dos o más trabajos. ¿A cuántos trabajadores debe entrevistar para satisfacer los requisitos?

63. La proporción de contadores públicos que cambiaron de empresa en los últimos tres años se debe calcular con un margen de 3%. Es necesario utilizar el nivel de confianza de 95%; además, un estudio que se realizó hace varios años reveló que el porcentaje de contadores públicos que cambió de compañía en tres años fue de 21.
- **a.** Para actualizar el estudio, ¿cuál es el número de expedientes de contadores públicos que se deben estudiar?

b. ¿Con cuántos contadores públicos es necesario ponerse en contacto si no se cuenta con estimadores anteriores de la proporción poblacional?

64. Como parte de una revisión anual de sus cuentas, un corredor selecciona una muestra aleatoria de 36 clientes. Al revisar sus cuentas, calculó una media de 32 000 dólares, con una desviación estándar de 8 200 dólares. ¿Cuál es el intervalo de confianza de 90% del valor medio de las cuentas de la población de clientes?

65. El Registro Nacional de Control de peso trata de obtener secretos de éxito de gente que ha perdido cuando menos 30 libras y mantuvo su peso por al menos un año. La dependencia reporta que de 2 700 registrados, 459 estuvieron en una dieta baja en carbohidratos (menos de 90 gramos al día).

a. Construya el intervalo de confianza de 95% de esta fracción.

b. ¿Es posible que el porcentaje de la población sea 18%?

c. ¿Qué tan grande debe ser la muestra para estimar la proporción dentro de 0.5%?

66. Cerca de las elecciones, un servicio de noticias por cable conduce una encuesta de opinión de 1 000 probables votantes. El resultado muestra que el contendiente republicano tiene una ventaja de 52% a 48%.

a. Construya el intervalo de confianza de 95% de la proporción que favorece al candidato republicano.

b. Calcule la probabilidad de que el candidato demócrata sea el líder real.

c. Repita el análisis anterior basándose en una muestra de 3 000 probables votantes.

67. Una muestra de 352 suscriptores de la revista *Wired* indicó que el tiempo medio invertido en el uso de internet es de 13.4 horas a la semana, con una desviación estándar de 6.8 horas. Determine un intervalo de confianza de 95% del tiempo medio que pasan los suscriptores en la red.

68. El Tennessee Tourism Institute (TTI) planea hacer un muestreo de la información que proporcionen algunos visitantes que ingresan al estado para saber cuántos de ellos van a acampar. Los cálculos actuales indican que acampa 35% de los visitantes. ¿De qué tamaño debe ser la muestra para calcular la proporción de la población con un nivel de confianza de 95% y un error admisible de 2%?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

69. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.

a. Construya el intervalo de confianza de 95% del precio de venta medio de las casas.

b. Construya el intervalo de confianza de 95% de la distancia media de la casa al centro de la ciudad.

c. Construya el intervalo de confianza de 95% de la proporción de casas con garage.

d. Para reportar sus hallazgos, redacte un informe de negocios a Gary Loftus, presidente de la Cámara de Comercio de Goodyear.

70. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Asuma que la información de 2012 representa una muestra.

a. Construya el intervalo de confianza de 95% de la cantidad media de cuadrangulares por equipo.

b. Construya el intervalo de confianza de 95% de la cantidad media de errores que cometió cada equipo.

c. Construya el intervalo de confianza de 95% de la cantidad media de robos de base de cada equipo.

71. Consulte los datos de los autobuses del Distrito Escolar Buena.

a. Construya el intervalo de confianza de 95% del mantenimiento medio de los autobuses.

b. Construya el intervalo de confianza de 95% del millaje medio de los autobuses.

c. Redacte un informe de negocios para el oficial estatal de transporte para reportar sus resultados.

REPASO DE LOS CAPÍTULOS 8 y 9

El capítulo 8 comenzó con la descripción de las razones por las que es necesario el muestreo. Se hacen muestreos porque es imposible estudiar cada elemento o individuo que integran algunas poblaciones. Resultaría muy costoso y consumiría demasiado tiempo, por ejemplo, ponerse en contacto con todos los ejecutivos de bancos de Estados Unidos y registrar sus ingresos anuales. Asimismo, el muestreo a veces destruye el producto; por ejemplo, un fabricante de medicamentos no puede probar las propiedades de cada tableta elaborada, pues no le quedaría nada para vender. Por consiguiente, para calcular un parámetro poblacional, se selecciona una muestra de la población. Una muestra forma parte de la población, pero debe tenerse cuidado en garantizar que cada miembro de la población tenga la misma oportunidad de que se le elija; de otra manera, las conclusiones pueden estar sesgadas. Es posible aplicar diversos métodos de muestreo, como el *muestreo aleatorio simple, sistemático, estratificado* y *por conglomerados.*

Sin importar qué método de muestreo elija, pocas veces un estadístico de la muestra es igual al parámetro poblacional correspondiente. Por ejemplo, la media de una muestra casi nunca es exactamente la misma que la media de la población; la diferencia entre el estadístico muestral y el parámetro poblacional es el *error de muestreo.*

En el capítulo 8 se demostró que, al seleccionar todas las muestras posibles de determinado tamaño de una población y calcular la media de estas, el resultado es exactamente igual a la media poblacional; también, que la dispersión en la distribución de las medias muestrales es igual a la desviación estándar de la población dividida entre la raíz cuadrada del tamaño de la muestra. Este resultado recibe el nombre de *error estándar de la media.* Existe menos dispersión en la distribución de las medias muestrales que en las poblacionales; además, conforme se incrementa el número de observaciones en cada muestra, se reduce la variación en la distribución del muestreo.

El teorema central del límite es el fundamento de la inferencia estadística; este establece que si la población de la que se seleccionan las muestras sigue la distribución de probabilidad normal, la distribución de las medias muestrales también seguirá dicha distribución. Si la población no es normal, se aproximará a la distribución de probabilidad normal conforme se incremente el tamaño de la muestra.

En el capítulo 9 se explicaron los estimadores puntuales y los estimadores por intervalo. Un estimador puntual es un solo valor que se utiliza para calcular un parámetro de la población. Un estimador por intervalo es un conjunto de valores en el que se espera que se presente el parámetro de la población; por ejemplo, con base en una muestra, se calcula que el ingreso anual medio de los pintores profesionales de casas de Atlanta, Georgia (la población), es de 45 300 dólares. Dicho estimador recibe el nombre de *estimador puntual.* Si establece que la media de la población probablemente se encuentre en el intervalo de 45 200 a 45 400 dólares, dicho estimador se denomina *estimador por intervalo.* Los dos puntos extremos ($45 200 y $45 400) son los *límites de confianza* de la media poblacional. Se describió el procedimiento para establecer un intervalo de confianza para medias grandes y pequeñas, así como para proporciones muestrales. En este capítulo también se expuso un método para determinar el tamaño necesario de una muestra con base en la dispersión de la población, el nivel de confianza deseado y la precisión deseada del estimador o margen de error.

PROBLEMAS

1. Un estudio reciente indicó que las mujeres tomaron un promedio de 8.6 semanas sin goce de sueldo después del nacimiento de su hijo. Suponga que esta distribución sigue la distribución normal de probabilidad, con una desviación estándar de 2.0 semanas, y considere una muestra de 35 mujeres, quienes recién regresaron a trabajar después del nacimiento de su hijo. ¿Cuál es la probabilidad de que la media de esta muestra sea de al menos 8.8 semanas?

2. El gerente de Tee Shirt Emporium informa que la cantidad media de camisas vendidas a la semana es de 1 210, con una desviación estándar de 325, y que las ventas se rigen por la distribución normal. ¿Cuál es la probabilidad de seleccionar una muestra de 25 semanas y encontrar que la media de la muestra es de 1 100 o menos?

3. El dueño de Gulf Stream Cafe pretende calcular el número medio de clientes que almuerzan diariamente. Una muestra de 40 reveló una media de 160 al día, con una desviación estándar de 20 al día; construya el intervalo de confianza de 98% del número medio de clientes diarios.

4. El gerente de la sucursal local de Hamburger Express desea calcular el tiempo medio que los clientes esperan en la ventanilla de servicio para el automóvil. Una muestra de 20 clientes esperó un tiempo medio de 2.65 minutos, con una desviación estándar de 0.45 minutos; construya el intervalo de confianza de 90% del tiempo medio de espera.

5. Defiance Tool and Die tiene 293 oficinas de ventas en todo el mundo. El vicepresidente de ventas estudia el uso de sus máquinas copiadoras; una muestra aleatoria de seis oficinas de ventas reveló la siguiente cantidad de copias sacadas la semana pasada en una oficina seleccionada.

826	931	1 126	918	1 011	1 101

Construya un intervalo de confianza de 95% de la cantidad media de copias por semana.

6. John Kleman es anfitrión del programa de noticias KXYZ Radio 55 AM de Chicago. Durante el programa matutino, él pide a los radioescuchas que se comuniquen y comenten sobre las noticias nacionales y locales. Esta mañana, John se quiso enterar de la cantidad de horas diarias que ven televisión los niños menores de 12 años. Las últimas cinco personas que se comunicaron informaron que, la noche anterior, sus hijos vieron la televisión la siguiente cantidad de horas:

3.0	3.5	4.0	4.5	3.0

¿Es razonable construir un intervalo de confianza a partir de estos datos para indicar la cantidad media de horas diarias que vieron televisión? Si la respuesta es afirmativa, ¿por qué no sería apropiado un intervalo de confianza?

7. Desde siempre, Widgets Manufacturing, Inc., produce 250 artículos al día. Hace poco, el nuevo propietario compró una máquina para fabricar más piezas por día. Una muestra de la producción de 16 días reveló una media de 240 unidades, con una desviación estándar de 35; construya el intervalo de confianza de la cantidad media de partes producidas al día. ¿Parece razonable concluir que se incrementó la producción media diaria? Justifique sus conclusiones.
8. Un fabricante de baterías para teléfono celular desea calcular la vida útil de su producto (en miles de horas): El estimador debe estar dentro de las 0.10 (100) horas. Asuma un nivel de confianza de 95% y que la desviación estándar de la vida útil de la batería es 0.90 (900) horas. Determine el tamaño de la muestra que se requiere.
9. El gerente de una tienda de artículos para hacer mejoras domésticas desea calcular la cantidad media de dinero que se gasta en la tienda. El estimador debe tener un valor con un margen inferior a 4.00 dólares, con un nivel de confianza de 95%. El gerente no conoce la desviación estándar de las cantidades que se han gastado; no obstante, si calcula que el rango va de 5.00 hasta 155.00 dólares, ¿de qué tamaño debe ser la muestra que necesita?
10. En una muestra de 200 residentes de Georgetown County, 120 informaron que creen que el impuesto predial en el condado es muy alto; construya el intervalo de confianza de 95% de la proporción de residentes que creen eso. ¿Parece razonable concluir que la mayoría de los contribuyentes considera que el impuesto predial es muy alto?
11. En los últimos tiempos, el porcentaje de consumidores que adquieren un vehículo nuevo por internet ha sido tan alto que a los distribuidores locales les preocupa el efecto de esta situación en su negocio. La información que se requiere constituye un estimador de la proporción de compras por internet. ¿De qué tamaño debe ser la muestra de compradores para que el estimador se encuentre a dos puntos porcentuales, con un nivel de confianza de 98%? Informes recientes indican que 8% de los vehículos se compra por internet.
12. Desde siempre, la proporción de adultos mayores de 24 años que fuman ha sido de 0.30; hace poco se publicó y transmitió por radio y televisión mucha información de que el tabaquismo no beneficia a la salud. Una muestra de 500 adultos reveló que solo 25% de los entrevistados fumaba; construya el intervalo de confianza de 98% de la proporción de adultos que fuma actualmente. ¿Sería razonable concluir que la proporción de adultos que fuman ha cambiado?
13. El auditor del estado de Ohio necesita un estimador de la proporción de residentes que juegan regularmente a la lotería estatal. De acuerdo con registros anteriores, alrededor de 40% juega con regularidad, pero el auditor quiere información actualizada. ¿De qué tamaño debe ser la muestra para que el estimador se encuentre a tres puntos porcentuales, con un nivel de confianza de 98%?

CASO

Century National Bank

Repase la descripción del Century National Bank, localizada al final del repaso de los capítulos 1 a 4. Cuando el señor Selig asumió el cargo como presidente de Century hace algunos años, apenas comenzaba el uso de las tarjetas de débito. A él le gustaría actualizarse en el uso de estas tarjetas; construya el intervalo de confianza de 95% de la proporción de clientes que las utiliza. ¿Es razonable concluir que más de la mitad de los clientes utiliza tarjeta de débito con base en el intervalo de confianza? Redacte un breve reporte interpretando los resultados.

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. Si cada elemento de la población tiene la misma oportunidad de ser seleccionado, estamos ante un ________. **1.** ________
2. La diferencia entre la media poblacional y la media muestral recibe el nombre de ________. **2.** ________
3. El ________ es la desviación estándar de la distribución de la media muestral. **3.** ________
4. Si aumenta el tamaño de la muestra, la varianza de la media muestral ________ (se reducirá, aumentará, no cambiará). **4.** ________
5. Un solo valor utilizado para calcular el parámetro de una población recibe el nombre de ________. **5.** ________
6. Un rango de valores dentro del cual se espera que se ubique el parámetro de la población recibe el nombre de ________. **6.** ________
7. ¿Cuál de los siguientes factores *no* afecta la amplitud de un intervalo de confianza (tamaño de la muestra, variación en la población, nivel de confianza, tamaño de la población)? **7.** ________
8. La fracción de una población que tiene una característica particular recibe el nombre de ________. **8.** ________
9. ¿Cuál de los siguientes elementos no es una característica de la distribución *t* (con sesgo positivo, continua, media de cero, basada en grados de libertad)? **9.** ________
10. ¿Qué valor se utiliza para determinar el tamaño de muestra requerido de una proporción cuando no se dispone de un estimador de la proporción de la población? **10.** ________

Parte 2: Problemas

1. Los estadounidenses pasan un promedio (media) de 12.2 minutos (al día) en la ducha. La distribución de tiempos es normal, con una desviación estándar de la población de 2.3 minutos. ¿Cuál es la probabilidad de que el tiempo medio por día de una muestra de 12 estadounidenses sea de 11 minutos o menos?
2. Un estudio reciente de 26 residentes de Conway, SC, reveló que habían vivido en su domicilio actual un promedio de 9.3 años. La desviación estándar de la muestra es de dos años.
   a. ¿Cuál es la media poblacional?
   b. ¿Cuál es el mejor estimador de la media poblacional?
   c. ¿Cuál es el error estándar del estimador?
3. Un reciente reporte federal indicó que 27% de los niños entre dos y cinco años de edad comen verduras cuando menos cinco veces a la semana. ¿Qué tan grande debe ser una muestra para calcular la proporción real de la población dentro de 2% con un nivel de confianza de 98%? Asegúrese de usar la información contenida en el reporte federal.
4. La dependencia Philadelphia Area Transit Authority desea calcular la proporción de trabajadores que laboran en el centro de la ciudad que utilizan transporte público para llegar a sus trabajos. Una muestra de 100 empleados reveló que 64 usan el transporte público; construya el intervalo de confianza de 95% de la proporción de la población.

Pruebas de hipótesis de una muestra

10

A DOLE PINEAPPLE, INC., le preocupa la posibilidad de que las latas de 16 onzas de piña rebanada contengan exceso de producto. Suponga que la desviación estándar del proceso es de 0.03 onzas. El departamento de control de calidad tomó una muestra aleatoria de 50 latas y comprobó que la media aritmética del peso era de 16.05 onzas. ¿Puede concluir que el peso medio es mayor a 16 onzas con un nivel de significancia de 5%? Determine el valor *p* (vea el ejercicio 26 y el OA10-5).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA10-1 Definir una hipótesis.

OA10-2 Explicar el proceso de prueba de una hipótesis.

OA10-3 Aplicar el procedimiento de los seis pasos para probar una hipótesis.

OA10-4 Distinguir entre las pruebas de hipótesis de una y dos colas.

OA10-5 Llevar a cabo una prueba de hipótesis de una media poblacional.

OA10-6 Calcular e interpretar el valor *p*.

OA10-7 Utilizar un estadístico *t* para probar una hipótesis.

OA10-8 Calcular la probabilidad de un error tipo II.

ESTADÍSTICA EN ACCIÓN

LASIK es un procedimiento quirúrgico de 15 minutos de duración que utiliza un rayo láser para modificar la forma de la córnea con el fin de mejorar la visión. Las investigaciones demuestran que casi 5% de las cirugías presenta complicaciones, como deslumbramientos, visión borrosa, corrección excesiva o insuficiente de la visión, y su pérdida. Desde una perspectiva estadística, las investigaciones someten a prueba una hipótesis nula acerca de que la cirugía no mejorará la visión frente a la hipótesis alternativa de que la cirugía sí la mejorará. Los datos de la muestra de la cirugía LASIK indican que 5% de los casos presenta complicaciones; este porcentaje representa un índice de error tipo I. Cuando una persona decide someterse a la cirugía, espera rechazar la hipótesis nula; pero en 5% de los casos futuros, esta expectativa no se cumplirá. (Fuente: *American Academy of Ophtalmology Journal*, vol. 16, núm. 43).

Introducción

En el capítulo 8 se inició el estudio de la inferencia estadística. Se describió cómo seleccionar una muestra aleatoria para estimar el valor de un parámetro poblacional. Por ejemplo, se seleccionó una muestra de cinco empleados de Spence Sprockets para determinar la cantidad de años de servicio de cada trabajador entrevistado, se calculó la media de los años de servicio de la muestra y se utilizó para estimar la media de los años de servicio de todo el personal. En otras palabras, se estimó un parámetro poblacional a partir de un estadístico de la muestra.

En el capítulo 9 se prosiguió con el estudio de la inferencia estadística mediante la construcción de un intervalo de confianza; el cual es un conjunto de valores en el que se encuentra el parámetro de la población. En este capítulo, en lugar de crear un conjunto de este tipo, se expone un procedimiento para *probar* la validez de un enunciado relativo a un parámetro poblacional. He aquí algunos ejemplos de enunciados por probar:

- La velocidad media de los automóviles que pasan por la señal de 150 millas de la carretera West Virginia Turnpike es de 68 millas por hora.
- La cantidad media de millas que recorre una Chevy TrailBlazer arrendada durante tres años es de 32 000 millas.
- El tiempo medio que una familia estadounidense habita una vivienda en particular es de 11.8 años.
- En 2013, el salario medio inicial para un graduado de un programa de negocios de cuatro años era de 51 541 dólares. Esto representó un incremento de 2.2% con respecto a 2011.
- De acuerdo con el Libro Azul de Kelley (**www.kbb.com**), un Ford Edge rinde en promedio 19 millas por galón en la ciudad.
- El costo medio para remodelar una oficina en casa es 10 500 dólares.

En este capítulo y algunos de los siguientes se cubren las pruebas de hipótesis estadísticas. Primero se definen los términos *hipótesis estadística* y *pruebas de hipótesis estadísticas;* después se muestran los pasos para llevar a cabo una prueba de hipótesis estadística, y a continuación se aplican pruebas de hipótesis para medias y proporciones. En la última sección del capítulo se describen los posibles errores del muestreo en las pruebas de hipótesis.

¿Qué es una hipótesis?

OA10-1
Definir una hipótesis.

Una hipótesis es una declaración relativa a una población. Por lo tanto, los datos se recaban para verificar lo razonable de un enunciado. Para comenzar, se debe definir la palabra *hipótesis.* En el sistema legal estadounidense, se presume que una persona es inocente hasta que se prueba su culpabilidad. Un jurado plantea como hipótesis que una persona a la que se le imputa un crimen es inocente, y después revisa la evidencia para evaluar si existen pruebas suficientes para sostener que es inocente o culpable. De forma similar, un paciente visita al médico y reporta varios síntomas. Con base en ellos, el médico indicará ciertos exámenes de diagnóstico; enseguida, de acuerdo con los síntomas y los resultados, determina el tratamiento.

En el análisis estadístico se establece una afirmación, una hipótesis, después se recogen datos que se utilizan para probar la aserción. Entonces, una hipótesis estadística se define como:

HIPÓTESIS Afirmación relativa a un parámetro de la población sujeta a verificación.

En la mayoría de los casos, la población es tan grande que no es viable estudiarla por completo; por ejemplo, no sería posible contactar a todos los analistas de sistemas de Estados Unidos para preguntarles su ingreso mensual. Del mismo modo, el departamento de control de calidad de Cooper Tire no puede verificar todos los neumáticos que la empresa produce para ver si duran más de 60 000 millas.

Como se observó en el capítulo 8, una opción para medir o entrevistar a toda la población es tomar una muestra de esta; por lo tanto, así se pone a prueba una declaración para determinar si la muestra apoya o no la declaración en lo concerniente a la población.

¿Qué es la prueba de hipótesis?

OA10-2
Explicar el proceso de prueba de una hipótesis.

Los términos *prueba de hipótesis* y *probar una hipótesis* se utilizan indistintamente. La prueba de hipótesis comienza con una afirmación, o suposición, sobre un parámetro de la población, como la media poblacional. Como ya se indicó, esta afirmación recibe el nombre de *hipótesis.* Por ejemplo, el monto mensual medio de las comisiones de los vendedores de tiendas al menudeo de aparatos electrónicos, como h.h. gregg, es de 2 000 dólares. No es posible entrar en contacto con todos los vendedores de h.h. gregg para comprobar que la media en realidad sea esa. El costo de localizar y entrevistar a cada vendedor de aparatos electrónicos de h.h. gregg en Estados Unidos sería exorbitante. Para probar la validez de la afirmación ($\mu = \$2\ 000$) se debe seleccionar una muestra de la población de vendedores de aparatos electrónicos de h.h. gregg, calcular el estadístico muestral y, con base en ciertas reglas de decisión, rechazar o aceptar la hipótesis. Una media muestral de 1 000 dólares al mes es mucho menos que 2 000 dólares, y muy probablemente sería rechazada. Sin embargo, suponga que la media de la muestra es de 1 995 dólares. ¿Es posible atribuir la diferencia de cinco dólares entre ambas medias al error de muestreo? ¿O dicha diferencia resulta estadísticamente significativa?

PRUEBA DE HIPÓTESIS Procedimiento basado en evidencia de la muestra y la teoría de la probabilidad para determinar si la hipótesis es una afirmación razonable.

Procedimiento de seis pasos para probar una hipótesis

OA10-3
Aplicar el procedimiento de seis pasos para probar una hipótesis.

Existe un procedimiento de seis pasos que sistematiza la prueba de una hipótesis; al llegar al quinto paso, se está en posibilidades de rechazarla o no; sin embargo, la prueba de hipótesis, como la emplean los especialistas en estadística, no define que algo es verdadero de la forma en que un matemático *demuestra* un enunciado. Más bien, proporciona un tipo de *prueba más allá de toda duda razonable,* como en el sistema judicial. De ahí que existan reglas específicas de evidencia, o procedimientos. En el siguiente diagrama se muestran los pasos; analice con detalle cada uno de ellos.

Paso 1		Paso 2		Paso 3		Paso 4		Paso 5		Paso 6
Se establecen las hipótesis nula y alternativa	→	Se selecciona un nivel de significancia	→	Se identifica el estadístico de prueba	→	Se formula la regla de decisión	→	Se toma una muestra y se decide	→	Se interpreta el resultado

Paso 1: se establecen las hipótesis nula (H_0) y alternativa (H_1)

El primer paso consiste en establecer la hipótesis que se debe probar; esta recibe el nombre de **hipótesis nula** y se designa H_0 (se lee "*H* subíndice cero"). La letra mayúscula *H* representa la hipótesis, y el subíndice cero implica que "no hay diferencia". Por lo general se incluye un término *no* en la hipótesis nula, que significa "no hay cambio". Por ejemplo, una hipótesis nula se refiere a que la cantidad media de millas que recorre cada neumático con cinturón de acero no es diferente de 60 000; entonces, se escribiría H_0: $\mu = 60\ 000$. En términos generales, la hipótesis nula se formula para realizar una prueba: o se rechaza o no se rechaza. Es una afirmación que no se rechaza a menos que la información de la muestra ofrezca evidencia convincente de que es falsa.

Cabe hacer hincapié en que, si la hipótesis nula no se rechaza con base en los datos de la muestra, no es posible decir que la hipótesis nula sea verdadera; en otras palabras, el hecho de no rechazar una hipótesis no prueba que H_0 sea verdadera, sino que *no se rechaza* H_0. Probar sin lugar a dudas que la hipótesis nula es verdadera implica conocer el parámetro poblacional; para determinarlo, habría que probar, entrevistar o contar cada elemento de la población, lo cual no resulta factible. La alternativa consiste en tomar una muestra de la población.

También debe destacarse que con frecuencia la hipótesis nula inicia con las expresiones: "No existe diferencia *significativa* entre…" o "La resistencia media del vidrio a los impactos no es *signi-*

ficativamente diferente de…". Al seleccionar una muestra de una población, el estadístico de la muestra es numéricamente distinto del parámetro poblacional hipotético; por ejemplo, suponga que la hipótesis de la resistencia de un platón de vidrio a los impactos es de 70 psi, y que la resistencia media de una muestra de 12 platones de vidrio es de 69.5 psi. Se debe tomar la decisión con la diferencia de 0.5 psi. ¿Se trata de una diferencia real, es decir, una diferencia significativa, o la diferencia entre el estadístico de la muestra (69.5) y el parámetro poblacional hipotético (70.0) es aleatoria y se debe al error de muestreo? Según se dijo, la respuesta a esta pregunta implica una prueba de significancia, que recibe el nombre *prueba de hipótesis.* Una hipótesis nula es:

HIPÓTESIS NULA Enunciado relativo al valor de un parámetro poblacional que se formula con el fin de probar evidencia numérica.

La **hipótesis alternativa** describe lo que se concluirá si se rechaza la hipótesis nula; se representa H_1 (se lee "*H* subíndice uno") y también se le conoce como *hipótesis de investigación.* La hipótesis alternativa se acepta si la información de la muestra ofrece suficiente evidencia estadística para rechazar la hipótesis nula.

HIPÓTESIS ALTERNATIVA Enunciado que se acepta si los datos de la muestra ofrecen suficiente evidencia para rechazar la hipótesis nula.

En el siguiente ejemplo se aclaran los términos hipótesis nula y alternativa. En un artículo reciente se indicó que el tiempo de uso medio de los aviones comerciales estadounidenses es de 15 años. Para llevar a cabo una prueba estadística relacionada con esta afirmación, el primer paso consiste en determinar las hipótesis nula y alternativa. La hipótesis nula representa el estado actual o reportado; se escribe: H_0: $\mu = 15$. La hipótesis alternativa se refiere al hecho de que la afirmación no es verdadera, es decir, H_1: $\mu \neq 15$. Es necesario recordar que, sin que importe la manera de plantear el problema, la *hipótesis nula siempre incluirá el signo de igual*; este signo (=) nunca aparecerá en la hipótesis alternativa porque es la afirmación que se va a probar, y es necesario un valor específico para incluir en los cálculos. Se recurre a la hipótesis alternativa solo si la información sugiere que la hipótesis nula es falsa.

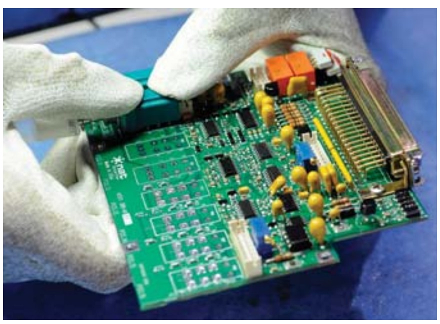
Paso 2: se selecciona un nivel de significancia

Después de establecer las hipótesis nula y alternativa, el siguiente paso consiste en determinar el nivel de significancia.

NIVEL DE SIGNIFICANCIA Probabilidad de rechazar la hipótesis nula cuando es verdadera.

El nivel de significancia se expresa con la letra griega alfa, α. En ocasiones también se conoce como *nivel de riesgo*; este quizá sea un término más adecuado porque se trata del riesgo que se corre al rechazar la hipótesis nula cuando es verdadera.

No existe ningún nivel de significancia que se aplique a todas las pruebas. Se toma la decisión de utilizar el nivel de 0.05 (con frecuencia se expresa como nivel de 5%), nivel de 0.01, nivel de 0.10 o cualquier otro entre 0 y 1. Se acostumbra elegir el nivel de 0.05 en el caso de proyectos de investigación relacionados con consumidores; el nivel de 0.01 en relación en el del control de calidad, y el de 0.10, en el de las encuestas políticas. Todo investigador debe elegir el nivel de significancia *antes* de formular una regla de decisión y recopilar los datos de la muestra.



Para ilustrar cómo es posible rechazar una hipótesis verdadera, suponga que una empresa fabricante de computadoras personales utiliza una gran cantidad de tarjetas con circuitos impresos. Los proveedores participan en una licitación y el que presenta la cotización más baja obtiene el contrato. Suponga que este especifica que el departamento de control de calidad del fabricante de computadoras tomará una muestra de los envíos que llegan; si más de 6% de las tarjetas de la muestra no cumple con las normas, el envío se rechaza. La hipótesis nula consiste en que el envío de tarjetas contiene 6% o menos tarjetas que no satisfacen las normas; la hipótesis alternativa consiste en que más de 6% de las tarjetas están defectuosas.

Se recibió un embarque de cuatro tarjetas de circuitos de Allied Electronics, y el departamento de control de calidad seleccionó una muestra aleatoria de 50 tarjetas de circuito para probarlas; de estas, cuatro, es decir, 8%, no cumplía con las normas. El embarque se rechazó en virtud de que excedía el máximo de 6% de tarjetas que no cumplían con las normas. Si en realidad el envío no cumplía con las normas, fue acertada la decisión de devolver las tarjetas al proveedor; no obstante, suponga que las cuatro tarjetas elegidas de la muestra de 50 eran las únicas que no cumplían con las normas en un envío de 4 000 tarjetas. Entonces, solo 0.1% se encontraba defectuoso (4/4 000 = 0.001). En este caso, menos de 6% de todo el embarque no satisfacía las normas, y rechazarlo fue un error. En términos de la prueba de hipótesis, se rechaza la hipótesis nula de que el embarque cumplía con las normas cuando no debió rechazarse; al hacerlo, se incurrió en un error tipo I. La probabilidad de cometer este tipo de error es α.

ERROR TIPO I Rechazar la hipótesis nula, H_0, cuando es verdadera.

La probabilidad de cometer otro tipo de error, conocido como error tipo II, se expresa con la letra griega beta (β).

ERROR TIPO II Aceptar la hipótesis nula cuando es falsa.

La empresa que fabrica computadoras personales cometería un error del tipo II si, sin que lo sepa el fabricante, un embarque de tarjetas de Allied Electronics contiene 15% de tarjetas que no cumplen con las normas, y aun así lo acepta. ¿Cómo puede suceder esto? Suponga que dos de las 50 tarjetas (4%) no son aceptables, mientras que 48 de 50 lo son. De acuerdo con el procedimiento mencionado, como la muestra contiene menos de 6% de tarjetas que no cumplen con las normas, el embarque se acepta. ¡Puede suceder que, *por azar*, las 48 tarjetas que contiene la muestra sean las únicas aceptables en todo el embarque, que consta de miles de tarjetas!

En retrospectiva, el investigador no puede estudiar cada elemento o individuo de la población; por lo tanto, existe la posibilidad de que se presenten dos clases de error: tipo I (se rechaza la hipótesis nula cuando en realidad debe aceptarse) y tipo II (se acepta la hipótesis nula cuando en realidad debe rechazarse).

Con frecuencia se hace referencia a la probabilidad de cometer estos dos posibles errores como *alfa* (α; es decir, la probabilidad de cometer un error tipo I) y *beta* (β; es decir, la probabilidad de cometer un error tipo II).

En la tabla de la derecha se resumen las decisiones que el investigador puede tomar y sus posibles consecuencias.

	Investigador	
Hipótesis nula	**No rechaza H_0**	**Rechaza H_0**
H_0 es verdadera	Decisión correcta	Error tipo I
H_0 es falsa	Error tipo II	Decisión correcta

Paso 3: se identifica el estadístico de prueba

Hay muchos estadísticos de prueba. En este capítulo se utilizan z y t como estadísticos de prueba; en otros capítulos aparecen estadísticos de prueba como F y χ^2, conocida como *ji*-cuadrada.

ESTADÍSTICO DE PRUEBA Valor, determinado a partir de la información de la muestra, para determinar si se rechaza la hipótesis nula.

La prueba de hipótesis de la media (μ), cuando se conoce σ o el tamaño de la muestra es grande, es el estadístico de prueba z que se calcula de la siguiente manera:

PRUEBA DE LA MEDIA CUANDO SE CONOCE σ

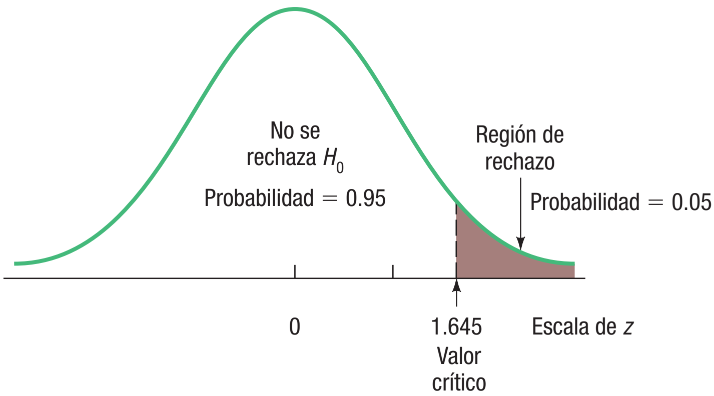
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad [10.1]$$

El valor z se basa en la distribución muestral de $\overline{X}$, que sigue la distribución normal cuando la muestra es razonablemente grande, con una media $(\mu_{\overline{X}})$ igual a μ y una desviación estándar $\sigma_{\overline{X}}$, igual a $\sigma/\sqrt{n}$. Por consiguiente, con la fórmula [10.1] se determina si la diferencia entre $\overline{X}$ y μ es significativa desde una perspectiva estadística al determinar el número de desviaciones estándar a las que se encuentra $\overline{X}$ de μ.

Paso 4: se formula la regla de decisión

Una regla de decisión es un enunciado sobre las condiciones específicas en que se rechaza la hipótesis nula y aquellas en las que no es así. La región o área de rechazo define la ubicación de todos los valores que son tan grandes o tan pequeños que la probabilidad de que ocurran en una hipótesis nula verdadera es muy remota.

En la gráfica 10.1 se presenta la región de rechazo de una prueba de significancia que se efectuará más adelante en este capítulo.



GRÁFICA 10.1 Distribución muestral del estadístico z; prueba de una cola a la derecha; nivel de significancia de 0.05

En la gráfica se observa lo siguiente:

- El área en que se acepta la hipótesis nula se localiza a la izquierda de 1.645 (en breve se explicará la forma de obtener este valor).
- El área de rechazo se encuentra a la derecha de 1.645.
- Se aplica una prueba de una sola cola (este hecho también se explicará más adelante).
- Se eligió el nivel de significancia 0.05.
- La distribución muestral del estadístico z es normal.
- El valor 1.645 separa las regiones en que se rechaza la hipótesis nula y en la que se acepta.
- El valor de 1.645 es el **valor crítico**.

VALOR CRÍTICO Punto de división entre la región en que se rechaza la hipótesis nula y aquella en la que se acepta.

Paso 5: se toma una muestra y se decide

El quinto paso en la prueba de hipótesis consiste en calcular el estadístico de prueba, compararla con el valor crítico y tomar la decisión de rechazar o no la hipótesis nula. De acuerdo con la gráfica 10.1, si a partir de la información de la muestra se calcula que z tiene un valor de 2.34, se rechaza la hipótesis nula con el nivel de significancia 0.05. La decisión de rechazar H_0 se tomó porque 2.34 se localiza en la región de rechazo; es decir, más allá de 1.645; así, es poco probable que un valor z tan alto se deba al error de muestreo (azar).

Si el valor calculado hubiera sido de 1.645 o menos, suponga 0.71, la hipótesis nula no se habría rechazado; un valor calculado tan bajo no se atribuye al azar, es decir, al error de muestreo. Como se indicó, en la prueba de hipótesis solo es posible una de las dos decisiones: la hipótesis nula se rechaza o no se rechaza.

ESTADÍSTICA EN ACCIÓN

Durante la Segunda Guerra Mundial, los encargados de la planeación militar de los aliados necesitaban cálculos aproximados de la cantidad de tanques alemanes. No era confiable la información que proporcionaban los métodos de espionaje tradicionales, y, en cambio, los métodos estadísticos probaron ser muy valiosos; por ejemplo, el espionaje y el reconocimiento llevaron a los analistas a calcular que durante junio de 1941 se produjeron 1 550 tanques; sin embargo, mediante los números de serie de los tanques capturados y el análisis estadístico, los encargados de la planeación militar calcularon que solo se produjeron 244. La cantidad real de tanques producidos, de acuerdo con los registros de producción alemanes, fue de 271; el cálculo a través del análisis estadístico resultó ser mucho más preciso. Un tipo de análisis similar se empleó para calcular la cantidad de tanques iraquíes que fueron destruidos en la guerra conocida como Tormenta del Desierto.

Sin embargo, como la decisión se basa en la muestra, siempre es posible cometer cualquiera de los dos errores de decisión. Es posible cometer un error tipo I cuando la hipótesis nula se rechaza cuando no debía ser así; o es posible cometer un error tipo II cuando la hipótesis nula no se rechaza y debía haberlo sido. Por fortuna, se puede elegir la probabilidad de cometer un error tipo I, α (alfa), y calcular las probabilidades asociadas con un error tipo II, β (beta).

Paso 6: se interpreta el resultado

El paso final en el procedimiento de prueba de hipótesis es interpretar los resultados. El proceso no termina con el valor de un estadístico para una muestra ni con la decisión de rechazar o no la hipótesis nula. ¿Qué se puede decir o reportar con base en los resultados de la prueba estadística? A continuación se presentan dos ejemplos:

- Un reportero investigador de un periódico de Colorado revela que el ingreso medio mensual de las tiendas de abarrotes en el estado es de 130 000 dólares. Usted decide realizar una prueba de hipótesis para verificar este reporte; las hipótesis nula y alternativa son:

$$H_0: \mu = \$130\,000$$
$$H_1: \mu \neq \$130\,000$$

  La muestra de las tiendas de abarrotes proporciona la media muestral y la desviación estándar, y uno debe calcular el estadístico z. Los resultados de la prueba de hipótesis se traducen en la decisión de no rechazar la hipótesis nula. ¿Cómo se interpreta el resultado? Tenga cuidado con esto, porque al no rechazar la hipótesis nula, no probó que esta fuera verdadera. Basándose en la muestra de datos, la diferencia entre la media muestral y la media poblacional hipotética no fue lo bastante grande como para rechazar la hipótesis nula.
- En un reciente discurso a los estudiantes, el decano del Colegio de Negocios reportó que la deuda media en tarjetas de crédito para los universitarios es de 3 000 dólares. Usted decide realizar una prueba de hipótesis para investigar la verdad de la afirmación. Las hipótesis nula y alternativa son:

$$H_0: \mu = \$3\,000$$
$$H_1: \mu \neq \$3\,000$$

  La muestra de los estudiantes universitarios proporciona la media muestral y la desviación estándar, y uno debe calcular el estadístico z. Los resultados de la prueba de hipótesis se traducen en la decisión de rechazar la hipótesis nula. ¿Cómo interpreta el resultado? La evidencia no sustenta la afirmación del decano; usted rechazó la hipótesis nula con una probabilidad determinada de cometer un error tipo I, α. Basándose en los datos de la muestra, la cantidad media de la deuda de los estudiantes en tarjetas de crédito es distinta a 3 000 dólares.

RESUMEN DE LOS PASOS DE LA PRUEBA DE HIPÓTESIS

1. Se establecen las hipótesis nula (H_0) y alternativa (H_1).
2. Se selecciona un nivel de significancia, es decir, α.
3. Se identifica el estadístico de prueba adecuado.
4. Se formula la regla de decisión con base en los pasos 1, 2 y 3 anteriores.
5. Se toma una decisión en lo que se refiere a la hipótesis nula con base en la información de la muestra.
6. Se interpreta el resultado de la prueba.

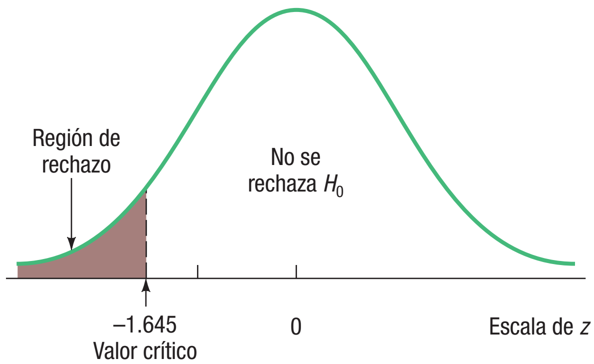
Antes de llevar a cabo una prueba de hipótesis, es importante diferenciar entre una prueba de significancia de una cola y una prueba de dos colas.

Pruebas de significancia de una y dos colas

OA10-4
Distinguir entre las pruebas de hipótesis de una y dos colas.

Consulte la gráfica 10.1; en la cual se describe una prueba de una cola. Se llama así porque la región de rechazo se localiza solo en una cola de la curva; en este caso, está en la derecha, o superior, de la curva. Para ilustrarlo, suponga que el departamento de empaque de General Foods Corporation se preocupa porque algunas cajas de Grape Nuts exceden considerablemente el peso. El cereal se empaca en cajas de 453 gramos, por lo que la hipótesis nula es H_0: $\mu \leq 453$ (se lee: "la media poblacional (μ) es igual o menor que 453"); por consiguiente, la hipótesis alternativa es H_0: $\mu > 453$ (se lee: "μ es mayor que 453"). Observe que el signo de desigualdad en la hipótesis alternativa ($>$) señala hacia la región de rechazo ubicada en la cola superior (vea la gráfica 10.1) y que la hipótesis nula incluye el signo igual; es decir, H_0: $\mu \leq 453$. La condición de igualdad siempre aparece en H_0 y jamás en H_1.

En la gráfica 10.2 se presenta un caso en el que la región de rechazo se encuentra en la cola izquierda (inferior) de la distribución normal. Como ejemplo, considere el problema de los fabricantes de automóviles; en este caso, las grandes compañías de renta de autos y otras empresas que compran grandes cantidades de neumáticos desean que duren un promedio de 60 000 millas en condiciones normales. Por consiguiente, rechazarán un envío de neumáticos si las pruebas revelan que la vida de estas es mucho menor al promedio. Con gusto aceptarán el envío si la vida media es mayor a 60 000 millas; sin embargo, esta posibilidad no les preocupa. Solo les interesa si cuentan con evidencias suficientes para concluir que los neumáticos tendrán un promedio de vida útil inferior a 60 000 millas; por lo tanto, la prueba se plantea de manera que satisfaga la preocupación de los fabricantes de automóviles respecto de que la *vida media de los neumáticos sea menor a 60 000 millas.* Este enunciado aparece en la hipótesis alternativa; en este caso, las hipótesis nula y alternativa se escriben H_0: $\mu \geq 60\,000$ y H_1: $\mu < 60\,000$.



GRÁFICA 10.2 Distribución muestral del estadístico z, prueba de cola izquierda, nivel de significancia de 0.05

Una manera para determinar la ubicación de la región de rechazo consiste en mirar en la dirección en la que señala el signo de desigualdad en la hipótesis alternativa ($<$ o $>$). En el problema del desgaste de las llantas, señala a la izquierda; por tanto, la región de rechazo se localiza en la cola izquierda.

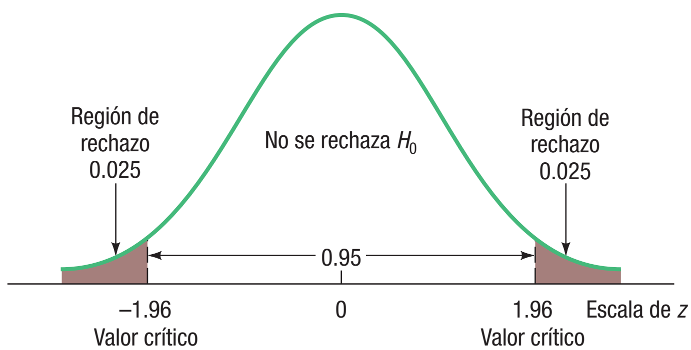
En resumen, una prueba es de *una cola* cuando la hipótesis alternativa, H_1, indica una dirección, como:

H_0: el ingreso medio anual de las corredoras de bolsa es *menor o igual a* 65 000 dólares.
H_1: el ingreso medio anual de las corredoras de bolsa es *mayor a* 65 000 dólares anuales.

Si no se especifica dirección alguna en la hipótesis alternativa, utilice una prueba de *dos colas.* Al cambiar el problema anterior (con fines de ilustración) se puede decir lo siguiente:

H_0: el ingreso medio anual de las corredoras de bolsa es de 65 000 dólares anuales.
H_1: el ingreso medio anual de las corredoras de bolsa *no es igual a* 65 000 dólares anuales.

Si se rechaza la hipótesis nula y se acepta H_1 en el caso de las dos colas, el ingreso medio puede ser significativamente mayor o inferior a 65 000 dólares anuales. Para dar cabida a ambas posibilidades, el área de 5% de rechazo se divide con equidad en las dos colas de la distribución muestral (2.5% cada una). En la gráfica 10.3 se presentan las dos áreas y los valores críticos; observe que el área total en la distribución normal es de 1.0000, que se calcula por medio de 0.9500 + 0.0250 + 0.0250.



GRÁFICA 10.3 Regiones de aceptación y rechazo de una prueba de dos colas con el nivel de significancia 0.05

Pruebas de la media de una población: se conoce la desviación estándar poblacional

OA10-5
Llevar a cabo una prueba de hipótesis de una media poblacional.

Prueba de dos colas

Un ejemplo mostrará los detalles del procedimiento para probar una hipótesis en seis pasos y el proceso subsecuente de toma de decisiones. También se desea usar una prueba de dos colas; es decir, *no interesa* si los resultados de la muestra son mayores o menores que la media poblacional propuesta. Lo que interesa es si esta es *diferente del* valor propuesto para la media poblacional. Como en el capítulo anterior, conviene que se inicie con un caso que proporcione un historial de datos sobre la población y, de hecho, se conozca la desviación estándar.

EJEMPLO

Jamestown Steel Company fabrica y arma escritorios y otros muebles para oficina en diferentes plantas en el oeste del estado de Nueva York. La producción semanal del escritorio modelo A325 en la planta de Fredonia tiene una distribución normal, con una media de 200 y una desviación estándar de 16. Hace poco, con motivo de la expansión del mercado, se introdujeron nuevos métodos de producción y se contrató a más empleados. El vicepresidente de fabricación pretende investigar si hubo algún *cambio* en la producción semanal del escritorio modelo A325; en otras palabras, ¿la cantidad media de escritorios que se produjeron en la planta de Fredonia es *diferente de* 200 escritorios semanales con el nivel de significancia 0.01?

SOLUCIÓN

En este ejemplo hay dos datos importantes: 1) la población de la producción semanal sigue una distribución normal y 2) la desviación estándar de esta distribución normal es de 16 escritorios por semana; por ello, es apropiado utilizar el estadístico z para resolver este problema. Aplique el procedimiento de prueba de hipótesis estadística para investigar si cambió el índice de producción de 200 escritorios semanales.

Paso 1: **Se establecen las hipótesis nula y alternativa.** La hipótesis nula es: "la media de la población es de 200". La hipótesis alternativa es: "la media es diferente de 200" o "la media no es de 200". Ambas se expresan de la siguiente manera:

$$H_0: \mu = 200$$
$$H_1: \mu \neq 200$$

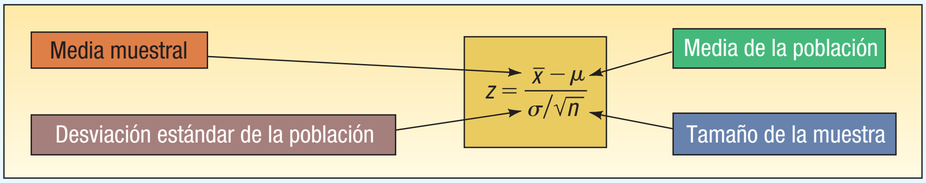
Esta es una *prueba de dos colas,* pues la hipótesis alternativa no indica dirección alguna; en otras palabras, no establece si la producción media es mayor o menor a 200. El vicepresidente solo desea saber si la tasa de producción es distinta de 200.

Antes de ir al paso 2 es preciso enfatizar dos puntos:

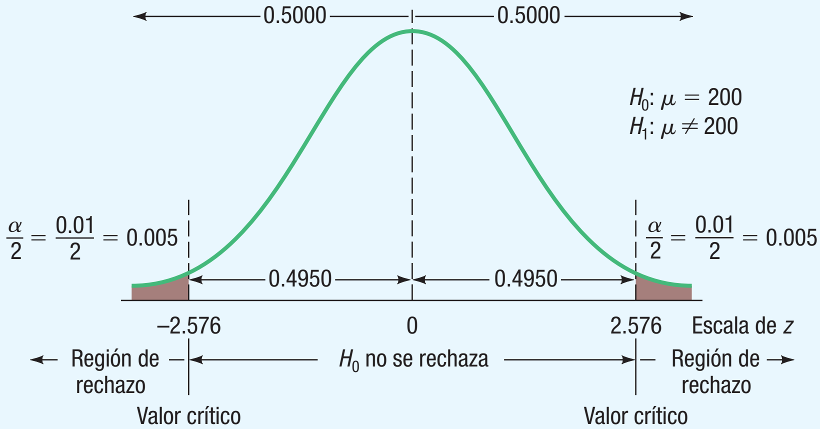
- La hipótesis nula tiene el signo de "igual". ¿Por qué? Porque el valor que se prueba siempre está en la hipótesis nula; por lógica, la hipótesis alternativa nunca contiene el signo de "igual".
- Tanto la hipótesis nula como la alternativa contienen letras griegas, en este caso μ, que es el símbolo de la media poblacional. Las pruebas de hipótesis **siempre** se refieren a parámetros poblacionales, nunca a estadísticos muestrales; dicho de otro modo, nunca verá el símbolo $\bar{x}$ como parte de la hipótesis nula ni de la hipótesis alternativa.

Paso 2: **Se selecciona un nivel de significancia.** En la descripción del ejemplo se utiliza el nivel de significancia 0.01; este es α, la probabilidad de cometer un error tipo I, que es la probabilidad de rechazar una hipótesis nula verdadera.

Paso 3: **Se identifica el estadístico de prueba.** El estadístico de prueba de una muestra grande es z (este tema se estudió lo suficiente en el capítulo 7). La transformación de los valores de producción en unidades estándar (z) permite que se les utilice no solo en este problema, sino en otros relacionados con la prueba de hipótesis. A continuación se repite la fórmula [10.1] para z y se identifican las diferentes letras.



Paso 4: Se formula la regla de decisión. La regla de decisión se formula al encontrar los valores críticos de *z*. Como se trata de una prueba de dos colas, la mitad de 0.01, o 0.005, se localiza en cada cola; por consiguiente, el área en la que no se rechaza H_0, que se ubica entre las dos colas, es 0.99. Use la tabla de distribución *t* de Student en el apéndice B.5, vaya al margen superior llamado "Nivel de significancia para pruebas de dos colas, α", seleccione la columna con $\alpha = 0.01$, y desplácese a la última fila, etiquetada como ∞, o grados infinitos de libertad. El valor *z* de la celda es 2.576. Para su conveniencia, el apéndice B.5 (tabla de distribución *t* de Student) se repite en las últimas páginas del libro. Todas las facetas de este problema se muestran en el diagrama de la gráfica 10.4.



GRÁFICA 10.4 Regla de decisión del nivel de significancia 0.01

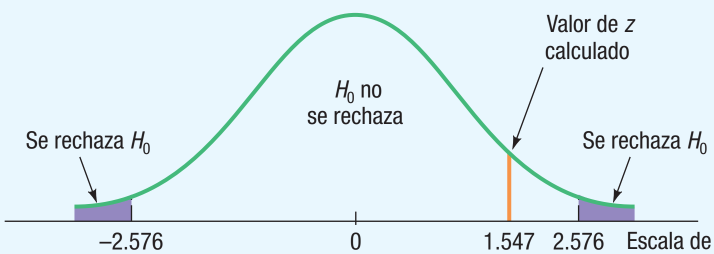
La regla de decisión es: si el valor *z* calculado no se encuentra entre -2.576 y 2.576, rechace la hipótesis nula. Si *z* se ubica entre -2.576 y 2.576, no rechace la hipótesis nula.

Paso 5: Se toma una muestra y se decide. Se toma una muestra de la población (producción semanal), se calcula *z*, se aplica la regla de decisión y se llega a la decisión de rechazar o no H_0. La cantidad media de escritorios que se produjeron el año anterior (50 semanas, pues la planta cerró dos semanas por vacaciones) es de 203.5. La desviación estándar de la población es de 16 escritorios semanales. Al calcular el valor *z* a partir de la fórmula [10.1], se obtiene:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{203.5 - 200}{16/\sqrt{50}} = 1.547$$

Como 1.547 no cae en la región de rechazo, no se rechaza H_0.

Paso 6: Se interpreta el resultado. No se rechaza la hipótesis nula, así que no se demuestra que la media poblacional (200 unidades a la semana) ha cambiado; dicho de otra manera, la diferencia entre la media poblacional de 200 por semana y la media muestral de 203.5 pudo deberse simplemente al azar. ¿Cuál es el reporte para el vicepresidente? La información de la muestra no indica que los nuevos métodos de producción resultaron en un cambio en el ritmo de producción de 200 escritorios a la semana.



¿Se demostró que el ritmo de montaje aún es de 200 a la semana? En realidad, no. Lo que se hizo, desde un punto de vista técnico, fue *no desaprobar la hipótesis nula.* No refutar la hipótesis de que la media poblacional es de 200 no es lo mismo que probar que necesariamente es verdadera. Como se sugiere en la introducción del capítulo, la conclusión es análoga a la del sistema jurídico estadounidense; para explicarlo, suponga que se acusa a una persona de un crimen, pero un jurado la absuelve. Si la persona es absuelta, se concluye que no había suficiente evidencia para probar su culpabilidad. El juicio no probó que el individuo era necesariamente inocente, sino que no había suficiente evidencia para probar su culpabilidad; eso evidencia las pruebas de hipótesis estadísticas cuando no se rechaza la hipótesis nula. La interpretación correcta consiste en que no se probó la falsedad de la hipótesis nula.

En este caso se eligió el nivel de significancia 0.01 antes de establecer la regla de decisión y tomar una muestra de la población; esta es la estrategia adecuada. El investigador debe establecer el nivel de significancia, pero debe determinarlo *antes* de reunir la evidencia de la muestra y no realizar cambios con base en la evidencia de esta.

¿Cómo se confronta el procedimiento de prueba de hipótesis, recién descrito, con el procedimiento de los intervalos de confianza que se estudió en el capítulo anterior? Al realizar la prueba de hipótesis en la producción de escritorios, se cambiaron las unidades de escritorios semanales a un valor *z.* Después se comparó el valor calculado del estadístico de prueba (1.547) con el de los valores críticos (-2.576 y 2.576). Como el valor calculado se localizó en la región de no rechazo de la hipótesis nula, se concluyó que la media poblacional podía ser de 200. Por otro lado, para aplicar el enfoque del intervalo de confianza, este se debía construir con la fórmula [9.1]. El intervalo iría de 197.671 a 209.329, el cual se calcula de la siguiente manera: $203.5 \pm 2.576(16/\sqrt{50})$. Observe que el valor poblacional propuesto, 200, se encuentra en este intervalo; de ahí que la media poblacional podría ser, razonablemente, dicho valor.

En general, H_0 se rechaza si el intervalo de confianza no incluye el valor hipotético; en caso contrario, no se rechaza H_0. Así, la *región de no rechazo* en una prueba de hipótesis equivale al valor poblacional propuesto en el intervalo de confianza.

AUTOEVALUACIÓN 10-1

Heinz, un fabricante de cátsup, utiliza una máquina para vaciar 16 onzas de su salsa en botellas. A partir de su experiencia de varios años con la máquina despachadora, la empresa sabe que la cantidad del producto en cada botella tiene una distribución normal con una media de 16 onzas y una desviación estándar de 0.15 onzas. Una muestra de 15 botellas llenadas durante la hora anterior reveló que la cantidad media por botella era de 16.017 onzas. Con el nivel de significancia 0.05, ¿sugiere la evidencia que la cantidad media despachada es diferente de 16 onzas?

(a) Establezca la hipótesis nula y la hipótesis alternativa.
(b) ¿Cuál es la probabilidad de cometer un error tipo I?
(c) Proporcione la fórmula del estadístico de prueba.
(d) Enuncie la regla de decisión.
(e) Determine el valor del estadístico de prueba.
(f) ¿Cuál es su decisión respecto de la hipótesis nula?
(g) Interprete, en un enunciado, el resultado de la prueba estadística.

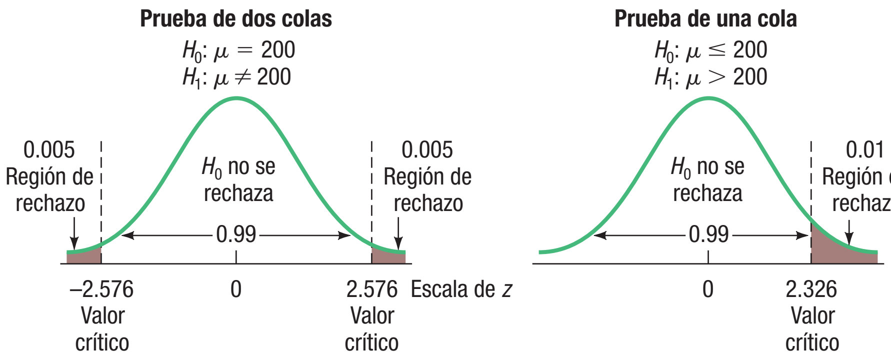
Prueba de una cola

En el ejemplo anterior se destacó el interés solo por informar al vicepresidente si ocurrió un cambio en la cantidad media de escritorios armados en la planta de Fredonia. No importaba si el cambio era un incremento o una disminución de la producción.

Para ilustrar la prueba de una cola, vea otro problema. Suponga que el vicepresidente desea saber si hubo un *incremento* de la cantidad de unidades que se armaron. ¿Puede concluir, debido al mejoramiento de los métodos de producción, que la cantidad media de escritorios que se ensamblaron en las pasadas 50 semanas fue superior a 200? Observe la diferencia al formular el problema; en el primer caso deseaba conocer si había una *diferencia* en la cantidad media armada; en cambio, ahora desea saber si hubo un *incremento.* Como se investigan diferentes cuestiones, se plantea la hipótesis de otra manera; la diferencia más importante se presenta en la hipótesis alternativa. Antes se enunció la hipótesis alternativa como “diferente de”; ahora se enuncia como “mayor que”. En símbolos:

Prueba de dos colas:	Prueba de una cola:
H_0: $\mu = 200$	H_0: $\mu \leq 200$
H_1: $\mu \neq 200$	H_1: $\mu > 200$

Los valores críticos en una prueba de una cola son diferentes a los de una prueba de dos colas en el mismo nivel de significancia. En el ejemplo anterior se dividió el nivel de significancia a la mitad y se colocó una mitad en la cola inferior y la otra en la cola superior. En una prueba de una cola, toda la región de rechazo se coloca en una cola (vea la gráfica 10.5).



GRÁFICA 10.5 Regiones de rechazo de las pruebas de una y dos colas; $\alpha = 0.01$

En el caso de la prueba de una cola, el valor crítico es de 2.326; utilice la tabla de distribución *t* de Student que se encuentra en el apéndice B.5 o en las últimas páginas del libro, ubíquese en el encabezado llamado “Nivel de significancia para pruebas de una cola, α”, seleccione la columna con $\alpha = 0.01$ y muévase a la última fila, etiquetada como ∞ o grados infinitos de libertad. El valor *z* en esta celda es 2.326.

OA10-6
Calcular e interpretar el valor *p*.

Valor *p* en la prueba de hipótesis

Cuando se desea probar una hipótesis, se compara el estadístico de prueba con un valor crítico. Se toma la decisión de rechazar la hipótesis nula o de no hacerlo; por ejemplo, si el valor crítico es de 1.96 y el valor calculado del estadístico de prueba es de 2.19, la decisión consiste en rechazar la hipótesis nula.

En años recientes, debido a la disponibilidad del software de computadora, con frecuencia se da información relacionada con la seguridad del rechazo o aceptación. Es decir, ¿cuánta confianza hay en el rechazo de la hipótesis nula? Este enfoque indica la probabilidad (en el supuesto de que la hipótesis nula sea verdadera) de obtener un valor del estadístico de prueba por lo menos tan extremo como el valor real que se obtuvo. Este proceso compara la probabilidad, denominada **valor *p*,** con el nivel de significancia; si el valor *p* es menor que este, H_0 se rechaza; si es mayor, H_0 no se rechaza.

VALOR *p* Probabilidad de observar un valor muestral tan extremo o más que el valor observado, si la hipótesis nula es verdadera.

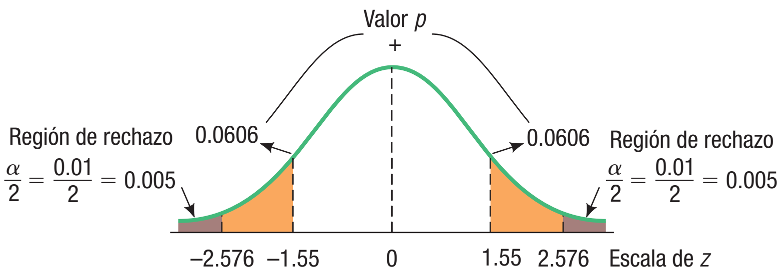
La determinación del valor *p* no solo da como resultado una decisión respecto de H_0, sino que brinda la oportunidad de observar la fuerza de la decisión. Un valor *p* muy pequeño, como 0.0001, indica que existe poca probabilidad de que H_0 sea verdadera; por otra parte, un valor *p* de 0.2033 significa que H_0 no se rechaza y que existe poca probabilidad de que sea falsa.

¿Cómo calcular el valor *p*? Se necesita usar la tabla de valores *z* que se incluye en el apéndice B.3 y en las últimas páginas del libro y, para utilizar esta tabla, se redondea el estadístico de prueba *z* a dos decimales. Para ilustrar cómo calcular un valor *p*, use el ejemplo en el que se probó la hipótesis nula relativa a que la cantidad de escritorios producidos a la semana en Fredonia fue de 200. No se rechazó la hipótesis nula, pues el valor *z* de 1.547 cayó en la región comprendida entre

ESTADÍSTICA EN ACCIÓN

Existe una diferencia entre *estadísticamente significativo* y *prácticamente significativo*. Para explicarlo, suponga que crea una nueva píldora para adelgazar y la prueba en 100 000 personas; y concluye que la persona común que toma la píldora durante dos años pierde una libra. ¿Cree usted que mucha gente se interesaría en tomarla para perder una libra? Los resultados de ingerir la nueva píldora fueron estadísticamente significativos, pero no prácticamente significativos.

-2.576 y 2.576. Se determinó no rechazar la hipótesis nula si el valor calculado de z caía en esta región. Al redondear 1.547 a 1.55 y con ayuda de la tabla, la probabilidad de hallar un valor z de 1.55 o más es de 0.0606, que se calcula mediante la diferencia de $0.5000 - 0.4394$; en otras palabras, la probabilidad de obtener una $\overline{X}$ mayor de 203.5 si $\mu = 200$ es de 0.0606. Para calcular el valor p es necesario concentrarse en la región menor a -1.55, así como en los valores superiores a 1.55 (pues la región de rechazo se localiza en ambas colas). El valor p de dos colas es de 0.1212, que se calcula así: 2(0.0606); el valor p de 0.1212 es mayor que el nivel de significancia 0.01 que se estableció al inicio, así que no se rechaza H_0. En la siguiente gráfica se muestran los detalles. Observe que para la prueba de hipótesis de dos colas, el valor p está representando por áreas en ambas colas de la distribución; entonces, el valor p se compara con facilidad con el nivel de significancia. Se aplica la misma regla de decisión en el caso de una prueba de una cola.



Un valor p es una manera de expresar la probabilidad de que H_0 sea falsa. Pero, ¿cómo se interpreta un valor *p?* Ya se mencionó que si el valor p es menor que el nivel de significancia, H_0 se rechaza; si es mayor que este, no se rechaza.; asimismo, si el valor p es muy grande, es probable que H_0 sea verdadera. Si el valor p es pequeño, quizá H_0 no lo es. Mediante el siguiente recuadro se pueden interpretar los valores *p.*

INTERPRETACIÓN DE LA IMPORTANCIA DE LA EVIDENCIA EN CONTRA DE H_0 Si el valor p es menor que

(a) 0.10, hay *cierta* evidencia de que H_0 no es verdadera.
(b) 0.05, hay evidencia *fuerte* de que H_0 no es verdadera.
(c) 0.01, hay evidencia *muy fuerte* de que H_0 no es verdadera.
(d) 0.001, hay evidencia *extremadamente fuerte* de que H_0 no es verdadera.

AUTOEVALUACIÓN 10-2

Consulte la autoevaluación 10.1.

(a) Suponga que se modifica el penúltimo enunciado para que diga: ¿Sugiere la evidencia que la cantidad media despachada es *mayor a* 16 onzas? Establezca la hipótesis nula y la hipótesis alternativa en estas condiciones.
(b) ¿Cuál es la regla de decisión en las nuevas condiciones definidas en el punto anterior?
(c) Una segunda muestra de 50 contenedores llenos reveló que la media es de 16.040 onzas. ¿Cuál es el valor del estadístico de prueba en esta muestra?
(d) ¿Cuál es su decisión respecto de la hipótesis nula?
(e) Interprete, en un solo enunciado, el resultado de la prueba estadística.
(f) ¿Cuál es el valor *p?* ¿Cuál es su decisión respecto de la hipótesis nula con base en el valor *p?* ¿Es la misma conclusión a la que se llegó en el punto (d)?

EJERCICIOS

Responda las siguientes preguntas en los ejercicios 1 a 4: a) ¿Es una prueba de una o de dos colas?; b) ¿Cuál es la regla de decisión?; c) ¿Cuál es el valor del estadístico de prueba? d) ¿Cuál es su decisión respecto de H_0?; e) ¿Cuál es el valor p? Interprete este valor.

1. Se selecciona una muestra de 36 observaciones de una población normal, la media muestral es de 49, y la desviación estándar de la población, 5. Lleve a cabo la prueba de hipótesis con el nivel de significancia 0.05.

$$H_0: \mu = 50$$
$$H_1: \mu \neq 50$$

2. Se selecciona una muestra de 36 observaciones de una población normal, la media muestral es de 12, y la desviación estándar de la población, 3. Lleve a cabo la prueba de hipótesis con un nivel de significancia 0.01.

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10$$

3. Se selecciona una muestra de 36 observaciones de una población normal, la media de la muestra es 21, y la desviación estándar de la población, 5. Lleve a cabo la prueba de hipótesis con el nivel de significancia 0.05.

$$H_0: \mu \leq 20$$
$$H_1: \mu > 20$$

4. Se selecciona una muestra de 64 observaciones de una población normal, la media de la muestra es 215, y la desviación estándar de la población, 15. Lleve a cabo la prueba de hipótesis, utilice el nivel de significancia 0.025.

$$H_0: \mu \geq 220$$
$$H_1: \mu < 220$$

En el caso de los ejercicios 5 a 8: a) establezca la hipótesis nula y la hipótesis alternativa; b) defina la regla de decisión; c) calcule el valor del estadístico de prueba; d) ¿cuál es su decisión respecto de H_0?; e) ¿cuál es el valor *p*? Interprete este valor.

5. El fabricante de neumáticos radiales con cinturón de acero X-15 para camiones señala que el millaje medio que cada uno recorre antes de que se desgasten las cuerdas es de 60 000 millas; la desviación estándar del millaje es de 5 000 millas. La Crosset Truck Company compró 48 neumáticos y comprobó que el millaje medio para sus camiones es de 59 500. ¿La experiencia de Crosset es diferente de lo que afirma el fabricante en el nivel de significancia 0.05?
6. El tiempo de espera de los clientes de la cadena de restaurantes MacBurger sigue una distribución normal, con una desviación estándar poblacional de 1 minuto. El departamento de control de calidad halló en una muestra de 50 clientes en Warren Road MacBurger que el tiempo medio de espera era de 2.75 minutos. Con el nivel de significancia 0.05, ¿puede concluir que el tiempo medio de espera sea menor a tres minutos?
7. Una encuesta nacional reciente determinó que los estudiantes de secundaria veían en promedio (media) 6.8 películas en DVD al mes, con una desviación estándar poblacional de 1.8. Una muestra aleatoria de 36 estudiantes universitarios reveló que la cantidad media de películas en DVD que vieron el mes previo fue de 6.2; con el nivel de significancia 0.05, ¿puede concluir que los estudiantes universitarios ven menos películas en DVD que los estudiantes de secundaria?
8. En el momento en que fue contratada como mesera en el Grumney Family Restaurant, a Beth Brigden le dijeron: "Puedes ganar en promedio 80 dólares al día en propinas". Suponga que la población de propinas diarias está normalmente distribuida, con una desviación estándar de 9.95 dólares. Durante los primeros 35 días de trabajo en el restaurante, la suma media de sus propinas fue de 84.85 dólares. Con el nivel de significancia 0.01, ¿la señorita Brigden puede concluir que gana un promedio superior a 80 dólares en propinas?

OA10-7
Utilizar un estadístico *t* para probar una hipótesis.

Prueba de la media poblacional: desviación estándar de la población desconocida

En el ejemplo anterior se conocía σ (la desviación estándar de la población); no obstante, en la mayoría de los casos, este valor se desconoce. Por consiguiente, σ debe basarse en estudios previos o calcularse por medio de la desviación estándar de la muestra, *s.* La del siguiente ejemplo no se conoce, por lo que se emplea la desviación estándar muestral para estimar σ.

Para determinar el valor del estadístico de prueba, se utiliza la distribución *t* y se modifica la fórmula [10.1] de la siguiente manera:

PRUEBA DE LA MEDIA; σ DESCONOCIDA

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} \quad \textbf{[10.2]}$$

con $n - 1$ grados de libertad, donde:

$\bar{x}$ representa la media de la muestra;
μ es la media poblacional hipotética;
s es la desviación estándar de la muestra;
n es el número de observaciones incluidas en la muestra.

Es una situación similar a cuando se construyeron intervalos de confianza en el capítulo anterior. Vea la sección "Desviación estándar poblacional σ desconocida", capítulo 9; en la gráfica 9.3 se resumió el problema. En estas condiciones, el procedimiento estadístico correcto consiste en sustituir la distribución normal estándar con la distribución *t*; cuyas principales características son:

- Es una distribución continua.
- Tiene forma de campana y es simétrica.
- Existe una familia de distribuciones *t*; cada vez que se cambia de grados de libertad, se crea una nueva distribución.
- Conforme se incrementa el número de grados de libertad, la forma de la distribución *t* se aproxima a la de la distribución normal estándar.
- La distribución *t* es plana, o más dispersa, que la distribución normal estándar.

En el siguiente ejemplo se muestran algunos detalles.

EJEMPLO

El departamento de quejas de McFarland Insurance Company informa que el costo medio para tramitar una queja es de 60 dólares. Una comparación en la industria demostró que esta cantidad es mayor que en las demás compañías de seguros, así que la empresa tomó medidas para reducir gastos y, para evaluar el efecto de tales medidas, el supervisor del departamento de quejas seleccionó una muestra aleatoria de 26 quejas atendidas y registradas para procesar cada una durante el mes anterior. La información de la muestra se presenta a continuación.

$45	$49	$62	$40	$43	$61
48	53	67	63	78	64
48	54	51	56	63	69
58	51	58	59	56	57
38	76				

Con el nivel de significancia 0.01, ¿es razonable concluir que el costo medio de atención de una queja ahora es menor a 60 dólares?

SOLUCIÓN

Se aplicará la prueba de hipótesis con el procedimiento de los seis pasos.

Paso 1: **Se establecen las hipótesis nula y alternativa.** La hipótesis nula consiste en que la media poblacional mínima es de 60 dólares. La hipótesis alternativa consiste en que la media poblacional es menor a 60 dólares. Las hipótesis nula y alternativa se expresan de la siguiente manera:

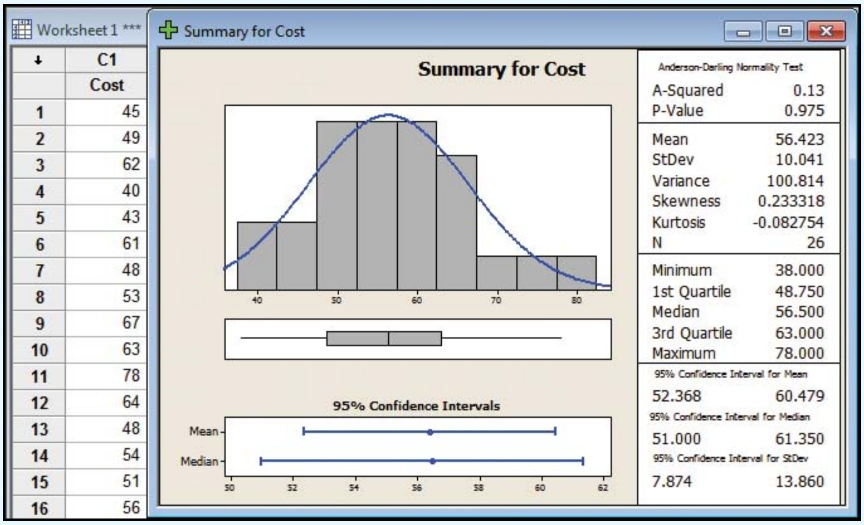
$$H_0: \mu \geq \$60$$
$$H_1: \mu < \$60$$

La prueba es de *una* cola, pues desea determinar si hubo una *reducción* en el costo. La desigualdad en la hipótesis alternativa señala la región de rechazo en la cola izquierda de la distribución.

Paso 2: **Se selecciona un nivel de significancia.** El nivel de significancia es 0.01.

Paso 3: **Se identifica el estadístico de prueba.** En este caso es la distribución *t.* ¿Por qué? Primero, porque resulta razonable concluir que la distribución del costo por queja sigue la distribución normal; lo cual se confirma a partir del histograma (vea la página siguiente), de la siguiente salida de Minitab. Observe la distribución normal superpuesta en la distribución de frecuencias.

Como no se conoce la desviación estándar de la población, esta se sustituye por la desviación estándar de la muestra. El valor del estadístico de prueba se calcula por medio de la fórmula [10.2]:



$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

Paso 4: **Se formula la regla de decisión.** Los valores críticos de t aparecen en el apéndice B.5, una parte del cual se reproduce en la tabla 10.1. El apéndice B.5 se repite también en las últimas páginas del libro. La columna extrema izquierda de la tabla se rotula *gl,* que representa los grados de libertad. El número de grados de libertad es el total de observaciones incluidas en la muestra menos el número de poblaciones muestreadas, lo cual se escribe $n - 1$. En este caso, el número de observaciones de la muestra es de 26, y se muestrea una población, así que hay $26 - 1 = 25$ grados de libertad. Para determinar el valor crítico, primero localice el renglón con los grados de libertad adecuados; este renglón se encuentra sombreado en la tabla 10.1; luego determine si la prueba es de una o de dos colas. En este caso, es una prueba de una cola, así que busque la sección de la tabla rotulada *una cola.* Localice la columna con el nivel de significancia elegido. En este ejemplo, el nivel de significancia es de 0.01. Desplácese hacia abajo por la columna rotulada "0.01" hasta intersecar el renglón con 25 grados de libertad, el valor es de 2.485. Como se trata de una prueba de una cola y la región de rechazo se localiza en la cola izquierda, el valor crítico es negativo. La regla de decisión consiste en rechazar H_0 si el valor de t es menor a -2.485.

TABLA 10.1 Parte de la tabla de la distribución t

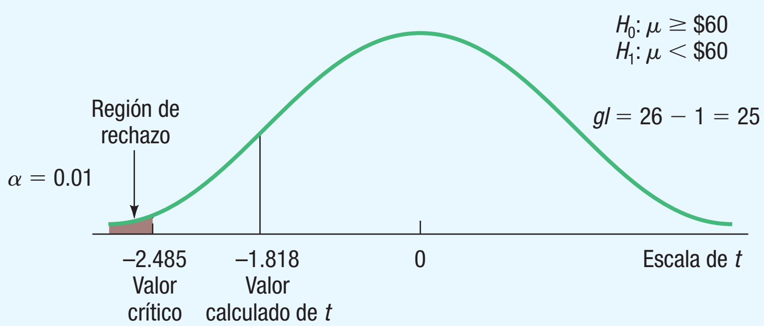
	Intervalos de confianza					
	80%	**90%**	**95%**	**98%**	**99%**	**99.9%**
	Nivel de significancia de una prueba de una cola, α					
	0.10	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
gl	**Nivel de significancia de una prueba de dos colas, α**					
	0.20	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	$\boxed{2.485}$	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646

Paso 5: **Se toma una muestra y se decide.** De acuerdo con la pantalla de Minitab, a la derecha del histograma, el costo medio por queja de la muestra de 26 observaciones es de 56.42 dólares, y la desviación estándar, de 10.04 dólares; al sustituirlos en la fórmula [10.2] y calcular el valor de *t:*

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\$56.42 - \$60}{\$10.04/\sqrt{26}} = -1.818$$

Como -1.818 se localiza en la región ubicada a la derecha del valor crítico de -2.485 (vea la tabla 10.6), la hipótesis nula no se rechaza en el nivel de significancia 0.01.

Paso 6: **Se interpreta el resultado.** La hipótesis nula no se rechaza. La muestra de quejas pudo haberse seleccionado de una población con un costo medio de 60 dólares por queja; en otras palabras, la diferencia de 3.58 dólares ($56.52 − $60) entre la media muestral y la media poblacional pudo deberse a un error de muestreo. Los resultados de la prueba no permiten al gerente del departamento de quejas concluir que las medidas para recortar costos han sido efectivas.



GRÁFICA 10.6 Región de rechazo, distribución *t*, nivel de significancia 0.01

En el ejemplo anterior, la media y la desviación estándar se calcularon con Minitab; en el siguiente ejemplo se muestran los detalles cuando se obtienen a partir de los datos de la muestra.

EJEMPLO

El estacionamiento de corta estancia del Aeropuerto Internacional de Myrtle Beach está cerca de la terminal, así que cuando alguien va a recoger a un pasajero solo tiene que caminar una corta distancia al área de recuperación de equipajes, que es un buen lugar para encontrarse. Para decidir si el estacionamiento tiene suficientes lugares, el gerente del aeropuerto necesita saber si el tiempo medio de permanencia es superior a 40 minutos. Una muestra de 12 clientes recientes mostró que estuvieron en el estacionamiento los siguientes lapsos de tiempo, en minutos.

55	49	53	47	39	27	64	48	48	53	37	56

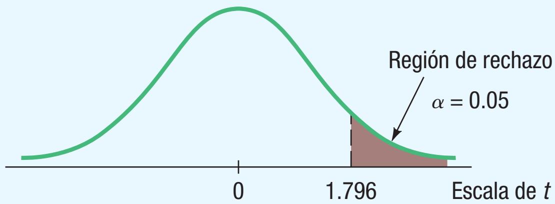
Con el nivel de significancia 0.05, ¿es razonable concluir que el tiempo medio en el estacionamiento es superior a 40 minutos?

SOLUCIÓN

En primera instancia, se establecen las hipótesis nula y alternativa. En este caso, la cuestión es si la media poblacional puede ser de más de 40 minutos; por tanto, es una prueba de una cola. Las dos hipótesis se establecen así:

$$H_0: \mu \leq 40$$
$$H_1: \mu > 40$$

Hay 11 grados de libertad, calculados mediante $n - 1 = 12\text{-}1 = 11$. El valor *t* crítico es 1.796, determinado recurriendo al apéndice B-5 para la prueba de una cola, usando a = 0.05 con 11 grados de libertad. La regla de decisión es: rechazar la hipótesis nula si el valor *t* calculado es superior a 1.796. Esta información se resume en la gráfica 10.7.



GRÁFICA 10-7 Región de rechazo, prueba de una cola, distribución t de Student, $\alpha = 0.05$

La media muestral se calculamos usando la fórmula [3.2] y la desviación estándar de la muestra con la fórmula [3.8]; la media muestral es 48 horas, y la desviación estándar de la muestra, 9.835 horas. Los detalles de los cálculos se muestran en la tabla 10.2

TABLA 10.2 Cálculo de la media muestral y la desviación estándar de los tiempos de estacionamiento

Cliente	x, minutos	$(x - \bar{x})^2$
1	55	49
2	49	1
3	53	25
4	47	1
5	39	81
6	27	441
7	64	256
8	48	0
9	48	0
10	53	25
11	37	121
12	56	64
Total	576	1 064

$$\bar{x} = \frac{\sum x}{n} = \frac{576}{12} = 48$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = \sqrt{\frac{1\,064}{12 - 1}} = 9.835$$

Ahora es posible calcular el valor t, usando la fórmula [10.2].

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{48 - 40}{9.835/\sqrt{12}} = -2.818$$

Se rechaza la hipótesis nula de que la media poblacional es menor o igual a 40 minutos porque el valor t calculado de 2.818 cae en el área a la derecha de 1.796; por tanto, el tiempo que los clientes pasan en el estacionamiento es mayor de 40 minutos. Este resultado indica que el aeropuerto necesita ampliar su estacionamiento.

AUTOEVALUACIÓN 10-3

La vida media de la batería de un reloj digital es de 305 días; las vidas medias de las baterías se rigen por la distribución normal. Hace poco se modificó un lote de estas para que tuvieran mayor duración; una muestra de 20 baterías modificadas exhibió una vida media de 311 días, con una desviación estándar de 12 días. ¿La modificación incrementó la vida media de la batería?

(a) Formule la hipótesis nula y la hipótesis alternativa.
(a) Muestre la gráfica de la regla de decisión; utilice el nivel de significancia 0.05.
(a) Calcule el valor de t. ¿Cuál es su decisión respecto de la hipótesis nula? Resuma sus resultados.

EJERCICIOS

9. Sean las siguientes hipótesis:

$$H_0: \mu \leq 10$$
$$H_1: \mu > 10$$

Se seleccionó una muestra aleatoria de 10 observaciones de una población normal, la media muestral fue de 12, y la desviación estándar de la muestra, de 3; utilice el nivel de significancia 0.05:

a. Formule la regla de decisión.
b. Calcule el valor del estadístico de prueba.
c. ¿Cuál es su decisión respecto de la hipótesis nula?

10. Sean las siguientes hipótesis:

$$H_0: \mu = 400$$
$$H_1: \mu \neq 400$$

En el caso de una muestra aleatoria de 12 observaciones seleccionada de una población normal, la media muestral fue de 407, y la desviación estándar de la muestra, de 6; utilice el nivel de significancia 0.01:
a. Formule la regla de decisión.
b. Calcule el valor del estadístico de prueba.
c. ¿Cuál es su decisión respecto de la hipótesis nula?

11. El gerente de ventas del distrito de las Montañas Rocallosas de Rath Publishing, Inc., editorial de textos universitarios, afirma que los representantes de ventas realizan en promedio 40 llamadas de ventas a profesores cada semana. Varios representantes señalan que el cálculo es muy bajo; una muestra aleatoria de 28 representantes revela que la cantidad media de llamadas que se realizó la semana pasada fue de 42, con una desviación estándar de la muestra de 2.1 llamadas. Con el nivel de significancia 0.05, ¿puede concluir que la cantidad media de llamadas semanales por vendedor es superior a 40?

12. La administración de White Industries analiza una nueva técnica para armar un carro de golf; la técnica actual requiere 42.3 minutos de trabajo en promedio. El tiempo medio de montaje de una muestra aleatoria de 24 carros, con la nueva técnica, fue de 40.6 minutos, y la desviación estándar, de 2.7 minutos. Con el nivel de significancia 0.10, ¿puede concluir que el tiempo de montaje con la nueva técnica es más breve?

13. El ingreso promedio por persona en Estados Unidos es de 40 000 dólares, y la distribución de ingresos sigue una distribución normal. Una muestra aleatoria de 10 residentes de Wilmington, Delaware, presentó una media de 50 000 dólares, con una desviación estándar de 10 000 dólares. Con el nivel de significancia 0.05, ¿existe suficiente evidencia para concluir que los residentes de Wilmington, Delaware, ganan más que el promedio nacional?

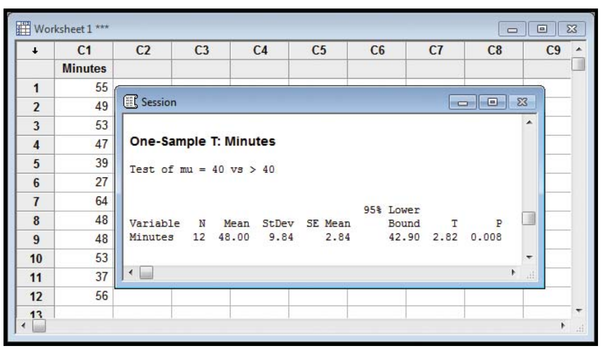
14. En la actualidad, la mayoría de quienes viajan por avión compra sus boletos en internet. De esta forma, los pasajeros evitan la preocupación de cuidar un boleto de papel; además, las aerolíneas ahorran. No obstante, en fechas recientes, las aerolíneas han recibido quejas relacionadas con los boletos, en particular cuando se requiere hacer un enlace para cambiar de línea. Para analizar el problema, una agencia de investigación independiente tomó una muestra aleatoria de 20 aeropuertos y recogió información relacionada con la cantidad de quejas que hubo sobre los boletos durante marzo. A continuación se presenta la información.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14	14	16	12	12	14	13	16	15	14
12	15	15	14	13	13	12	13	10	13

Con el nivel de significancia 0.05, ¿la agencia de investigación puede concluir que la cantidad media de quejas por aeropuerto es menor de 15 al mes?
a. ¿Qué suposición se requiere antes de llevar a cabo una prueba de hipótesis?
b. Ilustre la cantidad de quejas por aeropuerto en una distribución de frecuencias o en un diagrama de dispersión. ¿Es razonable concluir que la población se rige por una distribución normal?
c. Realice una prueba de hipótesis e interprete los resultados.

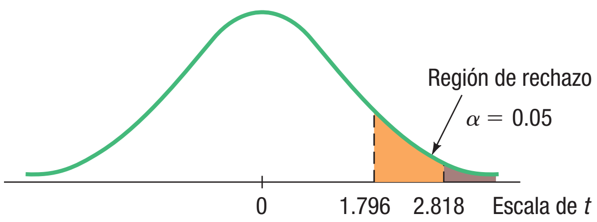
Solución con software



El sistema de software de estadística Minitab, que se utilizó en los capítulos precedentes y en la sección anterior, proporciona una forma eficaz de llevar a cabo una prueba de hipótesis de una cola para la media de la población. Los pasos para generar la pantalla se muestran en el apéndice C.

Una característica adicional de la mayoría de los paquetes de software consiste en que calculan el valor *p,* el cual proporciona más información sobre la hipótesis nula; dicho valor es la probabilidad de un valor *t* tan extremo como el que se calculó, si la hipótesis nula es verdadera. Usando el análisis de Minitab del ejemplo anterior del estacionamiento del aeropuerto, se estableció que el valor p de 0.008 es la probabilidad de un valor t de 2.82 o mayor, dada una media poblacional de 40. Así, la comparación del valor p con el nivel de significancia indica si la hipótesis nula se encontraba cerca de ser rechazada, si apenas se rechazó, etcétera.

Para mayor claridad, refiérase al diagrama. El valor p de 0.008 es el área sombreada, y el nivel de significancia es la totalidad de las áreas sombreadas. Como el valor p de 0.008 es menor que el nivel de significancia 0.05, la hipótesis nula se rechaza. Si el valor p hubiera sido mayor que el nivel de significancia, por decir, 0.06, 0.19 o 0.57, la hipótesis nula no se habría rechazado.



En el ejemplo anterior se estableció que la hipótesis alternativa era de un lado, y la cola superior (derecha) de la distribución t contenía la región de rechazo.

El valor p es el área a la derecha de 2.818 para una distribución t con 11 grados de libertad.

¿Y si se tratara de una prueba de dos colas, de modo que la región de rechazo quedara tanto en las colas superior e inferior? Esto es, en el ejemplo del estacionamiento, si H_1 se estableciera como $\mu \neq 40$, sería preciso reportar que el valor p es el área a la derecha de 2.818 más el valor a la izquierda de -2.818. Ambos valores son 0.008, así que el valor p es $0.008 + 0.008 = 0.016$.

¿Cómo se calcula un valor p sin una computadora? Para ilustrarlo, recuerde que, en el ejemplo relativo al tiempo de estacionamiento, se rechazó la hipótesis nula que indicaba $\mu \leq 40$ y se aceptó la hipótesis alternativa que indicaba $\mu > 40$. El nivel de significancia era de 0.05, así que, por lógica, el valor p es menor que 0.05. Para calcular el valor p con mayor precisión, vea el apéndice B.5 y localice la fila con 11 grados de libertad. El valor calculado de *t,* 2.818, se localiza entre 2.718 y 3.106 (parte del apéndice B.5 se reproduce en la tabla 10.3). El nivel de significancia de una cola correspondiente a 2.718 es 0.01, y en el caso de 3.106, es 0.005; por lo tanto, el valor p se encuentra entre 0.005 y 0.01. Se acostumbra indicar que el valor p es *menor* que el mayor de los dos niveles de significancia. Así, “el valor p es menor que 0.01”.

TABLA 10.3 Parte de la distribución *t* de Student

	Intervalos de confianza					
	80%	**90%**	**95%**	**98%**	**99%**	**99.9%**
gl	**Nivel de significancia de una prueba de una cola, α**					
	0.10	**0.05**	**.0025**	**0.01**	**0.005**	**0.0005**
	Nivel de significancia de una prueba de dos colas, α					
	0.20	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073

AUTOEVALUACIÓN 10-4

Se programa una máquina para llenar un frasco pequeño con 9.0 gramos de medicamento. Una muestra de ocho frascos arrojó las siguientes cantidades (en gramos) por botella.

9.2	8.7	8.9	8.6	8.8	8.5	8.7	9.0

¿Puede concluir que el peso medio es inferior a 9.0 gramos si el nivel de significancia es de 0.01?

(a) Formule la hipótesis nula y la hipótesis alternativa.
(b) ¿Cuántos grados de libertad existen?
(c) Establezca la regla de decisión.
(d) Calcule el valor de *t.* ¿Qué decide respecto de la hipótesis nula?
(e) Estime el valor *p.*

EJERCICIOS

15. Sean las siguientes hipótesis:

$$H_0: \mu \geq 20$$
$$H_1: \mu < 20$$

Una muestra aleatoria de cinco elementos dio como resultado los siguientes valores: 18, 15, 12, 19 y 21. Con el nivel de significancia 0.01, ¿puede concluir que la media poblacional es menor que 20?

a. Establezca la regla de decisión.
b. Calcule el valor del estadístico de prueba.
c. ¿Cuál es su decisión en lo que se refiere a la hipótesis nula?
d. Calcule el valor *p.*

16. Sean las siguientes hipótesis:

$$H_0: \mu = 100$$
$$H_1: \mu \neq 100$$

Mediante una muestra aleatoria de seis elementos se obtuvieron los siguientes valores: 118, 105, 112, 119, 105 y 111. Con el nivel de significancia 0.05, ¿puede concluir que la media poblacional es diferente de 100?

a. Establezca la regla de decisión.
b. Calcule el valor del estadístico de prueba.
c. ¿Cuál es su decisión en lo que se refiere a la hipótesis nula?
d. Calcule el valor *p.*

17. La cantidad de agua consumida al día por un adulto sano sigue una distribución normal, con una media de 1.4 litros. Una campaña de salud promueve el consumo de cuando menos 2.0 litros diarios; después de la campaña, una muestra de 10 adultos muestra el siguiente consumo en litros:

1.5	1.6	1.5	1.4	1.9	1.4	1.3	1.9	1.8	1.7

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Con el nivel de significancia 0.01, ¿se puede concluir que se ha elevado el consumo de agua? Calcule e interprete el valor *p.*

18. El cloro líquido que se agrega a las albercas para combatir las algas tiene una duración relativamente corta en las tiendas antes que pierda su eficacia. Los registros indican que la duración media de un frasco de cloro es de 2 160 horas (90 días). Como experimento, se agregó Holdlonger al cloro para saber si este incrementaba la duración del cloro; una muestra de nueve frascos de cloro arrojó los siguientes tiempos de duración (en horas) en las tiendas:

2 159	2 170	2 180	2 179	2 160	2 167	2 171	2 181	2 185

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Con el nivel de significancia 0.025, ¿incrementó el Holdlonger la duración del cloro en las tiendas? Calcule el valor *p.*

19. Un grupo de expertos en Washington, D.C., anuncia que el adolescente típico envió 50 mensajes de texto por día durante 2009. Para actualizar la estimación, usted contacta por teléfono a una muestra de adolescentes y les pregunta cuántos mensajes enviaron el día anterior; sus respuestas fueron:

51	175	47	49	44	54	145	203	21	59	42	100

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Con el nivel de significancia 0.05, ¿puede concluir que el número medio es mayor a 50? Estime el valor *p* y describa qué le revela.

20. Hugger Polls afirma que un agente realiza una media de 53 entrevistas extensas a domicilio a la semana. Se introdujo un nuevo formulario para las entrevistas, y Hugger desea evaluar su eficacia. La cantidad de entrevistas extensas por semana de una muestra aleatoria de agentes es:

53	57	50	55	58	54	60	52	59	62	60	60	51	59	56

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Con el nivel de significancia 0.05, ¿puede concluir que la cantidad media de entrevistas de los agentes es superior a 53 a la semana? Calcule el valor *p.*

Error tipo II

OA10-8 Calcular la probabilidad de un error tipo II.

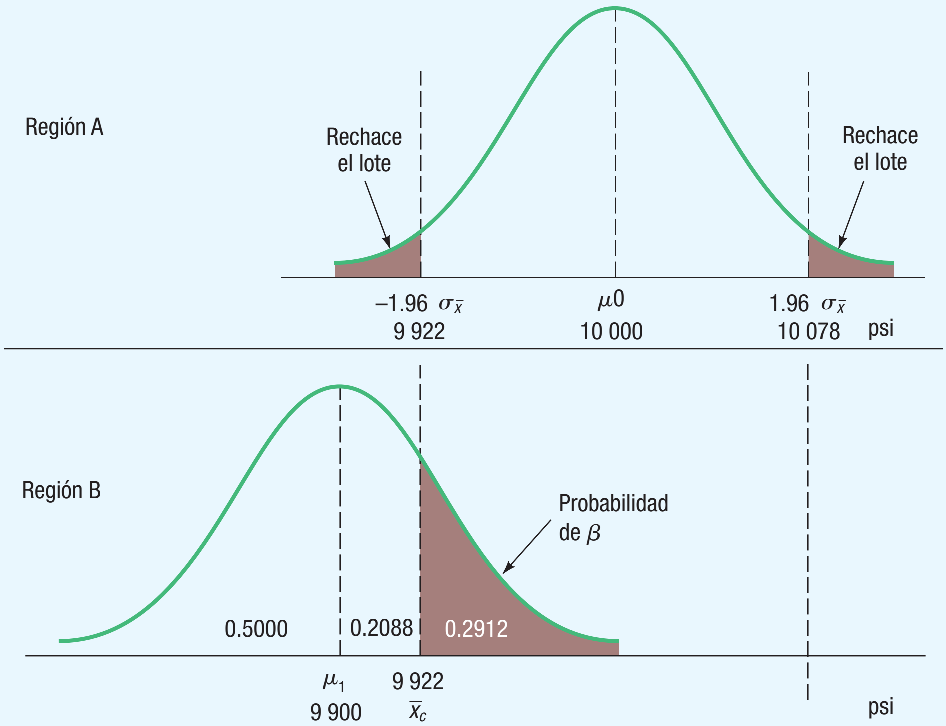
Recuerde que el nivel de significancia, identificado con el símbolo α, es la probabilidad de que la hipótesis nula se rechace cuando sea verdadera; esto recibe el nombre de error tipo I. Los niveles de significancia más comunes son 0.05 y 0.01, y los establece el investigador desde el inicio de la prueba.

En un caso de prueba de hipótesis también existe la posibilidad de que no se rechace una hipótesis nula cuando en realidad es falsa; es decir, se acepta una hipótesis nula falsa. Esto recibe el nombre de *error tipo II*, y su probabilidad se identifica con la letra griega beta (β). En los siguientes ejemplos se ilustran los detalles de la determinación del valor de β.

EJEMPLO

Western Wire Products compra barras de acero para hacer clavijas. La experiencia indica que la fuerza media de tensión de las cargas que llegan es de 10 000 psi, y que la desviación estándar, σ, es de 400 psi.

Con el fin de tomar una decisión sobre las cargas de barras de acero que llegan, el fabricante establece la siguiente regla para que el inspector de control de calidad se apegue a ella: “Tome una muestra de 100 barras de acero. Si la fuerza media ($\overline{x}$) se encuentra entre 9 922 y 10 078 psi, con un nivel de significancia de 0.05, acepte el lote; de lo contrario, debe rechazarlo”. En la gráfica 10.8, región A, se muestra el área en que se rechaza cada lote y en donde no se rechaza. La media de esta distribución se representa mediante μ_0. Las colas de la curva representan la probabilidad de cometer un error tipo I, es decir, rechazar el lote de barras de acero que ingresa cuando, en realidad, se trata de un buen lote, con una media de 10 000 psi.



GRÁFICA 10.8 Gráficas que muestran los errores tipo I y tipo II

Suponga que la media poblacional desconocida de un lote que llega, designada μ, es en realidad de 9 900 psi. ¿Cuál es la probabilidad de que el inspector de control de calidad no rechace la carga (error tipo II)?

SOLUCIÓN

La probabilidad de cometer un error tipo II, según se representa el área sombreada en la gráfica 10.8, región B, se calcula al determinar el área bajo la curva normal que se localiza sobre 9 922 libras. El cálculo de las áreas bajo la curva normal se analizó en el capítulo 7; he aquí un breve repaso: es necesario determinar primero la probabilidad de que la media muestral caiga entre 9 900 y 9 922. Después, se resta esta probabilidad de 0.5000 (que representa toda el área más allá de la media de 9 900) para llegar a la probabilidad de cometer un error tipo II.

El número de unidades estándar (valor de *z*) entre la media del lote que llega (9 900), designada μ_1 y $\overline{x}_c$, que representa el valor crítico de 9 922, se calcula de la siguiente manera:

ERROR TIPO II

$$z = \frac{\bar{x}_0 - \mu_1}{\sigma/\sqrt{n}} \qquad [10.3]$$

Si $n = 100$ y $\sigma = 400$, el valor de z es 0.55:

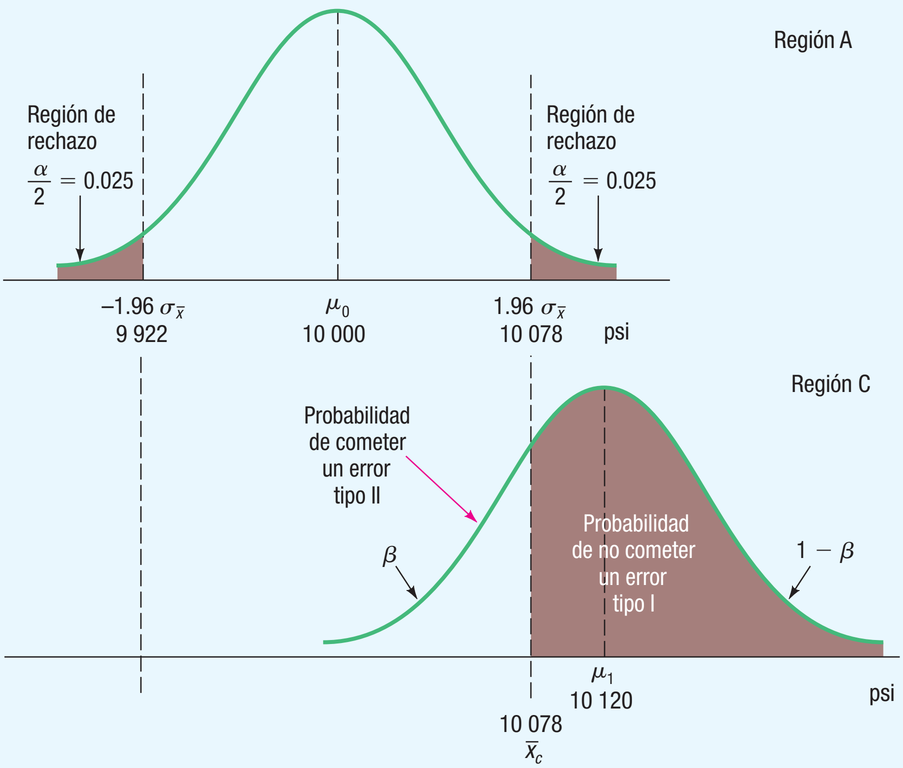
$$z = \frac{\bar{x}_c - \mu_1}{\sigma/\sqrt{n}} = \frac{9\,922 - 9\,900}{400/\sqrt{100}} = \frac{22}{40} = 0.55$$

El área bajo la curva entre 9 900 y 9 922 (un valor z de 0.55) es 0.2088. El área bajo la curva más allá de 9 922 libras es 0.5000 −0.2088, o 0.2912; tal es la probabilidad de cometer un error tipo II, es decir, de aceptar el ingreso de un lote de barras de acero cuando la media poblacional es de 9 900 psi.

En otra ilustración, en la gráfica 10.9, región C, se describe la probabilidad de aceptar un lote cuando la media poblacional es de 10 120. Para determinar la probabilidad:

$$z = \frac{\bar{x}_c - \mu_1}{\sigma/\sqrt{n}} = \frac{10\,078 - 10\,120}{400/\sqrt{100}} = -1.05$$

La probabilidad de que z sea menor que -1.05 es 0.1469, que se determina al calcular 0.5000 − 0.3531; por lo tanto, β, o la probabilidad de cometer un error tipo II, es 0.1469. Cuidado: si la diferencia entre μ_0 y μ_1 es relativamente pequeña, puede ocurrir la probabilidad de un error tipo II en ambas colas. Aquí no se considera tal eventualidad.



GRÁFICA 10.9 Errores tipo I y tipo II (otro ejemplo)

De acuerdo con los métodos que se ilustran en las gráficas 10.8 (región B) y 10.9 (región C), puede determinarse la probabilidad de aceptar una hipótesis como verdadera cuando en realidad es falsa para cualquier valor de μ_1.

Las probabilidades de cometer un error tipo II se muestran en la columna central de la tabla 10.4 para valores seleccionados de μ, dados en la columna de la izquierda. La columna derecha proporciona la probabilidad de no cometer un error tipo II, que también se conoce como la fuerza de una prueba.

TABLA 10.4 Probabilidades de cometer un error tipo II con $\mu_0 = 10\,000$ libras y medias alternativas seleccionadas, nivel de significancia 0.05

Media alternativa seleccionada (μ_1)	Probabilidad de cometer un error tipo II (β)	Probabilidad de no cometer un error tipo II ($1 - \beta$)
9 820	0.0054	0.9946
9 880	0.1469	0.8531
9 900	0.2912	0.7088
9 940	0.6736	0.3264
10 000	—*	—
10 060	0.6736	0.3264
10 100	0.2912	0.7088
10 120	0.1469	0.8531
10 180	0.0054	0.9946

* No es posible cometer un error tipo II cuando $\mu_1 = \mu_0$.

AUTOEVALUACIÓN 10-5

Con base en el ejemplo anterior, se supone que la media real de un lote de barras de acero que llega es de 10 180 psi. ¿Cuál es la probabilidad de que el inspector de control de calidad acepte las barras como si tuvieran una media de 10 000 psi? Parece poco probable que las barras de acero se rechacen si la fuerza de tensión es mayor que la especificada. No obstante, puede ser que la clavija tenga una doble función en un motor fuera de borda; tal vez esté diseñada para que no se desprenda si el motor golpea un objeto pequeño, aunque sí lo haga si golpea una roca. Por consiguiente, el acero no debe ser demasiado fuerte.

El área no sombreada en la gráfica 10.9 (región C) representa la probabilidad de aceptar por error la hipótesis que indica que la fuerza de tensión media de las barras de acero es de 10 000 psi. ¿Cuál es la probabilidad de cometer un error tipo II?

EJERCICIOS

21. Consulte la tabla 10.4 y el ejemplo anterior; si $n = 100$, $\sigma = 400$, $\bar{x}_C = 9\,922$ y $\mu_1 = 9\,880$, verifique que la probabilidad de cometer un error tipo II sea de 0.1469.

22. Consulte la tabla 10.4 y el ejemplo anterior; si $n = 100$, $\sigma = 400$, $\bar{x}_C = 10\,078$ y $\mu_1 = 10\,100$, verifique que la probabilidad de cometer un error tipo II sea de 0.2912.

RESUMEN DEL CAPÍTULO

I. El objetivo de la prueba de hipótesis consiste en verificar la validez de una afirmación relacionada con un parámetro de la población.

II. Los pasos para llevar a cabo una prueba de hipótesis son los siguientes:
- **A.** Se formula la hipótesis nula (H_0) y la hipótesis alternativa (H_1).
- **B.** Se selecciona el nivel de significancia.
  - **1.** El nivel de significancia es la probabilidad de rechazar una hipótesis nula verdadera.
  - **2.** Los niveles de significancia más frecuentes son 0.01, 0.05 y 0.10, pero es posible utilizar cualquier valor entre 0 y 1.00.
- **C.** Se selecciona el estadístico de prueba.
  - **1.** Un estadístico de prueba es un valor que se calcula a partir de la información de una muestra para determinar si se rechaza la hipótesis nula.
  - **2.** En este capítulo se consideraron dos estadísticos de prueba.
    - **a.** La distribución normal estándar se utiliza cuando la población sigue la distribución normal y se conoce la desviación estándar de la población.
    - **b.** La distribución *t* Student se emplea cuando la población sigue la distribución normal y se desconoce la desviación estándar de la población.

D. Se establece la regla de decisión.
   1. La regla de decisión indica la condición o las condiciones en que se rechaza la hipótesis nula.
   2. En una prueba de dos colas, la región de rechazo se divide uniformemente entre las colas izquierda y derecha de la distribución.
   3. En una prueba de una cola, toda la región de rechazo se encuentra en una sola cola.
E. Se selecciona una muestra, se calcula el valor del estadístico de prueba y se toma una decisión respecto de la hipótesis nula.
F. Se interpretan los resultados de la decisión.

III. Un valor *p* es la probabilidad de que el valor del estadístico de prueba sea tan extremo como el valor calculado cuando la hipótesis nula es verdadera.

IV. Al probar una hipótesis sobre la media de la población:
A. Si se conoce la desviación estándar de la población, σ, el estadístico de prueba es la distribución normal estándar, y se determina a partir de:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \qquad \textbf{[10.1]}$$

B. Si no se conoce la desviación estándar de la población, pero hay por lo menos 30 observaciones en la muestra, *s* se sustituye por σ. El estadístico de prueba es la distribución *t*, y su valor se determina de acuerdo con:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad \textbf{[10.2]}$$

Las principales características de la distribución *t* son:
   1. Es una distribución continua.
   2. Tiene forma de campana y es simétrica.
   3. Es plana o más amplia que la distribución normal estándar.
   4. Existe una familia de distribuciones *t,* según el número de grados de libertad.

V. Existen dos tipos de errores que se pueden presentar en una prueba de hipótesis.
A. Un error tipo I, cuando se rechaza una hipótesis nula verdadera.
   1. La probabilidad de cometer un error tipo I es igual al nivel de significancia.
   2. Esta probabilidad se designa con la letra griega σ.
B. Un error tipo II, cuando no se rechaza una hipótesis nula falsa.
   1. La probabilidad de cometer un error tipo II se designa con la letra griega β.
   2. La probabilidad de cometer un error tipo II se determina por medio de

$$z = \frac{\bar{x}_c - \mu_1}{\sigma/\sqrt{n}} \qquad \textbf{[10.3]}$$

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
H_0	Hipótesis nula	*H, subíndice cero*
H_1	Hipótesis alternativa	*H, subíndice uno*
$\alpha/2$	Nivel de significancia de dos colas	*Alfa sobre dos*
$\bar{x}_C$	Límite de la media muestral	*x barra, subíndice c*
μ_0	Media supuesta de la población	*Mu, subíndice cero*

EJERCICIOS DEL CAPÍTULO

23. De acuerdo con el presidente del sindicato local, el ingreso bruto medio de los plomeros en el área de Salt Lake City sigue la distribución de probabilidad normal con una media de 45 000 dólares y una desviación estándar de 3 000 dólares. Un reportaje de investigación reciente de KYAK TV reveló que el ingreso bruto medio de una muestra de 120 plomeros era de 45 500 dólares. ¿Es razonable concluir que el ingreso medio no es igual a 45 000 dólares en el nivel de significancia 0.10? Determine el valor *p*.

24. Rutter Nursery Company empaca su aserrín de pino en bolsas de 50 libras. Desde hace tiempo, el departamento de producción informa que la distribución de pesos de las bolsas se rige por una distribución normal y que la desviación estándar del proceso es de 3.0 libras por bolsa. Al final de cada

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

día, Jeff Rutter, gerente de producción, pesa 10 bolsas y calcula el peso medio de la muestra. Enseguida aparecen los pesos de 10 bolsas de la producción de hoy.

45.6	47.7	47.6	46.3	46.2	47.4	49.2	55.8	47.5	48.5

a. Con el nivel de significancia 0.01, ¿puede concluir Rutter que el peso medio de las bolsas es inferior a 50 libras?
b. Indique en un breve informe la razón por la que Rutter puede utilizar la distribución *z* como estadístico de prueba.
c. Calcule el valor *p*.

25. Una nueva compañía dedicada al control de peso, Weight Reducers International, anuncia que quienes ingresan perderán, en promedio, 10 libras las primeras dos semanas, con una desviación estándar de 2.8 libras; una muestra aleatoria de 50 personas que iniciaron el programa de reducción de peso reveló que el peso medio perdido fue de 9 libras. Con el nivel de significancia 0.05, ¿puede concluir que quienes ingresan a Weight Reducers perderán en promedio más de 10 libras? Determine el valor *p*.

26. A Dole Pineapple, Inc., le preocupa la posibilidad de que las latas de 16 onzas de piña contengan exceso de producto. Suponga que la desviación estándar del proceso es de 0.03 onzas. El departamento de control de calidad tomó una muestra aleatoria de 50 latas y comprobó que la media aritmética del peso era de 16.05 onzas. ¿Puede concluir que el peso medio es mayor a 16 onzas con un nivel de significancia de 5%? Determine el valor *p*.

27. De acuerdo con una encuesta reciente, los estadounidenses duermen un promedio de 7 horas por noche. Una muestra aleatoria de 50 estudiantes de West Virginia University reveló que la cantidad media de horas que durmieron la noche anterior fue de 6horas, 48 minutos (6.8 horas). La desviación estándar de la muestra fue de 0.9 horas. ¿Es razonable concluir que los estudiantes de West Virginia duermen menos que el estadounidense normal? Calcule el valor *p*.

28. Una agencia estatal de venta de bienes raíces, Farm Associates, se especializa en la venta de granjas en el estado de Nebraska. Sus registros indican que el tiempo medio de venta de una granja es de 90 días. Como consecuencia de las recientes sequías, la agencia cree que el tiempo medio de venta es superior a 90 días. Una encuesta reciente en 100 granjas de todo el estado mostró que el tiempo medio de venta fue de 94 días, con una desviación estándar de 22 días. Con el nivel de significancia 0.10, ¿aumentó el tiempo de venta?

29. De acuerdo con la Oficina del Censo, 3.13 personas residen en un típico hogar estadounidense. Una muestra de 25 hogares de las comunidades de retirados de Arizona mostró que el número medio de residentes por hogar era de 2.86 personas; la desviación estándar de esta muestra es de 1.20 residentes. Con el nivel de significancia 0.05, ¿es razonable concluir que el número medio de residentes en los hogares de las comunidades de retirados es menor a 3.13 personas?

30. Un artículo reciente en la revista *Vitality* informó que la cantidad media de tiempo de descanso semanal de los estadounidenses es de 40.0 horas. Usted piensa que la cifra es muy alta y decide llevar a cabo sus propias pruebas. En una muestra aleatoria de 60 hombres, descubre que la media es de 37.8 horas de descanso a la semana, con una desviación estándar de la muestra de 12.2 horas. Con el nivel de significancia 0.05, ¿puede concluir que la información del artículo es incorrecta? Determine el valor *p* y explique su significado.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

31. En años recientes, la tasa de interés de los créditos hipotecarios se redujo a menos de 6.0%; sin embargo, de acuerdo con un estudio llevado a cabo por la Junta de Gobernadores de la Reserva Federal de Estados Unidos, la tasa de los cargos a las tarjetas de crédito es superior a 14%. En la siguiente lista se muestra la tasa de los cargos aplicados a una muestra de 10 tarjetas de crédito.

14.6	16.7	17.4	17.0	17.8	15.4	13.1	15.8	14.3	14.5

Con el nivel de significancia 0.01, ¿resulta razonable concluir que la tasa media es superior a 14%?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

32. Un artículo reciente de *The Wall Street Journal* informó que en la actualidad la tasa hipotecaria es inferior a 6%. Una muestra de ocho bancos pequeños de la región central de Estados Unidos reveló las siguientes tasas hipotecarias (porcentuales):

4.8	5.3	6.5	4.8	6.1	5.8	6.2	5.6

Con el nivel de significancia 0.01, ¿puede concluir que la tasa hipotecaria a 30 años de los bancos pequeños es inferior a 6%? Calcule el valor *p*.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

33. Un estudio reciente reveló que el típico bebedor habitual estadounidense de café consume un promedio de 3.1 tazas al día. Una muestra de 12 personas de la tercera edad reveló que el día de ayer consumieron las siguientes cantidades de café, expresadas en tazas:

3.1	3.3	3.5	2.6	2.6	4.3	4.4	3.8	3.1	4.1	3.1	3.2

Con el nivel de significancia 0.05, ¿sugieren los datos que existe una diferencia entre el promedio nacional y la media de la muestra tomada de las personas de la tercera edad?

34. Hace poco se amplió el área de recuperación del hospital St. Luke, de Maumee, Ohio. Se esperaba que con la ampliación la cantidad media de pacientes al día fuera mayor de 25. Una muestra aleatoria de 15 días reveló las siguientes cantidades de pacientes.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

25	27	25	26	25	28	28	27	24	26	25	29	25	27	24

Con el nivel de significancia 0.01, ¿puede concluir que la cantidad media de pacientes al día es mayor a 25? Calcule el valor *p* e interprételo.

35. El sitio egolfsmith.com recibe un promedio de 6.5 devoluciones al día de compradores en línea; en el caso de una muestra de 12 días, recibió el siguiente número de devoluciones:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

0	4	3	4	9	4	5	9	1	6	7	10

Con el nivel de significancia 0.01, ¿puede concluir que la cantidad media de devoluciones es inferior a 6.5?

36. En temporadas recientes, las Ligas Mayores de Béisbol han sido criticadas por la duración de los juegos; un informe indica que el juego promedio dura 3 horas, 30 minutos. Una muestra de 17 juegos reveló los siguientes tiempos de juego (observe que los minutos se convirtieron en fracciones de hora, de manera que un juego que duró 2:24 horas, se expresa como 2.40 horas).

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2.98	2.40	2.70	2.25	3.23	3.17	2.93	3.18	2.80
2.38	3.75	3.20	3.27	2.52	2.58	4.45	2.45	

Con el nivel de significancia 0.05, ¿puede concluir que el tiempo medio de un juego es menor de 3.50 horas?

37. Watch Corporation de Suiza afirma que, en promedio, sus relojes jamás se atrasan o adelantan durante una semana. Una muestra de 18 relojes arrojó los siguientes adelantos (+) o atrasos (−) en segundos, por semana.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

−0.38	−0.20	−0.38	−0.32	+0.32	−0.23	+0.30	+0.25	−0.10
−0.37	−0.61	−0.48	−0.47	−0.64	−0.04	−0.20	−0.68	+0.05

Con el nivel de significancia 0.05, ¿es razonable concluir que el adelanto o atraso medio de tiempo de los relojes es de 0? Calcule el valor *p*.

38. En la tabla siguiente se muestran los índices de recuperación (porcentual) de un año de una muestra de 12 fondos mutualistas clasificados como fondos gravables del mercado monetario.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

4.63	4.15	4.76	4.70	4.65	4.52	4.70	5.06	4.42	4.51	4.24	4.52

Con el nivel de significancia 0.05, ¿es razonable concluir que los índices de recuperación son de 4.50%?

39. Muchos supermercados y grandes tiendas de menudeo, como Wal-Mart y K-Mart, instalaron sistemas de *autopago* con el fin de que los clientes registren sus artículos y los paguen. ¿Les gusta este servicio a los clientes? ¿Con qué frecuencia lo utilizan? Enseguida aparece la cantidad de clientes que utilizan el servicio en una muestra de 15 días en la tienda Wal-Mart en la carretera 544 en Surfside, Carolina del Sur.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

120	108	120	114	118	91	118	92	104	104
112	97	118	108	117					

Con el nivel de significancia 0.05, ¿es razonable concluir que la cantidad media de clientes que utiliza el sistema de autopago supera los 100 diarios?

40. En un año reciente, la tarifa media para viajar en avión de Charlotte, Carolina del Norte, a Seattle, Washington, con un boleto de descuento fue de 267 dólares. El mes anterior, una muestra aleatoria de 13 tarifas de descuento para viajes redondos en esta ruta arrojó los siguientes datos:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

\$321	\$286	\$290	\$330	\$310	\$250	\$270	\$280	\$299	\$265	\$291	\$275	\$281

Con el nivel de significancia 0.01, ¿puede concluir que la tarifa media se incrementó? Calcule el valor *p.*

41. El editor de *Celebrity Living* afirma que las ventas medias de revistas de personalidad en las que aparecen personajes como Angelina Jolie o Paris Hilton venden 1.5 millones de ejemplares a la semana; una muestra de 10 títulos comparables arroja ventas medias semanales de la semana pasada de 1.3 millones de ejemplares, con una desviación estándar de 0.9. Con el nivel de significancia 0.01, ¿estos datos contradicen lo que afirma el editor?

42. Un informe de Naciones Unidas muestra que el ingreso medio familiar de inmigrantes mexicanos en Estados Unidos es de 27 000 dólares al año. Una evaluación del FLOC (Farm Labor Organizing Committee) de 25 familias mexicanas reveló una media de 30 000 dólares, con una desviación estándar de 10 000 dólares. Con el nivel de significancia 0.01, ¿esta información discrepa con el informe de Naciones Unidas?

43. El costo de las bodas en Estados Unidos se disparó en los últimos años. Como resultado, muchas parejas optan por casarse en el Caribe. Un centro vacacional caribeño anunció en *Bride Magazine* que el costo de una boda caribeña era inferior a 10 000 dólares. Enseguida aparece una lista del costo total (en miles de dólares) de una muestra de ocho bodas caribeñas.

9.7	9.4	11.7	9.0	9.1	10.5	9.1	9.8

Con el nivel de significancia 0.05, ¿es razonable concluir que el costo medio de una boda es inferior a 10 000 dólares?

44. La American Water Works Association reporta que el uso de agua *per capita* en una casa familiar es de 69 galones por día. Legacy Rancho es un desarrollo residencial relativamente nuevo de cien viviendas; los constructores instalaron artefactos para utilizarla de forma más eficiente, como sanitarios de bajo consumo, y posteriormente condujeron una encuesta de las residencias. Respondieron 36 hogares, y la media muestral del consumo de agua por día fue de 64 galones, con una desviación estándar de 8.8 galones diarios. Con el nivel de significancia 0.10, ¿se tiene suficiente evidencia para concluir que los residentes de Legacy Rancho usan menos agua en promedio?

45. Una máquina expendedora de refresco de cola está programada para despachar 9.00 onzas de refresco por vaso, con una desviación estándar de 1.00 onzas. El fabricante de la máquina desea establecer el límite de control para que, en una muestra de 36, 5% de las medias de la muestra sea superior al límite de control superior, y 5%, inferior al límite de control inferior.

a. ¿En qué valor se debe programar el límite de control?

b. Si la media de la población cambia a 8.9, ¿cuál es la probabilidad de que el cambio no se detecte?

c. Si la media de la población cambia a 9.3, ¿cuál es la probabilidad de que el cambio no se detecte?

46. Los propietarios del centro comercial Franklin Park desean estudiar los hábitos de compra de sus clientes; de acuerdo con estudios anteriores, los propietarios tienen la impresión de que un comprador común invierte 0.75 horas en el centro comercial, con una desviación estándar de 0.10 horas. Hace poco, los propietarios del centro comercial incluyeron algunos restaurantes de especialidades diseñados para que los clientes pasen más tiempo en sus instalaciones. Se contrató a la empresa de consultoría Brunner and Swanson Marketing Enterprises para que evaluara los efectos de los restaurantes. Una muestra de 45 clientes mostró que el tiempo medio invertido en el centro comercial se incrementó a 0.80 horas.

a. Elabore una prueba de hipótesis para determinar si el tiempo medio invertido en el centro comercial ha cambiado; utilice el nivel de significancia 0.10.

b. Suponga que el tiempo medio de compras realmente aumentó de 0.75 a 0.79 horas. ¿Cuál es la probabilidad de cometer un error tipo II?

c. Cuando Brunner and Swanson comunicó a los dueños la información del punto anterior, estos pensaron que la probabilidad de cometer un error tipo II era muy alta, ¿cómo se puede reducir?

47. Considere las siguientes hipótesis nula y alternativa.

$$H_0: \mu \leq 50$$
$$H_1: \mu > 50$$

Suponga que la desviación estándar de la población es de 10, la probabilidad de cometer un error tipo I se establece en 0.01, y la probabilidad de cometer un error tipo II, en 0.30; si la media de la población cambia de 50 a 55, ¿de qué tamaño debe ser una muestra para satisfacer estos requisitos?

48. A partir de su experiencia, una compañía aseguradora calcula que el daño medio de un desastre natural en su área asciende a 5 000 dólares. Después de presentar varios planes para prevenir pérdidas, la empresa toma una muestra aleatoria de 200 asegurados y descubre que la cantidad media por reclamo fue de 4 800 dólares, con una desviación estándar de 1 300 dólares. Con el nivel de significancia 0.05, ¿resultaron eficaces los planes de prevención al reducir la media de los reclamos?

49. Una revista de abarrotes de circulación nacional informa que el consumidor habitual pasa 8.00 minutos en la fila de espera de la caja registradora. Una muestra de 24 clientes de una sucursal de Farmer Jack's reveló una media de 7.5 minutos con una desviación estándar de 3.2 minutos. Con el nivel de significancia 0.05, ¿es menor el tiempo de espera en esta tienda que el reportado por la revista?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

50. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.

- **a.** Un artículo reciente en el *Arizona Republic* indicó que el precio medio de venta de las casas en esta área es superior a 220 000 dólares. Con el nivel de significancia 0.01, ¿puede concluir que el precio medio de venta en el área de Goodyear es superior a 220 000 dólares? Determine el valor *p.*
- **b.** El mismo artículo informó que el tamaño medio es superior a 2 100 pies cuadrados. Con el nivel de significancia 0.01, ¿puede concluir que el tamaño medio de las casas que se vendieron en Goodyear es superior a 2 100 pies cuadrados? Determine el valor *p.*

51. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012.

- **a.** Lleve a cabo una prueba de hipótesis para determinar si el salario medio de los equipos fue distinto de 80.0 millones de dólares. Aplique el nivel de significancia 0.05.
- **b.** Con un nivel de significancia de 5%, lleve a cabo una prueba de hipótesis para determinar si la asistencia media fue superior a 2 000 000 por equipo.

52. Consulte los datos de los autobuses del Distrito Escolar Buena.

- **a.** Seleccione la variable del número de millas que recorrieron el mes anterior. Realice una prueba de hipótesis para determinar si el número medio de millas recorridas es igual a 840; utilice el nivel de significancia 0.01. Determine el valor *p* y explique lo que significa.
- **b.** Utilizando la variable "costo de mantenimiento", realice una prueba de hipótesis para determinar si el costo medio de mantenimiento es menor a 500 dólares con el nivel de significancia 0.05. Determine el valor *p* e interprete el resultado.

11 Pruebas de hipótesis de dos muestras

LA COMPAÑÍA GIBBS BABY desea comparar el aumento de peso de bebés que consumen sus productos equiparados con los que emplean el producto de su competidor. Una muestra de 40 bebés que consumen los productos Gibbs reveló un aumento de peso medio de 7.6 libras en sus primeros tres meses de vida, con una desviación estándar de la población es de 2.3 libras. Una muestra de 55 bebés que consumen la marca del competidor reveló un aumento medio de 8.1 libras, con una desviación estándar de la población de 2.9 libras. Con el nivel de significancia 0.05, ¿es posible concluir que los bebés que consumieron la marca Gibbs ganaron menos peso? (vea el ejercicio 3 y el OA11-1).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA11-1 Efectuar una hipótesis acerca de la igualdad entre dos medias poblacionales independientes, asumiendo que las desviaciones estándar de la población se conocen y son iguales.

OA11-2 Efectuar una prueba de hipótesis acerca de la igualdad entre dos medias poblacionales independientes cuando las desviaciones estándar de la población se desconocen.

OA11-3 Efectuar una prueba de hipótesis acerca de la diferencia en la media poblacional entre observaciones apareadas o dependientes.

OA11-4 Explicar la diferencia entre muestras dependientes e independientes.

Introducción

En el capítulo 10 se inició el estudio de las pruebas de hipótesis. Se describió su naturaleza y se realizaron algunas pruebas de hipótesis en las cuales se compararon los resultados de una sola muestra con un valor poblacional; es decir, se seleccionó una sola muestra aleatoria de una población y se realizó una prueba para ver si era razonable el valor propuesto de la población. Recuerde que en el capítulo 10 se seleccionó una muestra del número de escritorios ensamblados por semana en la Jamestown Steel Company para determinar si había un cambio en la tasa de producción. De modo similar, se muestreó el costo del procesamiento de reclamaciones de seguros para determinar si las medidas de recorte de costos resultaron en una media menor a la existente, que era de 60 dólares por reclamación. En ambos casos, se compararon los resultados estadísticos de una *sola* muestra con un parámetro de la población.

En este capítulo se amplía la idea de pruebas de hipótesis para dos muestras. Se seleccionan muestras aleatorias de dos poblaciones distintas para determinar si las medias son iguales. Algunas interrogantes por probar son:

1. ¿Hay alguna diferencia entre el valor medio de los bienes raíces residenciales que vendieron los agentes hombres y los que negociaron las mujeres en el sur de Florida?
2. ¿Los empleados de atención a clientes en Grabit Software, Inc., reciben más llamadas pidiendo asistencia por la mañana o por la tarde?
3. ¿Hay alguna diferencia entre el número de días de ausentismo de los trabajadores jóvenes (menores de 21 años de edad) y los trabajadores de la tercera edad (mayores de 60 años) en la industria de comida rápida?
4. ¿Hay un aumento de la tasa de producción si se toca música en el área de manufactura?

Este capítulo se inicia con el caso en el que se seleccionan muestras aleatorias de dos poblaciones independientes y se desea investigar si tienen la misma media.

ESTADÍSTICA EN ACCIÓN

La elección presidencial de Estados Unidos en 2000 fue una de las más cerradas de la historia. Los medios de información fueron incapaces de hacer una proyección del ganador y la decisión final, con recuentos y decisiones judiciales, tardó más de cinco semanas. Esta no fue la única elección en la cual hubo controversia; poco antes de la elección presidencial de 1936, el *New York Times* publicó el encabezado: "La encuesta de *Digest* da a Landon 32 estados: Landon va ganando 4-3". Sin embargo, Alfred Landon, de Kansas, no resultó electo presidente. En realidad, Roosevelt ganó por más de 11 millones de votos y recibió 523 votos en el Colegio Electoral. ¿Por qué el encabezado estuvo tan errado?

(*continúa*)

Pruebas de hipótesis de dos muestras: muestras independientes

OA11-1
Efectuar una hipótesis acerca de la igualdad entre dos medias poblacionales independientes, asumiendo que las desviaciones estándar de la población se conocen y son iguales.

Un especialista en planeación urbana de Florida desea saber si hay alguna diferencia entre los salarios medios por hora de los plomeros y los electricistas en el centro de ese estado. Un contador financiero quiere saber si la tasa de recuperación media de los fondos mutualistas de alto rendimiento es distinta de la tasa de recuperación media de los fondos mutualistas globales. En cada uno de estos casos hay dos poblaciones independientes. En el primero, los plomeros representan una población, y los electricistas, otra; en el segundo caso, los fondos mutualistas de alto rendimiento son una población, y los fondos mutualistas globales, otra.

Para despejar la duda en estos casos, se debe seleccionar una muestra aleatoria de cada población y calcular la media de las dos muestras. Si ambas medias poblacionales son iguales, es decir, si el salario medio por hora de los plomeros y los electricistas es igual, se esperaría que la *diferencia* entre estas fuese de cero. Pero, ¿qué pasaría si los resultados produjeran una diferencia distinta de cero? ¿La diferencia se debe a la casualidad o a que existe una disparidad real entre los salarios por hora? Para responder la pregunta se debe hacer una prueba de las medias de dos muestras.

Es necesario regresar a los resultados del capítulo 8. Recuerde que se demostró que una distribución de las medias suele aproximarse a la distribución normal. Es necesario, una vez más, suponer que una distribución de las medias de muestras seguirá una distribución normal; de esa manera, será posible demostrar en forma matemática que la distribución de las diferencias entre las medias muestrales de dos distribuciones normales también es normal.

Esta teoría se ejemplifica mediante un especialista en planeación urbana de Tampa, Florida; para iniciar, asuma que determinada información, que por lo general no está disponible, es cierta. Suponga que la población de plomeros tiene un salario medio de 30 dólares por hora y una desviación estándar de 5 dólares por hora, y que la población de electricistas tiene un salario medio de 29 dólares y una desviación estándar de 4.50 dólares. Ahora, a partir de esta información, es claro que las dos medias poblacionales no son iguales. Los plomeros ganan un dólar más por hora que los electricistas pero no se encontrará esta diferencia cada vez que alguien tome muestras de ambas poblaciones.

Suponga que selecciona una muestra aleatoria de 40 plomeros y otra de 35 electricistas, y que calcula la media de cada una. Después determina la diferencia entre las medias muestrales, la cual llama la atención porque si las poblaciones tienen la misma media, es de esperar que la diferencia entre las dos medias muestrales sea cero. Si hay alguna diferencia entre las medias poblacionales, debería existir una diferencia entre las medias muestrales.

Para comprender la teoría es preciso tomar varios pares de muestras, calcular la media de cada una, determinar la diferencia entre las medias muestrales y estudiar la distribución de dicha diferencia. En el capítulo 8 se estudió que la distribución de las medias muestrales sigue la distribución normal; en tal caso, la distribución de sus diferencias también debe seguirla. Este es el primer obstáculo.

El segundo se refiere a la media de la distribución de sus diferencias. Determinar que la media de esta distribución es cero implica que no hay diferencia entre las dos poblaciones; por otro lado, si la media de la distribución de las diferencias es igual a algún valor distinto de cero, ya sea positivo o negativo, se concluye que las dos poblaciones no tienen la misma media.

Para reportar algunos resultados concretos, recuerde al especialista en planeación urbana de Tampa, Florida. En la tabla 11.1 se indica el resultado de la selección de 20 muestras diferentes de 40 plomeros y 35 electricistas tras calcular la media de cada muestra y determinar la diferencia entre dos medias muestrales. En el primer caso, la muestra de 40 plomeros tiene una media de 29.80 dólares, y la de los electricistas es de 28.76 dólares. La diferencia entre las medias muestrales es de 1.04 dólares. Este proceso se repitió 19 veces más; observe que en 17 de los 20 casos, las diferencias son positivas porque la media de los plomeros es mayor que la de los electricistas. En dos casos, las diferencias son negativas porque la media de los electricistas es mayor que la de los plomeros. Las medias son iguales en un caso.

El obstáculo final es que se necesita saber algo acerca de la *variabilidad* de la distribución de las diferencias; en otras palabras, ¿cuál es la desviación estándar de esta distribución de las diferencias? En la teoría estadística se demuestra que cuando se tienen poblaciones independientes, como en este caso, la distribución de las diferencias tiene una varianza (desviación estándar elevada al cuadrado) igual a la suma de dos varianzas individuales. Esto significa que las varianzas de dos distribuciones muestrales se pueden sumar; en otras palabras, la varianza de la diferencia

(*continuación*)

El *Literary Digest* recopiló una muestra de votantes entre las listas de números telefónicos, registros automovilísticos y sus lectores. En 1936 no muchas personas tenían teléfono o automóvil; además, quienes leían el *Digest* solían ser más ricos y votaban por los republicanos. Por todo ello, la población de la muestra no representaba a la población de votantes. Un segundo problema fue la falta de respuestas. Se enviaron encuestas a más de 10 millones de personas y cerca de 2.3 millones las respondieron; sin embargo, no se tomó en cuenta si las personas que respondieron formaban una muestra representativa de los votantes.

Con las computadoras y los métodos modernos de encuestas, las muestras se seleccionan y verifican con cuidado para garantizar que sean representativas. ¿Qué sucedió con *Literary Digest*? Cerró poco después de la elección de 1936.

TABLA 11.1 Medias de muestras aleatorias de 20 plomeros y electricistas

Muestra	Plomeros	Electricistas	Diferencia
1	$29.80	$28.76	$1.04
2	30.32	29.40	0.92
3	30.57	29.94	0.63
4	30.04	28.93	1.11
5	30.09	29.78	0.31
6	30.02	28.66	1.36
7	29.60	29.13	0.47
8	29.63	29.42	0.21
9	30.17	29.29	0.88
10	30.81	29.75	1.06
11	30.09	28.05	2.04
12	29.35	29.07	0.28
13	29.42	28.79	0.63
14	29.78	29.54	0.24
15	29.60	29.60	0.00
16	30.60	30.19	0.41
17	30.79	28.65	2.14
18	29.14	29.95	−0.81
19	29.91	28.75	1.16
20	28.74	29.21	−0.47

entre medias muestrales ($\bar{x}_1 - \bar{x}_2$) es igual a la suma de la varianza de los plomeros y de la varianza de los electricistas.

VARIANZA DE LA DISTRIBUCIÓN DE LAS DIFERENCIAS ENTRE MEDIAS

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \qquad [11.1]$$

El término $\sigma^2_{\bar{x}_1 - \bar{x}_2}$ parece complejo, pero no es difícil interpretarlo. La parte σ^2 indica que es una varianza, y el subíndice, $\bar{x}_1 - \bar{x}_2$, que es una distribución de las diferencias de las medias muestrales.

Es posible representar esta ecuación en forma más práctica con la raíz cuadrada, de modo que se obtenga la desviación estándar de la distribución o “error estándar” de las diferencias; por último, se estandariza la distribución de las diferencias. El resultado se representa mediante la siguiente ecuación.

PRUEBA DE DOS MEDIAS DE MUESTRAS σ CONOCIDA

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \qquad [11.2]$$

Antes de presentar un ejemplo, repase las suposiciones necesarias para emplear la fórmula [11.2].

- Las dos poblaciones siguen distribuciones normales.
- Las dos muestras no deben estar relacionadas, es decir, deben ser independientes.
- Debe conocerse la desviación estándar de ambas medias poblacionales.

En el ejemplo siguiente se muestran los detalles de la prueba de hipótesis de dos medias poblacionales.

EJEMPLO

Los clientes de los supermercados FoodTown tienen dos opciones al pagar sus compras: pueden hacerlo en una caja registradora normal, operada por un cajero, o emplear el nuevo procedimiento *Fast Lane* (caja rápida). Cuando eligen la primera alternativa, un empleado de FoodTown registra cada artículo, lo pone en una banda transportadora pequeña de donde otro empleado lo toma y lo pone en una bolsa, la cual coloca en el carrito de víveres. En el procedimiento Fast Lane, el cliente registra cada artículo, lo pone en una bolsa y coloca las bolsas en el carrito. Este procedimiento está diseñado para reducir el tiempo que los clientes pierden en la fila de la caja.

En la sucursal de FoodTown de la calle Byrne se acaba de instalar un aparato de Fast Lane. La gerente de la tienda desea saber si el tiempo medio de pago con el método tradicional es mayor que con Fast Lane, para lo cual reunió la información siguiente sobre la muestra. El tiempo se mide desde el momento en que el cliente ingresa a la fila hasta que sus bolsas están en el carrito. De aquí que el tiempo incluye tanto la espera en la fila como el registro. ¿Cuál es el valor *p*?

Tipo de cliente	Media muestra	Desviación estándar de la población	Tamaño de la muestra
Tradicional	5.50 minutos	0.40 minuto	50
Fast Lane	5.30 minutos	0.30 minuto	100

SOLUCIÓN

Para responder la pregunta anterior se emplea el procedimiento de prueba de hipótesis de seis pasos.

Paso 1: Se establecen las hipótesis nula y alternativa. La hipótesis nula es que no hay diferencia entre los tiempos medios de pago de ambos grupos; en otras palabras, la diferencia de 0.20 minutos entre el tiempo medio de pago con el método tradicional y el tiempo medio de pago con Fast Lane se debe a la casualidad. La hipótesis alternativa es que el tiempo

ESTADÍSTICA EN ACCIÓN

¿Vive para trabajar o trabaja para vivir? Una encuesta reciente entre 802 trabajadores estadounidenses reveló que, entre quienes consideran su trabajo como una profesión, el número medio de horas que trabajan por día es de 8.7 y, entre quienes lo consideran como un empleo, el número medio es de 7.6.

medio de quienes utilizan el método tradicional es mayor. Si μ_s se refiere al tiempo medio de pago de la población de clientes tradicionales y μ_u al tiempo medio de pago de los clientes que emplean Fast Lane, las hipótesis nula y alternativa son:

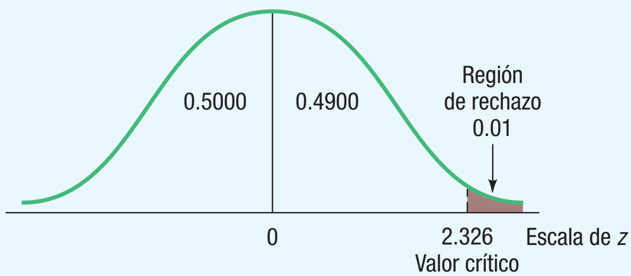
$$H_0: \mu_s \leq \mu_F$$
$$H_1: \mu_s > \mu_F$$

Paso 2: **Se selecciona un nivel de significancia.** Es decir, la probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera; la cual se determina antes de seleccionar la muestra o realizar algún cálculo. Los niveles de significancia 0.05 y 0.01 son los más comunes, pero también se emplean otros, como 0.02 y 0.10. En teoría, se puede seleccionar cualquier valor entre 0 y 1 como nivel de significancia; en este caso se seleccionó 0.01.

Paso 3: **Se identifica el estadístico de prueba.** En el capítulo 10 empleó la distribución normal estándar (o distribución *z*) y *t* como estadísticos de prueba; en este caso se usa la distribución *z* como el estadístico de prueba debido a que se asume que las desviaciones estándar de las dos poblaciones se conocen.

Paso 4: **Se formula la regla de decisión.** Esta se basa en las hipótesis nula y alternativa (es decir, prueba de una o dos colas), en el nivel de significancia y en el estadístico de prueba empleado. Se seleccionó el nivel de significancia 0.01 y la distribución *z* como el estadístico de prueba, y se desea determinar si el tiempo medio de pago es mayor con el método tradicional. Se formula la hipótesis alternativa que indica que el tiempo medio de pago de quienes emplean el método tradicional es mayor; de aquí, la región de rechazo se encuentra en la cola superior de la distribución normal (una prueba de una cola). Para determinar el valor crítico, vaya a la distribución de la *t* de Student (apéndice B.5). En los encabezados de la tabla se encuentra la fila etiquetada **"Nivel de significancia de la prueba de una cola";** ubíquese en ese punto y seleccione la columna para un alfa de 0.01. Vaya a la última fila con los grados de libertad infinitos. El valor crítico *z* es 2.326; así que la regla de decisión es rechazar la hipótesis nula si el valor el estadístico de prueba sobrepasa 2.326. En la gráfica 11-1 se representa la regla de decisión.

$$H_0: \mu_S \leq \mu_F$$
$$H_1: \mu_S > \mu_F$$



GRÁFICA 11.1 Regla de decisión para prueba de una cola con un nivel de significancia 0.01

Paso 5: **Se toma una muestra y se decide respecto de H_0.** Emplee la fórmula [11.2] para calcular el valor del estadístico de prueba.

$$z = \frac{\bar{X}_S - \bar{X}_F}{\sqrt{\frac{\sigma_S^2}{n_S} + \frac{\sigma_F^2}{n_F}}} = \frac{5.5 - 5.3}{\sqrt{\frac{0.40^2}{50} + \frac{0.30^2}{100}}} = \frac{0.2}{0.064031} = 3.123$$

El valor calculado, 3.123, es mayor que el valor crítico 2.326. La decisión es rechazar la hipótesis nula y aceptar la alternativa.

Paso 6: **Se interpreta el resultado.** La diferencia de 0.20 minutos entre el tiempo medio de pago con el método tradicional es demasiado grande para deberse a la casualidad; en conclusión, el método Fast Lane es más rápido.

¿Cuál es el valor *p* del estadístico de prueba? Recuerde que el valor *p* es la probabilidad de determinar un valor del estadístico de prueba así de excepcional cuando la hipótesis nula es verdadera. Para calcular el valor *p* es necesaria la probabilidad de un valor *z* mayor que 3.123. En el apéndice B.3 no aparece la probabilidad asociada con 3.123, el mayor valor disponible es 3.09; el área que corresponde a 3.09 es 0.4990. En este caso, el valor p es menor que 0.0010, calculado mediante $0.5000 - 0.4900$. ¡La conclusión es que hay muy pocas probabilidades de que la hipótesis nula sea verdadera!

En resumen, los criterios para emplear la fórmula [11.2] son:

1. *Las muestras son de poblaciones independientes*. Esto significa, por ejemplo, que el tiempo de pago de los clientes que emplean Fast Lane no está relacionado con el tiempo de pago de los demás clientes; es decir, el tiempo del señor Smith no afecta el tiempo de pago de ningún otro cliente.
2. *Ambas poblaciones siguen la distribución normal.* En el ejemplo de FoodTown, esto significa que la población de tiempos tanto en la fila estándar como en la de Fast Lane siguen la distribución normal.
3. *Las dos desviaciones estándar de las poblaciones se conocen*. En el ejemplo, la desviación estándar de la población de los tiempos de pago con Fast Lane fue 0.30 minutos, y la de los tiempos de pago tradicionales, 0.40 minutos. Emplee la fórmula [11.2] para determinar el valor del estadístico de prueba.

AUTOEVALUACIÓN 11-1

Tom Sevits, propietario de Appliance Patch, observó una diferencia en el total en dólares de las ventas entre los hombres y las mujeres que emplea como agentes de ventas. Una muestra de 40 días reveló que los hombres tienen una media de 1 400 dólares por concepto de venta de aparatos por día. En una muestra de 50 días, las mujeres tuvieron una media de 1 500 dólares por el mismo concepto. Suponga que la desviación estándar de los hombres es de 200 dólares, y la de las mujeres, de 250 dólares. Con el nivel de significancia 0.05, ¿puede el señor Sevits concluir que la cantidad media que venden por día las mujeres es mayor?

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuál es la regla de decisión?
(c) ¿Cuál es el valor del estadístico de prueba?
(d) ¿Cuál es su decisión respecto de la hipótesis nula?
(e) ¿Cuál es el valor p?
(f) Interprete el resultado.

EJERCICIOS

1. Considere una muestra de 40 observaciones de una población con una desviación estándar de la población de 5 y una media muestral de 102. Otra muestra de 50 observaciones de una segunda población tiene una desviación estándar de la población de 6 y una media muestral de 99. Realice la prueba de hipótesis siguiente con el nivel de significancia 0.04.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

   a. ¿Se trata de una prueba de una o de dos colas?
   b. Formule la regla de decisión.
   c. Calcule el valor del estadístico de prueba.
   d. ¿Cuál es su decisión respecto de H_0?
   e. ¿Cuál es el valor p?

2. Considere una muestra de 65 observaciones de una población con una desviación estándar de la población de 0.75 y una media muestral de 2.67. Otra muestra de 50 observaciones de una segunda población tiene una desviación estándar de la población de 0.66 y una media muestral de 2.59. Realice la prueba de hipótesis siguiente con el nivel de significancia 0.08.

$$H_0: \mu_1 \leq \mu_2$$
$$H_1: \mu_1 > \mu_2$$

   a. ¿Se trata de una prueba de una o de dos colas?
   b. Formule la regla de decisión.
   c. Calcule el valor del estadístico de prueba.
   d. ¿Cuál es su decisión respecto de H_0?
   e. ¿Cuál es el valor p?

Nota: Los siguientes ejercicios se resuelven mediante el procedimiento de prueba de hipótesis de seis pasos.

3. La compañía Gibbs Baby desea comparar el aumento de peso de bebés que consumen sus productos en comparación con los que emplean producto de su competidor. Una muestra de 40 bebés que consumen los productos Gibbs reveló un aumento de peso medio de 7.6 libras en sus primeros tres meses de vida, con una desviación estándar de la población de la muestra de 2.3 libras. Una muestra de 55 bebés que consumen la marca del competidor reveló un aumento medio de 8.1 libras, con una desviación estándar de la población de 2.9 libras. Con el nivel de significancia 0.05, ¿es posible concluir que los bebés que consumieron la marca Gibbs ganaron menos peso? Calcule el valor *p* e interprételo.

4. Como parte de un estudio de empleados corporativos, el director de recursos humanos de PNC, Inc., desea comparar entre la distancia que deben cubrir para ir al trabajo los empleados de su oficina del centro de Cincinnati y la que recorren quienes trabajan en el centro de Pittsburgh. Una muestra de 35 empleados de Cincinnati revela que viajan una media de 370 millas al mes; por su parte, una muestra de 40 empleados de Pittsburgh indica que viajan una media de 380 millas al mes. La desviación estándar de la población de los empleados de Cincinnati y Pittsburgh es de 30 y 26 millas, respectivamente. Con el nivel de significancia 0.05, ¿existe alguna diferencia entre el número medio de millas recorrido al mes entre los empleados de Cincinnati y los de Pittsburgh?

5. Se sospecha que la altura de las mujeres es un factor para tener partos difíciles; esto es, una mujer más bajita tiene más probabilidades de necesitar una cesárea. Un investigador médico encontró, en una muestra de 45 mujeres que habían tenido un parto normal, que su estatura media era de 61.4 pulgadas. Una segunda muestra de 39 mujeres que fueron sometidas a cesárea tuvo una estatura media de 60.6 pulgadas. Suponga que la población de estaturas relacionadas con los partos normales tiene una desviación estándar de 1.2 pulgadas, y la otra población tiene una desviación estándar de 1.1 pulgadas. ¿Eran más bajas las que tuvieron parto por cesárea? Utilice el nivel de significancia 0.05. Encuentre el valor *p* y explique lo que significa.

6. Mary Jo Fitzpatrick es la vicepresidenta de servicios de enfermería del hospital Luke's Memorial. Hace poco observó que, en las ofertas de trabajo para enfermeras sindicalizadas, los sueldos son más altos que para las no sindicalizadas; por tanto, tras investigar, reunió la siguiente información.

Grupo	Salario medio	Desviación estándar de la población	Tamaño de la muestra
Sindicalizado	\$20.75	\$2.25	40
No sindicalizado	\$19.80	\$1.90	45

¿Sería razonable que la señora Fitzpatrick concluyera que las enfermeras sindicalizadas ganan más? Utilice el nivel de significancia 0.02. ¿Cuál es el valor *p*?

OA11-2
Efectuar una prueba de hipótesis acerca de la igualdad entre dos medias poblacionales independientes cuando las desviaciones estándar de la población se desconocen.

Comparación de medias poblacionales con desviaciones estándar desconocidas

En la sección anterior utilizamos la distribución normal estándar y *z* como el estadístico de prueba para comprobar la hipótesis de que dos medias muestrales de poblaciones independientes eran iguales. La prueba de hipótesis presumió que las poblaciones estaban normalmente distribuidas y que se conocían las desviaciones estándar de la población. Sin embargo, en muchos casos no se conocen las desviaciones estándar de la población. Este problema se soluciona, al igual que en el caso de una muestra en el capítulo anterior, al sustituir la desviación estándar de la muestra (s) por la desviación estándar de la población (σ) (vea la fórmula [10.2] del capítulo 10).

Prueba de dos muestras agrupadas

En esta sección se describe otro método para comparar las medias muestrales de dos poblaciones independientes y determinar si las poblaciones muestreadas pueden tener, de forma razonable, la misma media. Dicho método *no* requiere que se conozcan las desviaciones estándar de las poblaciones; lo cual proporciona más flexibilidad cuando se investiga la diferencia entre las medias de las muestras. Hay dos diferencias importantes entre esta prueba y la descrita antes en este capítulo.

1. Se asume que las poblaciones muestreadas tienen desviaciones estándar iguales pero desconocidas; por tanto, las desviaciones estándar de las muestras se combinan o "agrupan".
2. Se utiliza la distribución t como el estadístico de prueba.

La fórmula para determinar el valor del estadístico de prueba t es similar a la fórmula [11.2], pero es necesario hacer otro cálculo: las dos desviaciones estándar de las muestras se agrupan para formar una sola estimación de la desviación estándar desconocida de la población. En esencia, se calcula una media ponderada de las dos desviaciones estándar de ambas muestras y este se emplea valor como una estimación de la desviación estándar desconocida de la población. Las ponderaciones son los grados de libertad que proporciona cada muestra. ¿Por qué es necesario agrupar las desviaciones estándar de las muestras? Al suponer que las dos poblaciones tienen desviaciones estándar iguales, la mejor estimación posible de ese valor es combinar o agrupar toda la información de las muestras que se tenga acerca del valor de la desviación estándar de la población.

La fórmula siguiente se emplea para agrupar las desviaciones estándar de las muestras; observe que participan dos factores: el número de observaciones en cada muestra y las propias desviaciones estándar de las muestras.

VARIANZA CONJUNTA

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} \quad [11.3]$$

donde:

s_1^2 es la varianza (desviación estándar elevada al cuadrado) de la primera muestra;
s_2^2 es la varianza de la segunda muestra.

El valor de t se calcula a partir de la ecuación siguiente.

PRUEBAS DE DOS MEDIAS DE DOS MUESTRAS σ DESCONOCIDAS

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad [11.4]$$

donde:

$\bar{x}_1$ es la media de la primera muestra;
$\bar{x}_2$ es la media de segunda muestra;
n_1 es el número de observaciones en la primera muestra;
n_1 es el número de observaciones en la segunda muestra;
s_p^2 es la estimación conjunta de la varianza de la población.

El número de grados de libertad de la prueba es el número total de elementos muestreados menos el número total de muestras. Como hay dos muestras, hay $n_1 + n_2 - 2$ grados de libertad.

En resumen, la prueba tiene tres requisitos o suposiciones.

1. Las poblaciones muestreadas siguen la distribución normal.
2. Las poblaciones muestreadas son independientes.
3. Las desviaciones estándar de las dos poblaciones son iguales.

En el ejemplo siguiente se explican los detalles de la prueba.

EJEMPLO

Owens Lawn Care, Inc., fabrica y ensambla podadoras de césped que envía a distribuidores instalados en Estados Unidos y Canadá. Se han propuesto dos procedimientos distintos para el montaje del motor al chasis de la podadora. La pregunta es: ¿existe una diferencia entre tales procedimientos con respecto al tiempo medio para montar los motores al chasis de las podadoras? El primer procedimiento lo desarrolló Herb Welles, un antiguo empleado de Owens (designado como procedimiento W), y el otro lo desarrolló William Atkins, vicepresidente de ingeniería de Owens (designado como procedimiento A). Para evaluar ambos métodos se decidió realizar un estudio de tiempos y movimientos. Se midió el tiempo de montaje en una muestra de cinco empleados según el método de Welles y seis con el de Atkins. Los resultados, en minutos, se registran a la derecha. ¿Hay alguna diferencia entre los tiempos medios de montaje? Utilice el nivel de significancia 0.10.

Welles (minutos)	Atkins (minutos)
2	3
4	7
9	5
3	8
2	4
	3

SOLUCIÓN

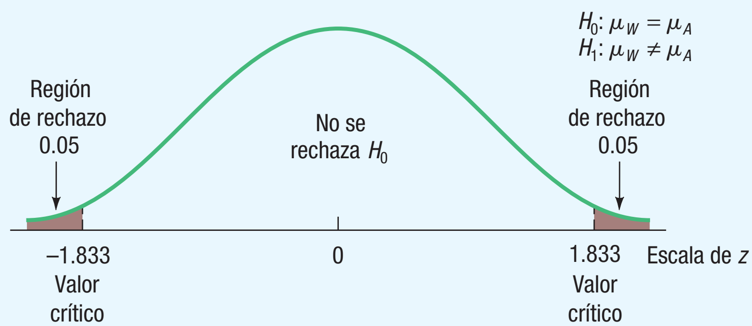
Al seguir el procedimiento de los seis pasos, la hipótesis nula establece que no hay diferencia entre los tiempos medios de montaje de ambos procedimientos. La hipótesis alternativa indica que sí existe una diferencia.

$$H_0\colon \mu_W = \mu_A$$
$$H_1\colon \mu_W \neq \mu_A$$

Las suposiciones son:

- Las observaciones incluidas en la muestra de Welles son *independientes* de las de la muestra de Atkins.
- Ambas poblaciones siguen la distribución normal.
- Ambas poblaciones tienen desviaciones estándar iguales.

¿Hay alguna diferencia entre los tiempos medios de ensamblado con los métodos de Welles y Atkins? Los grados de libertad son iguales al número total de elementos muestreados menos el número de muestras, en este caso, $n_W + n_A - 2$. Cinco trabajadores utilizaron el método de Welles y seis el de Atkins; por lo tanto, hay 9 grados de libertad, calculados así: $5 + 6 - 2$. Los valores críticos de t (del apéndice B.5) para $gl = 9$, una prueba de dos colas y el nivel de significancia de 0.10, son -1.833 y 1.833. La regla de decisión se ilustra en la gráfica 11.2. La hipótesis nula no se rechaza si el valor calculado de t se encuentra entre -1.833 y 1.833.



GRÁFICA 11.2 Regiones de rechazo, prueba de dos colas, $gl = 9$ y nivel de significancia 0.10

Se emplean tres pasos para calcular el valor de t.

Paso 1: Calcule las desviaciones estándar de las muestras. Para calcular las desviaciones estándar se utiliza la fórmula [3.11]; cuyos detalles se muestran a continuación.

Método de Welles		Método de Atkins	
x_W	$(x_W - \bar{x}_W)^2$	x_A	$(x_A - \bar{x}_A)^2$
2	$(2 - 4)^2 = 4$	3	$(3 - 5)^2 = 4$
4	$(4 - 4)^2 = 0$	7	$(7 - 5)^2 = 4$
9	$(9 - 4)^2 = 25$	5	$(5 - 5)^2 = 0$
3	$(3 - 4)^2 = 1$	8	$(8 - 5)^2 = 9$
2	$(2 - 4)^2 = 4$	4	$(4 - 5)^2 = 1$
20	34	3	$(3 - 5)^2 = 4$
		30	22

$$\bar{x}_w = \frac{\Sigma x_w}{n_w} = \frac{20}{5} = 4 \qquad \bar{x}_A = \frac{\Sigma x_A}{n_A} = \frac{30}{6} = 5$$

$$s_w = \sqrt{\frac{\Sigma(x_w - \bar{x}_w)^2}{n_w - 1}} = \sqrt{\frac{34}{5 - 1}} = 2.9155 \qquad s_A = \sqrt{\frac{\Sigma(x_A - \bar{x}_A)^2}{n_A - 1}} = \sqrt{\frac{22}{6 - 1}} = 2.0976$$

Paso 2: Agrupe las varianzas de las muestras. Emplee la fórmula [11.3] para agrupar las varianzas de las muestras (desviaciones estándar al cuadrado).

$$s_p^2 = \frac{(n_w - 1)s_w^2 + (n_A - 1)s_A^2}{n_w + n_A - 2} = \frac{(5 - 1)(2.9155)^2 + (6 - 1)(2.0976)^2}{5 + 6 - 2} = 6.2222$$

Paso 3: Determine el valor de *t*. El tiempo medio de montaje del método de Welles es de 4.00 minutos, determinado mediante $\bar{x}_w = 20/5$. El tiempo medio de montaje del método de Atkins es de 5.00 minutos, que se determinó mediante $\bar{x}_A = 30/6$. Mediante la fórmula [11.4] se calcula el valor de t.

$$t = \frac{\bar{x}_w - \bar{x}_A}{\sqrt{s_p^2\left(\frac{1}{n_w} + \frac{1}{n_A}\right)}} = \frac{4.00 - 5.00}{\sqrt{6.2222\left(\frac{1}{5} + \frac{1}{6}\right)}} = -0.662$$

La decisión es no rechazar la hipótesis nula porque -0.662 se encuentra en la región entre -1.833 y 1.833; en conclusión, los datos muestreados no mostraron una diferencia entre los tiempos medios necesarios para montar el motor en el chasís con ambos métodos.

El valor p también se determina mediante el apéndice B.5: localice la fila con 9 grados de libertad y utilice la columna de prueba de dos colas. Encuentre el valor t, sin considerar el signo, el cual está más cercano al valor calculado de 0.662. Es 1.383, que corresponde al nivel de significancia 0.20; así, aunque se hubiera utilizado el nivel de significancia de 20%, no habría rechazado la hipótesis nula de medias iguales. El valor p es mayor que 0.20.

En Excel se encuentra un procedimiento denominado "Prueba t: dos muestras si las varianzas son iguales" para realizar los cálculos de las fórmulas [11.3] y [11.4], así como para determinar las medias y varianzas de las muestras; los detalles del procedimiento se proporcionan en el apéndice C. Los datos se ingresan en las dos primeras columnas de la hoja de cálculo de Excel y se identifican como "Welles" y "Atkins"; a continuación se presenta la salida en pantalla. El valor de t, denominado "estadístico t", es -0.662, y el valor p de dos colas es 0.525. Como se esperaba, el valor p es mayor que el nivel de significancia 0.10. La conclusión es no rechazar la hipótesis nula.

welles and atkins.xlsx

	A	B	C	D	E	F
1	Welles	Atkins		Prueba *t*: dos muestras si las varianzas son iguales		
2	2	3				
3	4	7			*Welles*	*Atkins*
4	9	5		Media	4.000	5.000
5	3	8		Varianza	8.500	4.400
6	2	4		Observaciones	5.000	6.000
7		3		Varianza conjunta	6.222	
8				Diferencia media hipotética	0.000	
9				gl	9.000	
10				Estadístico t	-0.662	
11				P(T<=t) de una cola	0.262	
12				t crítica de una cola	1.833	
13				P(T<=t) de dos colas	0.525	
14				t crítica de dos colas	2.262	

AUTOEVALUACIÓN 11-2

El gerente de producción de Bellevue Steel, fabricante de sillas de ruedas, desea comparar entre el número de sillas de ruedas defectuosas producidas en el turno matutino y el del turno vespertino. Una muestra de la producción de seis turnos matutinos y ocho vespertinos reveló el siguiente número de defectos.

Matutino	5	8	7	6	9	7		
Vespertino	8	10	7	11	9	12	14	9

Con el nivel de significancia 0.05, ¿hay alguna diferencia entre el número medio de defectos por turno?

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuál es la regla de decisión?
(c) ¿Cuál es el valor del estadístico de prueba?
(d) ¿Cuál es su decisión respecto de la hipótesis nula?
(e) ¿Cuál es el valor p?
(f) Interprete el resultado.
(g) ¿Cuáles son las suposiciones necesarias de esta prueba?

EJERCICIOS

En los ejercicios 7 y 8: a) formule la regla de decisión, b) calcule la estimación conjunta de la varianza de la población, c) calcule el estadístico de prueba, d) tome una decisión respecto de la hipótesis nula y e) calcule el valor p.

7. Las hipótesis nula y alternativa son:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Una muestra aleatoria de 10 observaciones de una población reveló una media muestral de 23 y una desviación estándar de 4; una muestra aleatoria de 8 observaciones de otra población reveló una media muestral de 26 y una desviación estándar de la muestra de 5. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre las medias poblacionales?

8. Las hipótesis nula y alternativa son:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Una muestra aleatoria de 15 observaciones de la primera población reveló una media muestral de 350 y una desviación estándar de la muestra de 12; una muestra aleatoria de 17 observaciones de la segunda población reveló una media de 342 y una desviación estándar de la muestra de 15. Con el nivel de significancia 0.10, ¿hay alguna diferencia entre las medias poblacionales?

Nota: Los siguientes ejercicios se resuelven mediante el procedimiento de prueba de seis pasos.

9. A continuación se muestran los nombres de los 25 jugadores en el rol de apertura del equipo de los Yankees de Nueva York de las Ligas Mayores de Béisbol, temporada 2013, sus salarios y sus posiciones.

Jugador	Salario (dólares)	Posición
CC Sabathia	24 285 714	Lanzador
Hiroki Kuroda	15 000 000	Lanzador
Andy Pettitte	12 000 000	Lanzador
Mariano Rivera	10 000 000	Lanzador
Phil Hughes	7 150 000	Lanzador
Boone Logan	3 150 000	Lanzador
David Robertson	3 100 000	Lanzador
Joba Chamberlain	1 875 000	Lanzador
Shawn Kelley	935 000	Lanzador
Ivan Nova	575 600	Lanzador
David Phelps	512 425	Lanzador
Adam Warren	490 525	Lanzador
Lyle Overbay	850 000	Primera base
Robinson Cano	15 000 000	Segunda base

Jugador	Salario (dólares)	Posición
Kevin Youkilis	12 000 000	Tercera base
Francisco Cervelli	515 350	Receptor
Chris Stewart	515 100	Receptor
Brett Gardner	2 850 000	Jardinero central
Travis Hafner	2 000 000	Bateador designado
Vernon Wells	24 642 857	Jardinero izquierdo
Ben Francisco	1 200 000	Jardinero izquierdo
Ichiro Suzuki	6 500 000	Jardinero derecho
Brennan Boesch	1 500 000	Jardinero derecho
Jayson Nix	900 000	Parador en corto (*shortstop*)
Eduardo Nuñez	533 300	Parador en corto (*shortstop*)

Divida a los jugadores en dos grupos: lanzadores (P) y no lanzadores (jugadores de posición) y asuma que existen desviaciones poblacionales iguales para ambos. Pruebe la hipótesis de que los salarios medios de los lanzadores y los jugadores de posición son los mismos en el nivel de significancia 0.01.

10. En un estudio reciente se comparó el tiempo que pasan juntas las parejas en las que solo trabaja uno de los cónyuges con las parejas en las que ambos trabajan. De acuerdo con los registros que llevaron las esposas durante el estudio, la cantidad media de tiempo que pasan juntos viendo televisión las parejas en las que solo trabaja uno de los cónyuges fue 61 minutos por día, con una desviación estándar de 15.5 minutos; las parejas en las que ambos trabajan pasaron una media de 48.4 minutos viendo televisión, con una desviación estándar de 18.1 minutos. Con el nivel de significancia 0.01, ¿se puede concluir que, en promedio, las parejas en que solo trabaja uno de los cónyuges pasan más tiempo juntos viendo televisión? En el estudio había 15 parejas con un solo trabajador y 12 con ambos trabajadores.

11. Lisa Monnin es la directora de presupuestos de Nexos Media, Inc., y quiere comparar los gastos diarios en viáticos del personal de ventas con los gastos del personal de auditoría, para lo cual, recopiló la información siguiente sobre las muestras.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Ventas (dólares)	131	135	146	165	136	142	
Auditoría (dólares)	130	102	129	143	149	120	139

Con el nivel de significancia 0.10, ¿puede Monnin concluir que los gastos diarios medios del personal de venta son mayores que los del personal de auditoría? ¿Cuál es el valor *p*?

12. La Area Chamber of Commerce de Tampa Bay (Florida) quería saber si el salario semanal medio de las enfermeras era superior al de los maestros de escuela. Para esta investigación recopiló la información siguiente sobre las cantidades que ganó la semana pasada una muestra de maestros y enfermeras.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Maestros de escuela (dólares)	845	826	827	875	784	809	802	820	829	830	842	832
Enfermeras (dólares)	841	890	821	771	850	859	825	829				

¿Es razonable concluir que el salario semanal medio de las enfermeras es mayor? Utilice el nivel de significancia 0.01. ¿Cuál es el valor *p*?

Medias poblacionales con desviaciones estándar desiguales

En las secciones anteriores se supuso que las poblaciones tenían desviaciones estándar iguales; en otras palabras, como no se conocían las desviaciones estándar de las poblaciones, se suponían iguales. En muchos casos, esta es una suposición razonable, pero ¿qué sucede si no son iguales? En el capítulo siguiente se presenta un método formal para probar esta suposición de varianzas iguales; en el capítulo 16 se describe una prueba de hipótesis que no requiere ni de una varianza igual ni de la presunción de normalidad.

Si no es razonable suponer que las desviaciones estándar poblacionales son iguales, se emplea un estadístico muy similar a la fórmula [11.2]. Las desviaciones estándar de las muestras, s_1 y s_2, se emplean en lugar de las desviaciones estándar de las poblaciones respectivas; además, los grados de libertad se ajustan hacia abajo mediante una fórmula de aproximación compleja. El efecto es reducir el número de grados de libertad de la prueba, lo cual requiere un valor mayor del estadístico de prueba para rechazar la hipótesis nula.

La fórmula del estadístico *t* es:

ESTADÍSTICO DE PRUEBA DE MEDIAS SIN DIFERENCIA, VARIANZAS DESIGUALES

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \qquad \textbf{[11.5]}$$

Los grados de libertad estadística se determinan mediante:

GRADOS DE LIBERTAD PARA PRUEBA CON VARIANZA DESIGUAL

$$gl = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \qquad \textbf{[11.6]}$$

donde n_1 y n_2 son los tamaños muestrales respectivos, y s_1 y s_2, las desviaciones estándar de las muestras respectivas; si es necesario, esta fracción se redondea hacia abajo a un valor entero. En el ejemplo siguiente se ilustran los detalles.

EJEMPLO

El personal en un laboratorio de pruebas del consumidor que evalúa la absorción de las toallas de papel desea comparar entre dos conjuntos similares de toallas de dos marcas distintas. Sumerge una toalla de cada conjunto en un tubo con un fluido, se deja escurrir en una charola durante dos minutos y después se evalúa la cantidad de líquido que absorbió de la charola. Una muestra aleatoria de 9 toallas de papel de la primera marca absorbió las cantidades siguientes de líquido (en milímetros).

8	8	3	1	9	7	5	5	12

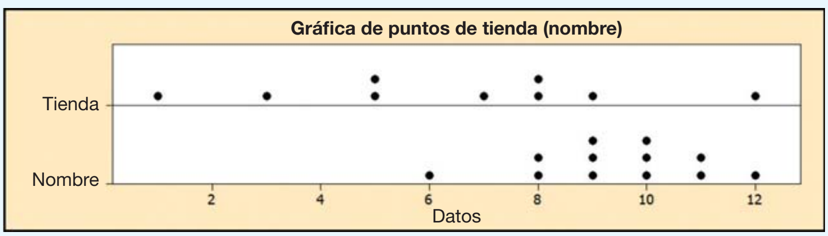
Una muestra aleatoria independiente de 12 toallas de la otra marca absorbió las cantidades siguientes de líquido (en milímetros).

12	11	10	6	8	9	9	10	11	9	8	10

Utilice el nivel de significancia 0.10 y pruebe si existe una diferencia entre las cantidades medias de líquido que absorbieron las toallas.

SOLUCIÓN

Para iniciar, se supone que las cantidades de líquido absorbido de las toallas de ambas marcas siguen la distribución de probabilidad normal. Como se desconocen las desviaciones estándar de las poblaciones, se empleará la distribución *t* como estadístico de prueba; además, no parece razonable suponer que las desviaciones estándar de las poblaciones son iguales. La cantidad de absorción en la primera marca varía de 1 ml a 12 ml; en el caso de la segunda, la cantidad de absorción varía de 6 ml a 12 ml. Es decir, existe más variación en la cantidad de absorción de la primera marca que de la segunda. La diferencia en la variación se observa en la gráfica de puntos siguiente que se obtuvo con Minitab. Los comandos del software para crear una gráfica de puntos en Minitab se dan en el apéndice B, apartado referente al capítulo 4.



Por lo tanto, se decide emplear la distribución *t* y suponer que las desviaciones estándar de las poblaciones no son iguales.

En el procedimiento de prueba de hipótesis de seis pasos, el primero es formular las hipótesis nula y alternativa; la primera es que no hay diferencia en la cantidad media de líquido que absorben ambos tipos de toallas; la segunda es que sí hay una diferencia.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

El nivel de significancia es 0.10, y el estadístico de prueba sigue la distribución *t*. Como no se desea suponer desviaciones estándar de las poblaciones iguales, los grados de libertad se ajustan con la fórmula [11.8]. Para ello es preciso determinar las desviaciones estándar de las muestras (el sistema

Minitab es útil para determinar rápidamente estos resultados). También se encontrará la tasa de absorción media, la cual se empleará en breve. Los tamaños muestrales respectivos son $n_1 = 9$ y $n_2 = 12$, y las desviaciones estándar respectivas, 3.32 ml y 1.621 ml.

Variable	n	Media	Desviación estándar
Tienda	9	6.444	3.321
Nombre	12	9.417	1.621

Al sustituir esta información en la fórmula [11.6]:

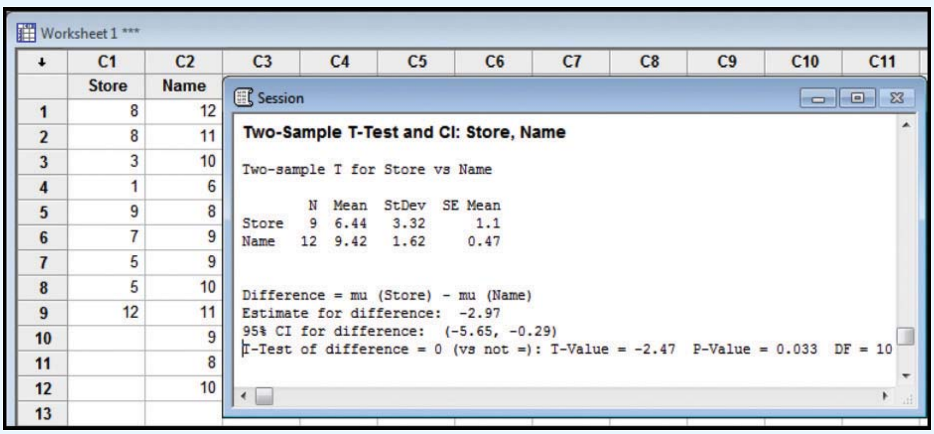
$$gl = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{[(3.321^2/9) + (1.621^2/12)]^2}{\dfrac{(3.321^2/9)^2}{9 - 1} + \dfrac{(1.621^2/12)^2}{12 - 1}} = \frac{1.4444^2}{0.1877 + 0.0044} = 10.86$$

La práctica común es redondear hacia abajo a un entero, por lo que se emplean 10 grados de libertad. Del apéndice B.5 con 10 grados de libertad, una prueba de dos colas y el nivel de significancia 0.10, los valores t críticos son -1.812 y 1.812. La regla de decisión es rechazar la hipótesis nula si el valor calculado de t es menor que -1.812 o mayor que 1.812.

Para determinar el valor del estadístico de prueba se emplea la fórmula [11.5]; recuerde que la cantidad de absorción de las toallas de papel de primera marca es 6.444 ml, y 9.417 ml de la otra.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{6.444 - 9.417}{\sqrt{\dfrac{3.321^2}{9} + \dfrac{1.621^2}{12}}} = -2.474$$

El valor calculado de t es inferior al valor crítico menor, por lo que la decisión es rechazar la hipótesis nula. Se concluye que la tasa de absorción media de las dos toallas no es la misma; a continuación se incluye la salida de Minitab para este ejemplo.



AUTOEVALUACIÓN 11-3

Con frecuencia para las compañías es útil saber quiénes son sus clientes y cómo llegaron a serlo. Una compañía de tarjetas de crédito quiere saber si el tarjetahabiente la solicitó por interés propio o si fue contactado por teléfono por un agente. La compañía obtuvo la información muestral siguiente respecto de los saldos al final del mes de ambos grupos.

Fuente	Media	Desviación de la muestra	Tamaño de la muestra
Solicitantes	$1 568	$356	10
Contactados	1 967	857	8

¿Es razonable concluir que el saldo medio de los tarjetahabientes que fueron contactados por teléfono es mayor que el de quienes solicitaron la tarjeta por cuenta propia? Suponga que las desviaciones estándar de las poblaciones no son iguales y utilice el nivel de significancia 0.05.

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuántos grados de libertad hay?
(c) ¿Cuál es la regla de decisión?
(d) ¿Cuál es el valor del estadístico de prueba?
(e) ¿Cuál es su decisión respecto de la hipótesis nula?
(f) Interprete el resultado.

EJERCICIOS

En los ejercicios 13 y 14 suponga que las poblaciones muestrales no tienen desviaciones estándar iguales y utilice el nivel de significancia 0.05: a) determine el número de grados de libertad, b) formule la regla de decisión, c) calcule el estadístico de prueba y d) tome su decisión acerca de la hipótesis nula.

13. Las hipótesis nula y alternativa son:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Una muestra aleatoria de 15 elementos de la primera población reveló una media de 50 y una desviación estándar de 5; una muestra de 12 elementos para la segunda población reveló una media de 46 y una desviación estándar de 15.

14. Las hipótesis nula y alternativa son:

$$H_0: \mu_1 \leq \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Una muestra aleatoria de 20 elementos de la primera población reveló una media de 100 y una desviación estándar de 15; una muestra de 16 elementos de la segunda población reveló una media de 94 y una desviación estándar de 8. Utilice el nivel de significancia 0.05.

15. Una encuesta reciente comparó el costo de adopción de niños a través de agencias públicas y privadas. En una muestra de 16 adopciones a través de una agencia pública, el costo medio fue 21 045 dólares, con una desviación estándar de 835 dólares. En una muestra de 18 adopciones de niños a través de una agencia privada, el costo medio fue 22 840 dólares, con una desviación estándar de 1 545 dólares. ¿Es posible concluir que el costo medio de adoptar niños es mayor a través de una agencia privada? Utilice el nivel de significancia de.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

16. Suponga que usted es un experto en la industria de la moda y desea reunir información para comparar la cantidad mensual que ganan las modelos que vistieron ropa de Liz Claiborne con respecto a las que modelaron ropa de Calvin Klein. La siguiente es la cantidad (en miles de dólares) que gana al mes una muestra de modelos de Liz Claiborne:

$5.0	$4.5	$3.4	$3.4	$6.0	$3.3	$4.5	$4.6	$3.5	$5.2
4.8	4.4	4.6	3.6	5.0					

La siguiente es la cantidad (en miles de dólares) que gana una muestra de modelos de Calvin Klein:

$3.1	$3.7	$3.6	$4.0	$3.8	$3.8	$5.9	$4.9	$3.6	$3.6
2.3	4.0								

¿Es razonable concluir que las modelos de Claiborne ganan más? Utilice el nivel de significancia 0.05 y suponga que las desviaciones estándar de las poblaciones no son iguales.

OA11-3
Efectuar una prueba de hipótesis acerca de la diferencia en la media poblacional entre observaciones apareadas o dependientes.

Pruebas de hipótesis de dos muestras: muestras dependientes

En el ejemplo de Owens Lawn Care Inc. se probó la diferencia entre las medias de dos muestras independientes. Se comparó el tiempo medio que se requiere para montar un motor según el método de Welles con el de Atkins. Las muestras eran *independientes*, lo que significa que la muestra de los tiempos de ensamblado del método de Welles no estaba de ninguna manera relacionada con la de los tiempos del método de Atkins.

Sin embargo, hay situaciones en que las muestras no son independientes; en otras palabras, las muestras son *dependientes* o están *relacionadas*. Por ejemplo, la compañía Nickel Savings and

Loan recurre a dos empresas, Schadek Appraisals y Bowyer Real State, para valuar los bienes raíces sobre los cuales se hacen los préstamos. Es importante que los avalúos de ambas empresas contemplen cantidades similares. Para revisar la consistencia de estas empresas, Nickel Savings selecciona en forma aleatoria 10 casas y pide a Schadek Appraisals y a Bowyer Real State que las valúen; es decir, de cada una se harán dos avalúos (uno de Schadek Appraisals y otro de Bowyer Real State), los cuales dependen, o están relacionados, con la casa seleccionada. A esto también se le conoce como **muestra apareada.**

El interés en la prueba de hipótesis recae en la distribución de las *diferencias* entre los valores de avalúo de cada casa. En este caso solo hay una muestra. En palabras más formales, se investiga si la media de la distribución de las diferencias entre los avalúos es cero. La muestra se compone de las *diferencias* entre los avalúos determinados por Schadek Appraisals y los de Bowyer Real State. Si ambas empresas reportan estimaciones similares, entonces algunas veces los avalúos de Schadek serán los de valor mayor y otras veces lo serán los de Bowyer Real State; sin embargo, la media de la distribución de las diferencias será cero. Por otro lado, si una de las empresas reporta de manera consistente los avalúos más altos, la media de la distribución de las diferencias no será cero.

Se empleará el símbolo μ_d para indicar la media poblacional de la distribución de las diferencias. Se supone que la distribución de las diferencias de la población sigue la distribución normal de manera aproximada, mientras que el estadístico de prueba sigue la distribución *t,* y su valor se calcula a partir de la fórmula siguiente:

PRUEBA *t* APAREADA

$$t = \frac{\overline{d}}{s_d/\sqrt{n}} \qquad \textbf{[11.7]}$$

Hay $n - 1$ grados de libertad y

$\overline{d}$ es la media de la diferencia entre las observaciones apareadas o relacionadas.
s_d es la desviación estándar de las diferencias entre las observaciones apareadas o relacionadas.
n es el número de observaciones apareadas.

La desviación estándar de las diferencias se calcula mediante la conocida fórmula de la desviación estándar, excepto que x se sustituye por d. La fórmula es:

$$s_d = \sqrt{\frac{\sum(d - \overline{d})^2}{n - 1}}$$

En el ejemplo siguiente se ilustra esta prueba.

EJEMPLO

Recuerde que como Nickel Savings and Loan desea comparar las dos compañías que contrata para valuar las casas, seleccionó una muestra de 10 propiedades y programa los avalúos de ambas empresas; los resultados, en miles de dólares, se muestran a la derecha.

Con el nivel de significancia 0.05, ¿se puede concluir que hay una diferencia entre los avalúos medios de las casas?

Casa	Schadek	Bowyer
1	235	228
2	210	205
3	231	219
4	242	240
5	205	198
6	230	223
7	231	227
8	210	215
9	225	222
10	249	245

SOLUCIÓN

El primer paso es formular las hipótesis nula y alternativa; en este caso es adecuada una alternativa de dos colas porque se quiere determinar si hay una *diferencia* entre los avalúos de las firmas. No existe interés en demostrar si una empresa en particular asigna a las propiedades mayor valor que la otra. La pregunta es si las diferencias en la muestra entre los avalúos pueden provenir de una población con una media de 0; en tal caso, se concluye que no hay diferencia entre los avalúos. Las hipótesis nula y alternativa son:

$$H_0: \mu_1 = 0$$
$$H_1: \mu_1 \neq 0$$

Hay 10 casas valuadas por ambas empresas, por lo que $n = 10$, y $gl = n - 1 = 10 - 1 = 9$. Se tiene una prueba de dos colas, y el nivel de significancia es 0.05. Para determinar el valor crítico consulte el apéndice B.5, y vea la fila con 9 grados de libertad hasta la columna de una prueba de dos colas y el nivel de significancia 0.05. El valor en la intersección es 2.262; el cual se presenta en el cuadro de la tabla 11.2. La regla de decisión es rechazar la hipótesis nula si el valor calculado de t es menor que -2.262 o mayor que 2.262. Estos son los detalles del cálculo.

Casa	Schadek	Bowyer	Diferencia, d	$(d - \bar{d})$	$(d - \bar{d})^2$
1	235	228	7	2.4	5.76
2	210	205	5	0.4	0.16
3	231	219	12	7.4	54.76
4	242	240	2	−2.6	6.76
5	205	198	7	2.4	5.76
6	230	223	7	2.4	5.76
7	231	227	4	−0.6	0.36
8	210	215	−5	−9.6	92.16
9	225	222	3	−1.6	2.56
10	249	245	4	−0.6	0.36
			46	0	174.40

$$\bar{d} = \frac{\sum d}{n} = \frac{46}{10} = 4.60$$

$$s_d = \sqrt{\frac{\sum(d - \bar{d})^2}{n - 1}} = \sqrt{\frac{174.4}{10 - 1}} = 4.402$$

Con la fórmula [11.7], el valor del estadístico de prueba es 3.305, determinado por

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{4.6}{4.402/\sqrt{10}} = \frac{4.6}{1.3920} = 3.305$$

Como el valor calculado de t se encuentra en la región de rechazo, se rechaza la hipótesis nula. La distribución de las diferencias de la población no tiene una media de 0; por tanto, se concluye que hay una diferencia entre los avalúos medios de las casas. Como la mayor diferencia (12 000 dólares) está en la tercera casa; quizás este sería un buen punto para iniciar una revisión más detallada.

Para determinar el valor p, consulte el apéndice B.5 y la sección de una prueba de dos colas. Busque en la fila con 9 grados de libertad y encuentre los valores de t que se aproximen al valor calculado. Para el nivel de significancia 0.01, el valor de t es 3.250. El valor calculado es mayor, pero menor que el valor de 4.781 que corresponde al nivel de significancia de 0.001; de aquí, el valor p es menor que 0.01. Esta información sombreada en la tabla 11.2.

TABLA 11.2 Parte de la distribución t del apéndice B.5

	Intervalos de confianza					
	80%	90%	95%	98%	99%	
	Nivel de significancia de una prueba de una cola					
gl	0.10	0.05	0.025	0.01	0.005	
	Nivel de significancia de una prueba de dos colas					
	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
[illegible]	[illegible]	2.132	2.776	3.747	4.604	8.610
[illegible]	[illegible]	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	$\boxed{2.262}$	2.821	$\boxed{3.250}$	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587

Valor p entre 0.01 y 0.001

Estadístico t crítico para 0.05

En Excel se encuentra un procedimiento denominado "Prueba *t*: dos muestras apareadas para medias" para realizar los cálculos de la fórmula [11.7]. La pantalla de salida de este procedimiento se presenta a continuación.

El valor calculado de *t* es 3.305, y el valor *p* de dos colas, 0.009. Como el valor *p* es menor que 0.05, se rechaza la hipótesis de que la media de la distribución de las diferencias entre los avalúos es cero; de hecho, este valor *p* se encuentra entre 0.01 y 0.001. Hay una pequeña probabilidad de que la hipótesis nula sea verdadera.

paired t test

	A	B	C	D	E	F	G
1	Home	Schadek	Bowyer		Prueba *t*: dos muestras si las varianzas son iguales		
2	1	235	228				
3	2	210	205			*Schadek*	*Bowyer*
4	3	231	219		Media	226.800	222.200
5	4	242	240		Varianza	208.844	204.178
6	5	205	198		Observaciones	10.000	10.000
7	6	230	223		Varianza conjunta	0.953	
8	7	231	227		Diferencia media hipotética	0.000	
9	8	210	215		gl	9.000	
10	9	225	222		Estadístico t	3.305	
11	10	249	245		P(T<=t) de una cola	0.005	
12					t crítica de una cola	1.833	
13					P(T<=t) de dos colas	0.009	
14					t crítica de dos colas	2.262	
15							

Comparación de muestras dependientes e independientes

OA11-4 Explicar la diferencia entre muestras dependientes e independientes.

Con frecuencia, los estudiantes principiantes confunden la diferencia entre las pruebas de muestras independientes (fórmula [11.4]) y las pruebas de muestras dependientes (fórmula [11.7]). ¿Cómo distinguir la diferencia entre muestras dependientes e independientes? Hay dos tipos de muestras dependientes: 1) las que se caracterizan por una medición, una intervención de algún tipo y después otra medición y 2) las que implican una relación o agrupación de las observaciones. Para explicarlo con más detalle:

1. El primer tipo de muestra dependiente se distingue por una medición seguida de una intervención de alguna clase y después otra medición. Esto se puede denominar un estudio de "antes" y "después". Dos ejemplos ayudarán a explicarlo mejor: suponga que desea demostrar que, al colocar bocinas en el área de manufactura y tocar música relajante, aumenta la producción; así, comienza con la selección de una muestra de trabajadores y una medición de sus resultados en las condiciones actuales. Después instala las bocinas en el área de producción y vuelve a medir la producción de los mismos trabajadores; por tanto, hay dos mediciones: antes de colocar las bocinas en el área de producción y después de hacerlo. La intervención es la colocación de las bocinas en el área de producción.

   Un segundo ejemplo comprende una empresa educativa que ofrece cursos diseñados para incrementar las calificaciones en los exámenes y la capacidad para leer (SAT). Suponga que la empresa quiere ofrecer un curso que ayudará a los alumnos de primer año de preparatoria a aumentar sus puntajes en el SAT; para iniciar, cada estudiante presenta el SAT en el primer año de preparatoria. Durante el verano, entre los años primero y último, participan en el curso que les proporciona consejos para presentar exámenes. Para finalizar, durante el otoño del último año de preparatoria, vuelven a presentarlo. Una vez más, el procedimiento se caracteriza por una medición (presentar el SAT como estudiante de primer año), una intervención (los talleres de verano) y otra medición (presentarlo durante su último año).
2. El segundo tipo de muestra dependiente se caracteriza por relacionar o aparear observaciones. En el ejemplo de Nickel Savings se ilustra una muestra dependiente de este tipo: se seleccionó

una propiedad y dos firmas la valuaron. Como segundo ejemplo, suponga que una psicóloga industrial desea estudiar las similitudes intelectuales de parejas recién casadas, para lo cual selecciona una muestra; después, administra una prueba de inteligencia estándar a ambos cónyuges para determinar la diferencia entre las calificaciones. Observe la relación que ocurrió: se comparan las calificaciones apareadas, o relacionadas, por un matrimonio.

¿Por qué se prefieren las muestras dependientes a las independientes? Cuando se emplean muestras dependientes se reduce la variación en la distribución del muestreo. Para ilustrar este ejemplo se utilizará nuevamente el caso de Nickel Savings and Loan. Suponga que se tienen dos muestras independientes de propiedades de bienes raíces para su avalúo y se realiza una prueba de hipótesis, con la fórmula [11.4]; las hipótesis nula y alternativa son:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Ahora hay dos muestras independientes de 10 cada una; así, el número de grados de libertad es $10 + 10 - 2 = 18$. Del apéndice B.5, en el nivel de significancia de 0.05, H_0 se rechaza si t es menor que -2.101 o mayor que 2.101.

Mediante Excel se encuentran las medias y las desviaciones estándar de dos muestras independientes, como se muestra en el apéndice C en la sección del capítulo 3 para determinar la media y la desviación estándar de las dos muestras independientes. Los comandos de Excel para encontrar la varianza agrupada y el valor del "estadístico t" están en la sección correspondiente al capítulo 11 del apéndice C; estos valores están sombreados.

Independent t test.xlsx

	A	B	C	D	E	F	G
1	Casa	Schadek	Bowyer		Prueba t para dos muestras pareadas		
2	1	235	228				
3	2	210	205			*Schadek*	*Bowyer*
4	3	231	219		Media	226.800	222.200
5	4	242	240		Varianza	208.844	204.178
6	5	205	198		Observaciones	10.000	10.000
7	6	230	223		Varianza conjunta	206.511	
8	7	231	227		Diferencia media hipotética	0.000	
9	8	210	215		gl	18.000	
10	9	225	222		Estadístico t	0.716	
11	10	249	245		P(T<=t) de una cola	0.242	
12					t crítica de una cola	1.734	
13	Media =	226.80	222.20		P(T<=t) de dos colas	0.483	
14	S =	14.45	14.29		t crítica de dos colas	2.101	

La media del avalúo de las 10 propiedades de Schadek es de 226 800 dólares, y la desviación estándar, de 14 500 dólares; la media de los avalúos de Bowyer Real State es de 222 000 dólares, y la desviación estándar, de 14 290 dólares (para facilitar los cálculos, se emplean miles de dólares en lugar de dólares). El valor de la estimación agrupada de la varianza a partir de la fórmula [11.3] es

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(10 - 1)(14.45^{2}) + (10 - 1)(14.29)^2}{10 + 10 - 2} = 206.50$$

De la fórmula [11.4], t es 0.716.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{226.8 - 222.2}{\sqrt{206.50\left(\frac{1}{10} + \frac{1}{10}\right)}} = \frac{4.6}{6.4265} = 0.716$$

El valor calculado de t (0.716) es menor que 2.101, de manera que la hipótesis nula no se rechaza; por tanto, no es posible demostrar que hay una diferencia entre los avalúos medios. ¡Esta no es la misma conclusión a la que se llegó antes! ¿Por qué? El numerador es el mismo que en la prueba de observaciones apareadas (4.6); sin embargo, el denominador es menor. En la prueba por pares el denominador es 1.3920 (vea los cálculos en la solución al ejemplo de Nickel Savings and Loan), pero en el caso de las muestras independientes, el denominador es 6.4265. La variación (o incertidumbre) mayor explica la diferencia entre valores t y entre decisiones estadísticas. El denominador mide el error estándar de la estadística. Cuando las muestras *no* se aparean, se presentan dos clases de variación: diferencias entre las dos empresas valuadoras y la diferencia en el valor del bien raíz. Las propiedades 4 y 10 tienen valores comparativamente altos, en tanto que el de la número 5 es relativamente bajo; esto demuestra lo diferentes que son los avalúos de las propiedades, pero lo que interesa en realidad es la diferencia entre las dos empresas valuadoras.

En suma, cuando es posible aparear las observaciones que miden las diferencias de una variable común, una prueba de hipótesis basada en muestras dependientes es más sensible para detectar una diferencia significativa que una prueba de hipótesis basada en muestras independientes. En el caso de comparar las valuaciones de Schadek Appraisals y Bowyer Real State, la prueba de hipótesis basada en muestras dependientes elimina la variación entre los valores de las propiedades y se concentra solo en las comparaciones entre las dos valuaciones de cada propiedad; pero hay una mala noticia: en la prueba de observaciones apareadas, los grados de libertad son la mitad de lo que serían si las muestras no se apareasen. En el ejemplo de bienes raíces, los grados de libertad disminuyen de 18 a 9 cuando las observaciones están apareadas; sin embargo, en la mayoría de los casos, este es un precio pequeño que se debe pagar por una prueba mejor.

AUTOEVALUACIÓN 11-4

La publicidad que realiza Sylph Fitness Center afirma que, al terminar su entrenamiento, las personas bajarán de peso. Una muestra aleatoria de ocho participantes recientes reveló los pesos que se muestran a la derecha antes y después de terminar el entrenamiento. Con el nivel de significancia 0.01, ¿se puede concluir que los participantes bajan de peso?

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuál es el valor crítico de *t*?
(c) ¿Cuál es el valor calculado de *t*?
(d) Interprete el resultado. ¿Cuál es el valor *p*?
(e) ¿Qué suposición necesita acerca de la distribución de las diferencias?

Nombre	Antes	Después
Hunter	155	154
Cashman	228	207
Mervine	141	147
Massa	162	157
Creola	211	196
Peterson	164	150
Redding	184	170
Poust	172	165

EJERCICIOS

17. Las hipótesis nula y alternativa son:

$$H_0: \mu_d \leq 0$$
$$H_1: \mu_d > 0$$

En la información muestral siguiente se registra el número de unidades defectuosas que producen los turnos matutino y vespertino en una muestra de cuatro días durante el mes anterior.

	Día			
	1	2	3	4
Turno matutino	10	12	15	19
Turno vespertino	8	9	12	15

Con el nivel de significancia 0.05, ¿se puede concluir que se producen más defectos en el turno vespertino?

18. Las hipótesis nula y alternativa son:

$$H_0: \mu_d = 0$$
$$H_1: \mu_d \neq 0$$

Las observaciones apareadas siguientes muestran el número de multas de tránsito por conducir a exceso de velocidad de los oficiales Dhondt y Meredith, de la South Carolina Highway Patrol, durante los últimos cinco meses.

	Cantidad de multas emitidas				
	Mayo	Junio	Julio	Agosto	Septiembre
Oficial Dhondt	30	22	25	19	26
Oficial Meredith	26	19	20	15	19

Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los números medios de multas que dieron los dos oficiales?

Nota: Los siguientes ejercicios se resuelven mediante el procedimiento de prueba de hipótesis de seis pasos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

19. La gerencia de Discount Furniture, cadena de mueblerías de descuento del noreste de Estados Unidos, diseñó un plan de incentivos para sus agentes de ventas. Para evaluar este plan innovador, se seleccionaron 12 vendedores al azar, y se registraron sus ingresos anteriores y posteriores al plan.

Vendedor	Antes	Después
Sid Mahone	\$320	\$340
Carol Quick	290	285
Tom Jackson	421	475
Andy Jones	510	510
Jean Sloan	210	210
Jack Walker	402	500
Peg Mancuso	625	631
Anita Loma	560	560
John Cuso	360	365
Carl Utz	431	431
A. S. Kushner	506	525
Fern Lawton	505	619

¿Hubo algún aumento significativo en el ingreso semanal de un vendedor debido al innovador plan de incentivos? Utilice el nivel de significancia 0.05. Calcule el valor *p* e interprételo.

20. Hace poco, el gobierno federal estadounidense otorgó fondos para un programa especial diseñado para reducir los delitos en áreas de alto riesgo. Un estudio de los resultados del programa en ocho de tales áreas en Miami, Florida, produjo los resultados siguientes.

	Número de delitos por área							
	A	B	C	D	E	F	G	H
Antes	14	7	4	5	17	12	8	9
Después	2	7	3	6	8	13	3	5

¿Hubo alguna disminución en el número de delitos desde la inauguración del programa? Utilice el nivel de significancia 0.01. Calcule el valor *p*.

RESUMEN DEL CAPÍTULO

I. Al comparar dos medias poblacionales se desea saber si pueden ser iguales.
- **A.** Se investiga si la distribución de la diferencia entre las medias puede tener una media de 0.
- **B.** Si se conocen las desviaciones estándar de las poblaciones, el estadístico de prueba sigue la distribución normal estándar.
  1. Ambas poblaciones siguen una distribución normal.
  2. Las muestras son de poblaciones independientes.
  3. La fórmula mediante la cual se calcula el valor z es

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \quad \textbf{[11.2]}$$

II. Si se desconocen las desviaciones estándar poblacionales, el estadístico de prueba que se utiliza para comparar dos medias es la distribución t.
- **A.** Ambas poblaciones deben seguir aproximadamente la distribución normal.
- **B.** Las poblaciones deben tener desviaciones estándar iguales.
- **C.** Las muestras son independientes.
- **D.** Para determinar el valor de t se requieren dos pasos.
  1. Como primer paso, las desviaciones estándar se agrupan de acuerdo con la fórmula siguiente:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad \textbf{[11.3]}$$

2. El valor de t se calcula a partir de la fórmula siguiente:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \textbf{[11.4]}$$

3. Los grados de libertad de la prueba son $n_1 + n_2 - 2$.

III. Si no es posible suponer que las desviaciones estándar de la población son iguales, los grados de libertad y la fórmula se ajustan para encontrar t.

A. Los grados de libertad se determinan mediante la fórmula siguiente:

$$gl = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \qquad \textbf{[11.6]}$$

B. El valor del estadístico de prueba se calcula a partir de la fórmula siguiente:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \textbf{[11.5]}$$

IV. Para muestras dependientes, se supone que la distribución de las diferencias apareadas entre las poblaciones tiene una media de 0.

A. Primero se calcula la media y la desviación estándar de las diferencias muestrales.

B. El valor del estadístico de prueba se calcula a partir de la fórmula siguiente:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \qquad \textbf{[11.7]}$$

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
s_p^2	Varianza conjunta de la muestra	*s subíndice p al cuadrado*
$\bar{x}_1$	Media de la primera muestra	*x barra subíndice 1*
$\bar{x}_2$	Media de la segunda muestra	*x barra subíndice 2*
$\bar{d}$	Media de la diferencia entre observaciones dependientes	*d barra*
s_d	Desviación estándar de la diferencia entre observaciones dependientes	*s subíndice d*

EJERCICIOS DEL CAPÍTULO

21. Un estudio reciente se enfocó en el número de veces que los hombres y las mujeres que viven solos compran comida para llevar en un mes. Asuma que la distribución de probabilidad normal y que las desviaciones estándar de la población son iguales. La información se resume a continuación.

Estadístico	Hombres	Mujeres
Media muestra	24.51	22.69
Desviación estándar de la población	4.48	3.86
Tamaño de la muestra	35	40

Con el nivel de significancia 0.01, ¿hay alguna diferencia entre el número medio de veces que los hombres y las mujeres piden comida para llevar en un mes? ¿Cuál es el valor p?

22. Clark Heter es un ingeniero industrial en Lyons Products, y le gustaría determinar si se producen más unidades en el turno nocturno que en el matutino. Suponga que la desviación estándar de la población del número de unidades producidas en el turno matutino es 21 y, en el nocturno, 28. Una muestra de 54 trabajadores del turno matutino reveló que el número medio de unidades producidas fue 345; una muestra de 60 trabajadores del turno nocturno reveló que el número medio de unidades producidas fue 351. Con el nivel de significancia 0.05, ¿es mayor el número de unidades producidas en el turno nocturno?

23. Fry Brothers Heating and Air Conditioning, Inc., emplea a Larry Clark y George Murnen para ofrecer por teléfono servicios de reparación de chimeneas y unidades de aire acondicionado a domicilio. Al propietario, Tom Fry, le gustaría saber si hay alguna diferencia entre los números medios de llamadas diarias. Suponga que la desviación estándar de la población de Larry Clark es 1.05 llamadas por día y, la de George Murnen, 1.23. Una muestra aleatoria de 40 días que se realizó el año anterior reveló que Larry Clark hace un promedio de 4.77 llamadas por día; por otra parte, en una muestra de 50 días, George Murnen realizó un promedio de 5.02 llamadas diarias. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los números medios de llamadas por día de los dos empleados? ¿Cuál es el valor *p*?

24. Un fabricante de café está interesado en saber si el consumo diario medio de personas que beben café regular es menor que el de quienes lo consumen descafeinado. Suponga que la desviación estándar de la población del primer grupo es 1.20 tazas por día y, en el caso del segundo, 1.36. Una muestra aleatoria de 50 consumidores de café regular reveló una media de 4.35 tazas por día; por otra parte, una muestra de 40 personas que beben café descafeinado reveló una media de 5.84 tazas por día. Utilice el nivel de significancia de 0.01. Calcule el valor *p*.

25. Una compañía de teléfonos celulares ofrece dos planes a sus suscriptores; cuando estos firman el contrato, se les pide alguna información demográfica. El ingreso anual medio de una muestra de 40 suscriptores al plan A es 57 000 dólares, con una desviación estándar de 9 200 dólares; en cuanto al plan B, el ingreso medio de una muestra de 30 suscriptores es de 61 000 dólares, con una desviación estándar de 7 100 dólares. Con el nivel de significancia 0.05, ¿es razonable concluir que el ingreso medio de quienes eligen el plan B es mayor? ¿Cuál es el valor *p*?

26. Un fabricante de computadoras ofrece una línea de ayuda para sus compradores, quienes pueden llamar las 24 horas de los siete días de la semana. Responder a estas llamadas de ayuda en forma oportuna es importante para la imagen de la compañía. Después de decirle al cliente que la solución del problema es importante, se le pregunta si este se relaciona con el software o con el hardware. El tiempo medio que emplea un técnico para resolver un problema de software es 18 minutos, con una desviación estándar de 4.2 minutos (esta información se obtuvo de una muestra de 35 llamadas supervisadas); por otra parte, en un estudio de 45 problemas de hardware, el tiempo medio que empleó el mismo técnico para resolver el problema fue de 15.5 minutos, con una desviación estándar de 3.9 minutos (esta información también se obtuvo de llamadas supervisadas). Con el nivel de significancia 0.05, ¿es más lento resolver problemas de software? ¿Cuál es el valor *p*?

27. Una de las preguntas más apremiantes en la industria de la música es: ¿las tiendas de pago en internet son competitivas frente a los servicios gratuitos para bajar música? Los datos recopilados durante los últimos 12 meses revelaron que, en promedio, 1.65 millones de hogares usaron iTunes, de Apple, con una desviación estándar de 0.56 millones de hogares; en contraste, durante los mismos 12 meses, un promedio de 2.2 millones de familias usaron WinMx (un servicio de descarga P2P gratuito) con una desviación estándar de la muestra de 0.30 millones. Suponga que las desviaciones estándar de las poblaciones no son iguales. Con el nivel de significancia 0.05, pruebe la hipótesis de que no hay diferencia entre los números medios de hogares que eligen cualquiera de los dos servicios de descarga de música.

28. Los negocios, en particular los de la industria de preparación de alimentos, como General Mills, Kellogg y Betty Crocker, dan cupones para fomentar la lealtad a su marca y estimular sus ventas. Se desea saber si los usuarios de cupones de papel son diferentes de los usuarios de cupones electrónicos (distribuidos por internet). En una encuesta se registró la edad de cada persona que usaba los cupones junto con el tipo de cupón (electrónico o de papel). La muestra de 35 usuarios de cupones electrónicos tenía una edad media de 33.6 años, con una desviación estándar de 10.9; en tanto que una muestra similar de 25 usuarios tradicionales de cupones de papel tenía una edad media de 39.5 años, con desviación estándar de 4.8. Suponga que las desviaciones estándar de las poblaciones no son iguales. Con el nivel de significancia 0.01, compruebe la hipótesis de que no hay diferencia entre las edades medias de los grupos de usuarios de cupones.

29. El propietario de hamburguesas Bun 'N' Run desea comparar las ventas por día en dos sucursales. El número medio de ventas de 10 días seleccionados al azar en la sucursal del lado norte fue 83.55, con una desviación estándar de 10.50; en contraparte, mediante una muestra aleatoria de 12 días en la sucursal del lado sur se reportó que el número medio de ventas fue 78.80, con una desviación estándar de 14.25. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los números medios de hamburguesas que venden las dos sucursales? ¿Cuál es el valor *p*?

30. El departamento de ingeniería de Sims Software, Inc., desarrolló dos soluciones químicas para aumentar la vida útil de los discos de computadora. Los discos que se trataron con la primera solución duraron 86, 78, 66, 83, 84, 81, 84, 109, 65 y 102 horas; los discos tratados con la segunda solución duraron 91, 71, 75, 76, 87, 79, 73, 76, 79, 78, 87, 90, 76 y 72 horas. Suponga que las desviaciones estándar de las poblaciones no son iguales. Con el nivel de significancia 0.10, ¿puede concluir que hay una diferencia entre las duraciones de los dos tipos de tratamientos?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

31. El centro comercial de descuento Willow Run tiene dos tiendas Haggar, una en la avenida Peach y la otra en la avenida Plum. Ambas están diseñadas de forma distinta, y el gerente de cada una afirma que su diseño maximiza las cantidades de artículos que los clientes compran por impulso. Una muestra de 10 clientes de la tienda de la avenida Peach reveló que gastaron las siguientes cantidades (en dólares), adicionales a lo planeado: 17.58, 19.73, 12.61, 17.79, 16.22, 15.82, 15.40, 15.86, 11.82 y 15.85; en contraste, una muestra de 14 clientes de la tienda de la avenida Plum reveló que gastaron las siguientes cantidades (en dólares), adicionales a lo planeado: 18.19, 20.22, 17.38, 17.96, 23.92. 15.87, 16.47, 15.96, 16.79, 16.74, 21.40, 20.57, 19.79 y 14.83. Con el nivel de significancia 0.01, ¿hay alguna diferencia entre las cantidades medias compradas por impulso en ambas tiendas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

32. El centro médico Grand Strand Family se diseñó para atender emergencias médicas menores de los habitantes del área de Myrtle Beach. Hay dos instalaciones, una en Little River Area y otra en Murrells Inlet. El departamento de control de calidad desea comparar los tiempos medios de espera de los pacientes en ambas ubicaciones. Las muestras de los tiempos de espera, en minutos, son:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Ubicación	Tiempo de espera											
Little River	31.73	28.77	29.53	22.08	29.47	18.60	32.94	25.18	29.82	26.49		
Murrells Inlet	22.93	23.92	26.92	27.20	26.44	25.62	30.61	29.44	23.09	23.10	26.69	22.31

Suponga que las desviaciones estándar de las poblaciones no son iguales. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los tiempos medios de espera?

33. El Commercial Bank and Trust Company estudia el uso de sus cajeros automáticos; en particular, se desea saber si los adultos jóvenes (menores de 25 años) emplean las máquinas más que los adultos de la tercera edad. Para investigar más, se seleccionaron muestras de clientes menores de 25 años de edad y mayores de 60, y se determinó el número de transacciones en cajeros automáticos que cada individuo seleccionado realizó el mes anterior (los resultados se muestran a continuación). Con el nivel de significancia 0.01, ¿se puede concluir que los clientes más jóvenes utilizan más los cajeros automáticos?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Menores de 25 años	10	10	11	15	7	11	10	9			
Mayores de 60 años	4	8	7	7	4	5	1	7	4	10	5

34. Dos veleros, el *Prada* (Italia) y el *Oracle* (Estados Unidos), compiten por la clasificación en la próxima carrera de la Copa América; lo hacen recorriendo una parte de la ruta varias veces. A continuación se muestran los tiempos de las muestras en minutos; suponga que las desviaciones estándar de las poblaciones no son iguales. Con el nivel de significancia 0.05, ¿puede concluir que hay una diferencia entre sus tiempos medios?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Velero	Tiempo (minutos)										
Prada (Italia)	12.9	12.5	11.0	13.3	11.2	11.4	11.6	12.3	14.2	11.3		
Oracle (Estados Unidos)	14.1	14.1	14.2	17.4	15.8	16.7	16.1	13.3	13.4	13.6	10.8	19.0

35. El fabricante de un reproductor MP3 desea saber si una reducción de 10% en el precio es suficiente para aumentar las ventas de su producto. Para saberlo con certeza, el propietario selecciona al azar ocho tiendas y vende el reproductor MP3 al precio reducido; por otra parte, elige siete tiendas de manera aleatoria y vende el aparato al precio normal. A continuación se presenta el número de unidades que se vendieron el mes anterior en las tiendas muestreadas. Con el nivel de significancia 0.01, ¿puede concluir el fabricante que la reducción de precio generó un aumento de ventas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Precio normal	138	121	88	115	141	125	96	
Precio reducido	128	134	152	135	114	106	112	120

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

36. A pesar de los semáforos, en Teton County ocurre cierto número de accidentes automovilísticos menores en varias intersecciones de alto riesgo. El departamento de tránsito afirma que una modificación del tipo de semáforos los reducirá. Los comisionados del condado acordaron poner en práctica un experimento; se eligieron ocho intersecciones al azar y se modificaron los semáforos. Los números de accidentes menores durante un periodo de seis meses antes y después de las modificaciones fueron:

	Número de accidentes							
	A	B	C	D	E	F	G	H
Antes de la modificación	5	7	6	4	8	9	8	10
Después de la modificación	3	7	7	0	4	6	8	2

Con el nivel de significancia 0.01, ¿es razonable concluir que la modificación redujo el número de accidentes de tránsito?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

37. Lester Hollar es vicepresidente de recursos humanos de una compañía manufacturera importante; en años recientes observó un aumento del absentismo, y considera que se relaciona con la salud general de los empleados. Hace cuatro años, en un intento para mejorar la situación, inició un programa de acondicionamiento físico en el cual los empleados se ejercitan durante la hora del almuerzo; para evaluar el programa, seleccionó una muestra aleatoria de ocho participantes y determinó el número de días que cada uno se ausentó del trabajo en los seis meses antes del inicio del programa de ejercicio y en los últimos seis meses. A continuación se presentan los resultados. Con el nivel de significancia 0.05, ¿se puede concluir que disminuyó el número de ausencias? Estime el valor *p*.

Empleado	Antes	Después
1	6	5
2	6	2
3	7	1
4	7	3
5	4	3
6	3	6
7	5	3
8	6	7

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

38. El presidente del American Insurance Institute desea comparar los costos anuales de los seguros para automóvil que ofrecen dos compañías. Selecciona una muestra de 15 familias, algunas con un solo conductor asegurado y otras con varios conductores adolescentes; entonces, le paga a cada familia una cuota para contactar a ambas compañías y pedir una estimación del costo del seguro. Para hacer comparables los datos, estandariza ciertas características, como la cantidad del deducible y los límites de la cobertura. La información muestral se reporta a continuación. Con el nivel de significancia 0.10, ¿se puede concluir que hay una diferencia en las cantidades estimadas?

Familia	Midstates Car Insurance	Gecko Mutual Insurance
Becker	$2 090	$1 610
Berry	1 683	1 247
Cobb	1 402	2 327
Debuck	1 830	1 367
DuBrul	930	1 461
Eckroate	697	1 789
German	1 741	1 621
Glasson	1 129	1 914
King	1 018	1 956
Kucic	1 881	1 772
Meredith	1 571	1 375
Obeid	874	1 527
Price	1 579	1 767
Phillips	1 577	1 636
Tresize	860	1 188

39. La inmobiliaria Fairfield Homes desarrolla dos áreas cerca de Pigeon Fork, Tennessee; a fin de probar estrategias publicitarias distintas, utiliza medios diferentes para llegar a los compradores potenciales. El ingreso familiar anual medio de 15 personas del primer desarrollo es de 150 000 dólares, con una desviación estándar de 40 000 dólares; una muestra correspondiente de 25 personas del segundo desarrollo obtuvo una media de 180 000 dólares, con una desviación estándar de 30 000 dólares. Suponga que las desviaciones estándar de las poblaciones son iguales. Con el nivel de significancia 0.05, ¿puede la inmobiliaria Fairfield concluir que las medias poblacionales son diferentes?

40. Una compañía de dulces probó el sabor de dos barras de chocolate, una con almendras y otra sin almendras; un panel de degustadores calificó las barras en una escala de 0 a 5, con el 5 indicando la mejor calificación en el sabor. Si la barra tenía almendras, se asignaba a la calificación el código 1; en caso contrario, el código 0. Asuma que las desviaciones estándar de las poblaciones son iguales. En el nivel de significancia 0.5, ¿muestran las calificaciones una diferencia entre las barras de chocolate que contenían almendras y las que no?

Calificación	Con/sin	Calificación	Con/sin
3	1	1	1
1	1	4	0
0	0	4	0
2	1	2	1
3	1	3	0
1	1	4	0

41. Mediante una investigación acerca de la eficacia de un jabón antibacterial para reducir la contaminación de una sala de operaciones se generó la tabla siguiente; en esta se reportan los niveles de contaminación antes y después de usarlo en cada quirófano. El jabón nuevo se probó en una muestra de ocho salas de operación en el área de Seattle durante el año anterior.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Sala de operaciones							
	A	B	C	D	E	F	G	H
Antes	6.6	6.5	9.0	10.3	11.2	8.1	6.3	11.6
Después	6.8	2.4	7.4	8.5	8.1	6.1	3.4	2.0

En el nivel de significancia 0.05, ¿se puede concluir que las mediciones de contaminación son menores después del uso del jabón nuevo?

42. Los datos que aparecen a la derecha sobre las tasas de recuperación anuales se recopilaron aleatoriamente de cinco tipos de acciones que se cotizan en la Bolsa de Valores de Nueva York (New York Stock Exchange, NYSE; también conocida como "el gran tablero") y cinco que lo hacen en NASDAQ. Suponga que las desviaciones estándar de las poblaciones son iguales. Con el nivel de significancia 0.10, ¿se puede concluir que las tasas de recuperación anuales son mayores en "el gran tablero"?

NYSE	NASDAQ
17.16	15.80
17.08	16.28
15.51	16.21
8.43	17.97
25.15	7.77

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

43. La ciudad de Laguna Beach opera dos estacionamientos públicos. El de Ocean Drive tiene capacidad para 125 automóviles, y el de Rio Rancho, para 130. Los planeadores urbanos consideran aumentar el tamaño de los estacionamientos y cambiar la estructura de las tarifas; para iniciar, la oficina de planeación desea conocer el número de automóviles que hay en los estacionamientos en diversas horas del día. Se encarga a un funcionario de planeación principiante la tarea de visitar ambos estacionamientos a horas aleatorias del día y la tarde para contar el número de vehículos estacionados en ellos (el estudio duró un mes). A continuación se presenta el número de automóviles en los estacionamientos durante 25 visitas al estacionamiento de Ocean Drive y 28 al de Rio Rancho. Suponga que las desviaciones estándar de las poblaciones son iguales.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Ocean Drive												
89	115	93	79	113	77	51	75	118	105	106	91	54
63	121	53	81	115	67	53	69	95	121	88	64	
Rio Rancho												
128	110	81	126	82	114	93	40	94	45	84	71	74
92	66	69	100	114	113	107	62	77	80	107	90	129
105	124											

¿Es razonable concluir que hay una diferencia entre los números medios de automóviles en cada estacionamiento? Utilice el nivel de significancia 0.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

44. La cantidad de ingresos que se gasta en vivienda es un componente importante del costo de vida; para los propietarios, los costos totales de vivienda incluyen pagos de la hipoteca, impuesto predial y de servicios (agua, calefacción, electricidad). Un economista seleccionó una muestra de 20 propietarios en Nueva Inglaterra, hace cinco años y en la actualidad, y después calculó los costos totales de vivienda como porcentaje del ingreso mensual. La información se reporta a continuación. ¿Es razonable concluir que el porcentaje es menor en la actualidad que hace cinco años?

Propietario	Hace cinco años	Actualmente	Propietario	Hace cinco años	Actualmente
1	17%	10%	11	35%	32%
2	20	39	12	16	32
3	29	37	13	23	21
4	43	27	14	33	12
5	36	12	15	44	40
6	43	41	16	44	42
7	45	24	17	28	22
8	19	26	18	29	19
9	49	28	19	39	35
10	49	26	20	22	12

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

45. La farmacia CVS, ubicada en la US 17 en Murrells Inlet ha sido una de las tiendas farmacéuticas minoristas más ocupadas de Carolina del Sur durante muchos años. Para capturar más negocios en el área, la gerencia abrió otra farmacia a unos 10 kilómetros, en SC 707. Después de unos meses, la gerencia decidió comparar el volumen de ventas de cada farmacia; una forma de medir esto es contar el número de autos en los estacionamientos en distintos días y horarios aleatorios; a continuación se reportan los resultados de la encuesta de los últimos tres meses del año. Para explicar, la primera observación fue el 2 de octubre a las 20:52; en ese momento había cuatro autos en el estacionamiento de la US 17, y nueve en el de la SC 707. En el nivel de significancia 0.05, ¿es razonable concluir que, basándose en el conteo de vehículos, la tienda de la US 17 tiene un nivel de ventas mayor que el de la SC 707?

		Conteo de vehículos	
Fecha	Hora	US 17	SC 707
Oct 2	20:52	4	9
Oct 11	19:30	5	7
Oct 15	22:08	9	12
Oct 19	11:42	4	5
Oct 25	15:32	10	8
Oct 26	11:02	9	15
Nov 3	11:22	13	7
Nov 5	19:09	20	3
Nov 8	15:10	15	14
Nov 9	13:18	15	11
Nov 15	22:38	13	11
Nov 17	18:46	16	12
Nov 21	15:44	17	8
Nov 22	15:34	15	3
Nov 27	21:42	20	6
Nov 29	9:57	17	13
Nov 30	17:58	5	9
Dic 3	19:54	7	13
Dic 15	18:20	11	6
Dic 16	18:25	14	15
Dic 17	11:08	8	8
Dic 22	21:20	10	3
Dic 24	15:21	4	6
Dic 25	20:21	7	9
Dic 30	14:25	19	4

46. Una de las metas de la instrucción financiera para niños es enseñarles a manejar sabiamente el dinero. Una de las preguntas es: ¿cuánto dinero tienen que manejar los niños? Un estudio reciente de Schnur Educational Research Associates muestreó al azar a 15 niños entre 8 y 10 años, y 18 niños entre 11 y 14 años, registrando sus mesadas. ¿Es razonable concluir que la mesada media recibida por los niños entre 11 y 14 años es mayor que la recibida por los niños entre los 8 y 10 años? Utilice el nivel de significancia 0.01. ¿Cuál es el valor *p*?

8-10 años	11-14 años	8-10 años	11-14 años
6	19	6	11
13	14	5	8
10	12	7	14
6	8	9	9
14	9	14	20
6	11	12	19
7	9		11
7	8		12
10	8		20

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

47. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.
   a. Con el nivel de significancia 0.05, ¿puede concluir que hay una diferencia entre los precios de venta medios de las casas con alberca y sin ella?
   b. Con el nivel de significancia 0.05, ¿concluye que hay una diferencia entre los precios de venta medio de las casas con cochera y sin ella?
   c. Con el nivel de significancia 0.05, ¿puede concluir que hay una diferencia entre los precios de venta medios de las casas que están en los municipios 1 y 2?

48. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012.
   a. Con el nivel de significancia 0.05, ¿puede concluir que hay una diferencia en el salario medio de los equipos en la Liga Americana en comparación con los de la Liga Nacional?
   b. Con el nivel de significancia 0.05, ¿concluye que hay una diferencia entre las asistencias medias como local de los equipos en la Liga Americana en comparación con los equipos de la Liga Nacional?
   c. Calcule la media y la desviación estándar del número de juegos que ganaron los 10 equipos con salarios más altos; haga lo mismo con los 10 equipos con salarios más bajos. Con el nivel de significancia 0.05, ¿hay una diferencia entre los números medios de juegos ganados de ambos grupos?

49. Consulte los datos de los autobuses del Distrito Escolar Buena. ¿Existe alguna diferencia entre los costos medios de mantenimiento de los que utilizan diesel comparados con los que utilizan gasolina? Aplique el nivel de significancia 0.05.

12 Análisis de la varianza

UNA COMPAÑÍA DE DISEÑO DE SOFTWARE está creando un buscador nuevo y más rápido; sin duda, este es más rápido, pero las pruebas iniciales indican que hay mayor variación en el tiempo para realizar la búsqueda. Una muestra de 16 búsquedas reveló que la desviación estándar del tiempo requerido fue de 0.22 segundos para el nuevo buscador y de 0.12 segundos para el actual. Con el nivel de significancia 0.05, ¿puede concluir que hay mayor variación en el tiempo de búsqueda del nuevo dispositivo? (vea el ejercicio 24 y el OA12-1).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA12-1 Aplicar la distribución *F* para probar una hipótesis de igualdad entre las varianzas de poblaciones.

OA12-2 Utilizar el enfoque ANOVA para probar que dos o más medias poblacionales son iguales.

OA12-3 Utilizar los intervalos de confianza para probar e interpretar las diferencias entre pares de medias poblacionales.

OA12-4 Utilizar una variable de bloqueo en un ANOVA de dos vías para probar una hipótesis de igualdad entre tres o más medias poblacionales.

OA12-5 Realizar un ANOVA de dos vías con interacción y describir los resultados.

Introducción

En este capítulo se continúa con el análisis de las pruebas de hipótesis. Recuerde que en los capítulos 10 y 11 se estudió la teoría general de las pruebas de hipótesis; se analizó el caso en que se seleccionó una muestra de una población; se utilizó la distribución z (distribución normal estándar) o la distribución t para determinar si era razonable concluir que la media poblacional era igual a un valor específico y se probó si dos medias poblacionales eran iguales. En este capítulo se amplía la idea de pruebas de hipótesis; se describe una prueba para varianzas y, después, una que compara de forma simultánea varias medias para determinar si provienen de poblaciones iguales.

Comparación de dos varianzas poblacionales

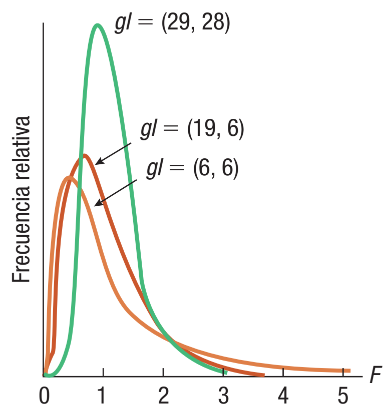
OA12-1
Aplicar la distribución F para probar una hipótesis de igualdad entre las varianzas de poblaciones.

En el capítulo 11 se probaron hipótesis sobre medias poblacionales iguales. Las pruebas diferían según los supuestos con respecto a si las desviaciones estándar o las varianzas de la población eran o no iguales. El supuesto sobre las varianzas poblacionales iguales también es importante en este capítulo; en esta sección se presenta una forma de probar este supuesto estadísticamente; esta se basa en la distribución F.

Distribución *F*

La distribución de probabilidad que se emplea en este capítulo es la distribución F, la cual debe su nombre a sir Ronald Fisher, uno de los pioneros de la estadística actual. Esta sirve como la distribución del estadístico de prueba en varias situaciones; con ella se pone a prueba si dos muestras provienen de poblaciones que tienen varianzas iguales, y también se aplica cuando se desea comparar varias medias poblacionales en forma simultánea; proceso que se denomina **análisis de la varianza (ANOVA).** En ambas situaciones, las poblaciones deben seguir una distribución normal, y los datos deben ser al menos de escala de intervalos.

¿Cuáles son las características de la distribución F?



1. **Existe una familia de distribuciones *F*.** Cada miembro se determina mediante dos parámetros: los grados de libertad del numerador y los grados de libertad del denominador. La forma de la distribución se ilustra en la gráfica de la derecha: hay una distribución F de la combinación de 29 grados de libertad del numerador (gl) y los 28 del denominador. Existe otra distribución F de los 19 grados en el numerador y los 6 del denominador. La distribución final que se muestra tiene 6 grados de libertad tanto en el numerador como en el denominador. Los grados de libertad se describen más adelante en este capítulo; observe que la forma de las curvas cambia cuando estos varían.
2. **La distribución *F* es continua.** Esto significa que supone un número infinito de valores entre cero y el infinito positivo.
3. **La distribución *F* no puede ser negativa.** El menor valor posible de F es 0.
4. **Tiene sesgo positivo.** La cola larga de la distribución es hacia el lado derecho. Cuando el número de grados de libertad aumenta, tanto en el numerador como en el denominador, la distribución se aproxima a ser normal.
5. **Es asintótica.** Cuando los valores de X aumentan, la curva F se aproxima al eje X pero nunca lo toca; este caso es similar al comportamiento de la distribución de probabilidad normal, descrito en el capítulo 7.

Comparación de dos varianzas poblacionales

La primera aplicación de la distribución F ocurre cuando se pone a prueba la hipótesis de que la varianza de una población normal es igual a la varianza de otra población normal. En los siguientes ejemplos se muestra el uso de la prueba:

- Una corporación de servicios de salud administra dos hospitales en Knoxville, Tennessee: St. Mary's North y St. Mary's South. En cada hospital, el tiempo medio de espera en la sala de emergencias es de 42 minutos; el administrador del hospital piensa que la sala de emergencias de St. Mary's North tiene más variación que la de St. Mary's South en el tiempo de espera.

- El índice de rendimiento medio de dos tipos de acciones comunes puede ser el mismo, pero quizá varíe más el índice de un tipo que el del otro. Una muestra de 10 acciones relacionadas con la tecnología y 10 acciones de compañías de servicios presentan el mismo índice de rendimiento medio, pero es probable que las acciones vinculadas a la tecnología varíen más.
- Un estudio del departamento de marketing de un periódico importante reveló que los hombres y las mujeres pasan casi la misma cantidad de tiempo por día navegando en internet; sin embargo, el mismo reporte indica que la variación del tiempo que los hombres pasaban por día casi duplicaba al de las mujeres.

La distribución F también sirve para probar suposiciones de algunas pruebas estadísticas. Recuerde que en el capítulo anterior se utilizó la prueba t para investigar si las medias de dos poblaciones independientes eran diferentes; para emplear esa prueba, algunas veces se supone que las varianzas de dos poblaciones normales son iguales. Vea la lista de suposiciones en la sección 11.4. La distribución F proporciona un medio para realizar una prueba considerando las varianzas de dos poblaciones normales.

Sin importar si se desea determinar si una población varía más que otra, o validar una suposición de una prueba estadística, primero se formula la hipótesis nula; la cual es que la varianza de una población normal, σ_1^2, es igual a la varianza de otra población normal, σ_2^2. La hipótesis alternativa podría ser que las varianzas difieren; en este caso, las hipótesis nula y alternativa son:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Para realizar la prueba, se selecciona una muestra aleatoria de n_1 observaciones de una población y una muestra aleatoria de n_2 observaciones de la segunda población. El estadístico de prueba se define como sigue.

ESTADÍSTICA DE PRUEBA PARA COMPARAR DOS VARIANZAS

$$F = \frac{s_1^2}{s_2^2} \quad \textbf{[12.1]}$$

Los términos s_1^2 y s_2^2 son las varianzas muestrales respectivas. Si la hipótesis nula es verdadera, el estadístico de prueba sigue la distribución F con $n_1 - 1$ y $n_2 - 1$ grados de libertad. Para reducir el tamaño de la tabla de valores críticos, la varianza *más grande* de la muestra se coloca en el numerador; de aquí, la razón F que se indica en la tabla siempre es mayor que 1.00; así, el valor crítico de la cola derecha es el único que se requiere. El valor crítico de F de una prueba de dos colas se determina dividiendo el nivel de significancia entre dos ($\alpha/2$) y después se consultan los grados de libertad apropiados en el apéndice B.6. Un ejemplo servirá de ilustración.

EJEMPLO

Lammers Limos ofrece servicio de transporte en limusina del ayuntamiento de Toledo, Ohio, al aeropuerto metropolitano de Detroit. Sean Lammers, presidente de la compañía, considera dos rutas: una por la carretera 25 y otra por la autopista I-75. Como Lammers desea estudiar el tiempo que tardaría en conducir al aeropuerto por cada camino y luego comparar los resultados, recopiló los siguientes datos muestrales, reportados en minutos. Usando el nivel de significancia 0.10, ¿hay alguna diferencia entre las variaciones de los tiempos de manejo por las dos rutas?

Carretera 25	Autopista I-75
52	59
67	60
56	61
45	51
70	56
54	63
64	57
	65

SOLUCIÓN

Los tiempos de manejo medios por ambos trayectos son casi iguales; el tiempo medio es de 58.29 minutos por la carretera 25 y de 59.0 minutos por la autopista I-75; sin embargo, al evaluar los tiempos de recorrido, Lammers también está interesado en la variación de ellos. El primer paso es calcular las dos varianzas muestrales; por lo tanto, se emplea la fórmula [3.11] para determinar la desviación estándar de cada muestra; para obtener la varianza muestral la desviación estándar se eleva al cuadrado.

Carretera 25

$$\bar{x} = \frac{\Sigma x}{n} = \frac{408}{7} = 58.29 \qquad s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{485.43}{7-1}} = 8.9947$$

Autopista I-75

$$\bar{x} = \frac{\Sigma x}{n} = \frac{472}{8} = 59.00 \qquad s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{134}{8-1}} = 4.3753$$

Según la medición de la desviación estándar, hay más variación en la carretera 25 que en la autopista I-75. Esto coincide con lo que Lammers sabe de las dos rutas; la de la carretera 25 tiene más semáforos, en tanto que la autopista I-75 es de acceso limitado; sin embargo, la trayectoria por la autopista I-75 tiene algunas millas más. Es importante que el servicio que ofrece sea tanto puntual como consistente, por lo que decide realizar una prueba estadística para determinar si en realidad existe una diferencia entre las variaciones de ambos caminos.

Se emplea el procedimiento de la prueba de hipótesis de seis pasos.

Paso 1: Se establecen las hipótesis nula y alternativa. La prueba es de dos colas debido a que se busca una diferencia entre las variaciones de las dos rutas. *No* se trata de demostrar que el tiempo que se emplea varía más por un camino que por el otro. Para este ejemplo, el subíndice 1 indica la información para la carretera US 25; el subíndice 2 indica la información para la autopista I-75.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Paso 2: Selecciona el nivel de significancia 0.10.

Paso 3: El estadístico de prueba apropiado sigue la distribución *F*.

Paso 4: Se obtiene el valor crítico del apéndice B.4, del cual se reproduce una parte en la tabla 12.1. Puesto que conduce una prueba de dos colas, el nivel de significancia en la tabla se establece en 0.05, determinado mediante $\alpha/2 = 0.10/2 = 0.05$. Hay $n_1 - 1 = 7 - 1 = 6$ grados de libertad en el numerador, y $n_2 - 1 = 8 - 1 = 7$ grados de libertad en el denominador. Para encontrar el valor crítico, recorra en forma horizontal la parte superior de la tabla *F* (tabla 12.1 o apéndice B.6) en el nivel de significancia 0.05 con 6 grados de libertad en el

TABLA 12.1 Valores críticos de la distribución F, $\alpha = 0.05$

Grados de libertad del denominador	Grados de libertad del numerador			
	5	6	7	8
1	230	234	237	239
2	19.3	19.3	19.4	19.4
3	9.01	8.94	8.89	8.85
4	6.26	6.16	6.09	6.04
5	5.05	4.95	4.88	4.82
6	4.39	4.28	4.21	4.15
7	3.97	3.87	3.79	3.73
8	3.69	3.58	3.50	3.44
9	3.48	3.37	3.29	3.23
10	3.33	3.22	3.14	3.07

numerador. Después vaya hacia abajo por esa columna hasta el valor crítico opuesto a 7 grados de libertad en el denominador; el cual es 3.87; por lo tanto, la regla de decisión es rechazar la hipótesis si la razón de las varianzas muestrales es mayor que 3.87.

Paso 5: Después calcule la razón de las dos varianzas muestrales, determine el valor del estadístico de prueba y tome una decisión respecto de la hipótesis nula. Observe que la fórmula [12.1] se refiere a las *varianzas* muestrales, pero se calcularon las *desviaciones estándar* de las muestras, las cuales se deben elevar al cuadrado para determinar las varianzas.

$$F = \frac{s_1^2}{s_2^2} = \frac{(8.9947)^2}{(4.3753)^2} = 4.23$$

La decisión es rechazar la hipótesis nula debido a que el valor F calculado (4.23) es mayor que el valor crítico (3.87).

Paso 6: Se concluye que hay una diferencia entre las variaciones de los tiempos de recorrido por las dos rutas. El señor Lammers querrá considerar esto en su programación.

La práctica habitual es determinar la razón F poniendo la mayor de las dos varianzas muestrales en el numerador; así, la razón F será al menos 1.00 y siempre será posible utilizar la cola derecha de la distribución F para que no se necesiten tablas F más extensas.

Surge una duda lógica: ¿es posible realizar pruebas de una cola? Por ejemplo, suponga que en el ejemplo anterior se sospecha que la varianza de los tiempos en la carretera 25, σ_1^2 es *mayor* que la varianza de los tiempos por la autopista I-75, σ_2^2. Las hipótesis nula y alternativa se deben plantear de la siguiente forma:

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

El estadístico de prueba se calcula como s_1^2/s_2^2. Observe que se designó *población 1* a la que se sospecha que tiene la varianza mayor; por lo tanto, s_1^2 aparece en el numerador. La razón F será mayor que 1.00, por lo que se puede utilizar la cola superior de la distribución F. Con estas condiciones no es necesario dividir el nivel de significancia a la mitad. Como en el apéndice B.6 solo se dan niveles de significancia de 0.05 y 0.10, estamos restringidos a estos niveles en el caso de pruebas de una cola y con 0.10 y 0.02 en el de pruebas de dos colas, a menos que se consulte una tabla más completa o se utilice software estadístico para calcular el estadístico F.

Excel tiene un procedimiento para realizar una prueba de varianzas; a continuación se presenta la salida en pantalla. El valor calculado de F es el mismo que se determinó con la fórmula [12.1].

Variance Test

	A	B	C	D	E	F	G
1	Carretera 25	A. interestatal 75		Prueba *F* para la varianza de dos muestras			
2	52	59			Carretera 25	A. interestatal 75	
3	67	60		Media	58.29	59.00	
4	56	61		Varianza	80.90	19.14	
5	45	51		Observaciones	7.00	8.00	
6	70	56		gl	6.00	7.00	
7	54	63		F	4.23		
8	64	57		P(F<=f) de una cola	0.04		
9		65		F crítica para prueba de una cola	3.87		
10							

AUTOEVALUACIÓN 12-1

Steele Electric Products, Inc. ensambla componentes eléctricos para teléfonos celulares. Durante los últimos 10 días, Mark Nagy ha promediado 9 productos rechazados, con una desviación estándar de 2 rechazos por día. Debbie Richmond promedió 8.5 productos rechazados, con una desviación estándar de 1.5 rechazos durante el mismo periodo. Con el nivel de significancia 0.05, ¿podría concluir que hay más variación en el número de productos rechazados por día de Mark?

EJERCICIOS

1. ¿Cuál es el valor crítico F de una muestra de seis observaciones en el numerador y cuatro en el denominador? Utilice una prueba de dos colas y el nivel de significancia 0.10.
2. ¿Cuál es el valor crítico F de una muestra de cuatro observaciones en el numerador y siete en el denominador? Utilice una prueba de una cola y el nivel de significancia 0.01.
3. Se dan las siguientes hipótesis.

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 \neq \sigma_2^2$$

   En una muestra aleatoria de ocho observaciones de la primera población resultó una desviación estándar de 10. Una muestra aleatoria de seis observaciones de la segunda población reveló una desviación estándar de 7. Con el nivel de significancia 0.02, ¿hay alguna diferencia entre las variaciones de cada población?
4. Se dan las siguientes hipótesis.

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

   En una muestra aleatoria de cinco observaciones de la primera población resultó una desviación estándar de 12. Una muestra aleatoria de siete observaciones de la segunda población reveló una desviación estándar de 7. Con el nivel de significancia 0.01, ¿varía más la primera población?
5. Arbitron Media Research, Inc., realiza un estudio sobre los hábitos de escuchar música en iPod de hombres y mujeres. Una parte del estudio incluyó el tiempo medio de escucha; se descubrió que el tiempo medio de escucha de los hombres era de 35 minutos por día. La desviación estándar de la muestra de los 10 hombres estudiados fue de 10 minutos por día. El tiempo medio de escucha de las 12 mujeres estudiadas también fue de 35 minutos, pero la desviación estándar muestral fue de 12 minutos. Con el nivel de significancia 0.10, ¿puede concluir que hay una diferencia entre las variaciones de los tiempos de escucha de los hombres y las mujeres?
6. Un corredor de bolsa de Critical Securities reportó que la tasa de rendimiento media de una muestra de 10 acciones de la industria petrolera era de 12.6%, con una desviación estándar de 3.9%. La tasa de rendimiento media de una muestra de 8 acciones de compañías de servicios fue de 10.9%, con una desviación estándar de 3.5%. Con el nivel de significancia 0.05, ¿puede concluir que varían más las acciones de la industria petrolera?

ANOVA: análisis de la varianza

OA12-2 Utilizar el enfoque ANOVA para probar que dos o más medias poblacionales son iguales.

La distribución F se utiliza para realizar una amplia variedad de pruebas de hipótesis; por ejemplo, cuando se prueba le igualdad de dos o más medias poblacionales, se utiliza la técnica del análisis de la varianza (ANOVA, *analysis of variance*) y el estadístico F se emplea como prueba.

Suposiciones en el análisis de la varianza (ANOVA)

El ANOVA para probar la igualdad de dos o más medias poblacionales requiere que los siguientes tres supuestos sean verdaderos:

1. Las poblaciones siguen la distribución normal.
2. Las poblaciones tienen desviaciones estándar iguales (σ).
3. Las poblaciones son independientes.

Cuando se cumplen estas condiciones, F se emplea como la distribución del estadístico de prueba.

¿Por qué es necesario estudiar el ANOVA? ¿Por qué no solo se emplea la prueba de las diferencias entre medias poblacionales, como se analizó en el capítulo anterior? Es posible comparar dos medias poblacionales a la vez. La razón más importante es la acumulación indeseable del error tipo I. Para ampliar la explicación, suponga cuatro métodos distintos (A, B, C y D) para capacitar personal para ser bomberos. La asignación de cada uno de los 40 prospectos del grupo de este año es aleatoria en cada método. Al final del programa de capacitación, a los cuatro grupos se les administra una prueba común para medir la comprensión de las técnicas contra incendios. La pregunta es: ¿existe una diferencia entre las calificaciones medias del examen de los cuatro grupos? La respuesta a esta pregunta permitirá comparar los cuatro métodos de capacitación.

Si emplea la distribución t para comparar las cuatro medias poblacionales, tendría que efectuar seis pruebas t distintas. Es decir, necesitaría comparar las calificaciones medias de los cuatro métodos como sigue: A frente a B, A frente a C, A frente a D, B frente a C, B frente a D y C frente a D.

Suponga que para cada prueba *t*, elegimos un $\alpha = 0.05$. Por lo tanto, la probabilidad de cometer el error tipo I, es decir, rechazar la hipótesis nula cuando es verdadera, es 0.05. El complemento es la probabilidad de 0.95 de no rechazar la hipótesis nula cuando es verdadera. Como se realizan seis pruebas separadas (independientes), la probabilidad de que *no* se tome una decisión incorrecta debido al error de muestreo en cualquiera de estas es:

$$P(\textit{Todas correctas}) = (0.95)(0.95)(0.95)(0.95)(0.95)(0.95) = 0.735$$

Para encontrar la probabilidad de al menos un error debido al muestreo, reste este resultado a 1; por lo tanto, la probabilidad de al menos una decisión incorrecta debida al muestreo es de $1 - 0.735 = 0.265$. El ANOVA permite comparar las medias poblacionales de forma simultánea con el nivel de significancia determinado; además, evita la acumulación del error tipo I asociado con probar muchas hipótesis.

El ANOVA se desarrolló inicialmente para aplicaciones en agricultura, y aún se emplean muchos de los términos relacionados con ese contexto; en particular, las diferentes poblaciones que se examinan se identifican con el término *tratamiento*; por ejemplo, se refiere a cómo se trató una extensión de terreno con un tipo particular de fertilizante. En la siguiente ilustración se aclara el término *tratamiento* y se muestra la aplicación de ANOVA.

EJEMPLO

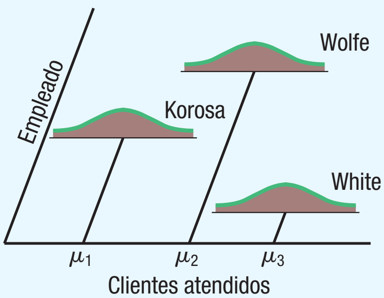
Joyce Kuhlman es gerente de un centro financiero regional y desea comparar la productividad, medida por el número de clientes atendidos, de tres empleados. Selecciona cuatro días en forma aleatoria y registra el número de clientes que atendió cada empleado; los resultados se muestran a la derecha.

Wolfe	White	Korosa
55	66	47
54	76	51
59	67	46
56	71	48

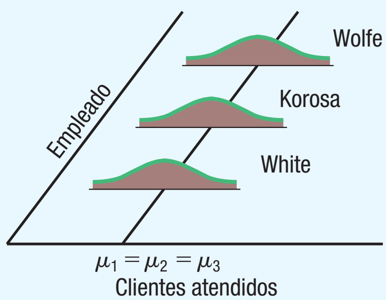
SOLUCIÓN

¿Hay alguna diferencia en el número medio de clientes atendidos? En la gráfica 12.1 se ilustra cómo pueden aparecer las poblaciones si hubiera una diferencia en las medias del tratamiento. Observe que las poblaciones siguen la distribución normal y la variación en cada población es la misma; sin embargo, las medias *no* son iguales.

Suponga que las poblaciones son iguales; es decir, que no hay una diferencia entre las medias (tratamiento). Esta igualdad, que se muestra en la gráfica 12.2, indica que las medias poblacionales son iguales. Observe de nuevo que las poblaciones siguen la distribución normal, y que la variación en cada una de las poblaciones es la misma.



GRÁFICA 12.1 Caso en el que las medias del tratamiento son diferentes



GRÁFICA 12.2 Caso en el que las medias del tratamiento son iguales

La prueba ANOVA

¿Cómo funciona la pruebaANOVA? Recuerde que se desea determinar si varias medias muestrales provienen de una sola población o de poblaciones con medias diferentes; en realidad, estas medias muestrales se comparan mediante sus varianzas. Para explicar esto, recuerde las suposiciones que requiere un ANOVA. Una de estas es que las desviaciones estándar de las diversas poblaciones normales deben ser las mismas; este requisito se aprovecha en la prueba ANOVA. La estrategia es estimar la varianza de la población (desviación estándar al cuadrado) de dos formas para después determinar la razón de dichas estimaciones. Si esta razón es aproximadamente 1 entonces, por lógica, las dos estimaciones son iguales, y se concluye que las medias poblacionales no lo son. La distribución *F* sirve como un árbitro para indicar en qué instancia la razón de las varianzas muestrales es mucho mayor que 1 para haber ocurrido por casualidad.

Consulte el ejemplo que se incluye en la sección anterior. El gerente desea determinar si hay una diferencia entre los números medios de clientes atendidos. Para iniciar, determine la media global de las 12 observaciones; la cual es de 58, calculada de $(55 + 54 + \cdots + 48)/12$. Después, en cada una de estas encuentre la diferencia entre el valor particular y la media global. Cada una de estas diferencias se eleva al cuadrado se suma. El término resultante se llama **variación total.**

VARIACIÓN TOTAL Suma de las diferencias entre cada observación y la media global elevadas al cuadrado.

En nuestro ejemplo, la variación total es de 1 082, determinada por $(55 - 58)^2 + (54 - 58)^2 + \cdots + (48 - 58)^2$.

Luego, esta se divide en dos partes: la que se debe a los **tratamientos** y la que es **aleatoria.** Para encontrarlas se determina la media de cada tratamiento (recuerde que la primera fuente de variación se debe a los tratamientos).

VARIACIÓN DE TRATAMIENTO Suma de las diferencias entre la media de cada tratamiento y la media total o global elevadas al cuadrado.

En el ejemplo, la variación debida a los tratamientos es la suma de las diferencias al cuadrado entre la media de cada empleado y la media global. Este término es 992; para calcularlo, primero se encuentra la media de cada uno de los tres tratamientos. La media de Wolfe es 56, determinada por $(55 + 54 + 59 + 56)/4$; las otras son 70 y 48, respectivamente. La suma de los cuadrados debida a los tratamientos es:

$$(56 - 58)^2 + (56 - 58)^2 + \cdots + (48 - 58)^2 = 4(56 - 58)^2 + 4(70 - 58)^2 + 4(48 - 58)^2 = 992$$

Si existe una variación considerable entre las medias de los tratamientos, es lógico que este término sea grande. Si las medias son similares, este término será un valor bajo; el menor posible es cero. Esto ocurrirá cuando todas las medias de los tratamientos sean iguales.

A la otra fuente de variación se le conoce como componente **aleatoria,** o componente de error.

VARIACIÓN ALEATORIA Suma de las diferencias entre cada observación y su media de tratamiento elevadas al cuadrado.

En el ejemplo, este término es la suma de las diferencias al cuadrado entre cada valor y la media de ese empleado en particular. La variación de error es 90.

$$(55 - 56)^2 + (54 - 56)^2 + \cdots + (48 - 48)^2 = 90$$

El estadístico de prueba, que es la razón de las dos estimaciones de la varianza poblacional, se determina a partir de la siguiente ecuación:

$$F = \frac{\text{Estimación de la varianza poblacional basada en las diferencias entre las medias muestrales}}{\text{Estimación de la varianza poblacional basada en la variación dentro de la muestra}}$$

La primera estimación de la varianza poblacional parte de los tratamientos, es decir, de la diferencia *entre* las medias. Esta es $992/2$. ¿Por qué se dividió entre 2? Recuerde del capítulo 3 que, para encontrar una varianza muestral (vea la fórmula [3.11]), se divide entre el número de observaciones menos uno; en este caso hay tres tratamientos, por lo que se divide entre 2. La primera estimación de la varianza poblacional es $992/2$.

La estimación de la varianza *dentro* de los tratamientos es la variación aleatoria divida entre el número total de observaciones menos el número de tratamiento; es decir 90/(12 − 3). De aquí, la segunda estimación de la varianza poblacional es 90/9. En realidad es una generalización de la fórmula [11.5], en la cual se agruparon las varianzas muestrales de dos poblaciones.

El paso final es tomar la razón de ambas estimaciones.

$$F = \frac{992/2}{90/9} = 49.6$$

Como esta razón es muy distinta a 1, se concluye que las medias de los tratamientos no son iguales. Hay una diferencia entre los números medios de clientes atendidos por cada uno de los tres empleados.

A continuación se presenta otro ejemplo, el cual abarca muestras de tamaños diferentes.

EJEMPLO

Desde hace algún tiempo las aerolíneas comenzaron a reducir sus servicios, como alimentos y bocadillos durante sus vuelos, y empezaron a cobrar por el equipaje. Un grupo de cuatro aerolíneas contrató a Brunner Marketing Research, Inc., para encuestar a sus pasajeros sobre el nivel de satisfacción en un vuelo reciente. La encuesta incluyó preguntas relacionadas con la adquisición de boletos, abordaje, servicio durante el vuelo, manejo del equipaje, comunicación del piloto, etcétera. Hicieron 25 preguntas con diversas respuestas posibles: excelente, bueno, regular o deficiente; una respuesta de excelente tiene una calificación de 4, bueno 3, regular 2 y deficiente 1. Estas respuestas se sumaron, de modo que la calificación final fue una indicación de la satisfacción con el vuelo. Entre mayor sea la calificación, mayor será el nivel de satisfacción con el servicio. La mayor calificación posible fue 100.

Brunner seleccionó y estudió al azar pasajeros de las cuatro aerolíneas. A continuación se muestra la información, ¿hay alguna diferencia entre los niveles de satisfacción medios con respecto a las cuatro aerolíneas? Use el nivel de significancia 0.01.

Northern	WTA	Pocono	Branson
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

SOLUCIÓN

Utilice el procedimiento de prueba de hipótesis de seis pasos.

Paso 1: **Se establecen las hipótesis nula y alternativa.** La hipótesis nula es que las calificaciones medias de las cuatro aerolíneas son iguales.

H_0: $\mu_N = \mu_w = \mu_p = \mu_B$

La hipótesis alternativa es que no todas las calificaciones medias son iguales.

H_1: No todas las calificaciones medias son iguales.

La hipótesis alternativa también se considera como "al menos dos calificaciones medias no son iguales".

Si no se rechaza la hipótesis nula, se concluye que no hay una diferencia entre las calificaciones medias de las cuatro aerolíneas. Si se rechaza H_0, se concluye que hay una diferencia en al menos un par de calificaciones medias, pero en este punto no se sabe cuál par o cuántos pares difieren.

Paso 2: **Se selecciona un nivel de significancia.** En este caso, 0.01.

Paso 3: **Se identifica el estadístico de prueba.** El estadístico de prueba sigue la distribución *F*.

Paso 4: **Se formula la regla de decisión.** Para determinar la regla de decisión, necesita el valor crítico; el del estadístico *F* aparece en el apéndice B.6. Los valores críticos del nivel de significancia 0.05 se encuentran en la primera página, y el nivel de significancia 0.01, en la segunda. Para utilizar esta tabla se deben conocer los grados de libertad del numerador y del denominador; los del numerador son iguales al número de tratamientos, designado k, menos 1, y los del denominador son el número total de observaciones, n, menos el número de tratamientos. En este ejemplo hay cuatro tratamientos y un total de 22 observaciones.

Grados de libertad del numerador $= k - 1 = 4 - 1 = 3$

Grados de libertad del denominador $= n - k = 22 - 4 = 18$

Consulte el apéndice B.6 y el nivel de significancia 0.01. Muévase horizontalmente por la parte superior de la página a tres grados de libertad del numerador; después vaya hacia abajo por esa columna hasta la fila con 18 grados de libertad. El valor en esta intersección es 5.09; por lo tanto, la regla de decisión es rechazar H_0 si el valor calculado de F es mayor que 5.09.

Paso 5: **Se toma una muestra, se realizan los cálculos y se decide.** Es conveniente que los cálculos del estadístico F se resuman en una **tabla ANOVA**; cuyo formato se muestra a continuación. En los paquetes de software estadístico también se emplea este formato.

Tabla ANOVA

Fuente de variación	Suma de cuadrados	gl	Media cuadrática	F
Tratamientos	SST	$k - 1$	SST/(k − 1) = MST	MST/MSE
Error	SSE	$n - k$	SSE/(n − k) = MSE	
Total	SS total	$n - 1$		

Hay tres valores, o suma de cuadrados, para calcular el estadístico de prueba F; estos se determinan al obtener SS total y SSE, después SST mediante una resta. El término SS total es la variación total, SST es la variación debida a los tratamientos, y SSE es la variación dentro de los tratamientos o el error aleatorio.

En general, el proceso se inicia al determinar SST total: la suma de las diferencias elevadas al cuadrado entre cada observación y la media global. La fórmula para determinar SS total es:

$$\text{SS total} = \Sigma(x - \bar{x}_G)^2 \qquad \textbf{[12.2]}$$

donde:

x es cada observación de la muestra;
$\bar{x}_G$ es la media global o total.

Enseguida se determina SSE o la suma de los errores elevados al cuadrado; es decir, la suma de las diferencias elevadas al cuadrado entre cada observación y su respectiva media de tratamiento. La fórmula para encontrar SSE es:

$$\text{SSE} = \Sigma(x - \bar{x}_c)^2 \qquad \textbf{[12.3]}$$

donde:

$\bar{x}_c$ es la media muestral del tratamiento c.

La SSE se calcula así:

$$\text{SSE} = \Sigma(x - \bar{x}_N)^2 + \Sigma(x - \bar{x}_w)^2 + \Sigma(x - \bar{x}_p)^2 + \Sigma(x - \bar{x}_B)^2$$

A la derecha se presentan los cálculos detallados de SS total y SSE de este ejemplo. Para determinar los valores (en este ejemplo hay 22) de SS total y SSE se comienza por calcular la media global o total; el cual es 1 664, por lo cual la media total es 75.64.

$$\bar{x}_G = \frac{1\,664}{22} = 75.64$$

Luego se encuentra la desviación de cada observación a la media total: se elevan al cuadrado estas desviaciones y se suma el resultado de las 22 observaciones; por ejemplo, el primer pasajero encuestado tenía una calificación de 94, y la media global o total es 75.64; por lo tanto, $(x - \bar{x}_G) = 94 - 75.64 = 18.36$. En

	Northern	WTA	Pocono	Branson	Total
	94	75	70	68	
	90	68	73	70	
	85	77	76	72	
	80	83	78	65	
		88	80	74	
			68	65	
			65		
Total de la columna	349	391	510	414	1 664
n	4	5	7	6	22
Media	87.25	78.20	72.86	69.00	75.64

Gran media

ESTADÍSTICA EN ACCIÓN

¿Alguna vez ha estado esperando que se desocupe un teléfono público y la persona que lo usa pareciera hablar sin parar? Existe evidencia de que la gente que habla por un teléfono público prolonga la conversación cuando alguien está esperando que lo desocupe. En una encuesta reciente en un centro comercial, los investigadores midieron el tiempo que 56 compradores pasaron hablando por teléfono cuando 1) estaban solos, 2) una persona estaba usando el teléfono de al lado y 3) una persona estaba usando un teléfono contiguo y alguien esperaba su turno. El estudio, que aplicó el ANOVA de una vía, demostró que el tiempo medio de uso del teléfono era significativamente menor cuando la persona estaba sola.

el caso del último pasajero, $(x - \bar{x}_G) = 65 - 75.64 = -10.64$. Los cálculos relativos a los otros pasajeros son:

Northern	WTA	Pocono	Branson
18.36	−0.64	−5.64	−7.64
14.36	−7.64	−2.64	−5.64
9.36	1.36	0.36	−3.64
4.36	7.36	2.36	−10.64
	12.36	4.36	−1.64
		−7.64	−10.64
		−10.64	

Después se eleva al cuadrado cada una de estas diferencias y se suman todos los valores; así, en el caso del primer pasajero:

$$(x - \bar{x}_G)^2 = (94 - 75.64)^2 = (18.36)^2 = 337.09$$

Por último, se suman todas las diferencias elevadas al cuadrado, como se indica en la fórmula [12.2]. El valor SS total es 1 485.10.

	Northern	WTA	Pocono	Branson	Total
	337.09	0.41	31.81	58.37	
	206.21	58.37	6.97	31.81	
	87.61	1.85	0.13	13.25	
	19.01	54.17	5.57	113.21	
		152.77	19.01	2.69	SS Total
			58.37	113.21	
			113.21		
Total	649.92	267.57	235.07	332.54	1 485.10

Para calcular el término SSE se encuentra la desviación entre cada observación y su media de tratamiento. En el ejemplo, la media del primer tratamiento (es decir, los pasajeros en Northern Airlines) es 87.25, determinada mediante $\bar{x}_N = 349/4$. El subíndice N se refiere a Northern Airlines.

El primer pasajero calificó a Northern con 94, por lo que $(x - \bar{x}_N) = (94 - 87.25) = 6.75$. El primer pasajero del grupo de TWA respondió con una calificación total de 75, por lo cual $(x - \bar{x}_w) = (75 - 78.20) = -3.2$. El detalle de todos los pasajeros es:

Northern	WTA	Pocono	Branson
6.75	−3.2	−2.86	−1
2.75	−10.2	0.14	1
−2.25	−1.2	3.14	3
−7.25	4.8	5.14	−4
	9.8	7.14	5
		−4.86	−4
		−7.86	

Todos estos valores (se muestran en la siguiente tabla) se elevan al cuadrado y después se suman.

	Northern	WTA	Pocono	Branson	Total
	45.5625	10.24	8.18	1	
	7.5625	104.04	0.02	1	
	5.0625	1.44	9.86	9	
	52.5625	23.04	26.42	16	
		96.04	50.98	25	
			23.62	16	SSE
			61.78		
Total	110.7500	234.80	180.86	68	594.41

Por lo tanto, el valor SSE es 594.41; es decir, $\Sigma(x - \bar{x}_c)^2 = 594.41$.

Por último, se determina SST (la suma de los cuadrados debida a los tratamientos) con la resta:

$$SST = SS\ Total - SSE \qquad \textbf{[12.4]}$$

En este ejemplo:

$$SST = SS\ Total - SSE = 1\,485.10 - 594.41 = 890.69$$

La tabla ANOVA se consulta para determinar el valor calculado de *F*. Los grados de libertad del numerador y del denominador son los mismos que en el paso 4 del ejemplo de la sección "La prueba ANOVA", donde se determinó el valor crítico de *F*. El término **media cuadrática** es otra expresión de la estimación de la varianza. La media cuadrática de tratamientos es SST dividido entre sus grados de libertad; el resultado es la **media cuadrática de tratamientos,** y se escribe MST. Calcule el **error cuadrático medio** de una manera similar; para ser precisos, divida SSE entre sus grados de libertad. Para completar el proceso y obtener *F*, divida MST entre MSE.

Inserte los valores particulares de *F* en una tabla ANOVA y calcule el valor de *F,* como se muestra a continuación.

Fuente de variación	Suma de cuadrados	*gl*	Media cuadrática	*F*
Tratamientos	890.69	3	296.90	8.99
Error	594.41	18	33.02	
Total	1 485.10	21		

El valor calculado de *F* es 8.99, mayor que el valor crítico de 5.09, por lo que la hipótesis nula se rechaza.

Paso 6: **Se interpreta el resultado.** La conclusión es que no todas las medias poblacionales son iguales; las calificaciones medias de las cuatro aerolíneas no son iguales. En este punto, los resultados del ANOVA solo muestran que al menos un par de calificaciones de satisfacción media no son las mismas entre las cuatro aerolíneas. No es posible mostrar estadísticamente qué aerolíneas difieren en satisfacción o cuáles tienen las mayores o menores calificaciones de satisfacción. Las técnicas para determinar cómo difieren las aerolíneas se presentan en la siguiente sección.

Los cálculos del ejemplo anterior se vuelven tediosos si la cantidad de observaciones en cada tratamiento es extensa; por fortuna, hay muchos paquetes de software para generar estos resultados. A continuación se utiliza Excel para calcular las estadísticas descriptivas y el ANOVA para el ejemplo anterior, con las calificaciones de aerolíneas y de pasajeros. Existen algunas diferencias sutiles entre la salida del software y los cálculos previos; estas se deben al redondeo.

Airline Anova

Northern	WTA	Pocono	Branson
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

ANOVA: un factor

Resumen

Grupos	Cuenta	Suma	Promedio	Varianza
Northern	4	349	87.250	36.917
WTA	5	391	78.200	58.700
Pocono	7	510	72.857	30.143
Branson	6	414	69.000	13.600

ANOVA

Fuente de variación	SS	df	MS	F	Valor P	F crítica
Entre grupos	890.684	3	296.895	8.99	0.0007	3.160
Dentro de grupos	594.407	18	33.023			
Total	1485.091	21				

Observe que en Excel se emplea el término "Entre grupos" para "Tratamientos", y "Dentro de grupos" para "Error"; sin embargo, tienen el mismo significado. El valor *p* es 0.0007; esta es la probabilidad de determinar un valor del estadístico de prueba de esta magnitud o mayor cuando la hipótesis nula es verdadera. En otras palabras, es la probabilidad de calcular un valor *F* mayor que 8.99 con 3 grados de libertad en el numerador y 18 grados de libertad en el denominador; por lo tanto, cuando se rechaza la hipótesis nula en este caso ¡hay una posibilidad muy remota de cometer un error tipo I!

AUTOEVALUACIÓN 12-2

Citrus Clean es un nuevo limpiador multiusos a prueba en el mercado, del cual se han colocado exhibidores en tres lugares distintos dentro de varios supermercados. A continuación se reporta la cantidad de botellas de 12 onzas que se vendieron en cada lugar del supermercado.

Cerca del pan	18	14	19	17
Cerca de la cerveza	12	18	10	16
Cerca de otros limpiadores	26	28	30	32

Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los promedios de botellas que se vendieron en los tres lugares?

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuál es la regla de decisión?
(c) Calcule los valores de SS total, SST y SSE.
(d) Elabore una tabla ANOVA.
(e) ¿Cuál es su decisión respecto de la hipótesis nula?

EJERCICIOS

7. La siguiente información es muestral; verifique la hipótesis de que las medias de tratamiento son iguales. Utilice el nivel de significancia 0.05.

Tratamiento 1	Tratamiento 2	Tratamiento 3
8	3	3
6	2	4
10	4	5
9	3	4

a. Formule las hipótesis nula y alternativa.
b. ¿Cuál es la regla de decisión?
c. Calcule los valores SST, SSE y SS total.
d. Elabore una tabla ANOVA.
e. Declare su decisión respecto de la hipótesis nula.

8. Se recolectaron las siguientes observaciones: 6 (tratamiento 1), 4 (tratamiento 2) y 5 (tratamiento 3). Con el nivel de significancia 0.05, verifique esta hipótesis: las medias de tratamiento son iguales.

Tratamiento 1	Tratamiento 2	Tratamiento 3
9	13	10
7	20	9
11	14	15
9	13	14
12		15
10		

a. Formule las hipótesis nula y alternativa.
b. ¿Cuál es la regla de decisión?
c. Calcule SST, SSE y SS total.
d. Elabore una tabla ANOVA.
e. Declare su decisión respecto de la hipótesis nula.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

9. Un inversionista en bienes raíces considera invertir en un centro comercial en los suburbios de Atlanta, Georgia, para lo cual evalúa tres terrenos. El ingreso familiar en el área circundante al centro comercial propuesto es un factor de particular importancia. Se selecciona una muestra aleatoria de cuatro familias cerca de cada centro comercial propuesto; a continuación se presentan los resultados de la muestra. Con el nivel de significancia 0.05, ¿el inversionista puede concluir que hay una diferencia entre los ingresos medios? Utilice el procedimiento de prueba de hipótesis habitual de cinco pasos.

Área de Southwyck (en miles de dólares)	Franklin Park (en miles de dólares)	Old Orchard (en miles de dólares)
64	74	75
68	71	80
70	69	76
60	70	78

10. La gerente de una compañía de software desea estudiar el número de horas que los directivos de diversas empresas utilizan sus computadoras de escritorio; el gerente seleccionó una muestra de cinco ejecutivos de cada una de tres industrias. Con el nivel de significancia 0.05, ¿puede la gerente concluir que hay una diferencia entre los promedios de horas por semana que se utilizan las computadoras en la industria?

Bancaria	Detallista	De seguros
12	8	10
10	8	8
10	6	6
12	8	8
10	10	10

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Inferencias sobre pares de medias de tratamiento

OA12-3
Utilizar los intervalos de confianza para probar e interpretar las diferencias entre pares de medias poblacionales.

Suponga que realiza el procedimiento ANOVA y toma la decisión de rechazar la hipótesis nula; esto permite concluir que no todas las medias de tratamiento son iguales. Algunas veces esta conclusión sería satisfactoria, pero otras se desea conocer cuáles medias de tratamiento difieren. En esta sección se proporcionan los detalles de prueba para saber cuáles medias de tratamiento difieren.

Recuerde que en el ejemplo respecto de las calificaciones que aplicaron los pasajeros de aerolíneas había una diferencia entre las medias de tratamiento; es decir, se rechazó la hipótesis nula y se aceptó la hipótesis alternativa. Si las calificaciones de los pasajeros no difieren, la pregunta es: ¿entre qué grupos difieren las medias de tratamiento?

Se dispone de varios procedimientos para responder; el más simple es emplear intervalos de confianza, es decir, la fórmula [9.2]. A partir del resultado del ejemplo anterior, observe que la calificación media muestral de los pasajeros del servicio de la aerolínea Northern es 87.25, mientras la de quienes califican el servicio de Branson es 69.00. ¿Existe suficiente disparidad para justificar la conclusión de que hay una diferencia significativa entre las calificaciones de satisfacción media de las dos aerolíneas?

La distribución t, descrita en los capítulos 10 y 11, sirve como base de esta prueba; recuerde que una de las suposiciones del ANOVA es que las varianzas poblacionales de todos los tratamientos son las mismas. Este valor común de la población es el **error cuadrático medio,** o MSE, y se determina mediante SSE$/(n - k)$. Un intervalo de confianza de la diferencia entre dos poblaciones se obtiene mediante:

INTERVALO DE CONFIANZA DE LA DIFERENCIA ENTRE LAS MEDIAS DE TRATAMIENTO

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \textbf{[12.5]}$$

donde:

- $\bar{x}_1$ es la media de la primera muestra;
- $\bar{x}_2$ es la media de la segunda muestra;
- t se obtiene del apéndice B.5. Los grados de libertad son iguales a $n - k$;
- MSE es el error cuadrático medio que se obtuvo de la tabla ANOVA [SSE$/(n - k)$];
- n_1 es el número de observaciones en la primera muestra;
- n_2 es el número de observaciones en la segunda muestra.

¿Cómo se determina si hay una diferencia entre las medias de tratamientos? Si el intervalo de confianza incluye cero, *no* existe diferencia entre ellas; por ejemplo, si el punto extremo izquierdo del intervalo de confianza tiene signo negativo y el punto extremo derecho tiene signo positivo, el intervalo incluye cero, y las dos medias no difieren; por lo tanto, si se desarrolla un intervalo de confianza a partir de la fórmula [12.5] y se tiene que la diferencia entre las medias muestrales fue 5.00, es decir, si $\bar{x}_1 - \bar{x}_2 = 5$ y $t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 12$, el intervalo de confianza estaría en un rango de 7.00 hasta 17.00. Para ponerlo en símbolos:

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 5.00 \pm 12.00 = -7.00 \text{ hasta } 17.00$$

Observe que en este intervalo se incluye el cero; por ello, se concluye que no hay una diferencia significativa entre las medias de tratamiento seleccionadas.

Por otro lado, si los puntos extremos del intervalo de confianza tienen el mismo signo, las medias de tratamiento difieren; por ejemplo, si $\bar{x}_1 - \bar{x}_2 = -0.35$ y $t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 0.25$, el intervalo de confianza variará de -0.60 hasta -0.10. Como -0.60 y -0.10 tienen el mismo signo, ambos negativos, cero no se encuentra en el intervalo y se concluye que estas medias de tratamiento difieren.

Use el ejemplo anterior de las aerolíneas para calcular el intervalo de confianza de la diferencia entre las calificaciones medias de los pasajeros de las aerolíneas Northern y Branson. Con un nivel de confianza de 95%, los puntos extremos del intervalo de confianza son 10.454 y 26.043.

$$(\bar{x}_N - \bar{x}_B) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (87.25 - 69.00) \pm 2.101\sqrt{33.023\left(\frac{1}{4} + \frac{1}{6}\right)}$$
$$= 18.25 \pm 7.793$$

donde:

$\bar{x}_N$ es 87.25;
$\bar{x}_B$ es 69.00;
t es 2.101: del apéndice B.5 con $(n - k) = 22 - 4 = 18$ grados de libertad;
MSE es 33.023: de la tabla ANOVA con $\text{SSE}/(n - k) = 594.4/18$;
N_N es 4;
N_B es 6.

El intervalo de confianza de 95% varía de 10.457 hasta 26.043. Los dos puntos extremos son positivos; de aquí se puede concluir que estas medias de tratamiento difieren de manera significativa; es decir, los pasajeros de Northern calificaron el servicio en gran medida diferente de los de Branson Airlines.

Las diferencias entre cada par de medias pueden obtenerse directamente con Minitab utilizando el ANOVA de una vía y el método de Fisher para comparar las medias; a continuación se muestra la salida de pantalla de Minitab.

```
Branson sustraído de:
           Bajo   Centro   Alto   -----+---------+---------+---------+----
Pocono    -2.860   3.857  10.574             (----*---)
WTA        1.889   9.200  16.511                (----*----)
Northern  10.457  18.250  26.043                      (----*----)
                                  -----+---------+---------+---------+----
                                     -15         0        15        30

Pocono sustraído de:
           Bajo   Centro   Alto   -----+---------+---------+---------+----
WTA       -1.726   5.343  12.412              (----*---)
Northern   6.826  14.393  21.960                    (----*----)
                                  -----+---------+---------+---------+----
                                     -15         0        15        30

WTA sustraído de:
          Bajo   Centro   Alto   -----+---------+---------+---------+----
Northern  0.951   9.050  17.149             (----*----)
                                 -----+---------+---------+---------+----
                                    -15         0        15        30
```

La salida de pantalla muestra los intervalos de confianza entre cada par de medias de tratamiento. Comenzando en la parte superior de la tabla, se muestran los intervalos de confianza para la diferencia entre la media de tratamiento de Branson y de cada una de las otras medias de tratamiento. A la izquierda, cada fila muestra el centro o diferencia entre las medias de tratamiento y los extremos inferior y superior del intervalo de confianza. Nuevamente, el cero no está presente en el intervalo, por lo que ambas medias de tratamiento son significativamente distintas. Observe que el intervalo de confianza para la diferencia entre Branson y Northern es la misma que la obtenida con los resultados más arriba; ambas medias de tratamiento son significativamente distintas.

A la derecha de la salida de la pantalla hay una gráfica en la cual se muestra el intervalo de confianza para la diferencia entre cada par de medias de tratamiento. El asterisco (*) indica la ubicación central de la diferencia entre las medias de tratamiento; los paréntesis que abren y cierran muestran los puntos extremos del intervalo. Utilizando la gráfica se puede ver fácilmente qué intervalos no incluyen el cero e indicar las medias de tratamiento que son significativamente distintas. La media de tratamiento de Branson es diferente de las medias de TWA y de Northern; la media de tratamiento de Pocono es distinta de la de Northern, y la de TWA es distinta de la de Northern.

Se debe destacar que esta investigación es un proceso que avanza por pasos; el primero es realizar la prueba ANOVA. Solo si se rechaza la hipótesis nula de que las medias de tratamiento son iguales se deberán analizar las medias de tratamiento individuales.

AUTOEVALUACIÓN 12-3

En los siguientes datos se registran las colegiaturas por semestre (en miles de dólares) de una muestra de cinco universidades privadas en la región noreste de Estados Unidos, cuatro en la región sureste y cinco en la región oeste. Con el nivel de significancia 0.05, ¿se puede concluir que hay una diferencia entre las colegiaturas medias de las diversas regiones?

Noreste (en miles de dólares)	Suroeste (en miles de dólares)	Oeste (en miles de dólares)
10	8	7
11	9	8
12	10	6
10	8	7
12		6

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuál es la regla de decisión?
(c) Elabore una tabla ANOVA. ¿Cuál es el valor del estadístico de prueba?
(d) ¿Cuál es su decisión respecto de la hipótesis nula?
(e) ¿Puede existir una diferencia significativa entre la colegiatura media en el noreste en comparación con la del oeste? Si la hay, desarrolle el intervalo de confianza de 95% de esa diferencia.

EJERCICIOS

11. Se recolectaron las siguientes observaciones: 3 (tratamiento 1), 5 (tratamiento 2) y 4 (tratamiento 3). Con el nivel de significancia 0.05, verifique esta hipótesis: las medias de tratamiento son iguales.

Tratamiento 1	Tratamiento 2	Tratamiento 3
8	3	3
11	2	4
10	1	5
	3	4
	2	

a. Formule las hipótesis nula y alternativa.
b. ¿Cuál es la regla de decisión?
c. Calcule SST, SSE y SS total.
d. Elabore una tabla ANOVA.
e. Declare su decisión respecto de la hipótesis nula.
f. Si se rechaza H_0, ¿puede concluir que los tratamientos 1 y 2 difieren? Utilice el nivel de confianza de 95%.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

12. Se recolectaron las siguientes observaciones: 6 (tratamiento 1), 10 (tratamiento 2) y 8 (tratamiento 3). Con el nivel de significancia 0.05, verifique esta hipótesis: las medias de tratamiento son iguales.

- **a.** Formule las hipótesis nula y alternativa.
- **b.** ¿Cuál es la regla de decisión?
- **c.** Calcule SST, SSE y SS total.
- **d.** Elabore una tabla ANOVA.
- **e.** Declare su decisión respecto de la hipótesis nula.
- **f.** Si rechaza H_0, ¿puede concluir que los tratamientos 2 y 3 difieren? Utilice el nivel de confianza de 95%.

Tratamiento 1	Tratamiento 2	Tratamiento 3
3	9	6
2	6	3
5	5	5
1	6	5
3	8	5
1	5	4
	4	1
	7	5
	6	
	4	

13. Una alumna en su último año en la carrera de contabilidad de la Midsouth State University tiene ofertas de trabajo de cuatro empresas de contabilidad pública; para estudiarlas a fondo, preguntó a una muestra de personas recién capacitadas cuántos meses trabajó cada una en la empresa antes de recibir un aumento salarial. La información muestral se corrió en Minitab con los siguientes resultados:

Análisis de varianza

Fuente	*gl*	SS	MS	F	P
Factor	3	32.33	10.78	2.36	0.133
Error	10	45.67	4.57		
Total	13	78.00			

Con el nivel de significancia 0.05, ¿hay una diferencia entre los números medios de meses antes que las empresas de contabilidad otorgaran un aumento a sus empleados?

14. Un analista de la bolsa de valores desea determinar si hay una diferencia entre las tasas de rendimiento medias de tres tipos de acciones: de compañías de servicios, detallistas y bancarias. Obtuvo los siguientes resultados:

```
Fuente              SS      MS       F       P
Factor      2    81.11   40.56   15.61   0.000
Error      14    36.37    2.60
Total      16   117.48

S = 1.612    R-Sq = 69.04%    R-Sq(adj) = 64.62%

Nivel              N    Media  Desviación estándar
Servicios          5   16.520   1.611
Detallistas        6   11.492   0.378
Bancarias          6   15.547   2.249

Acciones de servicios sustraídas de:
                   Inferior   Centro   Superior
Detallistas        -7.121   -5.028   -2.935
Bancarias          -3.066   -0.972    1.121

                       +---------+---------+---------+-------
Detallistas            (-----*-----)
Bancarias                           (-----*-----)
                       +---------+---------+---------+-------
                   -7.0      -3.5       0.0       3.5

Acciones detallistas sustraídas de:
                  Inferior   Centro  Superior
Bancarias          2.060    4.056   6.051

                       +---------+---------+---------+-------
Bancarias                                          (-----*----)
                       +---------+---------+---------+-------
                   -7.0      -3.5       0.0       3.5
```

a. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre las tasas de recuperación media de los tres tipos de acciones?
b. ¿Puede el analista concluir que hay una diferencia entre las tasas medias de rendimiento de las acciones de servicios y de detallistas? ¿De servicios y bancarias? ¿De bancarias y detallistas? Explique.

Análisis de la varianza de dos vías

OA12-4 Utilizar una variable de bloqueo en una prueba ANOVA de dos vías para probar una hipótesis de igualdad entre tres o más medias poblacionales.

En el ejemplo de la sección anterior, la variación total se dividió en dos categorías: entre los tratamientos y dentro de los tratamientos (a esta se le llamó error o variación aleatoria); en otras palabras, solo se consideraron dos fuentes de variación, la debida a los tratamientos y la debida a las diferencias aleatorias. En el ejemplo de las calificaciones de los pasajeros de aerolíneas puede haber otras causas de variación; tales factores pueden incluir, por ejemplo, la estación del año, el aeropuerto o el número de pasajeros en el vuelo.

El beneficio al considerar otros factores es que se reduce la varianza del error; es decir, si se reduce el denominador del estadístico *F* (al reducir la varianza del error o, de manera más directa, el término SSE), el valor de *F* será mayor, lo que ocasionará el rechazo de la hipótesis del tratamiento de medias iguales. En otras palabras, si se puede explicar más la variación, habrá menos "error". Un ejemplo aclarará la reducción de la varianza del error.

EJEMPLO

El director de Warren Area Transit Authority (WARTA) está pensando en ampliar el servicio de autobuses del suburbio de Starbrick al distrito comercial central de Warren. Se consideran cuatro rutas de Starbrick al centro de Warren: 1) por la Carretera 6, 2) por el West End, 3) por Hickory Street Bridge y 4) por la Ruta 59. El director realizó varias pruebas para determinar si había una diferencia entre los tiempos de recorrido medio por los cuatro trayectos; como habrá muchos conductores distintos, la prueba se diseñó para que cada conductor manejara a lo largo de todos ellos. A continuación se presenta el tiempo del recorrido, en minutos, de cada combinación conductor-ruta.

	Tiempo de recorrido de Starbrick a Warren (minutos)			
Conductor	**Carretera 6**	**West End**	**Hickory St.**	**Ruta 59**
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los tiempos de recorrido medios a lo largo de los cuatro caminos? Si elimina el efecto de los conductores, ¿hay alguna diferencia entre los tiempos de recorrido medios?

SOLUCIÓN

Para iniciar, realice una prueba de hipótesis con una prueba ANOVA de una vía; es decir, solo considere los cuatro trayectos. Con esta condición, la variación entre los tiempos del recorrido se debe a los tratamientos o es aleatoria. En este ejemplo, los subíndices corresponden a los tratamientos o rutas: 1 para la carretera 6, 2 para la West End, 3 para Hickory Street y 4 para la ruta 59. La hipótesis nula y la alternativa para comparar los tiempos de recorrido medios por los cuatro recorridos son:

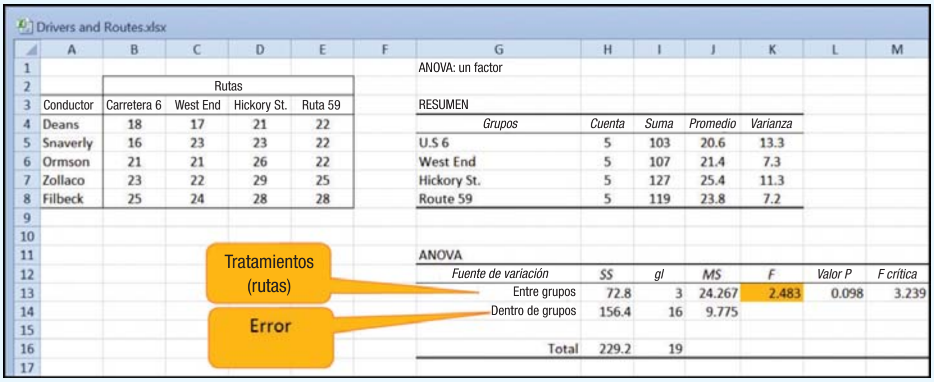
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: No todas las medias de tratamiento son iguales.

Hay cuatro alternativas, por lo cual los grados de libertad del numerador son $k - 1 = 4 - 1 = 3$. Hay 20 observaciones; por consiguiente, los grados de libertad del denominador son $n - k = 20 - 4 = 16$.

Del apéndice B.6, con el nivel de significancia 0.05, el valor crítico de F es 3.24. La regla de decisión es rechazar la hipótesis nula si el valor calculado de F es may or que 3.24.

Para realizar los cálculos emplee Excel. El valor calculado de F es 2.483, por lo que la decisión es no rechazar la hipótesis nula; en conclusión, no hay una diferencia entre los tiempos de recorrido medios a lo largo de las cuatro rutas; no hay una razón para seleccionar una de ellas como más rápida que las demás.



Drivers and Routes.xlsx

	Rutas			
Conductor	Carretera 6	West End	Hickory St.	Ruta 59
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

ANOVA: un factor

RESUMEN

Grupos	Cuenta	Suma	Promedio	Varianza
U.S 6	5	103	20.6	13.3
West End	5	107	21.4	7.3
Hickory St.	5	127	25.4	11.3
Route 59	5	119	23.8	7.2

ANOVA

Fuente de variación	SS	gl	MS	F	Valor P	F crítica
Entre grupos	72.8	3	24.267	2.483	0.098	3.239
Dentro de grupos	156.4	16	9.775			
Total	229.2	19				

De la salida de Excel anterior, los tiempos de recorrido medios a lo largo de los caminos fueron: 20.6 minutos por la Carretera 6, 21.4 minutos por West End, 25.4 minutos por Hickory Street, y 23.8 minutos por la Ruta 59; por lo tanto, es razonable atribuir estas diferencias a la casualidad. De la tabla ANOVA se observa que: SST es 72.8, SSE es 156.4 y SS total es 229.2.

En este ejemplo se consideró la variación debida a los tratamientos (rutas) y se tomó toda variación restante como aleatoria. Si se pudiera considerar el efecto de los diversos conductores, se podría reducir el término SSE, lo cual generaría un valor mayor de F. A la segunda variable de tratamiento, en este caso los conductores, se le conoce como **variable de bloqueo.**

VARIABLE DE BLOQUEO Una segunda variable de tratamiento que, cuando se incluye en el ANOVA, tendrá el efecto de reducir el término SSE.

En este caso, los conductores son la variable de bloqueo, y la razón F de la variable de tratamiento cambia al eliminar el efecto de los conductores del término SSE. Primero, es necesario determinar la suma de los cuadrados debida a los bloques.

En una prueba ANOVA de dos vías, la suma de los cuadrados debida a los bloques se determina mediante la siguiente fórmula.

$$\text{SSB} = k\sum(\overline{x}_D - \overline{x}_G)^2 \tag{12.6}$$

donde:

- k es el número de tratamientos;
- b es el número de bloqueos;
- $\overline{x}_b$ es la media muestral del bloque b;
- $\overline{x}_G$ es la media global o total.

A partir de los cálculos siguientes, se determina que las medias de los conductores respectivos son 19.5 minutos, 21 minutos, 22.5 minutos y 26.25 minutos. La media global es 22.8 minutos,

Tiempo de recorrido de Starbrick a Warren (en minutos)

Conductor	Carretera 6	West End	Hickory St.	Ruta 59	Sumas de los conductores	Medias de los conductores
Deans	18	17	21	22	78	19.50
Snaverly	16	23	23	22	84	21.00
Ormson	21	21	26	22	90	22.50
Zollaco	23	22	29	25	99	24.75
Filbeck	25	24	28	28	105	26.25

determinada por la suma del tiempo de recorrido de los 20 conductores (456 minutos) y su división entre 20.

Al sustituir esta información en la fórmula [12.6], se determina SSB, y la suma de los cuadrados debida a los conductores (la variable de bloqueo) es 119.7.

$$\begin{aligned} \text{SSB} &= k\sum(\bar{x}_b - \bar{x}_G)^2 \\ &= 4(19.5 - 22.8)^2 + 4(21.0 - 22.8)^2 + 4(22.5 - 22.8)^2 \\ &\quad + 4(24.75 - 22.8)^2 + 4(26.25 - 22.8)^2 \\ &= 119.7 \end{aligned}$$

Se utiliza el mismo formato en la tabla ANOVA de dos vías, como en el caso de una vía, excepto que hay una fila adicional para la variable de bloqueo; SS total y SST se calculan como se hizo antes, y SSB se determina con la fórmula [12.6]. El término SSE se calcula mediante una resta.

SUMA DE ERRORES CUADRÁTICOS, DOS VÍAS $\text{SSE} = \text{SS total} - \text{SST} - \text{SSB}$ **[12.7]**

Los valores de los diversos componentes de la tabla ANOVA se calculan de la siguiente manera.

Fuente de variación	Suma de los cuadrados	*gl*	Media cuadrática	*F*
Tratamientos	SST	$k - 1$	$\text{SST}/(k-1) = \text{MST}$	MST/MSE
Bloqueo	SSB	$b - 1$	$\text{SSB}/(b-1) = \text{MSB}$	MSB/MSE
Error	SSE	$(k-1)(b-1)$	$\text{SSE}/(k-1)(b-1) = \text{MSE}$	
Total	SS total	$n - 1$		

SSE se obtiene con la fórmula [12.7].

$$\text{SSE} = \text{SS total} - \text{SST} - \text{SSB} = 229.2 - 72.8 - 119.7 = 36.7$$

Fuente de variación	(1) Suma de los cuadrados	(2) Grados de libertad	(3) Media cuadrática (1)/(2)
Tratamientos	72.8	3	24.27
Bloques	119.7	4	29.93
Error	36.7	12	3.06
Total	229.2	19	

En este punto hay un desacuerdo: si el objetivo de la variable de bloqueo (los conductores en este ejemplo) fue solo reducir la variación del error, no se debe realizar una prueba de hipótesis de las diferencias entre las medias de los bloques. Es decir, si el objetivo era reducir el término MSE, no se debe probar una hipótesis respecto de la variable de bloqueo; por otro lado, quizá se desee dar a los bloques la misma condición que a los tratamientos y realizar una prueba de hipótesis. Este último caso, cuando los bloques son lo bastante importantes para considerarse un segundo factor, se conoce como **experimento de dos factores.** En muchos casos, la decisión no es clara; en este ejemplo lo importante es la diferencia entre los tiempos de recorrido de los diversos conductores, por lo que se realizará la prueba de hipótesis. Los dos conjuntos de hipótesis son:

1. H_0: Las medias de tratamiento son iguales ($\mu_1 = \mu_2 = \mu_3 = \mu_4$).
   H_1: Las medias de tratamiento no son iguales.
2. H_0: Las medias de los bloques son iguales ($\mu_D = \mu_S = \mu_O = \mu_F$).
   H_1: Los medias de los bloques no son iguales.

Primero se pondrá a prueba la hipótesis respecto de las medias de tratamiento. Hay $k - 1 = 4 - 1 = 3$ grados de libertad en el numerador y $(b - 1)(k - 1) = (5 - 1)(4 - 1) = 12$ grados de libertad en el denominador. Con el nivel de significancia 0.05, el valor crítico de F es 3.49. La hipótesis nula de que los tiempos medios para las cuatro rutas son iguales se rechaza si la razón F es mayor que 3.49.

$$F = \frac{\text{MST}}{\text{MSE}} = \frac{24.27}{3.06} = 7.93$$

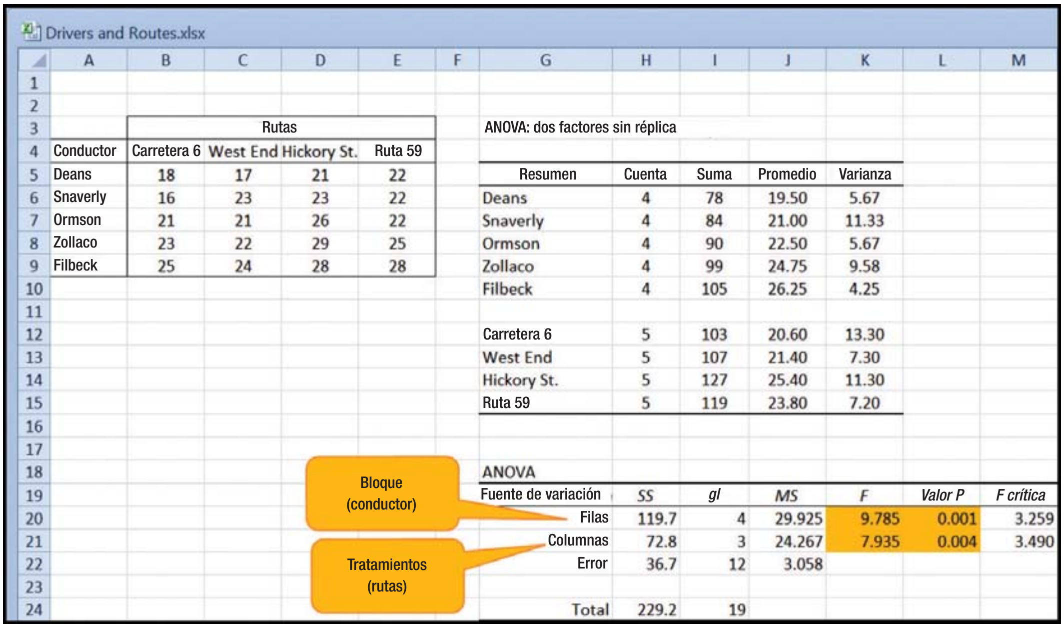
La hipótesis nula se rechaza y la hipótesis alternativa se acepta; en conclusión, el tiempo de recorrido medio no es el mismo para todos los trayectos. Sería recomendable que WARTA realizara algunas pruebas para determinar cuáles medias de tratamiento difieren.

Enseguida se prueba si el tiempo de recorrido es el mismo para los diversos conductores. Los grados de libertad en el numerador para los bloques son $b - 1 = 5 - 1 = 4$; los grados de libertad para el denominador son los mismos que antes: $(b - 1)(k - 1) = (5 - 1)(4 - 1) = 12$. La hipótesis nula de que las medias de los bloques son iguales se rechaza si la razón F es mayor que 3.26.

$$F = \frac{\text{MSB}}{\text{MSE}} = \frac{29.93}{3.06} = 9.78$$

Se rechaza la hipótesis nula y la hipótesis alternativa se acepta. El tiempo medio no es el mismo para los conductores; así, la gerencia de WARTA puede concluir, con base en los resultados de la muestra, que hay una diferencia en las rutas y en los conductores.

La hoja de cálculo de Excel tiene un procedimiento ANOVA de dos factores. A continuación se presenta la salida del ejemplo WARTA recién terminado; esta salida también incluye los valores p (el de la hipótesis nula respecto de los conductores es 0.001, y el de los trayectos, 0.004); los cuales confirman que las hipótesis nula de tratamientos y bloques se deberán rechazar debido a que el valor p es menor que el nivel de significancia.



	Rutas			
Conductor	Carretera 6	West End	Hickory St.	Ruta 59
Deans	18	17	21	22
Snaverly	16	23	23	22
Ormson	21	21	26	22
Zollaco	23	22	29	25
Filbeck	25	24	28	28

ANOVA: dos factores sin réplica

Resumen	Cuenta	Suma	Promedio	Varianza
Deans	4	78	19.50	5.67
Snaverly	4	84	21.00	11.33
Ormson	4	90	22.50	5.67
Zollaco	4	99	24.75	9.58
Filbeck	4	105	26.25	4.25
Carretera 6	5	103	20.60	13.30
West End	5	107	21.40	7.30
Hickory St.	5	127	25.40	11.30
Ruta 59	5	119	23.80	7.20

ANOVA

Fuente de variación	SS	gl	MS	F	Valor P	F crítica
Filas	119.7	4	29.925	9.785	0.001	3.259
Columnas	72.8	3	24.267	7.935	0.004	3.490
Error	36.7	12	3.058			
Total	229.2	19				

AUTOEVALUACIÓN 12-4

Rudduck Shampoo vende tres tipos de champú: para cabello seco, normal y graso. En la tabla siguiente se presentan las ventas, en millones de dólares, de los últimos cinco meses. Con el nivel de significancia 0.05, compruebe si las ventas medias difieren entre los tres tipos de champú o según el mes.

	Venta (millones de dólares)		
Mes	**Seco**	**Normal**	**Graso**
Junio	7	9	12
Julio	11	12	14
Agosto	13	11	8
Septiembre	8	9	7
Octubre	9	10	13

EJERCICIOS

En los ejercicios 15 y 16 realice una prueba de hipótesis para determinar si difieren las medias de bloqueo o de tratamiento. Con el nivel de significancia 0.05: a) formule las hipótesis nula y alternativa para los tratamientos, b) establezca la regla de decisión para los tratamientos y c) formule las hipótesis nula y alternativa para los bloques. También establezca la regla de decisión para los bloques, entonces d) calcule SST, SSB, SS total y SSE, e) elabore una tabla ANOVA y f) exponga su decisión respecto de los dos conjuntos de hipótesis.

15. Los siguientes datos corresponden a un ANOVA de dos factores con dos tratamientos y tres bloques.

	Tratamiento	
Bloque	**1**	**2**
A	46	31
B	37	26
C	44	35

16. Los siguientes datos corresponden a un ANOVA de dos factores con tres tratamientos y tres bloques.

	Tratamiento		
Bloque	**1**	**2**	**3**
A	12	14	8
B	9	11	9
C	7	8	8

17. Chapin Manufacturing Company opera 24 horas al día, cinco días a la semana y los trabajadores alternan turnos cada semana. La gerencia desea saber si hay una diferencia entre los números de unidades producidas por los empleados que trabajan en diversos turnos; se selecciona una muestra de cinco trabajadores y se registran las unidades producidas en cada turno. Con el nivel de significancia 0.05, ¿puede concluir que hay una diferencia entre las tasas de producción medias por turno o por empleado?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Unidades producidas		
Empleado	**Matutino**	**Vespertino**	**Nocturno**
Skaff	31	25	35
Lum	33	26	33
Clark	28	24	30
Treece	30	29	28
Morgan	28	26	27

18. En el área de Tulsa, Oklahoma, hay tres hospitales; en los siguientes datos se muestra el número de cirugías realizadas a pacientes externos el lunes, martes, miércoles, jueves y viernes durante la semana pasada. Con el nivel de significancia 0.05, ¿puede concluir que hay una diferencia entre los números medios de cirugías realizadas por cada hospital o por día de la semana?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Número de cirugías realizadas		
Día	**St. Luke's**	**St. Vincent**	**Mercy**
Lunes	14	18	24
Martes	20	24	14
Miércoles	16	22	14
Jueves	18	20	22
Viernes	20	28	24

ANOVA de dos vías con interacción

OA12-5 Realizar una prueba ANOVA de dos vías con interacción y describir los resultados.

En la sección anterior se estudiaron los efectos separados o independientes de dos variables o factores, con una variable de respuesta: el tiempo de recorrido. En el ejemplo, los dos factores fueron las rutas de autobuses y los conductores, y la respuesta fue el tiempo de recorrido. El análisis muestra dos resultados significativos: primero, los tiempos de recorrido medios entre trayectos, promediados con respecto a todos los conductores, son distintos; segundo, los tiempos de recorrido medios entre los cinco conductores, promediados con respecto a todas las rutas, son distintos. ¿Qué pueden explicar estas diferencias? Las diferencias entre recorridos pueden estar relacionadas simplemente a las distintas distancias de las mismas; en realidad, no se estudió la distancia de las rutas. Tal vez las diferencias se explican por la velocidad, en promedio, a la que manejan los conductores, independientemente del camino.

Existe otro efecto que no se ha considerado; se llama **efecto de interacción** entre la ruta y el conductor sobre el tiempo de recorrido. Esto es, las diferencias en el tiempo de recorrido pueden depender tanto del conductor como del trayecto; por ejemplo, es posible que uno de los conductores sea especialmente bueno conduciendo por una o más de las rutas. Tal vez un conductor sabe cronometrar con eficacia los semáforos o cómo evitar intersecciones muy congestionadas en uno o más recorridos. En este caso, el efecto combinado del conductor y la ruta también explica las diferencias entre los tiempos de recorrido medios; los resultados de la interacción entre el conductor y la ruta pueden proporcionar resultados interesantes.

INTERACCIÓN El efecto de un factor sobre una variable de respuesta difiere según el valor de otro factor.

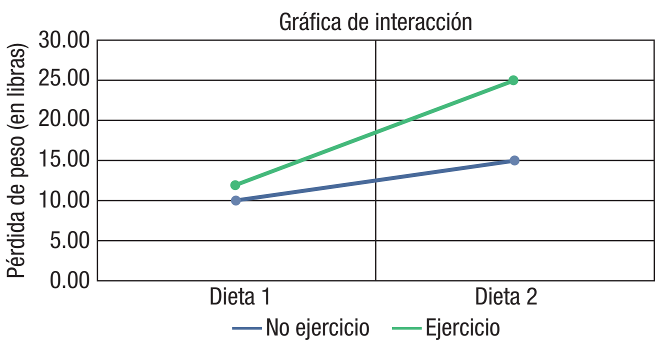
Gráficas de interacción

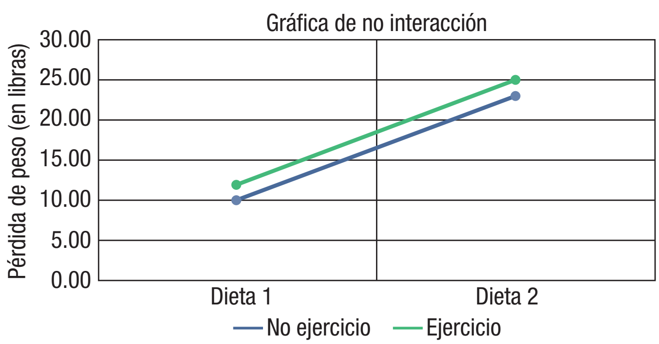
Un ejemplo cotidiano de interacción es el efecto de la dieta y el ejercicio en el peso corporal. Generalmente se acepta que el peso corporal (la variable de respuesta) pude verse afectado por dos factores: dieta y ejercicio; sin embargo, las investigaciones muestran que también existe un efecto combinado, o de *interacción,* entre la dieta y el ejercicio en la pérdida de peso; esto es, la cantidad de peso perdido será diferente y dependerá de la dieta *y* de que la gente haga ejercicio.

En la gráfica izquierda inferior de interacción se ilustra precisamente la interacción entre dieta y ejercicio. Primero se grafican mediante puntos, las pérdidas de peso medias para la dieta 1 y la dieta 2 para personas que no hacen ejercicio. Estos puntos están conectados con la línea azul y es clara la diferencia en la pérdida de peso para cada dieta. En segundo lugar, se grafican las pérdidas medias de peso para la dieta 1 y la dieta 2 para personas que hacen ejercicio. Estos puntos están conectados con la línea verde y, de nuevo, se aprecia una clara diferencia de pérdida de peso entre ambas dietas para la gente que hace ejercicio; en la gráfica también se muestra el efecto de interacción entre la dieta y el ejercicio en la pérdida de peso. Observe que las dos líneas no son paralelas. Para la dieta 1, la pérdida media de peso es mayor cuando la gente también hace ejercicio. Para la dieta 2, la pérdida media de peso también es mayor cuando la gente se ejercita, pero la reducción de peso es mucho mayor que para la dieta 1. Así que, ¿cuál es el efecto de la dieta y el ejercicio en la pérdida de peso? Depende de los efectos combinados, o de interacción, entre dieta y ejercicio.

¿Cómo se vería la gráfica de interacción si no existiera interacción? En la gráfica derecha inferior se muestra un análisis de dieta y ejercicio sin interacción.

En este caso, las líneas son paralelas; al comparar las medias, el efecto del ejercicio en la pérdida de peso para la dieta 1 y la dieta 2 es el mismo. La pérdida de peso estimada es de casi dos libras. Adicionalmente, el efecto de la dieta es el mismo ya sea que la gente se ejercite o no (aproximadamente 13 libras).





Prueba de interacción

Para probar el efecto de interacción se utiliza un ANOVA de dos vías con interacción. Para ilustrar, se retoma el ejemplo anterior referente a WARTA, pero replanteando el problema de factores en la administración; en esta ocasión, desean expandir su servicio de autobuses del centro de Warren hasta Starbrick. Hasta ahora, basándose en análisis estadístico, han concluido que existe una diferencia en el tiempo medio de recorrido a lo largo de las cuatro rutas propuestas, y en el tiempo medio de recorrido de los cinco conductores. Pero es posible que la combinación, o interacción, entre trayectos y conductores tenga un efecto significativo en el tiempo medio de recorrido.

En este análisis a ambas variables (ruta y conductor) se les denomina **factores**, mientras que "tiempo de recorrido" es la variable de **respuesta**. Para comprobar la interacción, los datos de la muestra deben replicarse en cada trayecto; en este caso, cada conductor recorre cada ruta tres veces, así que hay tres tiempos observados para cada combinación entre ruta y conductor. Esta información se resume en la siguiente hoja de cálculo de Excel.

	Rutas				Medias para los conductores	
	Carretera 6	West End	Hickory St	Ruta 59		
Deans	18	17	21	22		
	15	14	20	19		
	21	20	22	25	19.50	
Snaverly	16	23	23	22		
	19	19	24	20		
	13	25	22	24	20.83	
Ormson	21	21	26	22		
	19	23	24	24		
	14	25	28	20	22.25	
Zollaco	23	22	29	25		
	21	24	30	20		
	25	20	28	26	24.42	
Filbeck	25	24	28	28		
	24	25	29	30		Gran media
	26	23	27	26	26.25	
Medias para las rutas	20.00	21.67	25.40	23.53	22.65	

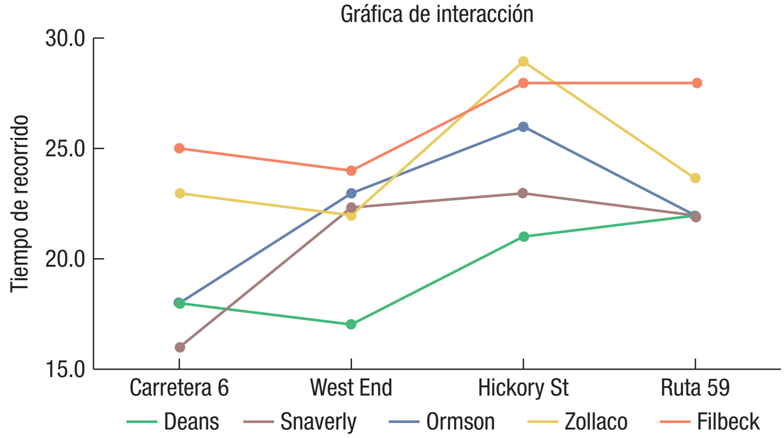
Para evaluar los efectos de interacción, es útil, como primer paso, graficar en puntos las medias de cada combinación entre conductor y ruta. Para la combinación Deans usando la carretera 6, la media es de 18 minutos, determinada por (18 + 15 + 21)/3. Para Fillbeck usando la ruta 59, la media es de 28 minutos, determinada por (28 + 30 + 26)/3. Las medias de las otras celdas se calculan en forma similar y los resultados se resumen en la siguiente tabla:

	Rutas			
Conductores	Carretera 6	West End	Hickory St	Ruta 59
Deans	18	17	21	22
Snaverly	16	22.33	23	22
Ormson	18	23	26	22
Zollaco	23	22	29	23.67
Filbeck	25	24	28	28

En la gráfica de interacción de la página siguiente se utiliza la información registrada en la tabla anterior. El eje vertical es el tiempo de recorrido en minutos. Las cuatro rutas aparecen en el eje horizontal, y cada línea traza los tiempos medios de recorrido para cada conductor en las cuatro rutas; por ejemplo, la línea verde reporta los tiempos de recorrido promedio para Deans en cada una de las cuatro rutas.

Con relación a la gráfica, ¿qué observaciones se pueden hacer con respecto a la interacción entre el conductor y la ruta en el tiempo de recorrido? Lo m**ás importante** es que las líneas no son paralelas; por lo tanto, existe un efecto de interacción entre el conductor y la ruta en el tiempo de recorrido; esto es, el tiempo de recorrido depende del efecto combinado del conductor y la ruta.

Observe las diferencias en los tiempos de recorrido: para la carretera 6, Snaverly tiene el menor, o más rápido, tiempo medio de recorrido; Deans tiene el menor tiempo medio de recorrido para las



rutas West End y Hickory Street; Zollaco tiene el tiempo promedio más lento para la ruta Hickory Street. Existen muchas otras observaciones que pueden conducir a la conclusión general de que el tiempo de recorrido está relacionado con los efectos combinados del conductor y la ruta. La pregunta crítica es si las interacciones observadas son significativas o si las diferencias se deben a la casualidad.

Prueba de hipótesis para detectar interacción

En el siguiente paso se deben realizar pruebas estadísticas para investigar aún más los efectos de interacciones posibles; en resumen, el estudio de los tiempos de recorrido plantea varias preguntas:

- ¿Hay alguna interacción entre rutas y conductores?
- ¿Son iguales los tiempos de recorrido de los conductores?
- ¿Son iguales los tiempos de recorrido de las rutas?

De las tres preguntas, la de mayor interés es la que se relaciona con la prueba de interacciones.

Estas ideas se formalizan en tres conjuntos de hipótesis:

1. H_0: No hay interacción entre conductores y rutas.
   H_1: Hay interacción entre conductores y rutas.
2. H_0: Las medias de los conductores son iguales.
   H_1: Las medias de los conductores *no* son iguales.
3. H_0: Las medias de las rutas son iguales.
   H_1: Las medias de las rutas *no* son iguales.

Cada hipótesis se prueba, como se hizo en la sección anterior, utilizando la distribución F. Las pruebas se resumen en la siguiente tabla ANOVA; la cual es similar al ANOVA de dos vías de la sección anterior, con la adición de la fuente de variación de la interacción. Además, el efecto del conductor es el **factor A** y el de la ruta es el **factor B**. Las hipótesis se prueban mediante el ya familiar estadístico F.

Fuente de variación	**Suma de cuadrados**	***gl***	**Media cuadrada**	***F***
Factor A (conductor)	SSA	$k - 1$	MSA = SSA/$(k - 1)$	MSA/MSE
Factor B (ruta)	SSB	$b - 1$	MSB = SSA/$(b - 1)$	MSB/MSE
Interacción	SSI	$(k - 1)(b - 1)$	MSI = SSI/$[(k - 1)(b - 1)]$	MSI/MSE
Error	SSE	$n - kb$	MSE = SSE/$(n - kb)$	
Total		$n - 1$		

Para probar la hipótesis para un ANOVA de dos vías con interacción se utiliza el ANOVA: dos factores con replicación en el análisis de datos, complemento de Excel. Los detalles para utilizar Excel se resumen en el apéndice C. En la tabla ANOVA que aparece en la página siguiente se muestran los resultados del análisis. Se utilizan los valores p para probar cada hipótesis. Con el nivel de significancia 0.05 se rechazan las hipótesis nulas si el valor p calculado es menor a 0.05.

ANOVA						
Fuente de variación	*SS*	*gl*	*MS*	*F*	*Valor p*	*F crítica*
Conductores	353.5667	4	88.39167	17.21916	0.0000	2.605975
Rutas	244.9833	3	81.66111	15.90801	0.0000	2.838745
Interaction	125.7667	12	10.48056	2.041667	0.0456	2.003459
Error	205.3333	40	5.133333			
Total	929.65	59				

Revisando los resultados de la prueba ANOVA, el valor *p* del efecto de interacción de 0.0456 es menor que nuestro nivel de significancia 0.05; por lo tanto, se rechaza la hipótesis nula de no interacción y se concluye que la combinación de ruta y conductor tiene un efecto significativo en la variable de respuesta (el tiempo de recorrido).

Un efecto significativo de interacción proporciona información importante acerca de los efectos combinados de las variables. Si la interacción está presente, entonces en el siguiente paso se realiza una prueba de diferencias en las medias de los factores a través de un ANOVA de una vía para cada nivel del otro factor. Este análisis se lleva cierto tiempo y esfuerzo, pero los resultados suelen ser esclarecedores.

El análisis continúa al realizar un ANOVA de una vía para cada ruta y probar la hipótesis H_0: los tiempos de los conductores son iguales. A continuación se registran los resultados.

Carretera 6; H_0 Los tiempos de los conductores son los mismos						
Fuente de variación	*SS*	*gl*	*MS*	*F*	*Valor p*	*F crítica*
Entre grupos	174	4	43.5	6.04167	0.010	3.478
Dentro de los grupos	72	10	7.2			
Total	246	14				
West End; H_0 Los tiempos de los conductores son los mismos						
Fuente de variación	*SS*	*gl*	*MS*	*F*	*Valor p*	*F crítica*
Entre grupos	88.6667	4	22.1667	4.05488	0.033	3.478
Dentro de los grupos	54.6667	10	5.46667			
Total	143.333	14				
Hickory; H_0 Los tiempos de los conductores son los mismos						
Fuente de variación	*SS*	*gl*	*MS*	*F*	*Valor p*	*F crítica*
Entre grupos	135.6	4	33.9	21.1875	0.000	3.478
Dentro de los grupos	16	10	1.6			
Total	151.6	14				
Ruta 59; H_0 Los tiempos de los conductores son los mismos						
Fuente de variación	*SS*	*gl*	*MS*	*F*	*Valor p*	*F crítica*
Entre grupos	81.0667	4	20.2667	3.23404	0.060	3.478
Dentro de los grupos	62.6667	10	6.26667			
Total	143.733	14				

Los resultados del ANOVA de una vía muestran que existen diferencias significativas en los tiempos medios de recorrido entre los conductores para cada ruta, excepto en la Ruta 59, con un valor *p* de 0.06. Mediante una revisión de la gráfica de interacción se revelan algunas de las diferencias; por ejemplo, para la ruta West End, en la gráfica se sugiere que Deans tiene el mejor tiempo medio de recorrido. Posteriores análisis estadísticos probarían pares de tiempos medios de recorridos para determinar las diferencias significativas entre los tiempos medios de recorrido de los conductores para cada ruta que tenga un valor *p* significativo.

AUTOEVALUACIÓN 12-5

Vea la siguiente tabla ANOVA.

ANOVA

Fuente de variación	SS	*gl*	MS	*F*	Valor *p*
Factor A	6.41	3	2.137	3.46	0.0322
Factor B	5.01	2	2.507	4.06	0.0304
Interacción	33.15	6	5.525	8.94	0.0000
Error	14.83	24	0.618		
Total	59.41	35			

Utilice el nivel de significancia 0.05 para responder las siguientes preguntas.

(a) ¿Cuántos niveles tiene el factor A? ¿Existe una diferencia significativa entre las medias del factor A? ¿Cómo lo sabe?

(b) ¿Cuántos niveles tiene el factor B? ¿Existe una diferencia significativa entre las medias del factor B? ¿Cómo lo sabe?

(c) ¿Cuántas observaciones hay en cada celda? ¿Existe alguna interacción significativa entre el factor A y el factor B sobre la variable de respuesta? ¿Cómo lo sabe?

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

19. Considere los siguientes datos muestrales de un análisis ANOVA de dos factores; existen dos niveles (pesado y ligero) del factor A (peso) y tres niveles (chico, mediano y grande) del factor B (talla); hay tres observaciones para cada combinación de talla y peso.

		Talla		
		Chico	**Mediano**	**Grande**
		23	20	11
	Pesado	21	32	20
		25	26	20
Peso				
		13	20	11
	Ligero	32	17	23
		17	15	8

Calcule un ANOVA con un software estadístico y utilice el nivel de significancia 0.05 para responder las siguientes preguntas.

a. ¿Hay alguna diferencia entre las medias de la talla?

b. ¿Hay alguna diferencia entre las medias del peso?

c. ¿Hay alguna interacción significativa entre peso y talla?

20. Considere la tabla ANOVA de dos vías parcialmente terminada. Suponga que hay cuatro niveles del factor A y tres niveles del factor B; el número de réplicas por celda es 5. Complete la tabla y realice pruebas para determinar si hay alguna diferencia significativa entre las medias del factor A, entre las medias del factor B o entre las medias de la interacción. Utilice el nivel de significancia 0.05 (se sugiere estimar los valores mediante la tabla *F*).

ANOVA

Fuente	*SS*	*gl*	MS	*F*
Factor A	75			
Factor B	25			
Interacción	300			
Error	600			
Total	1 000			

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

21. El distribuidor del *Wapakoneta Daily News*, periódico regional del suroeste de Ohio, considera tres tipos de máquinas expendedoras, o “anaqueles”. La gerencia desea saber si los modelos (J-1000, D-320 y UV-57) y su ubicación (dentro o fuera de los supermercados), afectan las ventas. A cada una de las seis tiendas similares les asignan de forma aleatoria una combinación de máquina y ubicación; en los datos de la página siguiente se muestra el número de periódicos vendidos durante cuatro días.

Ubicación/Máquina	J-1000	D-320	UV-57
Dentro	33, 40, 30, 31	29, 28, 33, 33	47, 39, 39, 45
Fuera	43, 36, 41, 40	48, 45, 40, 44	37, 32, 36, 35

a. Trace la gráfica de interacción. Con base en sus observaciones, ¿hay algún efecto de interacción? A partir de la gráfica, describa el efecto de interacción entre la máquina y su posición.
b. Calcule un ANOVA con un software estadístico y utilice el nivel de significancia 0.05 para probar los efectos de posición, máquina e interacción sobre las ventas; reporte los resultados estadísticos.
c. Compare las ventas medias dentro y fuera de cada máquina mediante técnicas estadísticas. ¿Cuál es su conclusión?

22. Una compañía importante está organizada en tres áreas funcionales: manufactura, marketing e I+D. Los empleados afirman que la compañía les paga a las mujeres menos que a los hombres en puestos similares; esta hizo una selección aleatoria de cuatro hombres y cuatro mujeres en cada área, y registró sus salarios semanales en dólares.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Área/Género	Femenino	Masculino
Manufactura	1 016, 1 007, 875, 968	978, 1 056, 982, 748
Marketing	1 045, 895, 848, 904	1 154, 1 091, 878, 876
I+D	770, 733, 844, 771	926, 1 055, 1 066, 1 088

a. Dibuje la gráfica de interacción. Con base en sus observaciones, ¿hay algún efecto de interacción? A partir de la gráfica, describa el efecto de la interacción del género y el área sobre el salario.
b. Calcule un ANOVA con un software estadístico y utilice el nivel de significancia 0.05 para probar los efectos del género, el área e interacción sobre el salario; reporte los resultados estadísticos.
c. Compare las ventas medias de hombres y mujeres en cada área mediante técnicas estadísticas. ¿Qué le recomendaría a la compañía?

RESUMEN DEL CAPÍTULO

I. Las características de la distribución *F* son:
  A. Es continua.
  B. Sus valores no pueden ser negativos.
  C. Tiene sesgo positivo.
  D. Hay una familia de distribuciones *F*. Cada vez que cambian los grados de libertad en el numerador o en el denominador, se crea una distribución nueva.

II. Con la distribución *F* se prueba si dos varianzas poblacionales son iguales.
  A. Las poblaciones muestreadas deben seguir la distribución normal.
  B. La mayor de las dos varianzas muestrales se coloca en el numerador, para forzar que la razón mínima sea 1.00.
  C. El valor de *F* se calcula con la siguiente ecuación:

$$F = \frac{s_1^2}{s_2^2} \qquad \textbf{[12.1]}$$

III. Un ANOVA de una vía se utiliza para comparar varias medias de tratamiento.
  A. Un tratamiento es una fuente de variación.
  B. Las suposiciones subyacentes al examen ANOVA son:
    1. Las muestras son de poblaciones que siguen una distribución normal.
    2. Las poblaciones tienen desviaciones estándar iguales.
    3. Las poblaciones son independientes.
  C. La información para determinar el valor de *F* se resume en una tabla ANOVA.
    1. La fórmula de SS total (el total de la suma de los cuadrados) es:

$$\text{SS total} = \sum(x - \bar{x}_G)^2 \qquad \textbf{[12.2]}$$

    2. La fórmula de SSE (la suma de los errores al cuadrado) es:

$$\text{SSE} = \sum(x - \bar{x}_c)^2 \qquad \textbf{[12.3]}$$

3. La fórmula de SST (el tratamiento de la suma de cuadrados) se determina por la resta:

$$\text{SST} = \text{SS Total} - \text{SSE} \tag{12.4}$$

4. Esta información se resume en la siguiente tabla y se determina el valor de *F*.

Fuente de variación	Suma de cuadrados	*gl*	Media cuadrática	*F*
Tratamientos	SST	$k - 1$	$\text{SST}/(k-1) = \text{MST}$	MST/MSE
Error	SSE	$n - k$	$\text{SSE}/(n-k) = \text{MSE}$	
Total	SS total	$n - 1$		

IV. Si se rechaza una hipótesis nula de tratamiento de medias iguales, los pares de medias diferentes se identifican a partir del intervalo de confianza siguiente.

$$(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \tag{12.5}$$

V. En un ANOVA de dos vías se considera una segunda variable de tratamiento.

A. La segunda variable de tratamiento se denomina variable de bloqueo.

B. Esta se determina con la siguiente ecuación:

$$\text{SSB} = k\sum(\bar{x}_b - \bar{x}_G)^2 \tag{12.6}$$

C. El término SSE, o suma de los errores al cuadrado, se determina a partir de la siguiente ecuación:

$$\text{SST} = \text{SS Total} - \text{SSE} - \text{SSB} \tag{12.7}$$

D. El estadístico *F* de la variable de tratamiento y de la variable de bloqueo se determina en la siguiente tabla:

Fuente de variación	Suma de cuadrados	*gl*	Media cuadrática	*F*
Tratamientos	SST	$k - 1$	$\text{SST}/(k-1) = \text{MST}$	MST/MSE
Bloques	SSB	$b - 1$	$\text{SSB}/(b-1) = \text{MSB}$	MSB/MSE
Error	SSE	$(k-1)(b-1)$	$\text{SSE}/[(k-1)(b-1)] = \text{MSE}$	
Total	SS total	$n - 1$		

VI. En un ANOVA de dos vías con observaciones repetidas se consideran dos variables de tratamiento y la interacción posible entre las variables. A continuación se muestra la tabla ANOVA completa, incluyendo las interacciones:

Fuente	Suma de cuadrados	*gl*	Media cuadrática	*F*
Factor A	SSA	$k - 1$	$\text{SSA}/(k-1) = \text{MSA}$	MSA/MSE
Factor B	SSB	$b - 1$	$\text{SSB}/(b-1) = \text{MSB}$	MSB/MSE
Interacción	SSI	$(k-1)(b-1)$	$\text{SSI}/[(k-1)(b-1)] = \text{MSI}$	MSI/MSE
Error	SSE	$n - kb$	$\text{SSE}/(n-kb) = \text{MSE}$	
Total	SS total	$n - 1$		

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
SS total	Suma del total de cuadrados	*Total de S S*
SST	Suma del tratamiento de cuadrados	*S S T*
SSE	Suma de los errores al cuadrado	*S S E*
MSE	Error cuadrático medio	*M S E*
SSB	Suma de los cuadrados debida al bloque	*S S B*
SSI	Suma de interacción de cuadrados	*S S I*

EJERCICIOS DEL CAPÍTULO

23. Un agente de bienes raíces del área costera de Georgia desea comparar la variación entre el precio de venta de casas con vista al mar y el de las ubicadas a tres cuadras del mar. Una muestra de 21 casas con vista al mar que se vendieron el año anterior reveló que la desviación estándar de los precios de venta fue de 45 600 dólares; una muestra de 18 casas, también vendidas el año previo, ubicadas de una a tres cuadras del mar, reveló que la desviación estándar fue de 21 330 dólares. Con el nivel de significancia 0.01, ¿puede concluir que hay más variación entre los precios de venta de las casas con vista al mar?

24. Una compañía de diseño de software está creando un buscador nuevo y más rápido; sin duda, este es más rápido, pero las pruebas iniciales indican que hay mayor variación en el tiempo para realizar la búsqueda. Una muestra de 16 búsquedas reveló que la desviación estándar del tiempo requerido fue de 0.22 segundos y 0.12 segundos. Con el nivel de significancia 0.05, ¿puede concluir que hay mayor variación en el tiempo de búsqueda del nuevo dispositivo?

25. En Jamestown, Nueva York, hay dos concesionarios Chevrolet. Las ventas mensuales medias en Sharkey Chevy y Dave White Chevrolet son más o menos iguales; sin embargo, Tom Sharkey, propietario de Sharkey Chevrolet, considera que sus ventas son más consistentes. A continuación se presenta el número de automóviles nuevos que Sharkey vendió en los últimos siete meses, y Dave White en los últimos ocho meses. ¿Concuerda con Sharkey? Utilice el nivel de significancia 0.01.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Sharkey	98	78	54	57	68	64	70	
Dave White	75	81	81	30	82	46	58	101

26. Se tomaron muestras aleatorias de cinco personas a partir de tres poblaciones; la suma del total de cuadrados fue 100 y la suma de cuadrados debida a los tratamientos fue 40.
a. Formule las hipótesis nula y alternativa.
b. ¿Cuál es la regla de decisión? Utilice el nivel de significancia 0.05.
c. Elabore una tabla ANOVA. ¿Cuál es el valor de *F*?
d. ¿Cuál es su decisión respecto de la hipótesis nula?

27. En una tabla ANOVA, MSE fue igual a 10; se seleccionaron muestras aleatorias de seis personas a partir de cuatro poblaciones y la suma del total de cuadrados fue 250.
a. Formule las hipótesis nula y alternativa.
b. ¿Cuál es la regla de decisión? Utilice el nivel de significancia 0.05.
c. Elabore una tabla ANOVA. ¿Cuál es el valor de *F*?
d. ¿Cuál es su decisión respecto de la hipótesis nula?

28. La siguiente es una tabla ANOVA parcial.

Fuente	Suma de cuadrados	*gl*	Media cuadrática	*F*
Tratamiento		2		
Error			20	
Total	500	11		

Complétela y responda las preguntas siguientes. Utilice el nivel de significancia 0.05.
a. ¿Cuántos tratamientos hay?
b. ¿Cuál es el tamaño total de la muestra?
c. ¿Cuál es el valor crítico de *F*?
d. Formule las hipótesis nula y alternativa.
e. ¿Cuál es su conclusión respecto de la hipótesis nula?

29. Una organización de consumidores desea saber si hay una diferencia entre los precios de un juguete en particular en tres tipos de tiendas. El precio del juguete se investigó en una muestra de 15 tiendas: cinco de descuento, cinco de artículos diversos y cinco departamentales; los resultados se muestran a continuación. Utilice el nivel de significancia 0.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Descuento	Variedad	Departamento
$12	$15	$19
13	17	17
14	14	16
12	18	20
15	17	19

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

30. Jacob Lee es un viajero frecuente entre Los Ángeles y San Francisco. El mes anterior registró los tiempos de vuelo en tres aerolíneas distintas; los resultados son:

Goust	Jet Red	Cloudtran
51	50	52
51	53	55
52	52	60
42	62	64
51	53	61
57	49	49
47	50	49
47	49	
50	58	
60	54	
54	51	
49	49	
48	49	
48	50	

a. Utilice el nivel de significancia 0.05 y el proceso de prueba de hipótesis de cinco pasos para comprobar si existen diferencias entre los tiempos medios de vuelo de las tres aerolíneas.
b. Desarrolle un intervalo de confianza de 95% para la diferencia entre las medias entre Goust y Cloudtran.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

31. Hay cuatro distritos en la ciudad de Maumee, y Andy North, jefe de la policía, desea determinar si hay una diferencia entre los números medios de delitos cometidos en cada uno; por lo tanto, registra el número de delitos reportados en cada distrito durante seis días. Con el nivel de significancia 0.05, ¿el jefe de la policía puede concluir que hay una diferencia entre los números medios de delitos entre los cuatro distritos?

Número de delitos			
Rec Center	Key Street	Monclova	Whitehouse
13	21	12	16
15	13	14	17
14	18	15	18
15	19	13	15
14	18	12	20
15	19	15	18

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

32. En un estudio del efecto de los comerciales de televisión sobre los niños de 12 años se midió el tiempo de su atención, en segundos; los comerciales fueron de ropa, alimentos y juguetes. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los lapsos de atención promedio de los niños con respecto a los diversos comerciales? ¿Existen diferencias significativas entre pares de promedios? ¿Recomendaría dejar de transmitir uno de los tres tipos de comerciales?

Ropa	Alimentos	Juguetes
26	45	60
21	48	51
43	43	43
35	53	54
28	47	63
31	42	53
17	34	48
31	43	58
20	57	47
	47	51
	44	51
	54	

33. Cuando únicamente se implican dos tratamientos, ANOVA y la prueba t de Student (capítulo 11) dan como resultado las mismas conclusiones; de igual forma, $t^2 = F$. Como ejemplo, suponga que se dividió al azar a 14 estudiantes de introducción a la historia en dos grupos, uno de 6 integrantes y el otro de 8; a un grupo se le asignó el curso tradicional de lectura, y al otro, un curso en internet. Al final, los integrantes de ambos contestaron un examen de 50 preguntas. En la lista de la derecha se incluye el número correcto de respuestas de cada grupo.

Lectura tradicional	Internet
19	32
17	28
23	31
22	26
17	23
16	24
	27
	25

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

a. Con las técnicas del análisis de la varianza, demuestre H_0 que las dos calificaciones medias son iguales; $\alpha = 0.05$.
b. Con la prueba t descrita en el capítulo 11 calcule t.
c. Interprete los resultados.

34. Hay cuatro talleres de hojalatería en Bangor, Maine, y todos afirman que dan servicio de manera eficiente a sus clientes; para comprobar si hay alguna diferencia en el servicio, se seleccionó a algunos clientes de manera aleatoria de cada taller y se registraron los tiempos de espera, en días. Los resultados en un paquete de software estadístico son:

Resumen				
Grupos	**Conteo**	**Suma**	**Promedio**	**Varianza**
Taller A	3	15.4	5.133333	0.323333
Taller B	4	32	8	1.433333
Taller C	5	25.2	5.04	0.748
Taller D	4	25.9	6.475	0.595833

ANOVA					
Fuente de variación	**SS**	***gl***	**MS**	***F***	**Valor *p***
Entre grupos	23.37321	3	7.791069	9.612506	0.001632
Dentro de grupos	9.726167	12	0.810514		
Total	33.09938	15			

¿Hay alguna evidencia que sugiera una diferencia entre los tiempos de espera medios en los cuatro talleres de hojalatería? Utilice el nivel de significancia 0.05.

35. Se ingresan los rendimientos de combustible de una muestra de 27 automóviles compactos, medianos y grandes en un paquete de software estadístico. Con el análisis de varianza se investiga si hay una diferencia entre los kilometrajes medios de los tres tipos de automóviles, ¿cuál es su conclusión? Utilice el nivel de significancia 0.01.

Resumen				
Grupos	**Conteo**	**Suma**	**Promedio**	**Varianza**
Compactos	12	268.3	22.35833	9.388106
Medianos	9	172.4	19.15556	7.315278
Grandes	6	100.5	16.75	7.303

A continuación se presentan resultados adicionales.

ANOVA					
Fuentes de variación	**SS**	***gl***	**MS**	***F***	**Valor *p***
Entre grupos	136.4803	2	68.24014	8.258752	0.001866
Dentro de grupos	198.3064	24	8.262766		
Total	334.7867	26			

36. En la producción de un componente para un avión se emplean tres líneas de ensamblado; para estudiar la tasa de producción, se elige una muestra aleatoria con periodos de seis horas por línea de ensamble y se registra el número de componentes producidos en cada línea durante estos periodos. Los resultados de un paquete de software estadístico se presentan en la página siguiente.

Resumen				
Grupos	**Conteo**	**Suma**	**Promedio**	**Varianza**
Línea A	6	250	41.66667	0.266667
Línea B	6	260	43.33333	0.666667
Línea C	6	249	41.5	0.7

ANOVA					
Fuente de variación	**SS**	***gl***	**MS**	***F***	**Valor *p***
Entre grupos	12.33333	2	6.166667	11.32653	0.001005
Dentro de grupos	8.166667	15	0.544444		
Total	20.5	17			

a. Utilice el nivel de significancia 0.01 para comprobar si hay alguna diferencia entre las producciones medias de las tres líneas de ensamblado.

b. Elabore un intervalo de confianza de 99% de la diferencia en las medias entre las líneas de producción B y C.

37. El servicio postal agrupa el correo como prioridad exprés, prioridad, primera clase o estándar. En un periodo de tres semanas, 18 envíos de cada tipo se enviaron desde el centro de distribución en Atlanta, Georgia, a Des Mones, Iowa. Se registró el tiempo total de entrega en días y se utilizó Minitab para realizar el ANOVA. Los resultados se presentan a continuación:

```
Fuente                         gl       SS        MS       F       P
Factor                          3    30.298    10.099   11.90   0.000
Error                          68    57.705     0.849
Total                          71    88.003

S = 0.9212  R-Sq = 34.43%  R-Sq(adj) = 31.54%

Nivel                           N    Mediana  Desv. est.
Prioridad exprés               18    1.0694    0.7652
Prioridad                      18    1.7210    1.0829
Primera clase                  18    2.5211    0.9097
Estándar                       18    2.6840    0.8991

Prioridad exprés, sustraído de:
                           Inferior   Centro    Superior
Prioridad                   0.0389    0.6516     1.2644
Primera clase               0.8390    1.4517     2.0645
Estándar                    1.0018    1.6146     2.2273

                          ------+---------+---------+---------+---
Prioridad                                (------*-----)
Primera clase                                   (------*-----)
Estándar                                          (-----*-----)
                          ------+---------+---------+---------+---
                              -1.0       0.0       1.0       2.0

Prioridad, sustraído de:
                           Inferior   Centro    Superior
Primera clase               0.1874    0.8001     1.4128
Estándar                    0.3502    0.9629     1.5757

                          ------+---------+---------+---------+---
Primera clase                              (-----*-----)
Estándar                                     (-----*-----)
                          ------+---------+---------+---------+---
                              -1.0       0.0       1.0       2.0

Primera clase, sustraído de:
                           Inferior   Centro    Superior
Estándar                   -0.4499    0.1628     0.7756

                          ------+---------+---------+---------+---
Estándar                             (-----*-----)
                          ------+---------+---------+---------+---
                              -1.0       0.0       1.0       2.0
```

Utilizando los resultados del ANOVA, compare los tiempos promedio de entrega para los cuatro distintos tipos de correo.

38. Usted emplea un filtro para bloquear el correo no deseado en su bandeja de entrada; registra el número de mensajes bloqueados por día de la semana y utiliza Minitab para efectuar el análisis siguiente. Los resultados son:

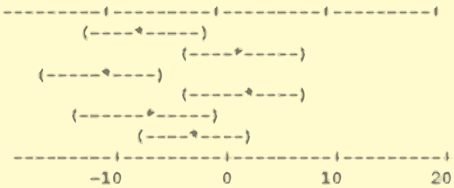
Fuente	gl	SS	MS	F	P
Factor	6	1242.5	207.1	6.66	0.000
Error	48	1493.1	31.1		
Total	54	2735.7			

S = 5.577 R–Sq = 45.42% R–Sq(adj) = 38.60

Nivel	N	Media	Desv. est.
Lunes	10	71.926	5.435
Martes	9	64.329	6.890
Miércoles	7	73.322	2.936
Jueves	8	60.539	4.268
Viernes	8	73.709	5.369
Sábado	5	64.436	6.729
Domingo	8	68.810	6.384

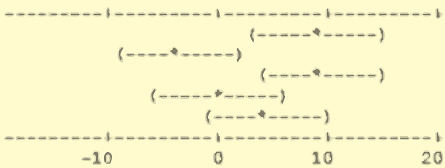
Lunes, sustraído de:

	Inferior	Centro	Superior
Martes	–12.749	–7.597	–2.444
Miércoles	–4.130	1.396	6.922
Jueves	–16.706	–11.387	–6.068
Viernes	–3.536	1.783	7.103
Sábado	–13.632	–7.490	–1.348
Domingo	–8.436	–3.117	2.203



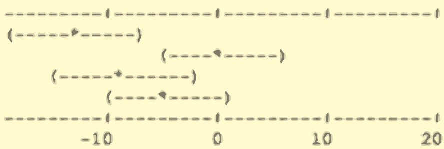
Martes, sustraído de:

	Inferior	Centro	Superior
Miércoles	3.341	8.993	14.644
Jueves	–9.239	–3.790	1.659
Viernes	3.931	9.380	14.829
Sábado	–6.148	0.107	6.362
Domingo	–0.969	4.480	9.929



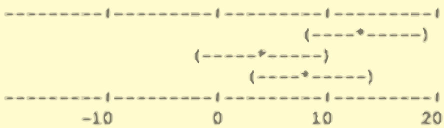
Miércoles, sustraído de:

	Inferior	Centro	Superior
Jueves	–18.587	–12.783	–6.979
Viernes	–5.416	0.388	6.191
Sábado	–15.452	–8.886	–2.319
Domingo	–10.316	–4.512	1.191



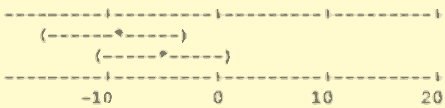
Jueves, sustraído de:

	Inferior	Centro	Superior
Viernes	7.563	13.170	18.777
Sábado	–2.496	3.897	10.290
Domingo	2.663	8.270	13.877



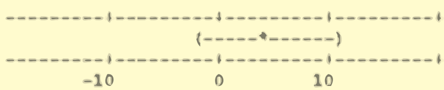
Viernes, sustraído de:

	Inferior	Centro	Superior
Sábado	–15.666	–9.273	–2.880
Domingo	–10.507	–4.900	0.707



Sábado, sustraído de:

	Inferior	Centro	Superior
Domingo	–2.020	4.373	10.766



Utilizando los resultados del ANOVA, compare el número de mensajes bloqueados por día de la semana.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

39. En Shank's, Inc., empresa publicitaria, se desea saber si el tamaño y el color de un anuncio genera respuestas diferentes en los lectores de revistas. A un grupo de lectores se le muestran anuncios con cuatro colores distintos y de tres tamaños diferentes; a cada lector se le pide asignar a cada combinación de tamaño y color una calificación entre 1 y 10. Suponga que las calificaciones siguen la distribución normal; la calificación de cada combinación se muestra en la siguiente tabla (por ejemplo, la calificación de un anuncio pequeño en color rojo es 2).

	Color del anuncio			
Tamaño del anuncio	**Rojo**	**Azul**	**Naranja**	**Verde**
Pequeño	2	3	3	8
Mediano	3	5	6	7
Grande	6	7	8	8

¿Hay alguna diferencia en la eficacia de un anuncio con base en su color y su tamaño? Utilice el nivel de significancia 0.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

40. En el área de Columbus, Georgia, hay cuatro restaurantes McBurger; en la siguiente tabla se muestran los números de hamburguesas que vendió cada uno de ellos en cada una de las últimas seis semanas. Con el nivel de significancia 0.05 y considerando el factor de la semana, ¿hay alguna diferencia entre los números medios que vendieron los cuatro restaurantes?

	Restaurante			
Semana	**Metro**	**Interestatal**	**Universidad**	**Río**
1	124	160	320	190
2	234	220	340	230
3	430	290	290	240
4	105	245	310	170
5	240	205	280	180
6	310	260	270	205

a. ¿Hay alguna diferencia entre las medias de tratamiento?
b. ¿Hay alguna diferencia entre las medias de bloqueo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

41. En la ciudad de Tucson, Arizona, se emplean personas para valuar las casas con el fin de establecer el impuesto predial. El administrador municipal envía a cada valuador a las mismas cinco casas y después compara los resultados; la información se presenta a continuación, en miles de dólares. ¿Puede concluir que hay una diferencia entre los avalúos, con $\alpha = 0.05$?

	Valuador			
Casa	**Zawodny**	**Norman**	**Cingle**	**Holiday**
A	$53.0	$55.0	$49.0	$45.0
B	50.0	51.0	52.0	53.0
C	48.0	52.0	47.0	53.0
D	70.0	68.0	65.0	64.0
E	84.0	89.0	92.0	86.0

a. ¿Hay alguna diferencia entre las medias de tratamiento?
b. ¿Hay alguna diferencia entre las medias de bloqueo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

42. El concesionario Martin Motors tiene tres automóviles de la misma marca y modelo. El director desea comparar el consumo de combustible de ellos (designados automóvil A, B y C) con cuatro tipos de gasolina. En cada ensayo se puso un galón de gasolina al tanque vacío de los automóviles y se condujeron hasta que se agotó; en la siguiente tabla se muestra el número de millas recorridas en cada ensayo.

	Distancia (millas)		
Tipos de gasolina	**Automóvil A**	**Automóvil B**	**Automóvil C**
Regular	22.4	20.8	21.5
Super regular	17.0	19.4	20.7
Sin plomo	19.2	20.2	21.2
Premium sin plomo	20.3	18.6	20.4

Con el nivel de significancia 0.05:

a. ¿Hay alguna diferencia entre los tipos de gasolina?

b. ¿Hay alguna diferencia entre los automóviles?

43. Una empresa de investigación desea comparar el rendimiento, en millas por galón, de gasolina regular, de grado medio y de Premium. Con base en el desempeño de los diversos automóviles, se seleccionan y tratan como bloques siete automóviles; por lo tanto, cada tipo de gasolina se probó en cada tipo de automóvil. Los resultados de los ensayos, en millas por galón, se muestran en la siguiente tabla. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre las gasolinas o entre los automóviles?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Automóvil	Regular	De grado medio	Premium
1	21	23	26
2	23	22	25
3	24	25	27
4	24	24	26
5	26	26	30
6	26	24	27
7	28	27	32

44. Tres cadenas de supermercados del área de Denver, Colorado, afirman tener los precios más bajos. Como parte de un estudio de investigación sobre la publicidad de los supermercados, el *Denver Daily News* realizó un estudio; primero seleccionó una muestra aleatoria de nueve artículos. Luego, verificó el precio de cada artículo seleccionado en cada una de las tres cadenas el mismo día. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los precios medios de los supermercados o de los artículos?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Artículo	Super$	Ralph's	Lowblaws
1	$1.12	$1.02	$1.07
2	1.14	1.10	1.21
3	1.72	1.97	2.08
4	2.22	2.09	2.32
5	2.40	2.10	2.30
6	4.04	4.32	4.15
7	5.05	4.95	5.05
8	4.68	4.13	4.67
9	5.52	5.46	5.86

45. A continuación se enumeran los pesos (en gramos) de una muestra de dulces M&M, clasificados según su color; utilice un paquete de software estadístico para determinar si hay alguna diferencia entre los pesos medios de los dulces de colores distintos. Emplee el nivel de significancia 0.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Rojo	Naranja	Amarillo	Café	Café claro	Verde
0.946	0.902	0.929	0.896	0.845	0.935
1.107	0.943	0.960	0.888	0.909	0.903
0.913	0.916	0.938	0.906	0.873	0.865
0.904	0.910	0.933	0.941	0.902	0.822
0.926	0.903	0.932	0.838	0.956	0.871
0.926	0.901	0.899	0.892	0.959	0.905
1.006	0.919	0.907	0.905	0.916	0.905
0.914	0.901	0.906	0.824	0.822	0.852
0.922	0.930	0.930	0.908		0.965
1.052	0.883	0.952	0.833		0.898
0.903		0.939			
0.895		0.940			
		0.882			
		0.906			

46. Hay cuatro estaciones de radio en Midland con formatos diferentes (rock pesado, música clásica, country/campirana e instrumental). Cada una de ellas se interesa por saber el número de minutos de música que transmite por hora. De una muestra de 10 horas de cada estación, la suma de las diferencias al cuadrado entre cada observación y la media para su respectiva estación de radio $\sum(x - \overline{x}_C)^2$ son:

Estación de rock pesado:	126.29	Estación de música country/campirana:	166.79
Estación de música clásica:	233.34	Estación instrumental:	77.57

La suma total de los cuadrados para los datos es: SS total = 1 099.61.

a. Determine SST.
b. Determine SSE.
c. Elabore una tabla ANOVA.
d. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre las medias de tratamiento?
e. Si la media de la estación de rock pesado es 51.32 y la media de la estación country/western es 50.85, determine si existe una diferencia utilizando el nivel de significancia 0.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

47. La American Accounting Association realizó un estudio para comparar los salarios semanales de hombres y mujeres empleados en el sector público o privado en contabilidad; se seleccionaron muestras de cinco hombres y cinco mujeres en cada grupo.

	Sector	
Género	**Público**	**Privado**
Hombres	$ 978	$1 335
	1 035	1 167
	964	1 236
	996	1 317
	1 117	1 192
Mujeres	$ 863	$1 079
	975	1 160
	999	1 063
	1 019	1 110
	1 037	1 093

a. Trace una gráfica de interacción de las medias de los hombres y las mujeres según el sector.
b. Calcule un ANOVA con un software estadístico y, utilizando el nivel de significancia 0.05, pruebe el efecto de interacción del género y el sector en los salarios.
c. Con base en los resultados del punto b, realice las pruebas de hipótesis adecuadas para detectar las diferencias entre las medias de los factores.
d. Interprete los resultados en un reporte breve.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

48. Robert Altoff es vicepresidente de ingeniería de un fabricante de máquinas lavadoras domésticas. Como parte del desarrollo de un producto nuevo, Altoff desea determinar el tiempo óptimo del ciclo de lavado; parte del desarrollo es estudiar la relación entre el detergente empleado (cuatro marcas) y la duración del ciclo de lavado (18, 20, 22 o 24 minutos). A fin de realizar el experimento se asignan 32 cargas estándar de ropa (con igual contenido de suciedad y pesos totales iguales) a las 16 combinaciones detergente-ciclo de lavado. Los resultados (en libras de suciedad eliminada) se muestran en la siguiente tabla.

	Tiemplo de ciclo (min)			
Marca de detergente	**18**	**20**	**22**	**24**
A	0.13	0.12	0.19	0.15
	0.11	0.11	0.17	0.18
B	0.14	0.15	0.18	0.20
	0.10	0.14	0.17	0.18
C	0.16	0.15	0.18	0.19
	0.17	0.14	0.19	0.21
D	0.09	0.12	0.16	0.15
	0.13	0.13	0.16	0.17

a. Trace una gráfica de interacción de las medias del detergente según el tiempo del ciclo.
b. Calcule un ANOVA con un software estadístico y, utilizando el nivel de significancia 0.05, pruebe el efecto de interacción de la marca y el tiempo del ciclo sobre la "suciedad eliminada".

c. Con base en los resultados del punto b, realice las pruebas de hipótesis apropiadas de las diferencias entre las medias de los factores.

d. Interprete los resultados en un reporte breve.

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

49. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.

a. Con el nivel de significancia 0.02, ¿hay alguna diferencia entre la variabilidad de los precios de venta de las casas que tienen alberca y las que no tienen?

b. Con el nivel de significancia 0.02, ¿hay alguna diferencia entre la variabilidad de los precios de venta de las casas con cochera y las que no tienen?

c. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los precios de venta medios de las casas de los cinco municipios?

50. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012.

a. Con el nivel de significancia 0.10, ¿hay alguna diferencia entre la variación de los salarios de los equipos entre los de la Liga Nacional y la Liga Americana?

b. Establezca una variable que clasifique la asistencia total (en millones de personas) a los juegos del equipo en tres grupos: menos de 2.0, de 2.0 hasta 3.0, y de 3.0 o más. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre los números medios de juegos ganados en los tres grupos?

c. Con la misma variable de asistencia que se estableció en el punto b, ¿hay alguna diferencia entre los promedios de bateo medios del equipo? Utilice el nivel de significancia 0.05.

d. Con la misma variable de asistencia que estableció en el punto b, ¿hay alguna diferencia entre los salarios medios de los tres grupos? Utilice el nivel de significancia 0.05.

51. Consulte los datos de los autobuses del Distrito Escolar Buena.

a. Realice una prueba de hipótesis para averiguar si los costos medios de mantenimiento de cada autobús son iguales. Utilice el nivel de significancia 0.01.

b. Realice una prueba de hipótesis para determinar si los números medios de millas que recorrió cada autobús son iguales. Utilice el nivel de significancia 0.05.

c. Desarrolle un intervalo de confianza de 95% de la disparidad en el costo promedio de mantenimiento entre los autobuses fabricados por Bluebird y Thompson.

REPASO DE LOS CAPÍTULOS 10 a 12

Esta sección es un repaso de los conceptos y términos importantes que se presentaron en los capítulos 10, 11 y 12. En el capítulo 10 se inició el estudio de la prueba de hipótesis (una afirmación acerca del valor del parámetro de una población). Una prueba de hipótesis estadística comienza con una afirmación respecto del valor del parámetro de la población en la hipótesis nula; esta se establece para realizar la prueba. Al completarla se debe rechazar o no la hipótesis nula; si se rechaza, se concluye que la hipótesis alternativa es verdadera. La hipótesis alternativa (también llamada hipótesis de investigación) se "acepta" solo si se demuestra que la hipótesis nula es falsa. La mayoría de las veces se desea probar la hipótesis alternativa.

En el capítulo 10 se seleccionaron muestras aleatorias de una sola población y se probó si era razonable que el parámetro de la población en estudio igualara un valor en particular; por ejemplo, para investigar si el tiempo medio de duración en el puesto de director ejecutivo en empresas importantes es de 12 años, se selecciona una muestra de personas en ese puesto, se calcula la media muestral y se compara con la población; la población individual en consideración es la duración de los directores ejecutivos de empresas importantes. Se describieron métodos para conducir la prueba cuando la desviación estándar de la población estaba disponible y cuando no lo estaba.

En el capítulo 11 se amplió la idea de prueba de hipótesis para verificar si dos muestras aleatorias independientes provenían de poblaciones con las mismas medias poblacionales (o iguales); por ejemplo, el St. Mathews Hospital opera una sala de urgencias en las zonas norte y sur de Knoxville, Tennessee; la pregunta de investigación es: ¿el tiempo de espera medio de los pacientes es igual en ambas salas? Para responder esta pregunta, se selecciona una muestra aleatoria de cada sala y se calculan las medias muestrales; se prueba la hipótesis nula (el tiempo de espera medio es el mismo en las dos salas); la hipótesis alternativa es que el tiempo medio de espera no es el mismo en las dos salas. Si se conocen las desviaciones estándar de cada población, se utiliza la distribución z como la del estadístico de prueba; en caso contrario, este sigue la distribución t.

El estudio del capítulo 11 también incluyó muestras *dependientes*, en cuyo caso se aplicó la prueba de la *diferencia pareada*; el estadístico de prueba es la distribución t. Un problema común de muestra pareada se presenta al registrar la presión arterial de individuos antes y después de la administración de medicamento para evaluar su eficacia. También se consideró el caso de probar dos proporciones poblacionales; por ejemplo, el gerente de producción desea comparar la proporción de defectos que se generan en el turno matutino con la del turno vespertino.

En el capítulo 11 se abordó la diferencia entre dos medias poblacionales. En el capítulo 12 se presentaron pruebas para varianzas y un procedimiento denominado *análisis de la varianza* o ANOVA; con el cual se determina de manera simultánea si varias poblaciones normales e independientes tienen la misma media. Este procedimiento se lleva a cabo mediante la comparación de las varianzas de las muestras aleatorias seleccionadas de estas poblaciones; se aplica el procedimiento habitual de prueba de hipótesis, pero se utiliza la distribución *F* como el estadístico de prueba. Con frecuencia, los cálculos son tediosos, por lo que se recomienda utilizar un paquete de software estadístico.

Como ejemplo del análisis de la varianza, se puede realizar una prueba para determinar si hay alguna diferencia entre las eficacias de cinco fertilizantes sobre el peso de mazorcas de maíz para hacer rosetas. A este tipo de análisis se le conoce como *ANOVA de un factor,* pues es posible obtener conclusiones acerca de un solo factor, denominado *tratamiento*. Si se desea obtener conclusiones respecto de los efectos simultáneos de más de un factor o variable, se utiliza el *ANOVA de dos factores*. En ambas pruebas (de uno y dos factores) se emplea la *distribución F* como la del estadístico de prueba para determinar si una población normal varía más que otra.

Una característica adicional del ANOVA de dos factores es la probabilidad de que existan interacciones entre estos. Hay una *interacción* si la respuesta a uno de los factores depende del nivel del otro; por fortuna, el ANOVA se amplía con facilidad para incluir una prueba de interacciones.

PROBLEMAS

En los problemas 1 a 6 establezca: a) las hipótesis nula y alternativa, b) la regla de decisión y c) la decisión respecto de la hipótesis nula. Interprete el resultado.

1. Se calibra una máquina para fabricar pelotas de tenis de modo que el rebote medio sea de 36 pulgadas cuando estas se dejen caer desde una plataforma a cierta altura. El supervisor sospecha que el rebote medio cambió y es menor que el establecido; para comprobarlo, se dejaron caer 42 pelotas desde la plataforma y la altura media del rebote fue de 35.5 pulgadas, con una desviación estándar de 0.9 pulgadas. Con el nivel de significancia 0.05, ¿puede el supervisor concluir que la altura del rebote medio es inferior a 36 pulgadas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2. Se piensa que los trabajadores de construcción de caminos no realizan un trabajo productivo durante un promedio de 20 minutos de cada hora; algunos afirmaban que el tiempo no productivo era aún mayor. Se realizó un estudio en un emplazamiento de construcción, con un cronómetro y otras formas de verificación de hábitos de trabajo; la verificación aleatoria de los trabajadores reveló los tiempos no productivos siguientes, en minutos, durante un periodo de una hora (sin incluir los descansos programados):

10	25	17	20	28	30	18	23	18

Con el nivel de significancia 0.05, ¿es razonable concluir que el tiempo no productivo medio es mayor a 20 minutos?

3. Se va a realizar una prueba que implica medir el poder de soporte medio de dos pegamentos para plástico. Primero se recubrió el extremo de un gancho pequeño con pegamento Epox y se sujetó a una hoja de plástico; cuando se secó, se agregó peso al gancho hasta que se separó de la hoja de plástico y se registró el peso. Este procedimiento se repitió hasta que se probaron 12 ganchos y después se siguió el mismo procedimiento con el pegamento Holdtite, pero solo se emplearon 10 ganchos. Los resultados de las muestras, en libras, aparecen a la izquierda.

	Epox	Holdtite
Media muestra	250	252
Desviación estándar muestra	5	8
Tamaño muestra	12	10

Con el nivel de significancia 0.01, ¿hay alguna diferencia entre el poder de soporte medio del pegamento Epox y el del pegamento Holdtite?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

4. Pittsburgh Paints desea probar un aditivo formulado para aumentar la vida de las pinturas empleadas en las condiciones calurosas y áridas del sureste de Estados Unidos. Se pintó la parte superior de una pieza de madera con la pintura normal, y en la parte inferior se usó pintura con el aditivo; se siguió el mismo procedimiento con un total de 10 piezas y después se sometió cada pieza a una luz brillante. Los datos (el número de horas que duró la pintura de cada pieza antes de desvanecerse más allá de cierto punto) son:

	Número de horas por muestra									
	A	B	C	D	E	F	G	H	I	J
Sin aditivo	325	313	320	340	318	312	319	330	333	319
Con aditivo	323	313	326	343	310	320	313	340	330	315

Con el nivel de significancia 0.05, determine si el aditivo es eficaz para prolongar la vida de la pintura.

5. Un distribuidor de refresco de cola de Buffalo, Nueva York, ofrece una oferta especial en empaques de 12 unidades, y se pregunta en qué parte de las tiendas de comestibles se debe colocar el refresco para captar más la atención. ¿Cerca de la puerta de acceso, en la sección de refrescos, en las cajas registradoras, o junto a la leche y otros productos lácteos? Cuatro tiendas con ventas totales similares cooperaron en un experimento; en una, los paquetes de 12 se colocaron cerca de la puerta de acceso; en otra, cerca de las cajas registradoras, y así sucesivamente. Las ventas se verificaron a horas específicas en cada tienda durante exactamente cuatro minutos. Los resultados son:

Cerca de la puerta	En la sección de refrescos	Cerca de las cajas registradoras	En la sección de lácteos
\$6	\$ 5	\$ 7	\$10
8	10	10	9
3	12	9	6
7	4	4	11
	9	5	
		7	

El distribuidor desea determinar si hay alguna diferencia entre las ventas medias del refresco en las cuatro ubicaciones de la tienda. Utilice el nivel de significancia 0.05.

6. Williams Corporation investiga los efectos de los antecedentes escolares en el desempeño de los empleados; una variable importante en este caso es el estado social autodefinido del empleado. En la compañía se registraron los volúmenes de ventas anuales (en miles de dólares) logrados por los empleados de ventas en cada una de las categorías que se muestran enseguida; realice un análisis completo de varianza de dos vías (con la probabilidad de interacciones) con los datos y describa qué sugieren sus resultados.

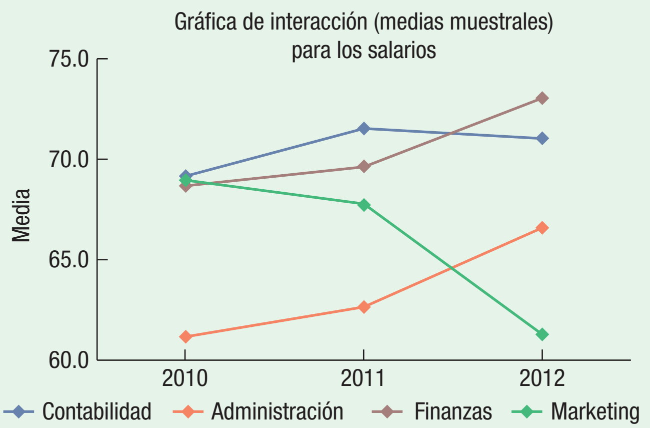
Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Tipo de escuela		
Estado social autodefinido	**De las ocho mejores**	**De gobierno**	**Privada pequeña**
Bajo	62, 61	68, 64	70, 70
Medio	68, 64	74, 68	62, 65
Alto	70, 71	57, 60	57, 56

7. Un supervisor de escuela revisa los salarios iniciales de exestudiantes (en miles de dólares). Se tomaron muestras durante tres años de cuatro especialidades (contabilidad, administración, finanzas y marketing).

Especialidad/Año	2010	2011	2012
Contabilidad	75.4, 69.8, 62.3	73.9, 78.8, 62.0	64.2, 80.8, 68.2
Administración	61.5, 59.9, 62.1	63.9, 57.6, 66.5	74.2, 67.5, 58.1
Finanzas	63.6, 70.2, 72.2	69.2, 72.5, 67.2	74.7, 66.4, 77.9
Marketing	71.3, 69.2, 66.4	74.0, 67.6, 61.7	60.0, 61.3, 62.5

a. En la gráfica de interacción de la derecha se muestra la información anterior. ¿Qué revela esta?
b. Escriba todos los pares de hipótesis nula y alternativa que aplicaría en un ANOVA de dos vías.
c. En la página siguiente se muestra la salida de un software estadístico. Utilice el nivel de significancia 0.05 para verificar interacciones.
d. Si lo considera adecuado, pruebe otras hipótesis con el nivel de significancia 0.05; en caso contrario, describa por qué no debe hacer las pruebas.



Fuente	gl	SS	MS	F	P
Especialidad	3	329.20	109.732	3.39	0.034
Año	2	7.32	3.659	0.11	0.894
Interacción	6	183.57	30.595	0.94	0.482
Error	24	777.29	32.387		
Total	35	1297.37			

CASOS

A. Century National Bank

Consulte la descripción del Century National Bank al final del repaso de los capítulos 1 a 4.

Con muchas opciones disponibles, los clientes ya no dejan que su dinero permanezca en una cuenta de cheques; durante muchos años, el saldo medio de estas fue de 1 600 dólares. ¿Indican los datos muestrales que el valor del saldo medio de la cuenta disminuyó a niveles inferiores de este valor?

En años recientes también se observó un aumento del uso de cajeros automáticos. Cuando el señor Selig asumió la responsabilidad del banco, el número medio de transacciones mensuales por cliente era 8; ahora él cree que aumentó a más de 10. En realidad, a la agencia de publicidad que prepara comerciales de televisión para el banco le gustaría usar este dato en el nuevo comercial que diseña. ¿Hay evidencia suficiente para concluir que el número medio de transacciones por cliente es mayor a 10 por mes? ¿Puede afirmar la agencia de publicidad que la media es superior a 9 al mes?

El banco tiene sucursales en cuatro ciudades distintas: Cincinnati, Ohio; Atlanta, Georgia; Louisville, Kentucky, y Erie, Pennsylvania. Al señor Selig le gustaría saber si hay alguna diferencia entre los saldos medios de las cuentas de cheques de las cuatro sucursales; de ser así, ¿en cuáles sucursales se presentan?

El señor Selig también tiene interés en los cajeros automáticos del banco, ¿hay alguna diferencia en el uso de estos en las sucursales? Asimismo, ¿los clientes con tarjetas de débito tienden a usar cajeros automáticos en forma distinta de los que no las tienen? ¿Hay alguna diferencia en el uso de los cajeros automáticos por parte de quienes tienen cuentas de cheques que pagan interés en comparación con las que no lo pagan? Prepare un reporte para el señor Selig que responda estas preguntas.

B. Bell Grove Medical Center

La señora Gene Dempsey es gerente del centro de atención de emergencia del Bell Grove Medical Center; una de sus responsabilidades es tener enfermeras suficientes para que se atienda con prontitud a los pacientes. Es muy estresante para los pacientes esperar mucho para recibir atención de emergencia, aunque sus necesidades no sean de vida o muerte. La señora Dempsey reunió la información siguiente respecto del número de pacientes durante las últimas semanas (el centro no atiende los fines de semana). ¿Parece que hay algunas diferencias en el número de pacientes atendidos el último día de la semana? Si hay diferencias, ¿cuáles días parecen ser los más ocupados? Redacte un breve reporte que resuma sus hallazgos.

Fecha	Día	Pacientes
9-29-06	Lunes	38
9-30-06	Martes	28
10-1-06	Miércoles	28
10-2-06	Jueves	30
10-3-06	Viernes	35
10-6-06	Lunes	35
10-7-06	Martes	25
10-8-06	Miércoles	22
10-9-06	Jueves	21
10-10-06	Viernes	32
10-13-06	Lunes	37
10-14-06	Martes	29
10-15-06	Miércoles	27
10-16-06	Jueves	28
10-17-06	Viernes	35
10-20-06	Lunes	37
10-21-06	Martes	26
10-22-06	Miércoles	28
10-23-06	Jueves	23
10-24-06	Viernes	33

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. Una afirmación acerca del valor de un parámetro poblacional que siempre incluye el signo de igual se llama ______________. **1.** __________________
2. La probabilidad de rechazar una hipótesis verdaderamente nula se denomina ______________. **2.** __________________
3. Asumiendo que la hipótesis nula es verdadera, la probabilidad de encontrar un valor de prueba estadística cuando menos tan extremo como el encontrado en la muestra se denomina ____________. **3.** __________________
4. Cuando se realiza una prueba de hipótesis con respecto a una sola media poblacional, la distribución z se utiliza como prueba estadística solo cuando no se conoce ______________. **4.** __________________

5. En una prueba de hipótesis para las medias de dos muestras, donde las desviaciones estándar de la población se desconocen, ¿qué se debe asumir con respecto a la forma de las poblaciones? 5. ________________
6. Un valor calculado a partir de información muestral que se utiliza para determinar si se rechaza la hipótesis nula se conoce como ____________. 6. ________________
7. En una prueba de dos colas, la región de rechazo está ____________ (toda en la cola superior, toda en la cola inferior, distribuida uniformemente entre ambas colas, o ninguna de las anteriores). 7. ________________
8. ¿Cuál de las siguientes no es una característica de la distribución *F*? Continua, con sesgo positivo, rango de $-\infty$ hasta ∞, familia de distribuciones. 8. ________________
9. Para realizar un ANOVA de una vía, los tratamientos deben ser ____________ (independientes, mutuamente excluyentes, continuos). 9. ________________
10. En un ANOVA de dos vías hay cuatro tratamientos y seis observaciones en cada tratamiento. ¿Cuáles son los grados de libertad de la distribución *F*? 10. ________________

Parte 2: Problemas

En el caso de los problemas 1 y 2, establezca las hipótesis nula y alternativa y la regla de decisión, tome una determinación con respecto a la hipótesis nula e interprete el resultado.

1. El administrador del Fort Fisher State Park, Carolina del Norte, piensa que el típico visitante de verano pasa en el parque más de 90 minutos. Una muestra de 18 turistas durante los meses de junio, julio y agosto reveló que el tiempo medio que estos permanecían en el parque era de 96 minutos, con una desviación estándar de 12 minutos. Con el nivel de significancia 0.01, ¿es razonable concluir que el tiempo medio de permanencia en el parque es mayor a 90 minutos?
2. ¿Existe alguna diferencia entre las millas promedio recorridas por semana de cada una de las dos compañías de taxis que operan en el área de Grand Strand? Mediante una investigación del periódico local *Sun News* se obtuvo la siguiente información muestral. Con el nivel de significancia 0.05, ¿es razonable concluir que hay una diferencia entre las millas promedio recorridas? Asuma que las varianzas de población son iguales.

Variable	Yellow Cab	Horse and Buggy Cab Company
Millas promedio	837	797
Desviación estándar	30	40
Tamaño de la muestra	14	12

3. A continuación se reportan los resultados de un ANOVA de una vía. Utilice el nivel de significancia 0.05.

ANOVA				
Fuente de variación	**SS**	***gl***	**MS**	***F***
Entre grupos	6.892202	2	3.446101	4.960047
Dentro de los grupos	12.50589	18	0.694772	
Total	19.3981	20		

Responda las siguientes preguntas.

a. ¿Cuántos tratamientos hay en el estudio?
b. ¿Cuál es el tamaño total de la muestra?
c. ¿Cuál es el valor crítico de *F*?
d. Escriba la hipótesis nula y la hipótesis alternativa.
e. ¿Cuál es su decisión con respecto a la hipótesis nula?
f. ¿Es posible concluir que las medias de tratamiento son diferentes?

13 Regresión lineal y correlación

UN SERVICIO DE VIAJES AÉREOS muestrea los vuelos de una aerolínea doméstica para explorar la relación entre sus tarifas y las distancias recorridas. Si existe una relación, ¿qué porcentaje de la variación en la tarifa está representado por la distancia? ¿Cuánto añade a la tarifa una milla extra? (vea el ejercicio 61 y los OA13-2, OA13-3 y OA13-5).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA13-1 Explicar el propósito del análisis de correlación.

OA13-2 Calcular un coeficiente de relación para probar e interpretar la relación entre dos variables.

OA13-3 Aplicar un análisis de regresión para estimar la relación lineal entre dos variables.

OA13-4 Evaluar la significancia de la pendiente de la ecuación de regresión.

OA13-5 Evaluar la capacidad de una ecuación de regresión para predecir utilizando el error estándar de estimación y el coeficiente de determinación.

OA13-6 Calcular e interpretar los intervalos de confianza y de predicción.

OA13-7 Utilizar una función de logaritmo para transformar una relación no lineal.

Introducción

En los capítulos 2 a 4 se abordó la *estadística descriptiva*. Los datos sin procesar se organizaron en una distribución de frecuencias, y se calcularon varias medidas de ubicación y dispersión para describir las características importantes de la distribución. En los capítulos 5 a 7 se describió la probabilidad y, a partir de algunos enunciados, se crearon distribuciones de probabilidad. En el capítulo 8 se inició el estudio de la *inferencia estadística,* donde se recolectó una muestra para estimar un parámetro poblacional, como la media poblacional o la proporción de la población. Además, se utilizaron los datos de la muestra para probar una inferencia o hipótesis acerca de una media poblacional o una proporción poblacional, la diferencia entre dos de estas medias, o si varias eran iguales. Todas las pruebas implicaron solo *una* variable de intervalo (o de razón), como la ganancia que se obtiene por la venta de un auto, el ingreso de los presidentes de un banco o el número de pacientes admitidos cada mes en un hospital.

Este capítulo se enfoca en el estudio de la relación entre dos variables de intervalo o de razón. En todos los campos de negocios, identificar y estudiar las relaciones entre variables puede proporcionar información sobre las formas de elevar las ganancias, los métodos para reducir los costos o las variables para predecir la demanda. Para comercializar sus productos y aumentar sus ventas, muchas empresas reducen sus precios a través de cupones y descuentos; en este ejemplo, lo que interesa es la relación entre dos variables: la reducción de precios y las ventas. Para recabar datos, una compañía puede probar en el mercado una variedad de métodos de reducción de precios y observar el comportamiento de las ventas. Usted encontrará que muchas relaciones entre ambas variables constituyen la base de la economía, tales como el precio y la demanda.

A manera de otro ejemplo, recuerde que en el capítulo 4 se utilizaron los datos del Applewood Auto Group para mostrar la relación entre dos variables con un diagrama de dispersión. Se graficó la ganancia por vehículo en el eje vertical, y la edad del comprador, en el eje horizontal. Vea la salida del software estadístico en el primer ejemplo del capítulo 4. En ese diagrama se observó que, conforme aumentaba la edad del comprador, la ganancia por vehículo también se incrementaba.

Otros ejemplos de relaciones entre dos variables son:

- ¿Existe alguna relación entre la cantidad que Healthtex gasta por mes en publicidad y sus ventas mensuales?
- ¿El número de metros cuadrados en una casa está relacionado con su costo de calefacción en enero?
- En un estudio de eficiencia de combustible, ¿existe una relación entre las millas por galón y el peso del auto?
- ¿Hay alguna relación entre el número de horas que estudian los alumnos para un examen y la calificación que obtienen?

En este capítulo se amplía esta idea; es decir, se desarrollan medidas numéricas para expresar la relación entre dos variables. ¿Es fuerte o débil la relación, o es directa o inversa? Además, se desarrolla una ecuación para expresar la relación entre variables, lo que permite estimar una variable con base en otra.

Para comenzar el estudio de las relaciones entre ambas variables, se examinan el significado y el propósito de un **análisis de correlación**. Después se desarrolla una ecuación matemática que permita estimar el valor de una variable con base en el valor de otra; este procedimiento se llama **análisis de regresión.** También se evaluará la capacidad de la ecuación para hacer estimaciones correctas.

ESTADÍSTICA EN ACCIÓN

El transbordador espacial *Challenger* explotó el 28 de junio de 1986. Una investigación para determinar la causa examinó a cuatro contratistas: Rockwell International, responsable del transbordador y motores; Lockheed Martin, del apoyo terrestre; Martin Marietta, de los tanques de combustible externos, y Morton Thiokol, de los cohetes aceleradores de combustible sólido. Después de varios meses de investigación se determinó que los empaques en forma de "O" producidos por Morton Thiokol eran responsables de la explosión. Un estudio de los precios de las acciones del contratista reveló algo interesante; el día del accidente, los títulos de Morton Thiokol bajaron 11.86% y los de los otros tres contratistas solo perdieron de 2% a 3%. ¿Es posible concluir que en los mercados financieros se anticipó el resultado de la investigación?

¿Qué es el análisis de correlación?

OA13-1
Explicar el propósito del análisis de correlación.

Cuando se estudia la relación entre dos variables en escala de intervalo (o de razón) es usual comenzar con un diagrama de dispersión porque este procedimiento proporciona una representación visual de la relación entre las variables. A continuación se calcula el coeficiente de correlación, que brinda una medida cuantitativa de la fuerza de la relación entre dos variables. Como ejemplo, suponga que el gerente de ventas de North American Copier Sales, que tiene una fuerza de ventas muy grande en Estados Unidos y Canadá, desea determinar si hay alguna relación entre el número de llamadas de ventas en un mes y la cantidad de copiadoras que se vendieron durante este. El gerente selecciona una muestra aleatoria de 15 representantes de ventas y determina el número de

TABLA 13.1 Número de llamadas de ventas y copiadoras vendidas por 15 vendedores

Representantes de ventas	Número de llamadas de ventas	Número de copiadoras vendidas
Brian Virost	96	41
Carlos Ramírez	40	41
Carol Saia	104	51
Greg Fish	128	60
Jeff Hall	164	61
Mark Reynolds	76	29
Meryl Rumsey	72	39
Mike Kiel	80	50
Ray Snarsky	36	28
Rich Niles	84	43
Ron Broderick	180	70
Sal Spina	132	56
Soni Jones	120	45
Susan Welch	44	31
Tom Keller	84	30

llamadas de ventas que cada uno hizo el mes anterior y las copiadoras que vendió. La información muestral se registra en la tabla 13.1.

Al revisar los datos parece haber una relación entre el número de llamadas de ventas y el de unidades vendidas, es decir, los vendedores que hicieron más llamadas de venta vendieron más unidades; sin embargo, la relación no es "perfecta" o exacta, por ejemplo, Soni Jones hizo menos llamadas de ventas que Jeff Hall, pero vendió más unidades.

Además de las técnicas de graficado expuestas en el capítulo 4, se desarrollarán mediciones numéricas para representar de manera más precisa la relación entre ambas variables: "llamadas de ventas" y "copiadoras vendidas". Este grupo de técnicas estadísticas se denomina **análisis de correlación.**

ANÁLISIS DE CORRELACIÓN Grupo de técnicas para medir la asociación entre dos variables.

La idea básica del análisis de correlación es reportar la asociación entre dos variables; por lo general, el primer paso es trazar los datos en un **diagrama de dispersión**. Mediante un ejemplo se ilustra cómo emplear un diagrama de dispersión.

EJEMPLO

North American Copier Sales vende copiadoras a empresas de todos los tamaños en Estados Unidos y Canadá. Hace poco ascendieron a la señora Marcy Bancer al puesto de gerente nacional de ventas. Los representantes de todo el país asistirán a la siguiente junta de ventas, y ella desea destacar la importancia de hacer una última llamada de ventas adicional cada día; por lo tanto, decide reunir información sobre la relación entre el número de llamadas de ventas y el de copiadoras vendidas, así que selecciona una muestra aleatoria de 15 representantes y determina el número de llamadas que hicieron el mes anterior y las copiadoras que vendieron. La información muestral se reporta en la tabla 13.1, ¿qué observaciones cabe hacer respecto de la relación entre el número de llamadas de ventas y la cantidad de copiadoras vendidas? Elabore un diagrama de dispersión para representar la información.

SOLUCIÓN

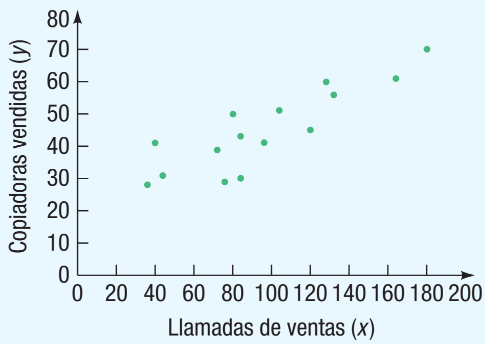
Con base en la información que se encuentra en la tabla 13.1, la señora Bancer sospecha que hay una relación entre el número de llamadas de venta hechas en un mes y la cantidad de copiadoras vendidas. Ron Broderick vendió más copiadoras el mes anterior, y realizó 180 llamadas o más; por otro lado, Ray Snarsky, Carlos Ramírez y Susan Welch hicieron el menor número de llamadas: 36, 40 y 44 y fueron quienes vendieron el menor número de copiadoras entre los representantes muestreados.

La implicación es que el número de copiadoras vendidas se relaciona con el de llamadas de ventas. Conforme aumenta el número de llamadas de venta, parece que la cantidad de copiadoras vendidas también lo hace. De este modo, el número de llamadas de ventas se considera la **variable independiente,** y el de copiadoras vendidas, la **variable dependiente.**

La variable independiente proporciona la base para estimar o predecir la variable dependiente; por ejemplo, se desea predecir el número esperado de copiadoras que se venderán si un representante realiza 100 llamadas de ventas. En los datos de muestra seleccionados al azar, la variable independiente, las llamadas de ventas, es un número aleatorio.

La variable dependiente es la que se desea predecir o estimar; también se puede describir como el resultado de un valor conocido de la variable independiente. La variable dependiente es aleatoria, esto es, por cada valor dado a la variable independiente, existen muchos posibles resultados para aquella.

Es práctica común situar la variable dependiente ("copiadoras vendidas") en el eje vertical o *Y* y la variable independiente ("número de llamadas de ventas") en el eje horizontal o *X*. Para elaborar un diagrama de dispersión de la información de North American Copier Sales, inicie con el primer representante de ventas, Brian Virost, quien hizo 96 llamadas el mes anterior y vendió 41 copiadoras, por lo cual $x = 96$ y $y = 41$. Para trazar esta información, a partir del origen vaya por el eje horizontal hasta el valor $x = 96$, después haga lo mismo en el eje vertical hasta $y = 41$ y marque un punto en la intersección. Continúe este proceso hasta trazar todos los datos pareados, como se muestra en la gráfica 13.1.



GRÁFICA 13.1 Diagrama de dispersión que representa las llamadas de ventas y las copiadoras vendidas

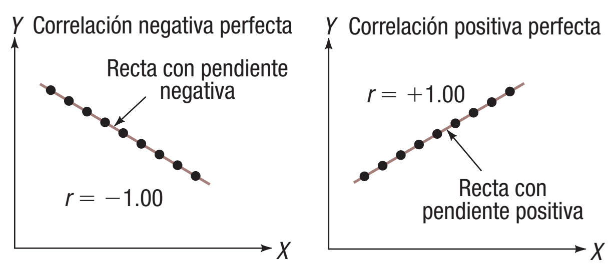
En el diagrama de dispersión se muestra en forma gráfica que los representantes que hacen más llamadas tienden a vender más copiadoras; es razonable que la señora Bancer, gerente nacional de ventas, diga a sus vendedores que, entre más llamadas de ventas hagan, se espera que vendan más copiadoras. Observe que, aunque parece haber una relación positiva entre ambas variables, no todos los puntos se ubican en una recta. En la siguiente sección se miden la fuerza y la dirección de esta relación entre dos variables, para determinar el coeficiente de correlación.

Coeficiente de correlación

OA13-2 Calcular un coeficiente de relación para probar e interpretar la relación entre dos variables.

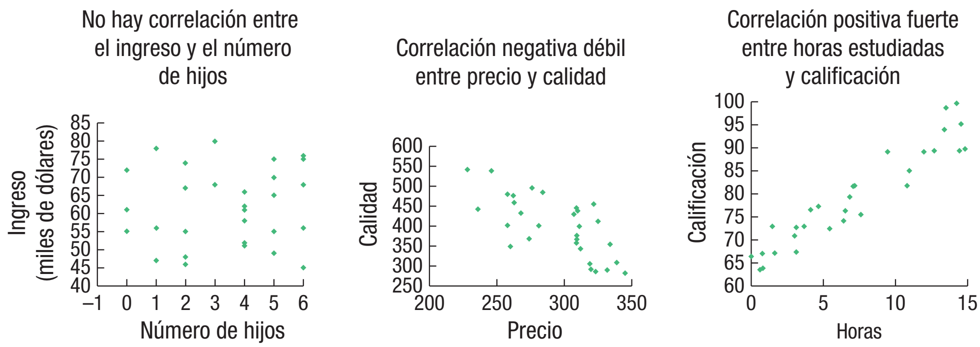
El **coeficiente de correlación,** creado por Karl Pearson alrededor de 1900, describe la fuerza de la relación entre dos conjuntos de variables en escala de intervalo o de razón. Se designa con la letra *r*, y con frecuencia se le conoce como *r de Pearson* o *coeficiente de correlación producto-momento*. Puede adoptar cualquier valor desde -1.00 hasta $+1.00$, inclusive. Un coeficiente de correlación de -1.00 o de $+1.00$ indica una *correlación perfecta*; por ejemplo, un coeficiente de correlación para el caso anterior que se calcule a $+1.00$ indicaría que el número de llamadas de ventas y la cantidad de copiadoras que vende cada representante están perfectamente relacionados en un sentido lineal positivo; por otra parte, un valor calculado de -1.00 revela que las llamadas de ventas y las copiadoras vendidas están perfectamente relacionados en un sentido lineal inverso. En la gráfica 13.2 se muestra cómo aparecería el diagrama de dispersión si la relación entre los dos conjuntos de datos fuera lineal y perfecta.

Si no hay ninguna relación entre los dos conjuntos de variables, la *r* de Pearson es cero. Un coeficiente de correlación *r* cercano a 0 (sea 0.08) indica que la relación lineal es muy débil. Se llega a la misma conclusión si $r = -0.08$. Los coeficientes de -0.91 y $+ 0.91$ tienen una fuerza igual; ambos indican una correlación muy fuerte entre las dos variables; por lo tanto, *la fuerza de la correlación no depende de la dirección (ya sea − o bien +)*.



GRÁFICA 13.2 Diagramas de dispersión con correlación negativa perfecta y correlación positiva perfecta

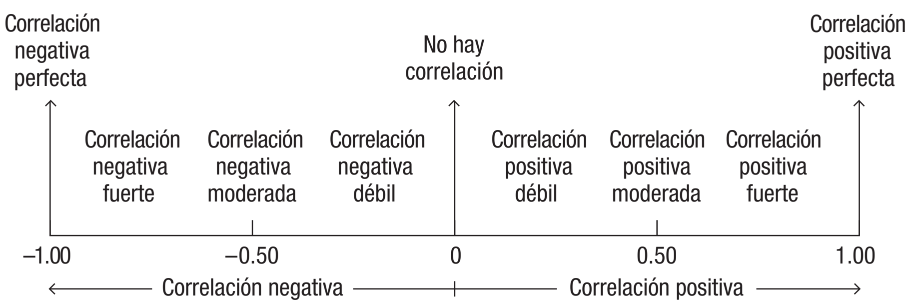
En la gráfica 13.3 se muestran los diagramas de dispersión cuando $r = 0$, una *r* débil (sea -0.23), y una *r* fuerte (sea $+0.87$). Observe que, si la correlación es débil, se presenta una dispersión considerable respecto de la recta trazada a través del centro de los datos. En el diagrama de dispersión que representa una



GRÁFICA 13.3 Diagrama de dispersión que representa correlación cero, débil y fuerte

fuerte relación, hay muy poca dispersión respecto de la recta; esto indica, en el ejemplo que se muestra en la gráfica, que las horas estudiadas constituyen un factor de pronóstico de la calificación en el examen.

En el siguiente diagrama se resumen la fuerza y la dirección del coeficiente de correlación.



COEFICIENTE DE CORRELACIÓN Medida de la fuerza de la relación lineal entre dos variables.

Las características del coeficiente de correlación se resumen a continuación.

CARACTERÍSTICAS DEL COEFICIENTE DE CORRELACIÓN

1. El coeficiente de correlación de la muestra se identifica con la letra minúscula *r*.
2. Muestra la dirección y fuerza de la relación lineal (recta) entre dos variables en escala de intervalo o de razón.
3. Varía desde -1 hasta $+1$, inclusive.
4. Un valor cercano a 0 indica que hay poca asociación entre las variables.
5. Un valor cercano a 1 indica una asociación directa o positiva entre las variables.
6. Un valor cercano a -1 indica una asociación inversa o negativa entre las variables.

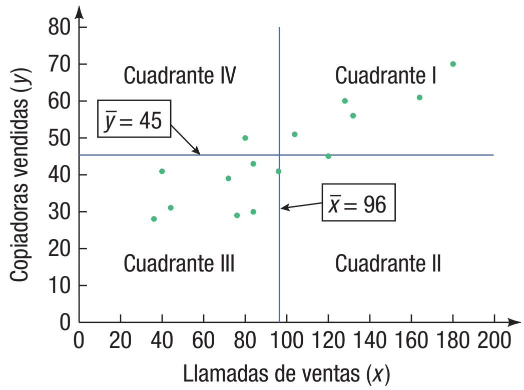
¿Cómo se determina el coeficiente de correlación? Como ejemplo, emplee los datos de North American Copier Sales, que se reportan en la tabla 13.1. Estos se repiten en la tabla 13.2 de la página siguiente para su conveniencia.

Comience con un diagrama de dispersión, similar al que se encuentra en la gráfica 13.2. Se traza una recta vertical con la media de los valores de *x* y una recta horizontal con la media de los de *y*. En la gráfica 13.4 (página siguiente) se agregó una recta en 96 llamadas ($\bar{x} = \sum x/n = 1\,440/15 = 96$) y una recta horizontal en 45.0 copiadoras ($\bar{y} = \sum y/n = 675/15{=}45.0$). Estas rectas pasan por el "centro" de los datos y dividen el diagrama de dispersión en cuatro cuadrantes. Considere mover el origen de (0, 0) a (96, 45).

Dos variables tienen una relación positiva cuando el número de copiadoras vendidas y las llamadas de ventas se encuentran arriba de la media; estos puntos se muestran en el cuadrante superior derecho (cuadrante I) de la gráfica 13.4. De manera similar, cuando el número de copiadoras vendidas es menor que la media, también lo es el número de llamadas de ventas; estos puntos se

TABLA 13.2 Número de llamadas de ventas y copiadoras vendidas por 15 vendedores

Representantes de ventas	Número de llamadas de ventas	Número de copiadoras vendidas
Brian Virost	96	41
Carlos Ramírez	40	41
Carol Saia	104	51
Greg Fish	128	60
Jeff Hall	164	61
Mark Reynolds	76	29
Meryl Rumsey	72	39
Mike Kiel	80	50
Ray Snarsky	36	28
Rich Niles	84	43
Ron Broderick	180	70
Sal Spina	132	56
Soni Jones	120	45
Susan Welch	44	31
Tom Keller	84	30
Total	1 440	675



GRÁFICA 13.4 Cálculo del coeficiente de correlación

ubican en el cuadrante inferior izquierdo de la gráfica 13.2 (cuadrante III); por ejemplo, la tercera persona de la lista en la tabla 13.2, Carol Saia, hizo 104 llamadas de ventas y vendió 51 copiadoras. Estos valores están arriba de sus medias respectivas, por lo que este punto se ubica en el cuadrante I, que es el cuadrante superior derecho; ella hizo 8 llamadas de ventas más que la media y vendió 6 copiadoras más que el número medio vendido. Tom Keller, el último nombre de la lista de la tabla 13.2, hizo 84 llamadas y vendió 30 copiadoras (ambos valores son menores que sus respectivas medias, por lo que este punto se ubica en el cuadrante inferior izquierdo); él hizo 12 llamadas menos y vendió 15 copiadoras menos que las medias respectivas. Las desviaciones del número medio de llamadas de ventas y del número medio de copiadoras vendidas de los 15 representantes de ventas se resumen en la tabla 13.3. La suma de los productos de las desviaciones de las medias respectivas es 6 672; es decir, el término $\sum(x - \bar{x})(y - \bar{y}) = 6\,672$.

En los cuadrantes superior derecho e inferior izquierdo, el producto de $(x - \bar{x})(y - \bar{y})$ es positivo debido a que ambos factores tienen el mismo signo. En el ejemplo, esto sucede con todos los representantes excepto Mike Kiel; quien realizó 80 llamadas de ventas (lo cual es menos que la

TABLA 13.3 Desviaciones de la media y sus productos

Representante de ventas	Llamadas de ventas (x)	Copiadoras vendidas (y)	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
Brian Virost	96	41	0	−4	0
Carlos Ramírez	40	41	−56	−4	224
Carol Saia	104	51	8	6	48
Greg Fish	128	60	32	15	480
Jeff Hall	164	61	68	16	1 088
Mark Reynolds	76	29	−20	−16	320
Meryl Rumsey	72	39	−24	−6	144
Mike Kiel	80	50	−16	5	−80
Ray Snarsky	36	28	−60	−17	1 020
Rich Niles	84	43	−12	−2	24
Ron Broderick	180	70	84	25	2 100
Sal Spina	132	56	36	11	396
Soni Jones	120	45	24	0	0
Susan Welch	44	31	−52	−14	728
Tom Keller	84	30	−12	−15	180
Totales	1 440	675	0	0	6 672

media) pero vendió 50 máquinas (lo cual es más que la media); por lo tanto, se espera que el coeficiente de correlación tenga un valor positivo.

Si las dos variables tienen una relación inversa, una estará arriba y la otra debajo de la media; en este caso, la mayoría de los puntos se ubican en los cuadrantes superior izquierdo e inferior derecho, es decir, en los cuadrantes II y IV. Ahora $(x - \bar{x})$ y $(y - \bar{y})$ tendrán signos opuestos, y su producto será negativo; por lo tanto, el coeficiente de correlación resultante es negativo.

¿Qué sucede si no hay una relación lineal entre ambas variables? Los puntos en el diagrama de dispersión aparecerán en los cuatro cuadrantes. Los productos negativos de $(x - \bar{x})(y - \bar{y})$ equilibran los productos positivos, por lo cual la suma es cero. Esto lleva al coeficiente de correlación cercano a cero; de esta manera, el término $\sum(x - \bar{x})(y - \bar{y})$ conduce la fuerza y el signo de la relación entre las dos variables.

El coeficiente de correlación no se ve afectado por las unidades de las dos variables; por ejemplo, si se hubieran empleado cientos de copiadoras vendidas en lugar del número real, el coeficiente de correlación sería el mismo. El coeficiente de correlación es independiente de la escala empleada si se divide el término $\sum(x - \bar{x})(y - \bar{y})$ entre las desviaciones estándar muestrales. También es independiente del tamaño muestral y se acota por los valores $+1.00$ y -1.00 si se divide entre $(n - 1)$.

Este razonamiento conduce a la siguiente fórmula:

COEFICIENTE DE CORRELACIÓN

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y} \qquad \textbf{[13.1]}$$

Para calcular el coeficiente de correlación, se utiliza la desviación estándar muestrales de 10 llamadas de ventas y 10 copiadoras vendidas. Se puede emplear la fórmula [3.12] para calcular la desviación estándar muestral o un paquete de software estadístico. Para los comandos específicos en Excel y Minitab, vea la sección “Comandos de software” en el apéndice C. He aquí la salida de pantalla de Excel; la desviación estándar del número de llamadas de ventas es 42.76, y del número de copiadoras vendidas, 12.89.

	A	B	C	D	E	F	G	H
1		Representante de ventas	Llamadas de ventas (x)	Copiadoras vendidas (y)			Llamadas de ventas (x)	Copiadoras vendidas (y)
2		Brian Virost	96	41		Media	96.00	45.00
3		Carlos Ramirez	40	41		Error estándar	11.04	3.33
4		Carol Saia	104	51		Mediana	84.00	43.00
5		Greg Fish	128	60		Moda	84.00	41.00
6		Jeff Hall	164	61		Desviación estándar	42.76	12.89
7		Mark Reynolds	76	29		Varianza de la muestra	1828.57	166.14
8		Meryl Rumsey	72	39		Curtosis	-0.32	-0.73
9		Mike Kiel	80	50		Sesgo	0.46	0.36
10		Ray Snarsky	36	28		Rango	144.00	42.00
11		Rich Niles	84	43		Mínimo	36.00	28.00
12		Ron Broderick	180	70		Máximo	180.00	70.00
13		Sal Spina	132	56		Suma	1440.00	675.00
14		Soni Jones	120	45		Cuenta	15.00	15.00
15		Susan Welch	44	31				
16		Tom Keller	84	30				
17		Total	1440	675				

Ahora se sustituyen estos valores en la fórmula [13.1] para determinar el coeficiente de correlación:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y} = \frac{6\,672}{(15 - 1)(42.76)(12.89)} = 0.865$$

¿Cómo se interpreta una correlación de 0.865? Primero, es positiva, por lo que se observa una relación directa entre el número de llamadas de ventas y la cantidad de copiadoras vendidas. Esto confirma el razonamiento basado en el diagrama de dispersión que se muestra en la gráfica 13.4. El valor de 0.865 es muy cercano a 1.00, y por ende se concluye que la asociación es fuerte.

Debe tener mucho cuidado con la interpretación; la correlación de 0.865 indica una asociación positiva lineal fuerte entre las variables. La señora Bancer acierta al motivar al personal de ventas para hacer llamadas adicionales debido a que el número de llamadas se relaciona con la cantidad de copiadoras que se venden; sin embargo, ¿más llamadas de ventas *ocasionan* más ventas? No, aquí no se ha demostrado la causa y el efecto, solo que las dos variables (“llamadas de ventas” y “copiadoras vendidas”) están relacionadas.

Si hay una relación fuerte (sea 0.97) entre dos variables, es factible suponer que un aumento o una disminución en una de ellas *causa* un cambio en la otra; por ejemplo, se puede demostrar que el consumo de cacahuates de Georgia y el consumo de aspirina tienen una correlación fuerte; sin embargo, esto no indica que un aumento del consumo de cacahuates *hizo* crecer el consumo de aspirina. De igual forma, los ingresos de profesores y el número de pacientes en instituciones psiquiátricas han aumentado en forma proporcional. Además, a medida que disminuye la población de burros, aumenta el número de grados doctorales otorgados. Las relaciones de este tipo se denominan **correlaciones espurias.** Lo que se puede concluir cuando se tienen dos variables con fuerte correlación es que hay una relación o asociación entre ambas variables, no que un cambio en una ocasiona una modificación en la otra.

EJEMPLO

El departamento de marketing de Applewood Auto Group piensa que los compradores más jóvenes adquieren vehículos que generan menos ganancias, contrario a lo que sucede en el caso de los compradores de mayor edad; por lo tanto, quisiera usar esta información como parte de una próxima campaña de publicidad para tratar de atraer a compradores mayores y obtener así más ganancias. Desarrolle un diagrama de dispersión que refleje la relación entre la ganancia que genera cada vehículo y la edad del comprador. Utilice un software estadístico para determinar el coeficiente de correlación. ¿Será este un elemento útil para la publicidad?

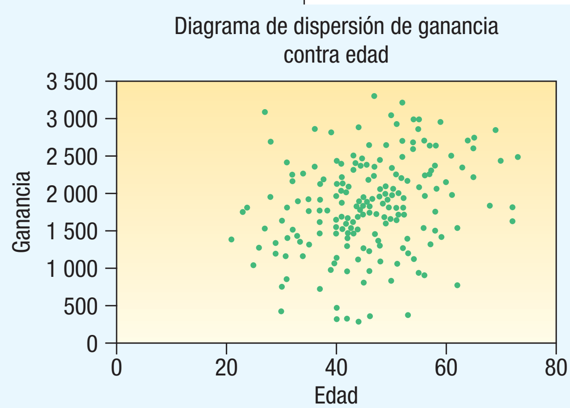
SOLUCIÓN

Utilizando el ejemplo de Applewood Auto Group, el primer paso es generar una gráfica de los datos mediante un diagrama de dispersión, tal como el que se muestra en la gráfica 13.5.

El diagrama de dispersión sugiere que existe una posible relación entre la edad y la ganancia; sin embargo, no parece que esta relación sea fuerte.

A continuación se calcula el coeficiente de correlación para evaluar la fuerza relativa de la relación. El software estadístico proporciona una forma sencilla de calcular el valor del coeficiente de correlación, como se muestra en la siguiente salida de Excel.

	Age	*Profit*
Age	1	
Profit	0.262	1



GRÁFICA 13.5 Diagrama de dispersión de los datos "ganancia" frente a "edad" de Applewood Auto Group

En el caso de estos datos, $r = 0.262$. He aquí la evaluación de la relación entre la edad del comprador y la ganancia que genera la venta de un auto:

1. La relación es positiva o directa. ¿Por qué? Porque el signo del coeficiente de correlación es positivo. Esto confirma que a medida que aumenta la edad del comprador, se eleva también la ganancia que genera la venta del vehículo.
2. La relación entre ambas variables es débil. En el caso de una relación positiva, los valores del coeficiente de correlación cercanos a uno indica relaciones más fuertes; en este caso, $r = 0.262$, lo cual es más cercana a cero, indicando que la relación no es muy fuerte.

No se recomienda que Applewood utilice esta información como parte de una campaña de publicidad para atraer a compradores mayores que dejen mayores ganancias.

AUTOEVALUACIÓN 13-1

Haverty's Furniture es un negocio familiar que vende a clientes minoristas en el área de Chicago desde hace muchos años. Tanto en radio como en televisión e internet, la compañía destaca sus precios bajos y fáciles términos de crédito; por tanto, el propietario desea analizar la relación entre las ventas y la suma de dinero que gastó en publicidad. A continuación se presenta la información de las ventas y los gastos publicitarios durante los últimos cuatro meses.

Mes	Gastos publicitarios (en millones de dólares)	Ingresos por ventas (en millones de dólares)
Julio	2	7
Agosto	1	3
Septtiembre	3	8
Octubre	4	10

(a) El propietario desea pronosticar las ventas con base en los gastos publicitarios. ¿Cuáles son las variables dependiente e independiente?
(b) Trace un diagrama de dispersión.
(c) Determine el coeficiente de correlación.
(d) Interprete la fuerza del coeficiente de correlación.

EJERCICIOS

1. Las siguientes observaciones muestrales se seleccionaron de manera aleatoria.

x	4	5	3	6	10
y	4	6	5	7	7

Determine el coeficiente de correlación e interprete la relación entre *x* y *y*.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2. Las siguientes observaciones muestrales se seleccionaron de manera aleatoria.

x	5	3	6	3	4	4	6	8
y	13	15	7	12	13	11	9	5

Determine el coeficiente de correlación e interprete la relación entre *x* y *y*.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

3. Bi-lo Appliance Super-Store tiene tiendas en varias áreas metropolitanas de Nueva Inglaterra. La gerente general de ventas planea transmitir un comercial de una cámara digital en estaciones de televisión locales antes del periodo de ventas que empezará el sábado y terminará el domingo; ella planea obtener la información de las ventas de la cámara digital durante el sábado y el domingo en las diversas tiendas y compararlas con el número de veces que se transmitió el anuncio en las estaciones de televisión. El propósito es determinar si hay alguna relación entre el número de veces que se transmitió el anuncio y las ventas de cámaras digitales. Los pares son:

Ubicación de la estación de TV	Número de transmisiones	Ventas de sábado a domingo (miles de dólares)
Providence	4	15
Springfield	2	8
New Haven	5	21
Boston	6	24
Hartford	3	17

a. ¿Cuál es la variable dependiente?
b. Trace un diagrama de dispersión.
c. Determine el coeficiente de correlación.
d. Interprete estas medidas estadísticas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

4. El departamento de producción de Celltronics International desea explorar la relación entre el número de empleados que trabajan en una línea de ensamblado parcial y la cantidad de unidades producida. Como experimento, se asignó a dos empleados al ensamblado parcial, cuyo desempeño fue de 15 productos durante un periodo de una hora. Después, cuatro empleados hicieron los ensamblados y su número fue de 25 durante un periodo idéntico. El conjunto completo de observaciones pareadas se muestra a la derecha.

La variable dependiente es la producción; es decir, se supone que el nivel de producción depende del número de empleados.

Número de ensambladores	Producción en una hora (unidades)
2	15
4	25
1	10
5	40
3	30

a. Trace un diagrama de dispersión.
b. Con base en el diagrama de dispersión, ¿parece haber alguna relación entre el número de ensambladores y la producción? Explique.
c. Calcule el coeficiente de correlación.

5. El consejo de la ciudad de Pine Bluffs considera aumentar el número de policías en un esfuerzo para reducir los delitos. Antes de tomar una decisión final, el ayuntamiento pide al jefe de policía realizar una encuesta en otras ciudades de tamaño similar para determinar la relación entre el número de policías y el de delitos reportados. El jefe de policía reunió la siguiente información muestral.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Ciudad	Policías	Número de delitos	Ciudad	Policías	Número de delitos
Oxford	15	17	Holgate	17	7
Starksville	17	13	Carey	12	21
Danville	25	5	Whistler	11	19
Athens	27	7	Woodville	22	6

a. ¿Cuáles son las variables dependiente e independiente? Pista: ¿Cuál de las siguientes afirmaciones tiene más sentido: las ciudades con más policías tienen menos crímenes o las ciudades con menos crímenes tienen más policías? Explique su respuesta.
b. Trace un diagrama de dispersión.
c. Determine el coeficiente de correlación.
d. Interprete el coeficiente de correlación. ¿Le sorprende que sea negativo?

6. El propietario de Maumee Ford-Mercury-Volvo desea estudiar la relación entre la antigüedad de un automóvil y su precio de venta. En la siguiente lista se observa una muestra aleatoria de 12 automóviles usados que vendió el concesionario durante el año anterior.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Automóvil	Antigüedad (años)	Precio de venta (miles de dólares)	Automóvil	Antigüedad (años)	Precio de venta (miles de dólares)
1	9	8.1	7	8	7.6
2	7	6.0	8	11	8.0
3	11	3.6	9	10	8.0
4	12	4.0	10	12	6.0
5	8	5.0	11	6	8.6
6	7	10.0	12	6	8.0

a. Trace un diagrama de dispersión.
b. Establezca el coeficiente de correlación.
c. Interprete el coeficiente de correlación. ¿Le sorprende que sea negativo?

Prueba de la importancia del coeficiente de correlación

Recuerde que la gerente de ventas de North American Copier Sales determinó que la correlación entre el número de llamadas de ventas y el de copiadoras vendidas era 0.865, lo que indicaba una asociación fuerte entre ambas variables; sin embargo, en la muestra había solo 15 vendedores. ¿Puede ser que la correlación entre la población sea cero? Esto significaría que la correlación de 0.865 se debió a la casualidad. En este ejemplo, la población es todo el personal de ventas de la empresa.

Resolver este dilema requiere una prueba para responder la pregunta obvia: ¿puede haber una correlación cero entre la población de la cual se seleccionó la muestra? En otras palabras, ¿el valor r calculado proviene de una población de observaciones pareadas con correlación cero? Para continuar la convención de usar letras griegas para representar un parámetro poblacional, ρ (se pronuncia “rho”) representará la correlación entre la población.

Continuaremos con el ejemplo de las llamadas de ventas y copiadoras vendidas, y se emplearán las mismas pruebas de hipótesis descritas en el capítulo 10. La hipótesis nula y la hipótesis alternativa son:

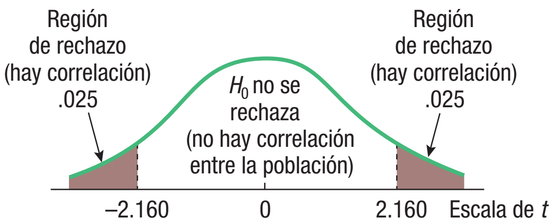
H_0: $\rho = 0$ (La correlación entre la población es cero).
H_1: $\rho \neq 0$ (La correlación entre la población es diferente de cero).

Por la forma en que se formula H_1, se sabe que la prueba es de dos colas.

La fórmula para t es:

PRUEBA *t* DEL COEFICIENTE DE CORRELACIÓN

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad \text{con } n - 2 \text{ grados de libertad} \qquad \textbf{[13.2]}$$



GRÁFICA 13.6 Regla de decisión en la prueba de hipótesis con el nivel de significancia 0.05 y 13 *gl*

Con el nivel de significancia 0.05, la regla de decisión en este caso indica que si el valor calculado de t se encuentra en el área entre $+2.160$ y -2.160, entonces no se rechaza la hipótesis nula. Para ubicar el valor crítico de 2.160, consulte el apéndice B.2 para $gl = n - 2 = 15 - 2 = 13$ (vea la gráfica 13.6).

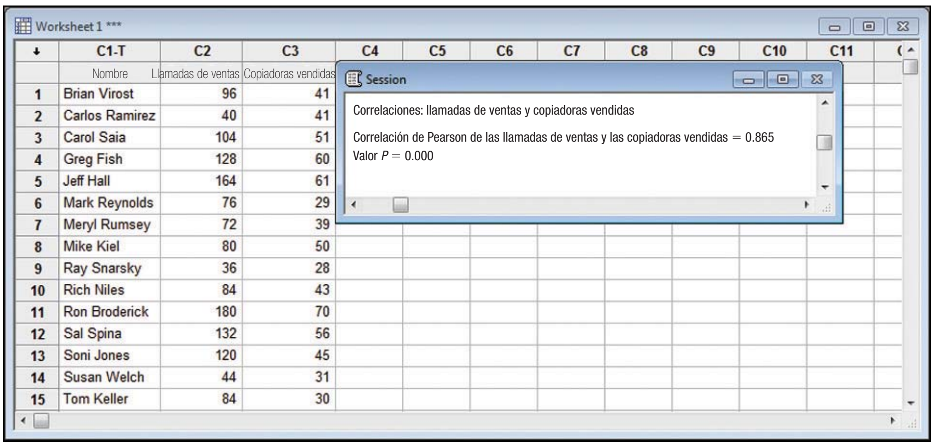
Si aplica la fórmula [13.2] al ejemplo de la relación entre número de llamadas de ventas y unidades vendidas:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.865\sqrt{15-2}}{\sqrt{1-0.865^2}} = 6.216$$

El valor t calculado está en la región de rechazo; por ello, H_0 se rechaza con el nivel de significancia 0.05. Esto significa que la correlación entre la población no es cero. Desde un punto de vista práctico, esto indica a la gerente de ventas que hay una correlación entre el número de llamadas de ventas y el de copiadoras vendidas en la población de vendedores.

La prueba de hipótesis también se interpreta en términos de valores p; los cuales representan la probabilidad de determinar un valor del estadístico de prueba más extremo que el calculado, cuando H_0 es verdadera. Para determinar el valor p, consulte la distribución t en el apéndice B.5 y ubique la fila de 13 grados de libertad (el estadístico de prueba es 6.216), encuentre el valor más cercano a 6.216 en dicha fila, considerando una prueba de dos colas. Observe que en una prueba de dos colas, con el nivel de significancia 0.001, el valor crítico es 4.221; como 6.216 es mayor que 4.221, se concluye que el valor p es menor a 0.001.

Tanto Minitab como Excel reportan la correlación entre dos variables; además, Minitab reporta el valor p de la prueba de hipótesis en que la correlación entre la población entre dos variables sea 0. A continuación se presenta la salida de Minitab.



Worksheet 1 ***

	C1-T Nombre	C2 Llamadas de ventas	C3 Copiadoras vendidas
1	Brian Virost	96	41
2	Carlos Ramirez	40	41
3	Carol Saia	104	51
4	Greg Fish	128	60
5	Jeff Hall	164	61
6	Mark Reynolds	76	29
7	Meryl Rumsey	72	39
8	Mike Kiel	80	50
9	Ray Snarsky	36	28
10	Rich Niles	84	43
11	Ron Broderick	180	70
12	Sal Spina	132	56
13	Soni Jones	120	45
14	Susan Welch	44	31
15	Tom Keller	84	30

Session

Correlaciones: llamadas de ventas y copiadoras vendidas

Correlación de Pearson de las llamadas de ventas y las copiadoras vendidas = 0.865
Valor P = 0.000

EJEMPLO

En el ejemplo que contiene la gráfica 13.5 se determinó que el coeficiente de relación entre la ganancia en la venta de un vehículo de Applewood Auto Group y la edad de la persona que compró dicho vehículo era de 0.262. Dado que el signo del coeficiente de correlación fue positivo, se concluyó que existía una relación directa entre ambas variables; sin embargo, debido a que la cifra de correlación era baja —esto es, cercana a cero—, se estableció que nada garantizaba que una campaña de publicidad dirigida a los compradores mayores generara una ganancia más grande. Es posible confirmar esta conclusión con una prueba de hipótesis que demuestre que el coeficiente de correlación es mayor a cero, con el nivel de significancia 0.05.

SOLUCIÓN

Para probar la hipótesis, se deben aclarar los temas de la muestra y la población. Asuma que los datos recolectados de los 180 vehículos vendidos por Applewood Group es una muestra de la población de *todos* los vehículos que la empresa vendió durante muchos años. La letra griega ρ es el coeficiente de relación entre la población, y r es el coeficiente de relación entre la muestra.

Ahora se establecen las hipótesis nula y alternativa. Hay que probar la hipótesis nula; es decir, que el coeficiente de correlación es igual o menor a cero. La hipótesis alternativa es que existe una correlación positiva entre ambas variables.

H_0: $\rho \leq 0$ (la correlación entre la población es negativa o igual a cero).
H_1: $\rho > 0$ (la correlación entre la población es positiva).

Esta es una prueba de una cola porque el interés es confirmar una asociación positiva entre las variables. El estadístico de prueba sigue la distribución t, con $n = 2$ grados de libertad, así que los grados de libertad son $180 - 2 = 178$; sin embargo, la cifra de 178 grados de libertad no aparece en el apéndice B.5 (el valor más cercano es 180, de modo que es el que se utilizará). La regla de decisión es rechazar la hipótesis nula si el valor calculado del estadístico de prueba es mayor a 1.653.

Se utiliza la fórmula [13.2] para encontrar el valor del estadístico de prueba.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.262\sqrt{180-2}}{\sqrt{1-0.262^2}} = 3.622$$

Al comparar el estadístico de prueba (3.622) con el valor crítico (1.653) se rechaza la hipótesis nula; en conclusión, el coeficiente de correlación de la muestra de 0.262 es demasiado grande como para provenir de una población sin correlación. Para decirlo en otras palabras, existe una correlación positiva entre la ganancia y la edad de la población.

El resultado es confuso y, en apariencia, contradictorio; por una parte se observa que el coeficiente de correlación no indica que haya una relación muy fuerte, y que el departamento de marketing de Applewood Auto Group no debería usar esta información para tomar decisiones promocionales y publicitarias. Por otra, la prueba de la hipótesis indicó que el coeficiente de correlación no es igual a cero y que existe una relación positiva entre la edad y la ganancia. ¿Cómo puede ser esto? Es necesario ser muy cuidadosos con la interpretación de los resultados de la prueba de la hipótesis, la cual revela un resultado estadísticamente significativo; sin embargo, este resultado no necesariamente sustenta una decisión práctica de comenzar una nueva campaña de marketing y promoción para los antiguos compradores. De hecho, el coeficiente de relación relativamente bajo indica que el resultado de otra campaña de marketing y promoción a antiguos compradores potenciales es, en el mejor de los casos, incierta.

AUTOEVALUACIÓN 13-2

Una muestra de 25 campañas para la alcaldía de ciudades de tamaño medio con poblaciones entre 50 000 y 250 000 habitantes demostró que la correlación entre el porcentaje de los votos recibidos y la cantidad gastada en la campaña por cada candidato fue 0.43. Con el nivel de significancia 0.05, ¿hay una asociación positiva entre las variables?

EJERCICIOS

7. Se dan las siguientes hipótesis.

$$H_0: \rho \leq 0$$
$$H_1: \rho > 0$$

Una muestra aleatoria de 12 observaciones pareadas indicó una correlación de 0.32. ¿Se puede concluir que la correlación entre la población es mayor que cero? Utilice el nivel de significancia 0.05.

8. Se dan las siguientes hipótesis.

$$H_0: \rho \geq 0$$
$$H_1: \rho < 0$$

Una muestra aleatoria de 15 observaciones pareadas tiene una correlación de −0.46. ¿Se puede concluir que la correlación entre la población es menor que cero? Utilice el nivel de significancia 0.05.

9. La Pennsylvania Refining Company estudia la relación entre el precio de la gasolina y el número de galones que vende. En una muestra de 20 gasolineras el martes previo, la correlación fue 0.78. Con el nivel de significancia 0.01, la correlación entre la población, ¿será mayor que cero?

10. Un estudio de 20 instituciones financieras de todo el mundo reveló que la correlación entre sus activos y las utilidades antes del pago de impuestos es 0.86. Con el nivel de significancia 0.05, ¿se puede concluir que hay una correlación positiva entre la población?

11. La asociación de pasajeros de aerolíneas estudió la relación entre el número de pasajeros en un vuelo en particular y su costo. Parece lógico que más pasajeros impliquen más peso y más equipaje, lo que a su vez incrementará el gasto de combustible. Con una muestra de 15 vuelos, la correlación entre el número de pasajeros y el gasto total del combustible fue 0.667. ¿Es razonable concluir que hay una asociación positiva entre las dos variables poblacionales? Utilice el nivel de significancia 0.01.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

12. La Student Government Association, de la Middle Carolina University, desea demostrar la relación entre el número de cervezas que beben los estudiantes y su contenido de alcohol en la sangre (CAS). Una muestra de 18 estudiantes participó en un estudio en el cual a cada uno se le asignó al azar un número de latas de cerveza de 12 onzas que debía beber; 30 minutos después de consumir su número asignado de cervezas un miembro de la oficina local del alguacil midió el contenido de alcohol en la sangre. La información muestral se registra a continuación.

Estudiante	Cervezas	CAS	Estudiante	Cervezas	CAS
1	6	0.10	10	3	0.07
2	7	0.09	11	3	0.05
3	7	0.09	12	7	0.08
4	4	0.10	13	1	0.04
5	5	0.10	14	4	0.07
6	3	0.07	15	2	0.06
7	3	0.10	16	7	0.12
8	6	0.12	17	2	0.05
9	6	0.09	18	1	0.02

Utilice un paquete de software estadístico para realizar lo siguiente.

a. Elabore un diagrama de dispersión del número de cervezas consumidas y el contenido de alcohol en la sangre. Comente sobre la relación. ¿Parece fuerte o débil, directa o inversa?

b. Determine el coeficiente de correlación.

c. Con el nivel de significancia 0.01, ¿es razonable concluir que hay una relación positiva entre el número de cervezas consumidas y el contenido de alcohol en la sangre de la población? ¿Cuál es el valor *p*?

OA13-3 Aplicar un análisis de regresión para estimar la relación lineal entre dos variables.

Análisis de regresión

En la sección anterior se evaluó la dirección y significancia de la relación lineal entre dos variables al encontrar el coeficiente de correlación. El análisis de regresión es otro método para examinar una relación lineal entre dos variables; en este se utilizan los conceptos básicos de correlación, pero proporciona mucho más información al expresar la relación lineal entre dos variables en forma de una ecuación. Al utilizarla se puede estimar el valor de la variable dependiente Y con base en un valor seleccionado de la variable independiente X. La técnica para desarrollar la ecuación y proporcionar las estimaciones se denomina **análisis de regresión.**

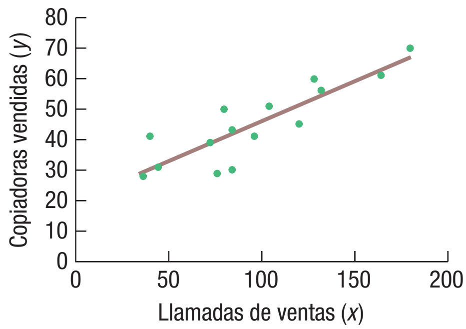
En la tabla 13.1 se reporta el número de llamadas de ventas y de unidades vendidas de una muestra de 15 representantes de ventas de North American Copier Sales. En la gráfica 13.1 se presenta esta información en un diagrama de dispersión; recuerde que se probó la significancia del coeficiente de correlación (r = 0.865) y se concluyó que existe una relación significativa entre ambas variables. Ahora se busca desarrollar una ecuación lineal que exprese la relación entre el número de llamadas de ventas (variable independiente) y las unidades vendidas (variable dependiente). A la ecuación de la recta para estimar Y con base en X se le denomina **ecuación de regresión.**

ECUACIÓN DE REGRESIÓN Expresa la relación lineal entre dos variables.

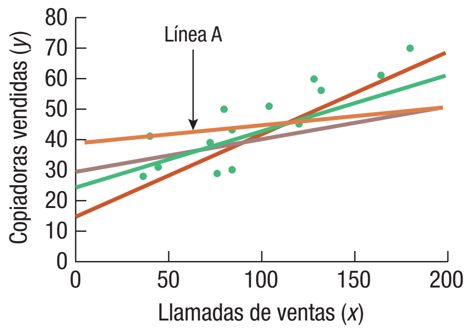
Principio de los mínimos cuadrados

En el análisis de regresión, el objetivo es utilizar los datos para trazar una línea que represente mejor la relación entre dos variables. El primer enfoque es utilizar un diagrama de dispersión para visualizar la posición de la línea.

El diagrama de dispersión de la gráfica 13.1 se reproduce en la gráfica 13.7 con una línea que une los puntos para ilustrar que una recta probablemente ajustaría los datos; sin embargo, cuando esta se traza con una regla hay una desventaja: en parte, su posición se basa en el criterio de la persona que la dibuja. Mediante las rectas trazadas a mano en la gráfica 13.8 se representan los criterios de cuatro personas; todas, excepto *A*, parecen razonables. Esto es, cada línea se centra entre los datos graficados; sin embargo, cada una generaría una estimación distinta de unidades vendidas para un número particular de llamadas de ventas.



GRÁFICA 13.7 Llamadas de ventas y copiadoras vendidas por 15 representantes de ventas



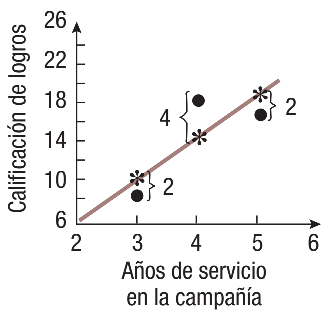
GRÁFICA 13.8 Cuatro rectas superpuestas en el diagrama de dispersión

Es preferible utilizar el método que resulta en una sola y mejor línea de regresión; este se denomina "principio de los mínimos cuadrados" y proporciona lo que comúnmente se conoce como recta del "mejor ajuste".

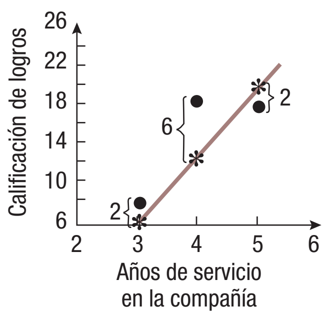
PRINCIPIO DE LOS MÍNIMOS CUADRADOS Determina una ecuación de regresión al minimizar la suma de los cuadrados de las distancias verticales entre los valores reales de *Y* y los que se pronosticaron.

Para ilustrar este concepto, se trazan los mismos valores en las tres gráficas siguientes; los puntos representan los de *y* (reales), y los asteriscos, los predichos de *y* para un valor dado de *x*. La recta de regresión de la gráfica 13.9 se determinó con el método de los mínimos cuadrados; la cual, es la recta de mejor ajuste porque la suma de los cuadrados de las desviaciones verticales respecto de sí misma es mínima. La primera gráfica ($x = 3$, $y = 8$) se desvía 2 unidades de la recta, calculada como $10 - 8$. El cuadrado de la desviación es 4; la desviación al cuadrado que se obtiene de la gráfica en $x = 4$, $y = 18$ es 16; la que se obtiene en $x = 5$, $y = 16$ es 4. La suma de las desviaciones al cuadrado es 24, calculada como $4 + 16 + 4$.

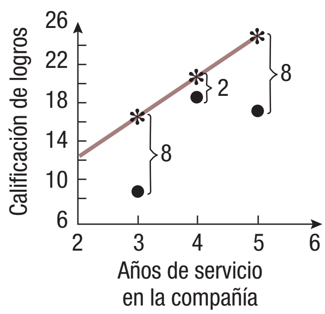
Suponga que las rectas de las gráficas 13.10 y 13.11 se trazaron con una regla. La suma de las desviaciones verticales al cuadrado que se encuentran en la gráfica 13.9 es 44; en el caso de la gráfica 13.10 es 132. Las dos sumas son mayores que la de la recta de la gráfica 13.9, la cual se determina mediante el método de los mínimos cuadrados.



GRÁFICA 13.9 Recta de mínimos cuadrados



GRÁFICA 13.10 Recta trazada con una regla



GRÁFICA 13.11 Recta diferente trazada con una regla

La ecuación de una recta tiene la forma

FORMA GENERAL DE LA ECUACIÓN DE REGRESIÓN LINEAL	$\hat{y} - a + bx$	**[13.3]**

donde:

- $\hat{y}$ (que se lee "*y* prima") es el valor de la estimación de la variable *y* para un valor *x* seleccionado;
- a es la intersección *y*. Es el valor estimado de *Y* cuando $x = 0$. En otras palabras, *a* es el valor estimado de *y* donde la recta de regresión cruza el eje *Y* cuando *x* es cero;
- b es la pendiente de la recta, o el cambio promedio en $\hat{y}$ por cada cambio de una unidad (ya sea aumento o reducción) de la variable independiente *x*;
- x es cualquier valor de la variable independiente que se seleccione.

La forma general de la ecuación de la regresión lineal es exactamente la misma que la ecuación de cualquier línea; es decir, *a* es la intersección con *Y* y *b* es la pendiente. El propósito de un análisis de regresión es calcular los valores de *a* y *b* para desarrollar una ecuación lineal que se ajuste mejor a estos.

Las fórmulas de *a* y *b* son:

PENDIENTE DE LA RECTA DE REGRESIÓN	$b = r\left(\frac{s_y}{s_x}\right)$	**[13.4]**

donde:

- r es el coeficiente de correlación;
- s_y es la desviación estándar de *y* (variable dependiente);
- s_x es la desviación estándar de *x* (variable independiente).

INTERSECCIÓN CON EL EJE *Y*	$a = \bar{y} - b\bar{x}$	**[13.5]**

donde:

- $\bar{y}$ es la media de *y* (variable dependiente).
- $\bar{x}$ es la media de *x* (variable independiente).

EJEMPLO

Recuerde el ejemplo de North American Copier Sales. La gerente de ventas reunió información sobre las llamadas realizadas y las copiadoras vendidas de una muestra de 15 representantes. Como parte de su presentación en la siguiente reunión, la señora Bancer desea proporcionar información específica acerca de la relación entre el número de llamadas y las ventas. Con el método de los mínimos cuadrados, determine una ecuación lineal que exprese la relación entre ambas variables. ¿Cuál es el número esperado de copiadoras vendidas de un representante que hizo 100 llamadas?

SOLUCIÓN

El primer paso para determinar la ecuación de regresión es encontrar la pendiente de la recta de regresión de mínimos cuadrados; es decir, se necesita el valor de *b*. Después de la fórmula [13.1] se determinó el coeficiente de correlación *r* (0.865), mientras que en la salida de Excel que aparece después de dicha fórmula se determinó la desviación estándar de la variable independiente *x* (42.76) y la desviación estándar de la variable dependiente *y* (12.89); los valores están insertados en la fórmula [13.4].

$$b = r\left(\frac{s_y}{s_x}\right) = 0.865\left(\frac{12.89}{42.76}\right) = 0.2608$$

Después se necesita encontrar el valor de *a*; para hacerlo, utilice el valor de *b* que acaba de calcular, así como las medias del número de llamadas de ventas y de la cantidad de copiadoras vendidas; estas también se encuentran en la pantalla de Excel antes mencionada. De la fórmula [13.5]:

$$a = \bar{y} - b\bar{x} = 45 - 0.2608(96) = 19.9632$$

Así, la ecuación de regresión es $\hat{y} = 19.9632 + 0.2608x$; por lo tanto, si un vendedor hace 100 llamadas, debería vender 46.0432 copiadoras, número que se determina por $\hat{y} = 19.9632 + 0.2608x = 19.9632 + 0.2608(100)$. El valor b de 0.2608 significa que por cada llamada de ventas adicional, el vendedor debería aumentar el número de copiadoras vendidas en aproximadamente 0.2608. En otras palabras, 20 llamadas de ventas adicionales en un mes generarán más o menos cinco copiadoras más vendidas, número determinado por 0.2608(20) = 5.216.

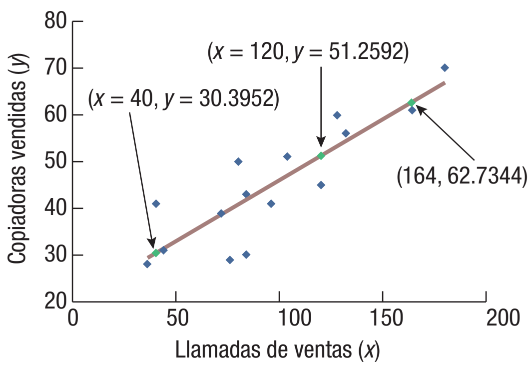
El valor a de 19.9632 es el punto donde la ecuación cruza el eje Y. Una traducción literal es que si no se hacen llamadas de ventas, es decir, $X = 0$, se venderán 19.9632 copiadoras. Observe que $x = 0$ está fuera del rango de valores incluidos en la muestra y, por lo tanto, no se debe emplear para estimar el número de copiadoras vendidas. Las llamadas de ventas varían de 36 a 180, por lo que las estimaciones se deben hacer dentro de ese rango.

Trazo de la recta de regresión

La ecuación de mínimos cuadrados, $\hat{y} = 19.9632 + 0.2608x$, se traza en el diagrama de dispersión. El quinto representante de ventas de la muestra es Jeff Hall, quien hizo 164 llamadas; su número estimado de copiadoras vendidas es $\hat{y} = 19.9632 + 0.2608(164) = 62.7344$. El punto $x = 164$ y $\hat{y} = 62.7344$ se ubica al moverse hasta 164 en el eje X y después, en sentido vertical, hasta 62.7344. Los puntos restantes en la ecuación de regresión se determinan al sustituir el valor particular de x en la ecuación de regresión y calcular $\hat{y}$. Todos los puntos se conectan para formar la recta (vea la gráfica 13.12).

La recta de regresión por mínimos cuadrados tiene algunas características interesantes y particulares. Primero, siempre pasa por el punto $(\bar{x}, \bar{y})$; para demostrar esto, se predice el número de

Representante de ventas	Llamadas de ventas (x)	Copiadoras vendidas (y)	Ventas estimadas ($\hat{y}$)
Brian Virost	96	41	45.0000
Carlos Ramírez	40	41	30.3952
Carol Saia	104	51	47.0864
Greg Fish	128	60	53.3456
Jeff Hall	164	61	62.7344
Mark Reynolds	76	29	39.7840
Meryl Rumsey	72	39	38.7408
Mike Kiel	80	50	40.8272
Ray Snarsky	36	28	29.3520
Rich Niles	84	43	41.8704
Ron Broderick	180	70	66.9072
Sal Spina	132	56	54.3888
Soni Jones	120	45	51.2592
Susan Welch	44	31	31.4384
Tom Keller	84	30	41.8704



GRÁFICA 13.12 Recta de regresión en el diagrama de dispersión

ESTADÍSTICA EN ACCIÓN

En finanzas, los inversionistas tienen interés en el intercambio entre ganancias y riesgo; una técnica para cuantificar este último es el análisis de regresión del precio accionario de una compañía (variable dependiente) y una medida promedio del mercado accionario (variable independiente). Con frecuencia se emplea el Índice 500 de Standard and Poor's (S&P) para estimar el mercado. El coeficiente de regresión, denominado beta en finanzas, muestra el cambio del precio de las acciones de una compañía ante un cambio de una unidad en el índice de S&P; por ejemplo, si una acción tiene un beta de 1.5, cuando el índice S&P aumenta 1%, su precio aumentará 1.5%. También sucede lo opuesto, si el índice S&P disminuye 1%, el precio de las acciones disminuirá 1.5%. Si su beta es 1.0, un cambio de 1% en el índice presentará un cambio de 1% en su precio; si es menor que 1.0, un cambio de 1% en el índice presenta un cambio menor a 1% del precio accionario.

copiadoras vendidas con el número medio de llamadas de ventas. En este ejemplo, el número medio de llamadas de ventas es 96, determinado por $\overline{x} = 1\,440/15$. El número medio de copiadoras vendidas es 45.0, que se calcula mediante $\overline{y} = 675/15$. Si $x = 96$ y luego se emplea la ecuación de regresión para encontrar el valor estimado de $\hat{y}$, el resultado es:

$$\hat{y} = 19.9632 + 0.2608(96) = 45$$

El número estimado de copiadoras vendidas es exactamente igual al número medio de copiadoras que realmente se vendieron. En este ejemplo sencillo se muestra que la recta de regresión pasa por el punto que representa a las dos medias; en este caso, la ecuación de regresión pasará por el punto $x = 96$ y $y = 45$.

Segundo, como se analizó antes en esta sección, no hay otra recta que pase por los datos donde la suma de las desviaciones al cuadrado sea menor. En otras palabras, el término $\sum(y - \hat{y})^2$ es menor cuando se aplica la ecuación de regresión por mínimos cuadrados que en cualquier otro caso. A continuación se emplea Excel para demostrar esta condición.

	A	B	C	D	E	F	G	H	I	J
1	Representante de ventas	Llamadas de ventas (x)	Copiadoras vendidas (y)	Ventas estimadas $\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$	y^*	$(y - y^*)^2$	y^{**}	$(y - y^{**})^2$
2	Brian Virost	96	41	45.0000	-4.0000	16.0000	44.4000	11.5600	41.6000	0.3600
3	Carlos Ramirez	40	41	30.3952	10.6048	112.4618	29.0000	144.0000	29.0000	144.0000
4	Carol Saia	104	51	47.0864	3.9136	15.3163	46.6000	19.3600	43.4000	57.7600
5	Greg Fish	128	60	53.3456	6.6544	44.2810	53.2000	46.2400	48.8000	125.4400
6	Jeff Hall	164	61	62.7344	-1.7344	3.0081	63.1000	4.4100	56.9000	16.8100
7	Mark Reynolds	76	29	39.7840	-10.7840	116.2947	38.9000	98.0100	37.1000	65.6100
8	Meryl Rumsey	72	39	38.7408	0.2592	0.0672	37.8000	1.4400	36.2000	7.8400
9	Mike Kiel	80	50	40.8272	9.1728	84.1403	40.0000	100.0000	38.0000	144.0000
10	Ray Snarsky	36	28	29.3520	-1.3520	1.8279	27.9000	0.0100	28.1000	0.0100
11	Rich Niles	84	43	41.8704	1.1296	1.2760	41.1000	3.6100	38.9000	16.8100
12	Ron Broderick	180	70	66.9072	3.0928	9.5654	67.5000	6.2500	60.5000	90.2500
13	Sal Spina	132	56	54.3888	1.6112	2.5960	54.3000	2.8900	49.7000	39.6900
14	Soni Jones	120	45	51.2592	-6.2592	39.1776	51.0000	36.0000	47.0000	4.0000
15	Susan Welch	44	31	31.4384	-0.4384	0.1922	30.1000	0.8100	29.9000	1.2100
16	Tom Keller	84	30	41.8704	-11.8704	140.9064	41.1000	123.2100	38.9000	79.2100
17	**Total**				**0.0000**	**587.1108**		**597.8000**		**793.0000**

En las columnas A, B, y C en la hoja de Excel anterior se duplicó la información muestral que se registra en la tabla 13.1. En la columna D se proporcionan los valores de las ventas estimadas, los valores $\hat{y}$, como se calculó antes.

En la columna E se calcularon los **residuales,** o los valores de error. Esta es la diferencia entre los reales y los pronosticados; es decir, la columna E es $(y - \hat{y})$. En el caso de Soni Jones,

$$\hat{y} = 19.9632 + 02608(120) = 51.2592$$

Su valor real es 45; por lo tanto, el residual, o error de estimación, es

$$(y - \hat{y} = (45 - 51.2592) = -6.2592$$

Este valor refleja que la cantidad del valor predicho de ventas está "fuera" del valor de ventas real.

Luego, en la columna F se elevan al cuadrado los residuales de cada vendedor y se obtiene el resultado. El total es 587.111.

$$\sum(y - \hat{y})^2 = 16.0000 + 112.4618 + \cdots + 140.9064 = 587.1108$$

Esta es la suma de las diferencias al cuadrado o el valor de los mínimos cuadrados; no hay otra recta que pase por estos 15 puntos de datos donde la suma de las diferencias al cuadrado sea menor.

Es posible demostrar el criterio de los mínimos cuadrados con dos ecuaciones arbitrarias cercanas a su ecuación y calcular la suma de las diferencias al cuadrado de estas ecuaciones. En la columna G se utilizó la ecuación $y^* = 18 + 275x$ para determinar el valor pronosticado; observe que esta ecuación es muy similar a la de mínimos cuadrados. En la columna H se determinan los residuos y se elevan al cuadrado. En el caso del primer vendedor, Brian Virost,

$$y^* = 18 + 0.275(96) = 44.4$$
$$(y - y^*)^2 = (41 - 44.4)^2 = 11.56$$

Este procedimiento se repite con los otros 14 representantes de ventas y se obtiene el total de los residuales al cuadrado. El resultado es 597.8, un valor mayor (597.8 es superior a 587.1108) que los residuales de la recta por mínimos cuadrados.

En las columnas I y J de la salida en pantalla se repite el proceso anterior para otra ecuación $y^{**} = 20 + 0.225x$. De nuevo, esta ecuación es similar a la de mínimos cuadrados. Los detalles de Brian Virost son:

$$y^{**} = 20 + 0.225x = 20 + 0.225(96) = 41.6$$
$$(y - y^{**})^2 = (41 + 41.6)^2 = 0.36$$

Este procedimiento se repite con los otros 14 representantes de ventas y se obtiene el total de los residuales. El resultado es 793, también mayor que los valores de los mínimos cuadrados.

¿Qué demuestra este ejemplo? La suma de los residuales al cuadrado $[\sum(y - \hat{y})^2]$ de la ecuación de los mínimos cuadrados es menor que la de otras rectas seleccionadas. No se encuentra una recta que pase por estos valores, donde la suma de los residuales al cuadrado es menor.

AUTOEVALUACIÓN 13-3

Consulte la autoevaluación 13.1, donde el propietario de Haverty's Furniture Company estudió la relación entre las ventas y la cantidad que gastaba en publicidad. El gasto en publicidad y los ingresos por las ventas, ambos en millones de dólares, de los últimos cuatro meses se repite a continuación.

Mes	Gastos de publicidad (millones de dólares)	Ganancias por ventas (millones de dólares)
Julio	2	7
Agosto	1	3
Septiembre	3	8
Octubre	4	10

(a) Determine la ecuación de regresión.
(b) Interprete los valores de a y b.
(c) Estime las ventas cuando se gastan tres millones de dólares en publicidad.

EJERCICIOS

13. Las siguientes observaciones muestrales se seleccionaron al azar.

x:	4	5	3	6	10
y:	4	6	5	7	7

a. Determine la ecuación de regresión.
b. Encuentre el valor de $\hat{y}$ cuando x es 7.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14. Las siguientes observaciones muestrales se seleccionaron al azar.

x:	5	3	6	3	4	4	6	8
y:	13	15	7	12	13	11	9	5

a. Determine la ecuación de regresión.
b. Encuentre el valor de $\hat{y}$ cuando x es 7.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

15. Bradford Electric Illuminating Company estudia la relación entre kilowatts-hora (miles) consumidos y la cantidad de habitaciones de una residencia privada familiar. Mediante una muestra aleatoria de 10 casas se reveló lo siguiente.

Número de habitaciones	Kilowatts-hora (miles)	Número de habitaciones	Kilowatts-hora (miles)
12	9	8	6
9	7	10	8
14	10	10	10
6	5	5	4
10	8	7	7

a. Determine la ecuación de regresión.
b. Encuentre el número de kilowatts-hora, en miles, de una casa de seis habitaciones.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

16. El señor James McWhinney, presidente de Daniel-James Financial Services, considera que hay una relación entre el número de contactos con sus clientes y la cantidad de ventas. Para probar esta

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

afirmación, el señor McWhinney reunió la siguiente información muestral; la columna x indica el número de contactos con sus clientes el mes anterior, la columna y indica el valor de las ventas (miles de dólares) el mismo mes por cada cliente muestreado.

Número de contactos, x	Ventas (miles de dólares), y	Número de contactos, x	Ventas (miles de dólares), y
14	24	23	30
12	14	48	90
20	28	50	85
16	30	55	120
46	80	50	110

Determine la ecuación de regresión.
Encuentre las ventas estimadas si se hicieron 40 contactos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

17. En un artículo reciente de *Bloomberg BusinessWeek* se enumeran las "Mejores pequeñas empresas". Lo que interesa aquí son los resultados actuales de las ventas e ingresos de ellas; se seleccionó una muestra de 12 empresas, y a continuación se reportan sus ventas e ingresos, en millones de dólares.

Compañía	Ventas (miles de dólares)	Ingresos (miles de dólares)	Compañía	Ventas (miles de dólares)	Ingresos (miles de dólares)
Papa John's International	$89.2	$4.9	Checkmate Electronics	$17.5	$ 2.6
Applied Innovation	18.6	4.4	Royal Grip	11.9	1.7
Integracare	18.2	1.3	M-Wave	19.6	3.5
Wall Data	71.7	8.0	Serving-N-Slide	51.2	8.2
Davidson & Associates	58.6	6.6	Daig	28.6	6.0
Chico's FAS	46.8	4.1	Cobra Golf	69.2	12.8

Sean las ventas la variable independiente, y los ingresos, la dependiente.

a. Trace un diagrama de dispersión.
b. Calcule el coeficiente de correlación.
c. Determine la ecuación de regresión.
d. Estime los ingresos de una compañía pequeña con ventas por 50 millones de dólares.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

18. Se realiza un estudio de fondos mutualistas para fines de inversión en varios de ellos. Este estudio en particular se enfoca en los activos y su desempeño a cinco años; la pregunta es: ¿se puede determinar la tasa de rendimiento a cinco años con base en los activos del fondo? Se seleccionaron nueve fondos mutualistas al azar, y sus activos y tasas de recuperación se muestran a continuación.

Fondo	Activos (en millones de dólares)	Rendimiento (%)	Fondo	Activos (en millones de dólares)	Rendimiento (%)
AARP High Quality Bond	$622.2	10.8	MFS Bond A	$494.5	11.6
Babson Bond L	160.4	11.3	Nichols Income	158.3	9.5
Compass Capital Fixed Income	275.7	11.4	T. Rowe Price Short-term	681.0	8.2
Galaxy Bond Retail	433.2	9.1	Thompson Income B	241.3	6.8
Keystone Custodian B-1	437.9	9.2			

a. Trace un diagrama de dispersión.
b. Calcule el coeficiente de correlación.
c. Escriba un reporte breve de sus resultados en los puntos b y c.
d. Determine la ecuación de regresión (utilice los activos como variable independiente).
e. Considere un fondo con 400.0 millones de dólares en ventas y determine la tasa de rendimiento a cinco años (en porcentaje).

19. Consulte el ejercicio 5. Asuma que la variable dependiente es el número de delitos.

a. Determine la ecuación de regresión.
b. Estime el número de delitos en una ciudad con 20 policías.
c. Interprete la ecuación de regresión.

20. Consulte el ejercicio 6.
a. Determine la ecuación de regresión.
b. Estime el precio de venta de un automóvil de 10 años.
c. Interprete la ecuación de regresión.

Probar la significancia de la pendiente

OA13-4
Evaluar la significancia de la pendiente de la ecuación de regresión.

En la sección anterior se mostró cómo encontrar la ecuación de la línea de regresión que mejor se ajusta a los datos; el método para encontrar la ecuación se basa en el *principio de los mínimos cuadrados.* El propósito de la ecuación de regresión es cuantificar una relación lineal entre dos variables.

A continuación se analiza la ecuación de regresión mediante una prueba de hipótesis para ver si la pendiente de la recta de regresión es distinta a cero. ¿Por qué es importante esto? Si es posible demostrar que la pendiente de la recta de la población es distinta de cero, entonces se puede concluir que al utilizar la ecuación de regresión aumenta la capacidad de predecir o pronosticar la variable dependiente con base en la variable independiente; si no es así, se concluye que no tiene caso utilizar la variable independiente como elemento de predicción. En otras palabras, si no se puede demostrar que la pendiente de la recta es distinta de cero, se podría utilizar la media de la variable dependiente como factor de predicción, en vez de usar la ecuación de regresión.

De acuerdo con el procedimiento de prueba de hipótesis que se expuso en el capítulo 10, las hipótesis nula y alternativa son:

$$H_0: \beta = 0$$
$$H_1: \beta \neq 0$$

La letra griega beta (β) se utiliza para representar la pendiente de la población de la ecuación de regresión, lo cual es consistente con la política de identificar los parámetros de población mediante las letras griegas. Se supone que la información respecto de North American Copier Sales (tabla 13.2) es una muestra. Es preciso tener cuidado aquí; recuerde que esta es solo una muestra, pero cuando se selecciona a un vendedor en particular se identifican dos tipos de información: cuántos clientes llamó y cuántas copiadores se vendieron; sin embargo, sigue siendo solo una muestra.

El valor de la pendiente se establece como b; esta se basa en una muestra y es una estimación de la pendiente de la población, identificada como "β". La hipótesis nula es que la pendiente de la ecuación de regresión de la población es cero; en tal caso, la recta de regresión es horizontal y no existe relación entre la variable independiente, X, y la variable dependiente, Y. En otras palabras, el valor de la variable dependiente es el mismo para cualquier valor de la variable independiente, y no ayuda a calcular el valor de la variable dependiente.

¿Qué ocurre si se rechaza la hipótesis nula? Si esta se rechaza y se acepta la hipótesis alternativa, se deduce que la pendiente de la recta de regresión de la población no es igual a cero; es decir, conocer el valor de la variable independiente permite estimar de mejor manera la variable dependiente. Para decirlo de otra forma, existe una relación significativa entre ambas variables.

Antes de probar la hipótesis, se utiliza un software estadístico para determinar los estadísticos de regresión necesarios. Se siguen utilizando los datos de North American Copier Sales que se registran en la tabla 13.2 y se utiliza Excel para realizar los cálculos. En la hoja de cálculo siguiente se muestran tres tablas a la derecha de los datos de la muestra.

	A	B	C	D	E	F	G	H	I	J
1	Representante de ventas	Llamadas de ventas (x)	Copiadoras vendidas (y)		RESUMEN DE SALIDA					
2	Brian Virost	96	41							
3	Carlos Ramirez	40	41		*Estadísticos de regresión*					
4	Carol Saia	104	51		Múltiple R	0.865				
5	Greg Fish	128	60		R cuadrada	0.748				
6	Jeff Hall	164	61		R cuadrada ajustada	0.728				
7	Mark Reynolds	76	29		Error estándar	6.720				
8	Meryl Rumsey	72	39		Observaciones	15				
9	Mike Kiel	80	50							
10	Ray Snarsky	36	28		ANOVA					
11	Rich Niles	84	43			*gl*	*SS*	*MS*	*F*	Significancia *F*
12	Ron Broderick	180	70		Regresión	1	1738.89	1738.89	38.5031	3.19277E-05
13	Sal Spina	132	56		Residual	13	587.11	45.1623		
14	Sani Jones	120	45		Total	14	2326			
15	Susan Welch	44	31							
16	Tom Keller	84	30			Coeficientes	Error estándar	Estadístico t	Valor P	
17					Intercepción	19.9800	4.389675533	4.55159	0.00054	
18					Llamadas de ventas (x)	0.2606	0.042001817	6.20509	3.2E-05	

1. En la parte de arriba están los *estadísticos de regresión*. Esta información se utilizará más adelante en este capítulo, pero observe que el valor "Múltiple R" (0.865) es conocido y que se trata del coeficiente de correlación calculado mediante la fórmula [13.1].
2. Enseguida está la tabla ANOVA; la cual se considera una herramienta útil para resumir la información de regresión. Se aborda más adelante en este capítulo, y se usará ampliamente en el siguiente, cuando se estudie la regresión múltiple.
3. Abajo, resaltada en azul, se encuentra la información necesaria para efectuar la prueba de hipótesis con respecto a la pendiente de la recta, la cual incluye el valor de la pendiente, que es 0.2606, y la intersección, que es 19.98 (observe que estos valores de la pendiente y la intersección son ligeramente distintos a los calculados en el ejemplo que aparece después de la fórmula [13.5]; estas pequeñas diferencias se deben al redondeo). En la columna a la derecha del coeficiente de regresión está la columna "Error estándar", el cual es un valor similar al error estándar de la media. Recuerde que el error estándar de la media reporta la variación entre las medias muestrales. En forma similar, estos errores estándar reportan la posible variación de los valores de la pendiente y de la intersección. El error estándar del coeficiente de la pendiente es 0.0420.

Para probar la hipótesis nula, se utiliza la distribución t con ($n - 2$ grados de libertad) y la siguiente fórmula:

PRUEBA DE LA PENDIENTE $$t = \left(\frac{b - 0}{s_b}\right) \quad \text{con } n - 2 \text{ grados de libertad} \qquad \textbf{[13.6]}$$

donde:

b es la estimación de la pendiente de la recta de regresión, calculada a partir de la información de la muestra;
s_b es el error estándar de la estimación de la pendiente, determinado también a partir de la información de la muestra.

El primer paso es establecer las hipótesis nula y alternativa, que son:

$$H_0: \beta \leq 0$$
$$H_1: \beta > 0$$

Observe que se trata de una prueba de una cola. Si no se rechaza la hipótesis nula, se concluye que la pendiente de la recta de regresión entre la población podría ser cero; esto significa que la variable independiente no tiene valor para mejorar la estimación de la variable dependiente. En este caso, esto quiere decir que conocer el número de llamadas de ventas que realizó un representante no sirve para predecir las ventas.

Si se rechaza la hipótesis nula y se acepta la alternativa, se concluye que la pendiente de la recta es mayor a cero; por lo tanto, la variable independiente es una ayuda para predecir la variable dependiente; por ello, conocer el número de llamadas de ventas que realizó un representante será útil para pronosticar las ventas que efectuó. También se sabe, porque se ha demostrado que la pendiente de la recta es mayor a cero −esto es, positiva-, que más llamadas se traducirán en la venta de más copiadoras.

La distribución t es el estadístico de prueba; hay 13 grados de libertad, determinados por $n - 2 = 15 - 2$. Se utiliza el nivel de significancia 0.05; del apéndice B.5 se determina que el valor crítico es 1.771. La regla de decisión es rechazar la hipótesis nula si el valor que se calculó con la fórmula [13.6] es mayor a 1.771. Se aplica la fórmula [13.6] para encontrar t.

$$t = \frac{b - 0}{s_b} = \frac{0.2606 - 0}{0.042} = 6.205$$

El valor calculado de 6.205 excede el valor crítico de 1.771, así que se rechaza la hipótesis nula y se acepta la hipótesis alternativa; en conclusión, la pendiente de la recta es mayor a cero. La variable independiente, que se refiere al número de llamadas de venta, es útil para obtener una mejor estimación de las ventas de copiadoras.

En la tabla también se proporciona información sobre el valor p de esta prueba; la celda está resaltada en color púrpura; por ello, es posible seleccionar el nivel de significancia, por ejemplo, 0.05 y comparar ese valor con el valor p. En este caso, el valor p calculado en la tabla se reporta como notación exponencial y es igual a 0.0000319, de modo que la decisión es rechazar la hipótesis nula. Una precaución importante es que los valores p que se reportan en el software estadístico suelen ser para una prueba de dos colas.

Antes de continuar, observe que en el ejemplo que aparece después de la fórmula [13.2], al realizar una prueba de hipótesis con respecto al coeficiente de correlación con estos mismos datos utilizando la fórmula [13.2], se obtuvo el mismo valor del estadístico t, $t = 6.205$; en realidad, cuando se comparan los resultados de un análisis de regresión lineal y correlación simple, las dos pruebas son equivalentes y siempre arrojarán exactamente los mismos valores t y p.

AUTOEVALUACIÓN 13-4

Consulte la autoevaluación 13.1, donde el propietario de Haverty´s Furniture estudió la relación entre las ventas y la cantidad que gastó en publicidad durante un mes; la cantidad de ventas es la variable dependiente, y el gasto en publicidad es la variable independiente. La ecuación de regresión en ese estudio fue $\hat{y} = 1.5 + 2.2x$ para una muestra de cinco meses. Realice una prueba de hipótesis para demostrar que existe una relación positiva entre la publicidad y las ventas (en el software estadístico, el error estándar del coeficiente de regresión es 0.42); utilice el nivel de significancia 0.05.

EJERCICIOS

21. Considere el ejercicio 5; la ecuación de regresión es $\hat{y} = 29.29 - 0.96x$, el tamaño de la muestra es 8, y el error estándar de la pendiente es 0.22. Aplique el nivel de significancia 0.05. ¿Se puede concluir que la pendiente de la recta de regresión es menor a cero?

22. Considere el ejercicio 6; la ecuación de regresión es $\hat{y} = 11.18 - 0.49x$, el tamaño de la muestra es 12, y el error estándar de la pendiente es 0.23. Aplique el nivel de significancia 0.05. ¿Se puede concluir que la pendiente de la recta de regresión es menor a cero?

23. Considere el ejercicio 17; la ecuación de regresión es $\hat{y} = 1.85 - 0.08x$, el tamaño de la muestra es 12, y el error estándar de la pendiente es 0.03. Aplique el nivel de significancia 0.05. ¿Se puede concluir que la pendiente de la recta de regresión es *distinta a cero*?

24. Considere el ejercicio 18; la ecuación de regresión es $\hat{y} = 9.9198 - 0.00039x$, el tamaño de la muestra es 9, y el error estándar de la pendiente es 0.0032. Aplique el nivel de significancia 0.05. ¿Se puede concluir que la pendiente de la recta de regresión es menor a cero?

Evaluación de la capacidad predictora de una ecuación de regresión

OA13-5 Evaluar la capacidad de una ecuación de regresión para predecir utilizando el error estándar de estimación y el coeficiente de determinación.

Error estándar de estimación

Los resultados del análisis de regresión de North American Copier Sales muestran una relación significativa entre el número de llamadas de ventas y la cantidad de ventas concretadas. Al sustituir el nombre de las variables en la ecuación, esta puede escribirse como:

$$\text{Número de copiadoras vendidas} = 19.9632 + 0.2608 \text{ (número de llamadas de ventas)}$$

La ecuación se puede usar para estimar el número de copiadoras vendidas por cada “número de llamadas de ventas” dentro del rango de los datos; por ejemplo, si el número de llamadas de ventas es 84, se puede predecir que el número de copiadoras vendidas es 41.8704, determinado por 19.9632 + 0.26.8(84); sin embargo, los datos muestran dos representantes con 84 llamadas de ventas de 30 y 43 copiadoras. Así que ¿la ecuación de regresión es un buen predictor del “número de copiadoras vendidas”?

El pronóstico perfecto, que implica encontrar el *resultado exacto,* es imposible en casi todas las disciplinas, incluyendo economía y negocios; por ejemplo:

- Una gran compañía electrónica, con instalaciones de producción en todo Estados Unidos, tiene un plan de opción en acciones para sus empleados. Suponga que existe una relación entre el número de años de empleo y las acciones que se poseen. Esta relación es posible porque, a medida que aumenta el número de los años de servicio, también se incrementa el número de acciones que un empleado gana. Es muy posible que los empleados que tienen hasta 20 años de servicio posean distinto número de acciones.

- Un desarrollador inmobiliario en el sudoeste de Estados Unidos estudió la relación entre el ingreso de los compradores y el tamaño, en pies cuadrados, de la casa que adquieren; su análisis muestra que conforme aumenta el ingreso de un comprador, se eleva también el tamaño de la casa que compra; sin embargo, no todos los compradores con un ingreso de 70 000 dólares comprarán una casa de exactamente el mismo tamaño.

Por ello, es necesario contar con una medida para describir cuán preciso es el pronóstico de Y con base en X, o a la inversa, qué tan inexacta puede ser la estimación. Esta medida se denomina **error estándar de estimación.** El error estándar de estimación, cuyo símbolo es $s_{y \cdot x}$. El subíndice $y \cdot x$ se interpreta como el error estándar de y para un valor dado de x. De hecho, es el mismo concepto que la desviación estándar que se analizó en el capítulo 3. La desviación estándar mide la dispersión respecto de la media; el error estándar de estimación la mide respecto de la recta de regresión para un valor dado de x.

ERROR ESTÁNDAR DE ESTIMACIÓN Medida de la dispersión de los datos respecto de la recta de regresión para un valor dado de x.

El error estándar de estimación se determina con la fórmula [13.7].

ERROR ESTÁNDAR DE ESTIMACIÓN

$$s_{y \cdot x} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - 2}} \quad \textbf{[13.7]}$$

El cálculo del error estándar de estimación requiere de la suma de las diferencias al cuadrado entre cada valor de y y su valor predicho, que se identifica como $\hat{y}$ en el numerador; esto se ilustra en la hoja de cálculo de la sección "Trazo de la recta de regresión"; observe que la celda G13 (el cual es un valor muy importante) constituye el numerador en el cálculo del error estándar de estimación.

$$s_{y \cdot x} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - 2}} = \sqrt{\frac{587.11}{15 - 2}} = 6.720$$

El error estándar de estimación puede calcularse mediante un software estadístico como Excel; dicho programa lo incluye en el análisis de regresión (resaltado en amarillo en la hoja de cálculo de la sección "Probar la significancia de la pendiente"), su valor es 6.720.

Cuando el error estándar de estimación es pequeño, los datos están relativamente cercanos a la recta de regresión, y mediante la ecuación de regresión se predice y con poco error; pero si dicho error es grande, significa que estos están muy dispersos respecto de la recta, y la ecuación no proporcionará una estimación precisa de y.

El coeficiente de determinación

El error estándar de estimación proporciona una medida relativa de la capacidad de predicción de una ecuación de regresión. En la próxima sección se utiliza para proporcionar información más específica con respecto a una predicción. En esta sección se explica otro estadístico que brindará una medida más interpretable de la capacidad de predicción de una ecuación de regresión; este se llama coeficiente de determinación, o R cuadrada.

COEFICIENTE DE DETERMINACIÓN Proporción de la variación total de la variable dependiente Y que se explica, o contabiliza, por la variación de la variable dependiente X.

El coeficiente de determinación es fácil de calcular. Es el coeficiente de correlación al cuadrado; por lo tanto, también se usa el término R cuadrada. En el caso de North American Copier Sales, el coeficiente de correlación de la relación entre el número de copiadoras vendidas y las llamadas de ventas es 0.865; si se calcula $(0.865)^2$, el coeficiente de determinación es 0.748. Observe las celdas azul ("Múltiple R") y verde ("R cuadrada") resaltadas en la hoja de cálculo de la sección "Probar la significancia de la pendiente". Para interpretar mejor el coeficiente de determinación, conviértalo a porcentajes; así, 74.8% de la variación del número de copiadoras vendidas se explica, o está representado por, la variación del número de llamadas de ventas.

¿Con cuánta exactitud predice la ecuación de regresión el número de copiadoras vendidas mediante el número de llamadas de ventas realizadas? Si fuera posible hacer predicciones perfectas, el coeficiente de determinación sería 100%; esto significaría que la variable independiente (el número de llamadas de ventas) explica, o representa, toda la variación del número de copiadoras vendidas. Un coeficiente de determinación de 100% se asocia con un coeficiente de correlación de $+1.0$ o -1.0. Consulte la gráfica 13.2, en la cual se muestra que una predicción perfecta se asocia con una relación lineal impecable, donde todos los puntos de los datos forman una recta perfecta en un diagrama de dispersión. El análisis muestra que solo 74.8% de la variación del número de copiadoras vendidas se explica por la variación del número de llamadas que se realizaron. Es claro que estos datos no forman una línea perfecta; en vez de eso, se diseminan alrededor de la recta de regresión de mínimos cuadrados que mejor se ajusta, y las predicciones no son exactas. En la próxima sección se utiliza el error estándar de estimación para proporcionar información más específica con respecto al error asociado con el empleo de la ecuación de regresión para hacer predicciones.

AUTOEVALUACIÓN 13-5

Consulte la autoevaluación 13.1, donde el propietario de la Haverty's Furniture Company estudió la relación entre la cantidad que gastó en publicidad y los ingresos por ventas durante el mes previo. La cantidad de ventas es la variable dependiente, y el gasto en publicidad es la variable independiente.
(a) Determine el error estándar de estimación.
(b) Determine el coeficiente de determinación.
(c) Interprete el coeficiente de determinación.

EJERCICIOS

Se sugiere utilizar un paquete de software como Excel para realizar los cálculos.

25. Considere el ejercicio 5; determine el error estándar de estimación y el coeficiente de determinación. Interprete el coeficiente de determinación.
26. Considere el ejercicio 6; determine el error estándar de estimación y el coeficiente de determinación. Interprete el coeficiente de determinación.
27. Considere el ejercicio 15; determine el error estándar de estimación y el coeficiente de determinación. Interprete el coeficiente de determinación.
28. Considere el ejercicio 16; determine el error estándar de estimación y el coeficiente de determinación. Interprete el coeficiente de determinación.

Relaciones entre el coeficiente de correlación, el coeficiente de determinación y el error estándar de estimación

En páginas anteriores se analizó el error estándar de estimación, el cual mide la cercanía entre los valores reales y la recta de regresión; cuando este es pequeño, las dos variables están muy relacionadas. En el cálculo del error estándar, el término clave es

$$\sum(y - \hat{y})^2$$

si el valor de este término es pequeño, el error estándar también lo será.

El coeficiente de correlación mide la fuerza de la asociación lineal entre dos variables. Cuando los puntos del diagrama de dispersión aparecen cerca de la recta, se observa que el coeficiente de correlación tiende a ser grande; todo ello indica que el error estándar de estimación y el coeficiente de correlación están inversamente relacionados. A medida que aumenta la fuerza de la relación lineal entre dos variables, aumenta el coeficiente de correlación y disminuye el error estándar de estimación.

También se observó que el cuadrado del coeficiente de correlación es el coeficiente de determinación, el cual mide el porcentaje de la variación de Y que se explica por la variación de X.

La tabla ANOVA se considera un medio conveniente para mostrar la relación entre estas tres medidas. Observe la porción que se resalta en amarillo en la hoja de cálculo que aparece en la página siguiente; esta tabla se estructura de manera similar al análisis de la tabla de la varianza que se desarrolló en el capítulo 12. En ese capítulo, la variación total se dividió en los dos elementos que la componen: los *tratamientos* y el *error aleatorio*. El concepto es similar en el análisis de regresión.

	A	B	C	D	E	F	G	H	I	J
1	Representante de ventas	Llamadas de ventas (x)	Copiadoras vendidas (y)		RESUMEN DE SALIDA					
2	Brian Virost	96	41							
3	Carlos Ramirez	40	41		*Estadísticos de regresión*					
4	Carol Saia	104	51		Múltiple R	0.865				
5	Greg Fish	128	60		R cuadrada	0.748				
6	Jeff Hall	164	61		R cuadrada ajustada	0.728				
7	Mark Reynolds	76	29		Error estándar	6.720				
8	Meryl Rumsey	72	39		Observaciones	15				
9	Mike Kiel	80	50							
10	Ray Snarsky	36	28		ANOVA					
11	Rich Niles	84	43			*gl*	SS	MS	F	Significancia F
12	Ron Broderick	180	70		Regresión	1	1738.89	1738.89	38.50312548	3.19277E-05
13	Sal Spina	132	56		Residual	13	587.11	45.162308		
14	Sani Jones	120	45		Total	14	2326			
15	Susan Welch	44	31							
16	Tom Keller	84	30			Coeficientes	Error estándar	Estadístico t	Valor P	
17					Intercepción	19.9800	4.389675533	4.5515893	0.000543565	
18					Llamadas de ventas (x)	0.2606	0.042001817	6.2050887	3.19277E-05	

La variación total se divide en dos componentes: 1) el que se deriva de la *regresión* (a su vez, explicada por la variable independiente) y 2) el *error* o *residual*, que es la variación inexplicable. Estas tres fuentes de varianza (total, regresión y residual) se identifican en la primera columna de la siguiente tabla ANOVA. La columna con el encabezado "*gl*" se refiere a los grados de libertad asociados a cada categoría; el número total de grados de libertad es $n - 1$, y el número de grados de libertad de la regresión es 1, pues solo hay una variable independiente; el número de grados de libertad asociados con el término de error es $n - 2$. El término "SS" que se ubica en medio de la tabla ANOVA se refiere a la suma de los cuadrados. Observe que el total de los grados de libertad es igual a la suma de los grados de libertad de la regresión y del residual (error), y que la suma total de los cuadrados es igual a la suma de los cuadrados de la suma de la regresión y el residuo (error); esto se aplica a cualquier tabla ANOVA.

La suma de cuadrados ANOVA se calcula como sigue:

$$\text{Suma de regresión de los cuadrados} = \text{SSR} = \sum(\hat{y} - \bar{y})^2 = 1\,738.89$$
$$\text{Suma del residual o error de los cuadrados} = \text{SSE} = \sum(y - \hat{y})^2 = 587.211$$
$$\text{Suma total de los cuadrados} = \text{SS total} = \sum(y - \bar{y})^2 = 2\,326.0$$

Recuerde que el coeficiente de determinación se define como el porcentaje de la variación total (SS total) explicado por la ecuación de regresión (SSR). El valor R-cuadrado (r^2) puede ser validado mediante la tabla ANOVA.

COEFICIENTE DE DETERMINACIÓN $$r^2 = \frac{\text{SSR}}{\text{SS Total}} = 1 - \frac{\text{SSE}}{\text{SS Total}} \quad \textbf{[13.8]}$$

Utilizando los valores que se registran en la tabla ANOVA, el coeficiente de determinación es 1 738.89/2 326.0 = 0.748; por lo tanto, a mayor variación de la variable dependiente (SS total) explicada por la variable independiente (SSR), más alto será el coeficiente de determinación.

El coeficiente de determinación también puede expresarse en términos de la variación del residuo o error:

$$r^2 = 1 - \frac{\text{SSR}}{\text{SS Total}} = 1 - \frac{587.11}{2\,326.0} = 1 - 0.252 = 0.748$$

Como se ilustra en la fórmula [13.8], el coeficiente de determinación y la suma del residuo o error de los cuadrados están inversamente relacionados: mientras más alta sea la variación inexplicable (o error) como porcentaje de la variación total, menor será el coeficiente de determinación. En este caso, 25.2% de la variación total de la variable dependiente es una variación residual o error.

La observación final que relaciona el coeficiente de relación, el coeficiente de determinación y el error estándar de estimación muestra la relación entre el error estándar de estimación y la SSE; al sustituir [SSE suma de los cuadrados de residuo o error = SSE = $\sum(y - \hat{y})^2$] en la fórmula del error estándar de estimación, resulta:

ERROR ESTÁNDAR DE ESTIMACIÓN $$s_{y \cdot x} = \sqrt{\frac{\text{SSE}}{n - 2}} \quad \textbf{[13.9]}$$

En suma, el análisis de regresión proporciona dos estadísticos para evaluar la capacidad de predicción de una ecuación de regresión: el error estándar de estimación y el coeficiente de determinación. Al reportar los resultados de un análisis de regresión es necesario explicar claramente los hallazgos, en especial cuando se emplean los resultados para hacer predicciones de la variable dependiente. El reporte siempre debe incluir un enunciado con respecto al coeficiente de determinación para que el lector del reporte pueda conocer la precisión relativa de la predicción. Se requiere un reporte objetivo del análisis estadístico para que los lectores puedan tomar sus propias decisiones.

EJERCICIOS

29. Con la siguiente tabla ANOVA:

Fuente	*gl*	SS	MS	*F*
Regresión	1	1 000.0	1 000.0	26.00
Error	13	500.0	38.46	
Total	14	1 500.0		

a. Encuentre el coeficiente de determinación.
b. Si hay una relación directa entre las variables, ¿cuál es el coeficiente de correlación?
c. Determine el error estándar de estimación.

30. En el primer examen de Estadística, el coeficiente de determinación entre las horas estudiadas y la calificación obtenida fue de 80% y el error estándar de estimación fue de 10; había 20 estudiantes en la clase. Elabore una tabla ANOVA para efectuar el análisis de regresión de horas estudiadas como un predictor de la calificación obtenida en el primer examen de Estadística.

Estimaciones de intervalo de predicción

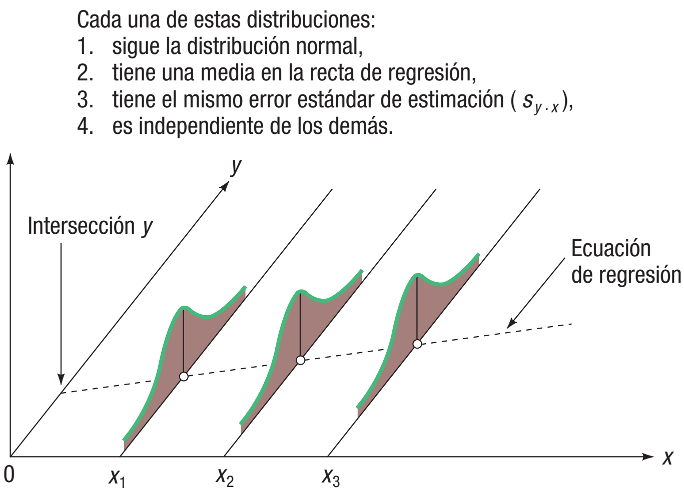
OA13-6 Calcular e interpretar los intervalos de confianza y de predicción.

El error estándar y el coeficiente de determinación son dos estadísticos que proporcionan una evaluación general de la capacidad de una ecuación de regresión para predecir una variable dependiente. Otra forma de reportar tal capacidad es analizar un valor declarado de la variable independiente; por ejemplo, se puede predecir el número de copiadoras vendidas (y) en el caso de un valor seleccionado de número de llamadas de ventas realizadas (x). En realidad, es posible calcular el intervalo de confianza del valor pronosticado de la variable dependiente para un valor seleccionado de la variable independiente.

Suposiciones subyacentes a la regresión lineal

Antes de presentar los intervalos de confianza, deben revisarse las suposiciones para aplicar de forma apropiada la regresión lineal. En la gráfica 13.13 se ilustran dichas suposiciones.

1. Para cada valor de x, existen valores de y correspondientes; los cuales siguen la distribución normal.
2. Las medias de estas distribuciones normales están en la recta de regresión.



GRÁFICA 13.13 Suposiciones de la regresión en forma gráfica

ESTADÍSTICA EN ACCIÓN

En ciertos estudios se reporta que, tanto en el caso de hombres como de mujeres, los considerados bien parecidos ganan salarios mayores que quienes no lo son. Además, en los hombres hay una correlación entre estatura y salario. Por cada pulgada adicional de estatura, un hombre puede esperar ganar 250 dólares más al año; por lo tanto, un individuo que mide 6′6″ recibe un "bono" de 3 000 dólares respecto de otro que mida 5′6″. Además, tener sobrepeso o estar muy delgado también se relaciona con los ingresos, en particular entre las mujeres. Un estudio de mujeres jóvenes demostró que 10% de las que más pesaba ganaba en promedio 6% menos que sus contrapartes más delgadas.

3. Todas las desviaciones estándar de estas distribuciones normales son iguales. La mejor estimación de esta desviación estándar común es el error estándar de la estimación ($s_{y \cdot x}$).
4. Los valores y son estadísticamente independientes. Esto significa que, al seleccionar una muestra, una x particular no depende de ningún otro valor de x; tal suposición es de particular importancia cuando los datos se recopilan durante cierto periodo. En esas situaciones, los errores de un periodo particular con frecuencia están correlacionados con los de otros periodos.

Recuerde, del capítulo 7, que si las observaciones siguen una distribución normal, la media más (o menos) una desviación estándar comprenderá 68% de estas, la media más (o menos) dos desviaciones estándar comprenderá 95% del total, y la media más (o menos) tres desviaciones estándar las comprenderá virtualmente a todas. Existe la misma relación entre los valores anticipados $\hat{y}$ y el error estándar de estimación ($s_{y \cdot x}$).

1. $\hat{y} \pm s_{y \cdot x}$ incluirá 68% de las observaciones.
2. $\hat{y} \pm 2s_{y \cdot x}$ incluirá 95% de las observaciones.
3. $\hat{y} \pm 3s_{y \cdot x}$ incluirá virtualmente todas las observaciones.

Ahora se relacionan estas suposiciones con la empresa North American Copier Sales, donde se estudió la relación entre el número de llamadas de ventas y el de copiadoras vendidas. Si se traza una recta paralela 6.72 unidades por arriba de la recta de regresión y otra 6.72 unidades por debajo de la recta de regresión, cerca de 68% de los puntos se encontraría entre ambas rectas. De manera similar, entre una recta 13.44 [$2s_{y \cdot x} = 2(6.72)$] unidades arriba de la recta de regresión y otra 13.44 unidades debajo de la recta de regresión se incluirá alrededor de 95% de los valores.

Como una verificación muy aproximada, consulte la columna "E" en la hoja de cálculo de Excel de la sección "Trazo de la recta de regresión". Cuatro de las 15 desviaciones sobrepasan un error estándar de estimación; es decir, las desviaciones de Carlos Ramírez, Mark Reynolds, Mike Keil y Tom Keller sobrepasan el valor de 6.72 (un error estándar). Todos los datos están a menos de 13.44 unidades de la línea de regresión; en resumen, 11 de las 15 desviaciones de la muestra están dentro de un error estándar de la recta de regresión y todas están dentro de dos errores estándar. Este es un resultado bastante bueno para ser una muestra relativamente pequeña.

Construcción de intervalos de confianza y de predicción

Cuando se utiliza una ecuación de regresión, se pueden hacer dos predicciones distintas para un valor seleccionado de la variable independiente; las diferencias son sutiles pero muy importantes, y están relacionadas con las suposiciones que se explicaron en la sección anterior. Recuerde que para cada valor seleccionado de la variable independiente (X), la variable dependiente (Y) es una variable aleatoria que está distribuida normalmente con una media $\hat{Y}$. Cada distribución de Y tiene una desviación estándar igual al error estándar de estimación del análisis de regresión.

El primer intervalo se denomina **intervalo de confianza**; este se utiliza cuando la ecuación de regresión se emplea para predecir el valor *medio* de Y para una x dada. Por ejemplo, se puede usar un intervalo de confianza para estimar el salario medio de todos los ejecutivos en la industria minorista con base en sus años de experiencia. Para determinar el intervalo de confianza del valor medio de y para una x dada, la fórmula es:

INTERVALO DE CONFIANZA DE LA MEDIA DE *Y*, DADA *X*

$$\hat{y} \pm ts_{y \cdot x}\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}} \qquad \textbf{[13.10]}$$

El segundo tipo de estimación se denomina **intervalo de predicción**; este se utiliza cuando la ecuación de regresión se emplea para predecir una y individual ($n = 1$) para un valor dado de x. Por ejemplo, para hacer una estimación del salario de ejecutivo minorista en particular con 20 años de experiencia. Para determinar el intervalo de predicción de una estimación individual para una x dada, la fórmula es:

INTERVALO DE PREDICCIÓN DE *Y*, DADA *X*

$$\hat{y} \pm ts_{y \cdot x}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}} \qquad \textbf{[13.11]}$$

EJEMPLO

De nuevo, el ejemplo de la compañía North American Copier Sales. Determine un intervalo de confianza de 95% para todos los representantes de ventas que hacen 25 llamadas y un intervalo de predicción para Sheila Baker, representante de la Costa Oeste que hizo 25 llamadas.

SOLUCIÓN

Emplee la fórmula [13.10] para determinar un intervalo de confianza; en la tabla 13.4 se incluyen los totales necesarios y se repite la información de la tabla 13.2.

TABLA 13.4 Cálculos necesarios para determinar el intervalo de confianza y el intervalo de predicción

Representante de ventas	Llamadas de ventas (x)	Ventas de copiadoras (y)	$(x - \bar{x})$	$(x - \bar{x})^2$
Brian Virost	96	41	0	0
Carlos Ramirez	40	41	−56	3 136
Carol Saia	104	51	8	64
Greg Fish	128	60	32	1 024
Jeff Hall	164	61	68	4 624
Mark Reynolds	76	29	−20	400
Meryl Rumsey	72	39	−24	576
Mike Kiel	80	50	−16	256
Ray Snarsky	36	28	−60	3 600
Rich Niles	84	43	−12	144
Ron Broderick	180	70	84	7 056
Sal Spina	132	56	36	1 296
Sani Jones	120	45	24	576
Susan Welch	44	31	−52	2 704
Tom Keller	84	30	−12	144
Total	1 440	675	0	25 600

El primer paso es determinar el número de copiadoras que se espera que venda un representante si hace 50 llamadas; el cual es 33.0032, determinado por

$$\hat{y} = 19.9632 + 0.2608x = 19.9632 + 0.2608(50) = 33.0032$$

Para encontrar el valor t, primero es necesario conocer el número de grados de libertad; en este caso, son $n - 2 = 15 - 2 = 13$, con un nivel de confianza de 95%. Para encontrar el valor de t, desplácese hacia abajo a la izquierda de la columna del apéndice B.5 a 13 grados de libertad, y después muévase por la columna con el nivel de confianza de 95%. El valor de t es 2.160.

En la sección anterior se calculó que el error estándar de estimación era de 6.720. Sea $x = 50$, y de la tabla 13.4, el número medio de llamadas de ventas es 96.0(1 440/15) y $\sum(x - \bar{x})^2 = 25\,600$. Estos valores se sustituyen en la fórmula [13.10] para determinar el intervalo de confianza.

$$\text{Intervalo de confianza} = \hat{y} \pm t s_{y \cdot x} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}}$$

$$= 33.0032 \pm 2.160(6.720) \sqrt{\frac{1}{15} + \frac{(50 - 96)^2}{25\,600}}$$

$$= 33.0032 \pm 5.6090$$

Por lo tanto, el intervalo de confianza de 95% de todos los representantes de ventas que hacen 50 llamadas es desde 27.3942 hasta 38.6122; para interpretar esto, redondee los valores. Si un representante hace 50 llamadas, debería vender 33 copiadoras. Es probable que las ventas varíen de 27.4 a 38.6 unidades.

Suponga que se desea estimar el número de copiadoras que vendió Sheila Baker, quien hizo 50 llamadas. El intervalo de predicción de 95% se determina como sigue:

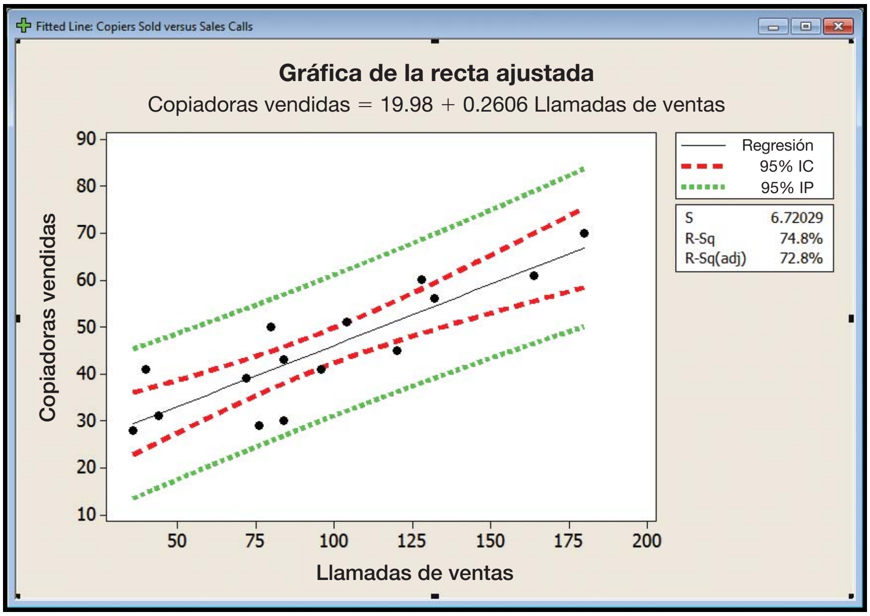
$$\text{Intervalo de predicción} = \hat{y} \pm ts_{y\cdot x}\sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{\sum(x - \overline{x})^2}}$$

$$= 33.0032 \pm 2.160(6.720)\sqrt{1 + \frac{1}{15} + \frac{(50 - 96)^2}{25\,600}}$$

$$= 33.0032 \pm 15.5612$$

Así, el intervalo es desde 17.442 hasta 48.5644 copiadoras. Se concluye que el número de copiadoras que venderá un representante, como Sheila Baker, que haga 50 llamadas estará aproximadamente entre 17.4 y 48.6. Este intervalo es muy grande; es mucho mayor que el de todos los representantes que hicieron 50 llamadas; sin embargo, es lógico que deba haber más variación en la estimación de ventas de un individuo que de un grupo.

En la siguiente gráfica de Minitab se muestra la relación entre la recta de regresión (en el centro), el intervalo de confianza (en color rojo) y el intervalo de predicción (en color verde). Las bandas del intervalo de predicción siempre están más alejadas de la recta de regresión que las del intervalo de confianza; asimismo, a medida que los valores de x se alejan del número medio de llamadas (96), en cualquier dirección, las bandas de los intervalos de confianza y de predicción se ensanchan. Esto se debe al numerador del término de la derecha debajo del radical en las fórmulas [13.10] y [13.11]; es decir, cuando el término aumenta, también lo hacen los anchos de ambos intervalos; en otras palabras, las estimaciones son menos precisas cuando hay un alejamiento, en cualquier dirección, de la media de la variable independiente.



Es conveniente destacar otra vez la distinción entre un intervalo de confianza y uno de predicción; el primero se refiere a todos los casos con un valor dado de x (su valor se calcula por medio de la fórmula [13.10]); el segundo, a un caso particular de un valor dado de x (su valor se determina mediante la fórmula [13.11]). El intervalo de predicción siempre será más ancho debido al 1 adicional debajo del radical en la segunda ecuación.

AUTOEVALUACIÓN 13-6

Consulte los datos muestrales en la autoevaluaciones 13.1, donde el propietario de Haverty's Furniture estudió la relación entre las ventas y la cantidad que gastó en publicidad. La información de las ventas de los últimos cuatro meses se repite en la página siguiente.

Mes	Gastos publicitarios (en millones de dólares)	Ingresos por ventas (en millones de dólares)
Julio	2	7
Agosto	1	3
Septiembre	3	8
Octubre	4	10

La ecuación de regresión calculada fue $\hat{y} = 1.5 + 2.2x$, y el error estándar, 0.9487; ambas variables se reportan en millones de dólares. Determine el intervalo de confianza de 90% para el mes común en el cual se gastaron tres millones de dólares en publicidad.

EJERCICIOS

31. Consulte el ejercicio 13.
   a. Determine el intervalo de confianza 0.95 para la media pronosticada cuando $x = 7$.
   b. Establezca el intervalo de predicción 0.95 para un individuo proyectado cuando $x = 7$.

32. Consulte el ejercicio 14.
   a. Determine el intervalo de confianza 0.95 para la media pronosticada cuando $x = 7$.
   b. Encuentre el intervalo de predicción 0.95 para una predicción individual cuando $x = 7$.

33. Consulte el ejercicio 15.
   a. Determine el intervalo de confianza 0.95, en miles de kilowatts-hora, de la media de todas las casas con seis habitaciones.
   b. Encuentre el intervalo de predicción 0.95, en miles de kilowatts-hora, de una casa en particular con seis habitaciones.

34. Consulte el ejercicio 16.
   a. Determine el intervalo de confianza 0.95, en miles de dólares, de la media de todo el personal de ventas que hace 40 contactos.
   b. Encuentre el intervalo de predicción 0.95, en miles de dólares, para un vendedor en particular que hace 40 contactos.

Transformación de datos

OA13-7 Utilizar una función de logaritmo para transformar una relación no lineal.

El análisis de regresión describe la relación entre dos variables; un requisito es que esta relación sea lineal, y lo mismo aplica para el coeficiente de correlación, el cual mide la fuerza de una relación *lineal* entre dos variables. Pero, ¿qué pasa si la relación no es lineal? El remedio es volver a hacer la escala de una o ambas variables para que la nueva relación sea lineal; por ejemplo, en vez de utilizar los valores actuales de la variable independiente *y*, se crearía otra variable independiente calculando el logaritmo a la base 10 de *y*, Log(*y*); a este cálculo se le llama transformación. Otras transformaciones comunes son: obtener la raíz cuadrada, calcular el recíproco, o elevar al cuadrado una o ambas variables.

Así, dos variables pueden estar estrechamente relacionadas aunque su relación no sea lineal. Tenga cuidado cuando interprete el coeficiente de correlación o una ecuación de regresión; estos estadísticos pueden indicar que no existe una relación lineal, pero quizá haya una relación de alguna otra forma no lineal o curvilínea.

EJEMPLO

GroceryLand Supermarkets es una cadena regional de tiendas de abarrotes con más de 300 locales ubicados en el medio oeste de Estados Unidos. El director corporativo de marketing de la compañía desea estudiar el efecto del precio en las ventas semanales de las botellas de dos litros de su marca privada de refresco de cola dietético. Los objetivos del estudio son:

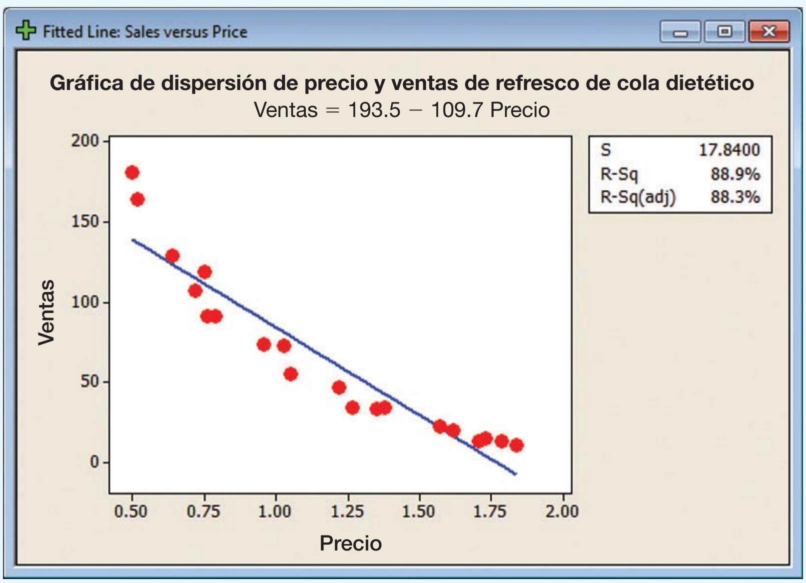
1. Determinar si existe una relación entre el precio de venta y las ventas semanales y, en su caso, saber si la relación es directa o indirecta, fuerte o débil.
2. Determinar si el efecto del precio aumenta o disminuye en términos de ventas y si es posible predecir efectivamente las ventas con base en el precio.

SOLUCIÓN

Para comenzar el proyecto, el director de marketing se reúne con el vicepresidente de ventas y otros miembros de la compañía. Deciden que sería razonable aumentar el precio de la botella de dos litros de su marca privada de refresco de cola dietético desde 0.50 hasta 2.00 dólares. Para recabar los datos necesarios para analizar la relación entre el precio y las ventas, el director selecciona una muestra aleatoria de 20 tiendas y después asigna al azar un precio de venta entre 0.50 dólares y 2.00 dólares en cada una. El director contacta a cada uno de los 20 gerentes de las tiendas incluidas en el estudio para informarles el precio y pedirles que reporten las ventas del producto al final de la semana. Los resultados se reportan en la siguiente tabla; por ejemplo, la tienda número 17 vendió 181 botellas de dos litros de refresco a 0.50 dólares cada una.

Datos de ventas y precio de GroceryLand			Datos de ventas y precio de GroceryLand		
Número de tienda	Precio (en dólares)	Ventas	Número de tienda	Precio (en dólares)	Ventas
17	0.50	181	30	0.76	91
121	1.35	33	127	1.79	13
227	0.79	91	266	1.57	22
135	1.71	13	117	1.27	34
6	1.38	34	132	0.96	74
282	1.22	47	120	0.52	164
172	1.03	73	272	0.64	129
296	1.84	11	120	1.05	55
143	1.73	15	194	0.72	107
66	1.62	20	105	0.75	119

Para examinar la relación entre precio y ventas se utiliza el análisis de regresión, estableciendo a "Precio" como la variable independiente y a "Ventas" como la variable dependiente; el análisis proporciona información importante con respecto a la relación entre estas, y se resume en la siguiente salida de Minitab.



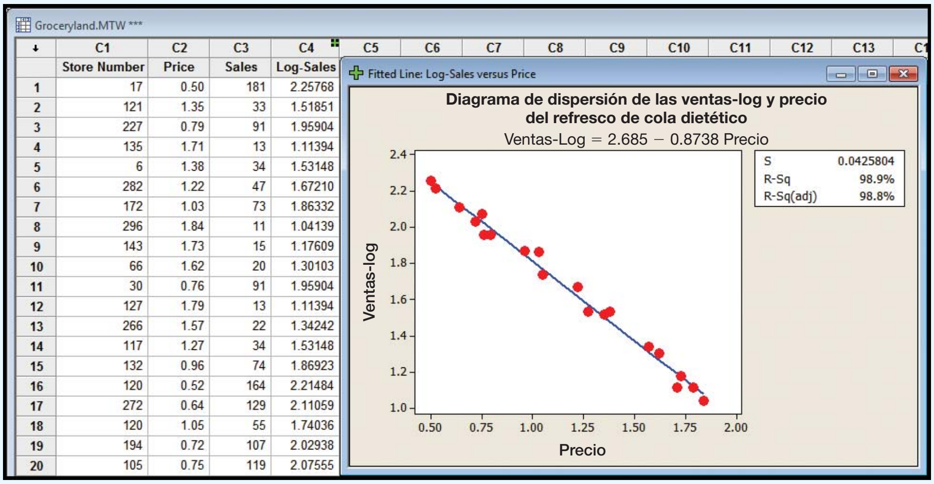
De esta salida de pantalla se concluye lo siguiente:

1. La relación entre ambas variables es inversa o indirecta. A medida que se incrementa el "Precio" del refresco, las "Ventas" del producto disminuyen; esto era de esperarse, dada la teoría económica básica de precio y demanda.
2. Existe una fuerte relación entre ambas variables; el coeficiente de determinación es 88.9%, así que 88.9% de la variación de "Ventas" está representado por la variación en "Precio". A partir del coeficiente de determinación, el de correlación se calcula como la raíz cuadrada del coefi-

ciente de determinación; en este caso, 0.889, o 0.943. El signo del coeficiente de correlación es negativo porque las ventas son inversamente relacionadas con el precio; por lo tanto, el coeficiente de correlación es -0.943.

3. Antes de continuar con este resumen de conclusiones, es preciso analizar con cuidado el diagrama de dispersión y el trazo de la línea de regresión. La suposición de una relación lineal es tenue; por lo tanto, si la relación es lineal, los puntos de los datos deberían estar distribuidos tanto por encima como por debajo de la línea en todo el rango de la variable independiente; sin embargo, en el caso de los precios más altos y más bajos, estos se encuentran encima de la línea de regresión. Para los precios de venta que están en el medio, la mayoría queda por debajo de la línea de regresión; así que la ecuación de regresión lineal no describe efectivamente la relación entre “Precio” y “Ventas”. Se necesita una transformación de los valores para crear otra relación lineal.

Al transformar una de las variables, se puede cambiar la relación no lineal entre estas a una lineal; entre las posibles opciones, el director de marketing decide transformar la variable dependiente, “Ventas”, sacando el logaritmo a la base 10 de cada valor de esta. Observe la nueva variable, “Ventas-Log”, que se muestra en el análisis siguiente; ahora, el análisis de regresión utiliza “Ventas-Log” como la variable dependiente y “Precio” como la variable independiente. Este análisis se reporta en la siguiente salida de pantalla.



	Store Number	Price	Sales	Log-Sales
1	17	0.50	181	2.25768
2	121	1.35	33	1.51851
3	227	0.79	91	1.95904
4	135	1.71	13	1.11394
5	6	1.38	34	1.53148
6	282	1.22	47	1.67210
7	172	1.03	73	1.86332
8	296	1.84	11	1.04139
9	143	1.73	15	1.17609
10	66	1.62	20	1.30103
11	30	0.76	91	1.95904
12	127	1.79	13	1.11394
13	266	1.57	22	1.34242
14	117	1.27	34	1.53148
15	132	0.96	74	1.86923
16	120	0.52	164	2.21484
17	272	0.64	129	2.11059
18	120	1.05	55	1.74036
19	194	0.72	107	2.02938
20	105	0.75	119	2.07555

¿Qué conclusión se deriva del análisis de regresión utilizando la transformación de la variable dependiente “Ventas”?

1. Al transformar dicha variable, se eleva el coeficiente de determinación de 0.889 a 0.989. Así que ahora “Precio” explica casi toda la variación en “Ventas-Log”.
2. Compare este resultado con el diagrama de dispersión antes que se transformara la variable dependiente. Los datos transformados parecen ajustarse mucho mejor al requisito de la relación lineal; observe que los puntos de estos se encuentran encima y debajo de la línea de regresión en el rango de “Precio”.
3. La ecuación de regresión es $\hat{y} = 2.685 - 0.8738x$. El signo del valor de la pendiente es negativo, lo que confirma la asociación inversa entre las variables. La nueva ecuación se utiliza para estimar las ventas y estudiar el efecto de los cambios en el precio; por ejemplo, si se decide vender la botella de dos litros de refresco a 1.25 dólares, la variable “Ventas-Log” predicha es:

$$\hat{y} = 2.685 - 0.8738x = 2.658 - 0.8738(1.25) = 1.593$$

Recuerde que la ecuación de regresión ahora predice el logaritmo, base 10, de “Ventas”; por lo tanto, es necesario deshacer la transformación obteniendo el antilogaritmo de 1.593, que es $10^{1.593}$, o 39.174. Así, si se establece el precio del refresco en 1.25 dólares, las ventas semanales predichas son 39 botellas; si se eleva el precio a 2.00 dólares, la ecuación de regresión predecirá un valor de 0.9374. Mediante el antilogaritmo, $10^{.9374}$, la disminución predicha en las ventas es 8.658 o, redondeando, 9 botellas de dos litros por semana. Es claro que a medida que el precio aumenta, las ventas disminuyen. Esta relación será muy útil a GroceryLand cuando tome decisiones sobre los precios para este producto.

EJERCICIOS

35. Con la siguiente muestra de cinco observaciones, trace un diagrama de dispersión usando *x* como la variable independiente y *y* como la variable dependiente; calcule el coeficiente de correlación. ¿La relación entre las variables parece lineal? Intente elevar al cuadrado la variable *x* y después desarrolle otro diagrama de dispersión y determine el coeficiente de correlación. Resuma su análisis.

x	−8	−16	12	2	18
y	58	247	153	3	341

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

36. Cada mes de abril, se lleva a cabo The Masters, uno de los torneos de golf más prestigiosos del PGA golf Tour, en Augusta, Georgia. En 2013, 60 jugadores recibieron premios en efectivo. El ganador de 2013, Adam Scott, de Australia, ganó un premio de 1 440 000 dólares; Miguel Cabrera terminó en segundo lugar, con un premio de 864 000 dólares; Jason Day finalizó en tercer lugar, con un premio de 544 000 dólares. Los datos se resumen brevemente en la tabla siguiente. Cada jugador tiene tres variables correspondientes: posición final (o rango), puntuación y premio (en dólares). El archivo completo está en las bases de datos disponibles en el sitio web del libro, www.mhhe.com/uni/lind_ae16e, etiquetados como Ex13-36. Se desea estudiar la relación entre puntuación y premio.

Posición	Jugador	Puntuación	Premio (en dólares)
1	Scott	279	$1 440 000
2	Cabrera	279	864 000
3	Day	281	544 000
Empate 4	Woods	283	352 000
Empate 4	Leishman	283	352 000
.	.	.	.
.	.	.	.
Empate 54	Bradley	297	18 320
Empate 54	Lyle	297	18 320
Empate 54	Mickelson	297	18 320
Empate 54	Piercy	297	18 320
58	Na	301	17 920
59	Peterson	302	17 760
60	Petterson	304	17 600

a. Usando “Puntuación” como la variable independiente, y “Premio” como la variable dependiente, desarrolle un diagrama de dispersión. ¿La relación parece ser lineal? ¿Parece razonable que a medida que “Puntuación” aumenta, “Premio” disminuye?
b. ¿Qué porcentaje de la variación en “Premio” está representado por “Puntuación”?
c. Calcule la nueva variable, “Premio-Log”, calculando el logaritmo a la base 10 de “Premio”. Desarrolle un diagrama de dispersión con “Premio-Log” como variable dependiente y “Puntuación” como la variable independiente.
d. Desarrolle una ecuación de regresión y calcule el coeficiente de determinación utilizando “Premio-Log” como la variable dependiente.
e. Compare el coeficiente de determinación en los puntos b y d. ¿Cuál es su conclusión?
f. Escriba la ecuación de regresión desarrollada en el punto d; si un jugador tuvo un total de 280 puntos en las cuatro rondas, ¿cuánto esperaría usted que dicho jugador ganara?

RESUMEN DEL CAPÍTULO

I. Un diagrama de dispersión es una herramienta gráfica para representar la relación entre dos variables.
  A. La variable dependiente se representa a escala en el eje *Y,* y es la variable que se debe estimar.
  B. La variable independiente se representa a escala en el eje *X,* y es la variable que se emplea como estimador.

II. El coeficiente de correlación mide la fuerza de la asociación lineal entre dos variables.
  A. Las dos variables deben estar al menos en la escala de medición de intervalo.
  B. El coeficiente de correlación varía desde −1.00 hasta 1.00.
  C. Si la correlación entre dos variables es cero, no hay asociación entre ellas.
  D. Un valor de 1.00 indica una correlación positiva perfecta, y uno de −1.00 muestra una correlación negativa perfecta.

E. Un signo positivo indica que hay una relación directa entre las variables, y un signo negativo, que hay una relación inversa.

F. Se designa con la letra *r,* y se determina mediante la siguiente ecuación:

$$r = \frac{\sum(x - \overline{x})(y - \overline{y})}{(n - 1)s_x s_y} \quad \textbf{[13.1]}$$

G. Para probar la hipótesis de que una correlación poblacional es distinta de cero. Se utiliza el siguiente estadístico:

$$t = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}} \quad \text{con } n - 2 \text{ grados de libertad} \quad \textbf{[13.2]}$$

III. En el análisis de regresión, se estima una variable con base en otra variable.

A. La que se estima es la variable dependiente.

B. La que se utiliza para hacer la estimación es la variable independiente.

1. La relación entre las variables debe ser lineal.

2. Las dos variables, independiente y dependiente, deben estar a escala de intervalo o de razón.

3. Con el criterio de mínimos cuadrados se determina la ecuación de regresión.

IV. La recta de regresión de mínimos cuadrados es de la forma $\hat{y} = a + bx$.

A. $\hat{y}$ es el valor estimado de y para un valor seleccionado de x.

B. a es la constante o intersección.

1. Es el valor de $\hat{y}$ cuando $x = 0$.

2. a se calcula con la siguiente ecuación.

$$a = \overline{y} - b\overline{x} \quad \textbf{[13.5]}$$

C. b es la pendiente de la recta ajustada.

1. Muestra la cantidad de cambio de $\hat{y}$ ante un cambio de una unidad en x.

2. Un valor positivo de b indica una relación directa entre las dos variables, y un valor negativo, una relación inversa.

3. El signo de b y el signo de r, el coeficiente de correlación, siempre son iguales.

4. b se calcula con la siguiente ecuación.

$$b = r\left(\frac{S_y}{S_x}\right) \quad \textbf{[13.4]}$$

D. x es el valor de la variable independiente.

V. En el caso de una ecuación de regresión, se prueba la pendiente para saber su significancia.

A. Se prueba la hipótesis de que la pendiente de la recta en la población es 0.

1. Si no se rechaza la hipótesis nula, se concluye que no hay relación entre las dos variables.

2. La prueba es equivalente a la que se realiza para el coeficiente de correlación.

B. Al probar la hipótesis nula con respecto a la pendiente, el estadístico de prueba con $n - 2$ grados de libertad.

$$t = \left(\frac{b - 0}{s_b}\right) \quad \textbf{[13.6]}$$

VI. El error estándar de estimación mide la variación alrededor de la recta de regresión.

A. Está en las mismas unidades que la variable dependiente.

B. Se basa en las desviaciones al cuadrado de la recta de regresión.

C. Los valores pequeños indican que los puntos se conectan estrechamente en la recta de regresión.

D. Se calcula utilizando la siguiente fórmula:

$$s_{y \cdot x} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - 2}} \quad \textbf{[13.7]}$$

VII. El coeficiente de determinación es la fracción de la variación de una variable dependiente que se explica por la variación de la variable independiente.

A. Varía de 0 a 1.0.

B. Es el cuadrado del coeficiente de correlación.

C. Se calcula a partir de la siguiente fórmula:

$$r^2 = \frac{\text{SSR}}{\text{SS Total}} = 1 - \frac{\text{SSE}}{\text{SS Total}} \quad \textbf{[13.8]}$$

VIII. La inferencia respecto de la regresión lineal se basa en las siguientes suposiciones.

A. Para un valor dado de x, los valores de Y están normalmente distribuidos respecto de la recta de regresión.

B. La desviación estándar de cada una de las distribuciones normales es la misma para todos los valores de *x,* y se estima mediante el error estándar de estimación.
C. Las desviaciones de la recta de regresión son independientes, sin un patrón debido al tamaño o la dirección.

IX. Hay dos tipos de estimaciones de intervalo.
A. En un intervalo de confianza, el valor medio de *y* se estima para un valor dado de *x*.
1. Se calcula a partir de la fórmula.

$$\hat{y} \pm t s_{y \cdot x} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}} \qquad \textbf{[13.10]}$$

2. El nivel de confianza, el tamaño del error estándar de estimación, el tamaño de la muestra y el valor de la variable independiente influyen en el ancho del intervalo.
B. En un intervalo de predicción, el valor individual de *y* se estima para un valor dado de *x*.
1. Se calcula a partir de la siguiente fórmula.

$$\hat{y} \pm t s_{y \cdot x} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}} \qquad \textbf{[13.11]}$$

2. La diferencia entre las fórmulas [13.10] y [13.11] es el 1 debajo del radical.
a. El intervalo de predicción será más amplio que el nivel de confianza.
b. El intervalo de predicción también se basa en el nivel de confianza, el tamaño del error estándar de estimación, el tamaño de la muestra y el valor de la variable independiente.

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
$\sum xy$	Suma de los productos de *x* y *y*	*Suma x y*
ρ	Coeficiente de correlación en la población	*Rho*
$\hat{y}$	Valor estimado de *Y*	*y prima*
$s_{y \cdot x}$	Error estándar de estimación	*s subíndice y punto x*
r^2	Coeficiente de determinación	*r cuadrada*

EJERCICIOS DEL CAPÍTULO

37. Una aerolínea comercial seleccionó una muestra aleatoria de 25 vuelos y determinó que la correlación entre el número de pasajeros y el peso total, en libras, del equipaje almacenado en el compartimiento para ello es 0.94. Con el nivel de significancia 0.05, ¿se puede concluir que hay una asociación positiva entre ambas variables?

38. Un sociólogo afirma que el éxito de los estudiantes en la universidad (medido por su promedio) se relaciona con el ingreso familiar; en una muestra de 20 estudiantes, el coeficiente de correlación es 0.40. Con el nivel de significancia 0.01, ¿se puede concluir que hay una correlación positiva entre las variables?

39. Un estudio que realizó la Agencia de Protección Ambiental en 12 automóviles reveló una correlación de 0.47 entre el tamaño del motor y sus emisiones. Con el nivel de significancia 0.01, ¿se puede concluir que hay una asociación positiva entre estas variables? ¿Cuál es el valor *p*? Interprete los resultados.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

40. Un hotel de los suburbios obtiene su ingreso bruto de la renta de sus instalaciones y de su restaurante. Los propietarios tienen interés en conocer la relación entre el número de habitaciones ocupadas por noche y el ingreso por día en el restaurante. En la siguiente tabla se presenta una muestra de 25 días (de lunes a jueves) del año previo que indica el ingreso del restaurante y la cantidad de habitaciones ocupadas.

Día	Ingreso	Habitaciones ocupadas	Día	Ingreso	Habitaciones ocupadas
1	$1 452	23	4	$1 470	39
2	1 361	47	5	1 456	37
3	1 426	21	6	1 430	29

(*continúa*)

(*continuación*)

Día	Ingreso	Habitaciones ocupadas	Día	Ingreso	Habitaciones ocupadas
7	$1 354	23	17	$1 348	19
8	1 442	44	18	1 450	38
9	1 394	45	19	1 431	44
10	1 459	16	20	1 446	47
11	1 399	30	21	1 485	43
12	1 458	42	22	1 405	38
13	1 537	54	23	1 461	51
14	1 425	27	24	1 490	61
15	1 445	34	25	1 426	39
16	1 439	15			

Utilice un paquete de software estadístico para responder las siguientes preguntas.

a. ¿Parece que aumenta el ingreso por desayunos a medida que aumenta el número de habitaciones ocupadas? Trace un diagrama de dispersión para apoyar su conclusión.

b. Determine el coeficiente de correlación entre las dos variables. Interprete el valor.

c. ¿Es razonable concluir que hay una relación positiva entre ingreso y habitaciones ocupadas? Utilice el nivel de significancia 0.10.

d. ¿Qué porcentaje de la variación de los ingresos del restaurante se contabilizan por el número de habitaciones ocupadas?

41. En la siguiente tabla se muestra el número de automóviles (en millones) vendidos en Estados Unidos durante varios años y el porcentaje de ellos que fabricó la compañía General Motors.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Año	Automóviles vendidos (millones)	Porcentaje de General Motors	Año	Automóviles vendidos (millones)	Porcentaje de General Motors
1950	6.0	50.2	1985	15.4	40.1
1955	7.8	50.4	1990	13.5	36.0
1960	7.3	44.0	1995	15.5	31.7
1965	10.3	49.9	2000	17.4	28.6
1970	10.1	39.5	2005	16.9	26.9
1975	10.8	43.1	2010	11.6	19.1
1980	11.5	44.0			

Utilice un paquete de software estadístico para responder las siguientes preguntas.

a. ¿El número de automóviles vendidos se relaciona de forma directa o indirecta con el porcentaje del mercado de la General Motors? Trace un diagrama de dispersión para apoyar su conclusión.

b. Determine el coeficiente de correlación entre las dos variables e interprete el valor.

c. ¿Es razonable concluir que hay una asociación negativa entre ambas variables? Utilice el nivel de significancia 0.01.

d. ¿Cuánta variación del mercado de la General Motors se contabiliza debido a la variación del número de automóviles vendidos?

42. En una muestra de 32 grandes ciudades de Estados Unidos, la correlación entre el número medio de pies cuadrados por empleado de oficina y la renta mensual media en el distrito comercial del centro es -0.363. Con el nivel de significancia 0.05, ¿se puede concluir que hay una asociación negativa entre las dos variables poblacionales?

43. A continuación se muestran los números de los puntos anotados y permitidos por cada uno de los 32 equipos de la Liga Nacional de Futbol de Estados Unidos durante la temporada 2012.

Equipo	Conferencia	Puntos anotados	Puntos permitidos	Equipo	Conferencia	Puntos anotados	Puntos permitidos
Chicago Bears	NFC	375	277	Atlanta Falcons	NFC	419	299
Green Bay Packers	NFC	433	336	New Orleans Saints	NFC	461	454
Minnesota Vikings	NFC	379	348	Tampa Bay Buccaneers	NFC	389	394
Detroit Lions	NFC	372	437	Carolina Panthers	NFC	357	363
Seattle Seahawks	NFC	412	245	New York Giants	NFC	429	344
San Francisco 49ers	NFC	397	273	Washington Redskins	NFC	436	388
St. Louis Rams	NFC	299	348	Dallas Cowboys	NFC	376	400

(*continúa*)

(*continuación*)

Equipo	Conferencia	Puntos anotados	Puntos permitidos	Equipo	Conferencia	Puntos anotados	Puntos permitidos
Arizona Cardinals	NFC	250	357	New York Jets	AFC	281	375
Philadelphia Eagles	NFC	280	444	Denver Broncos	AFC	481	289
Cincinnati Bengals	AFC	391	320	San Diego Chargers	AFC	350	350
Baltimore Ravens	AFC	398	344	Oakland Raiders	AFC	290	443
Pittsburgh Steelers	AFC	336	314	Kansas City Chiefs	AFC	211	425
Cleveland Browns	AFC	302	368	Houston Texans	AFC	416	331
New England Patriots	AFC	557	331	Indianapolis Colts	AFC	357	387
Miami Dolphins	AFC	288	317	Tennessee Titans	AFC	330	471
Buffalo Bills	AFC	344	435	Jacksonville Jaguars	AFC	255	444

Asumiendo que son datos simples, responda a las siguientes preguntas (se recomienda usar un software estadístico como ayuda).

a. ¿Cuál es el coeficiente de correlación entre estas variables? ¿Le sorprende que la asociación sea negativa? Interprete sus resultados.

b. Encuentre el coeficiente de determinación. ¿Qué le dice con respecto a la relación?

c. Con el nivel de significancia 0.05, ¿puede concluir que existe una asociación negativa entre "puntos anotados" y "puntos permitidos"?

d. Con el nivel de significancia 0.05, ¿puede concluir que existe una asociación negativa entre "puntos anotados" y "puntos permitidos" para cada conferencia?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

44. The Cotton Mill es una cadena de tiendas finas de ropa para dama, localizadas principalmente al sudoeste de Estados Unidos; debido a su reciente éxito, la directiva de The Cotton Mill está planeando expandirse, ubicando nuevas tiendas en otras regiones del país. Se le ha pedido al director de planeación que estudie la relación entre las ventas anuales y el tamaño de la tienda; como parte del estudio, el director selecciona una muestra de 25 tiendas y determina el tamaño de cada una en pies cuadrados, así como sus ventas del año previo. A continuación se presentan los datos muestrales (se sugiere el uso de un software estadístico).

Tamaño de la tienda (miles de pies cuadrados)	Ventas (millones de dólares)	Tamaño de la tienda (miles de pies cuadrados)	Ventas (millones de dólares)
3.7	9.18	0.4	0.55
2.0	4.58	4.2	7.56
5.0	8.22	3.1	2.23
0.7	1.45	2.6	4.49
2.6	6.51	5.2	9.90
2.9	2.82	3.3	8.93
5.2	10.45	3.2	7.60
5.9	9.94	4.9	3.71
3.0	4.43	5.5	5.47
2.4	4.75	2.9	8.22
2.4	7.30	2.2	7.17
0.5	3.33	2.3	4.35
5.0	6.76		

a. Trace un diagrama de dispersión; use el tamaño de la tienda como la variable independiente. ¿Parece existir una relación entre las dos variables? ¿Es positiva o negativa?

b. Determine el coeficiente de relación y el coeficiente de determinación. ¿La relación es fuerte o débil? ¿Por qué?

c. Con el nivel de significancia 0.05, ¿es posible concluir que existe una correlación positiva significativa?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

45. El fabricante de equipo para ejercicio Cardio Glide desea estudiar la relación entre el número de meses desde la compra de un aparato y el tiempo que se utilizó el aparato la semana anterior.

Persona	Meses con el equipo	Horas de uso	Persona	Meses con el equipo	Horas de uso
Rupple	12	4	Massa	2	8
Hall	2	10	Sass	8	3
Bennett	6	8	Karl	4	8
Longnecker	9	5	Malrooney	10	2
Phillips	7	5	Veights	5	5

a. Trace la información en un diagrama de dispersión; suponga que las horas de uso son la variable dependiente. Comente la gráfica.
b. Determine el coeficiente de correlación e interprete el resultado.
c. Con el nivel de significancia 0.01, ¿existe una asociación negativa entre las variables?

46. La siguiente ecuación de regresión se calculó a partir de una muestra de 20 observaciones:

$$\hat{y} = 15 - 5x$$

el resultado para SSE fue 100, y para SS total, 400.
a. Determine el error estándar de estimación.
b. Encuentre el coeficiente de determinación.
c. Determine el coeficiente de correlación (precaución: ¡cuidado con el signo!).

47. Los planeadores urbanos piensan que las ciudades más grandes están pobladas por residentes de más edad; para investigar la relación, reunieron datos sobre la población y la media de edad en 10 grandes ciudades.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Ciudad	Población (en millones)	Edad media
Chicago, IL	2.833	31.5
Dallas, TX	1.233	30.5
Houston, TX	2.144	30.9
Los Angeles, CA	3.849	31.6
New York, NY	8.214	34.2
Philadelphia, PA	1.448	34.2
Phoenix, AZ	1.513	30.7
San Antonio, TX	1.297	31.7
San Diego, CA	1.257	32.5
San Jose, CA	0.930	32.6

a. Trace estos datos en un diagrama de dispersión, con la edad media como la variable dependiente.
b. Encuentre el coeficiente de correlación.
c. Se realizó un análisis de regresión, y la ecuación de regresión resultante es Edad (media) = 31.4 + 0.272 Población. Interprete el significado de la pendiente.
d. Estime la media de edad en una ciudad de 2.5 millones de habitantes.
e. La siguiente es una fracción de la salida del software de la regresión. ¿Qué le dice esto?

```
Predictor      Coef    SE Coef       T       P
Constante   31.3672     0.6158   50.94   0.000
Población    0.2722     0.1901    1.43   0.190
```

f. Pruebe la significancia de la pendiente (con el nivel 0.10) e interprete el resultado. ¿Existe una relación significativa entre ambas variables?

48. Emily Smith decide comprar un auto que consuma poco combustible. Considera varios vehículos, con base en el costo estimado de compra y la edad del vehículo.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Vehículo	Costo estimado	Edad
Honda Insight	$5 555	8
Toyota Prius	$17 888	3
Toyota Prius	$9 963	6
Toyota Echo	$6 793	5
Honda Civic Hybrid	$10 774	5
Honda Civic Hybrid	$16 310	2
Chevrolet Prizm	$2 475	8
Mazda Protege	$2 808	10
Toyota Corolla	$7 073	9
Acura Integra	$8 978	8
Scion xB	$11 213	2
Scion xA	$9 463	3
Mazda3	$15 055	2
Mini Cooper	$20 705	2

a. Trace estos datos en un diagrama de dispersión, con “costo estimado” como la variable dependiente.

b. Calcule el coeficiente de correlación.
c. Se realizó un análisis de regresión y la ecuación de regresión resultante es Costo estimado = 18 358 − 1 534 Edad. Interprete el significado de la pendiente.
d. Calcule el costo de un auto de cinco años.
e. La siguiente es una fracción de la salida del software de la regresión. ¿Qué le dice esto?

```
Factor de pronóstico      Coef    SE Coef        T         P
Constante                18358       1817    10.10     0.000
Población              −1533.6      306.3    −5.01     0.000
```

f. Pruebe la significancia de la pendiente (con el nivel 0.10) e interprete el resultado. ¿Existe una relación significativa entre ambas variables?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

49. La National Highway Association estudia la relación entre el número de licitadores en un proyecto para una carretera y la mejor licitación (menor costo) del proyecto. De interés particular resulta saber si el número de licitadores aumenta o disminuye la cantidad de la oferta ganadora.

Proyecto	Número de licitadores, *x*	Oferta ganadora (millones de dólares), *y*	Proyecto	Número de licitadores, *x*	Oferta ganadora (millones de dólares), *y*
1	9	5.1	9	6	10.3
2	9	8.0	10	6	8.0
3	3	9.7	11	4	8.8
4	10	7.8	12	7	9.4
5	5	7.7	13	7	8.6
6	10	5.5	14	7	8.1
7	7	8.3	15	6	7.8
8	11	5.5			

a. Determine la ecuación de regresión e interprétela. ¿Más licitadores tienden a aumentar o a disminuir la cantidad de la oferta ganadora?
b. Estime la cantidad de la oferta ganadora si se hubieran presentado siete licitadores.
c. Se desea construir una nueva entrada en la carretera de cuota de Ohio. Se presentaron siete licitadores; determine un intervalo de predicción de 95% de la oferta ganadora.
d. Determine el coeficiente de determinación e interprete su valor.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

50. El señor William Profit estudia compañías que se hacen públicas por primera vez; le interesa en particular la relación entre el tamaño de la oferta y el precio por acción. Una muestra de 15 compañías que recién se hicieron públicas reveló la siguiente información.

Compañía	Tamaño (en millones de dólares), *x*	Precio por acción, *y*	Compañía	Tamaño (en millones de dólares), *x*	Precio por acción, *y*
1	9.0	10.8	9	160.7	11.3
2	94.4	11.3	10	96.5	10.6
3	27.3	11.2	11	83.0	10.5
4	179.2	11.1	12	23.5	10.3
5	71.9	11.1	13	58.7	10.7
6	97.9	11.2	14	93.8	11.0
7	93.5	11.0	15	34.4	10.8
8	70.0	10.7			

a. Determine la ecuación de regresión.
b. Realice una prueba para determinar si la pendiente de la línea de regresión es positiva.
c. Establezca el coeficiente de determinación. ¿Considera que el señor Profit debe estar satisfecho con el tamaño de la oferta como variable independiente?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

51. Bardi Trucking Co., ubicada en Cleveland, Ohio, hace entregas en la región de los Grandes Lagos, en las partes sur y norte. El presidente, Jim Bardi, estudia la relación entre la distancia de recorrido de un embarque y el tiempo, en días, que dura en llegar a su destino. Para investigar esta cuestión, el señor Bardi seleccionó una muestra aleatoria de 20 embarques del mes anterior; la distancia de envío es la variable independiente y el tiempo de envío es la variable dependiente. Los resultados son los siguientes:

Embarque	Distancia (millas)	Tiempo de envío (días)	Embarque	Distancia (millas)	Tiempo de envío (días)
1	656	5	11	862	7
2	853	14	12	679	5
3	646	6	13	835	13
4	783	11	14	607	3
5	610	8	15	665	8
6	841	10	16	647	7
7	785	9	17	685	10
8	639	9	18	720	8
9	762	10	19	652	6
10	762	9	20	828	10

a. Trace un diagrama de dispersión; con base en estos datos, ¿parece haber una relación entre la cantidad de millas que debe recorrer el embarque y el tiempo que tarda en llegar a su destino?
b. Determine el coeficiente de correlación. ¿Es posible concluir que hay una correlación positiva ente la distancia y el tiempo? Utilice el nivel de significancia 0.05.
c. Establezca e interprete el coeficiente de determinación.
d. Determine el error estándar de estimación.
e. ¿Recomendaría aplicar la ecuación de regresión para predecir el tiempo de envío? ¿Por qué sí o por qué no?

52. Super Markets, Inc., considera ampliarse hasta el área de Scottsdale, Arizona. Usted es el director de planeación y debe presentar un análisis de la ampliación propuesta al comité de operación de la junta de directores; como parte de su propuesta, necesita incluir información sobre la cantidad que gastan por mes en abarrotes las personas de la región. Tal vez deba incluir información sobre la relación entre la cantidad gastada en abarrotes y el ingreso. Su asistente reunió la siguiente información muestral.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Hogar	Cantidad gastada	Ingreso mensual
1	$ 555	$4 388
2	489	4 558
⋮	⋮	⋮
39	1 206	9 862
40	1 145	9 883

a. Sea la cantidad gastada la variable dependiente y el ingreso mensual la variable independiente. Trace un diagrama de dispersión con un paquete de software estadístico.
b. Determine la ecuación de regresión e interprete el valor de la pendiente.
c. Determine el coeficiente de correlación. ¿Puede concluir que es mayor que cero?

53. En la siguiente tabla se muestra información sobre el precio por acción y el dividendo de una muestra de 30 compañías.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Compañía	Precio por acción	Dividendo
1	$20.00	$ 3.14
2	22.01	3.36
⋮	⋮	⋮
29	77.91	17.65
30	80.00	17.36

a. Calcule la ecuación de regresión usando el precio de venta con base en el dividendo anual e interprete el valor de la pendiente.
b. Pruebe la significancia de la pendiente.
c. Encuentre el coeficiente de determinación e interprete su valor.
d. Determine el coeficiente de correlación. Con el nivel de significancia 0.05, ¿puede concluir que su valor es mayor que cero?

54. Un empleado de carreteras realizó un análisis de regresión de la relación entre el número de accidentes fatales en zonas de construcción y la cantidad de desempleados en el estado. La ecuación de regresión es Accidentes fatales $= 12.7 + 0.000114$ (Desempleados). Algunos datos adicionales se muestran en la página siguiente.

```
Factor de pronóstico          Coef       SE Coef       T       P
Constante                   12.726         8.115    1.57   0.134
Desempleados            0.00011386    0.00002896    3.93   0.001

Análisis de la varianza
Fuente             GL           SS            MS       F       P
Regresión           1        10354         10354   15.46   0.001
Error residual     18        12054           670
Total              19        22408
```

a. ¿Cuántos estados había en la muestra?
b. Determine el error estándar de estimación.
c. Encuentre el coeficiente de determinación.
d. Determine el coeficiente de correlación.
e. Con el nivel de significancia 0.05, ¿sugiere la evidencia que hay una asociación positiva entre los accidentes fatales y el número de desempleados?

55. El siguiente es un análisis de regresión que relaciona el valor actual de mercado (en dólares) con el tamaño en pies cuadrados de casas de Green County, Tennessee. La ecuación de regresión es: Valor $= -37\ 186 + 65.0$ Tamaño.

```
Coeficiente de pronóstico          Coef        SE Coef        T        P
Constante                        −37186           4629    −8.03    0.000
Tamaño                           64.993          3.047    21.33    0.000

Análisis de la varianza
Fuente           GL             SS              MS          F        P
Regresión         1    13548662082     13548662082     454.98    0.000
Error residual   33      982687392        29778406
Total            34    14531349474
```

a. ¿Cuántas casas había en la muestra?
b. Calcule el error estándar de estimación.
c. Calcule el coeficiente de determinación.
d. Calcule el coeficiente de correlación.
e. Con el nivel de significancia 0.05, ¿la evidencia sugiere una asociación positiva entre el valor de mercado de las casas y el tamaño de la casa (en pies cuadrados)?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

56. En la tabla de la derecha se muestra el interés porcentual anual del capital (rentabilidad) y el crecimiento porcentual anual medio de las ventas de ocho compañías aeroespaciales y de defensa.

Compañía	Rentabilidad	Crecimiento
Alliant Techsystems	23.1	8.0
Boeing	13.2	15.6
General Dynamics	24.2	31.2
Honeywell	11.1	2.5
L-3 Communications	10.1	35.4
Northrop Grumman	10.8	6.0
Rockwell Collins	27.3	8.7
United Technologies	20.1	3.2

a. Calcule el coeficiente de correlación y realice una prueba de hipótesis para determinar si es razonable concluir que la correlación poblacional es mayor que cero. Utilice el nivel de significancia 0.05.
b. Elabore la ecuación de regresión de la rentabilidad con base en el crecimiento. ¿Es posible concluir que la pendiente de la recta de regresión es negativa?
c. Utilice un paquete de software estadístico para determinar el residual de cada observación. ¿Qué compañía tiene el mayor residual?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

57. En los siguientes datos se registra el precio al menudeo de 12 computadoras portátiles, seleccionadas al azar, junto con sus velocidades de procesador correspondientes en gigahertz.

Computadora	Velocidad	Precio	Computadora	Velocidad	Precio
1	2.0	$2 017	7	2.0	$2 197
2	1.6	922	8	1.6	1 387
3	1.6	1 064	9	2.0	2 114
4	1.8	1 942	10	1.6	2 002
5	2.0	2 137	11	1.0	937
6	1.2	1 012	12	1.4	869

a. Elabore una ecuación lineal que sirva para describir cómo depende el precio de la velocidad del procesador.
b. Con base en su ecuación de regresión, ¿hay alguna computadora que parezca tener, de manera particular, un precio menor o mayor?
c. Calcule el coeficiente de correlación entre dos variables. Con el nivel de significancia 0.05 realice una prueba de hipótesis para determinar si la correlación de la población puede ser mayor que cero.

58. Una cooperativa de compras para el consumidor probó el área de calefacción efectiva de 20 calentadores eléctricos distintos, con consumos, en vatios, diferentes. Los resultados son los siguientes.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Calentador	Varios	Área	Calentador	Varios	Área
1	1 500	205	11	1 250	116
2	750	70	12	500	72
3	1 500	199	13	500	82
4	1 250	151	14	1 500	206
5	1 250	181	15	2 000	245
6	1 250	217	16	1 500	219
7	1 000	94	17	750	63
8	2 000	298	18	1 500	200
9	1 000	135	19	1 250	151
10	1 500	211	20	500	44

a. Calcule la correlación entre consumo en vatios y área de calefacción. ¿Existe una relación directa o indirecta?
b. Realice una prueba de hipótesis para determinar si es razonable que el coeficiente sea mayor que cero. Utilice el nivel de significancia 0.05.
c. Elabore la ecuación de regresión del calentamiento efectivo con base en el consumo en vatios.
d. ¿Qué calentador parece la “mejor compra” con base en el tamaño del residuo?

59. Un entrenador de perros investiga la relación entre el tamaño del can (peso en libras) y su consumo alimentario diario (medido en tazas estándar). El resultado de una muestra de 18 observaciones es el siguiente:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Can	Peso	Consumo	Can	Peso	Consumo
1	41	3	10	91	5
2	148	8	11	109	6
3	79	5	12	207	10
4	41	4	13	49	3
5	85	5	14	113	6
6	111	6	15	84	5
7	37	3	16	95	5
8	111	6	17	57	4
9	41	3	18	168	9

a. Calcule el coeficiente de correlación. ¿Es razonable concluir que la correlación entre la población es mayor que cero? Utilice el nivel de significancia 0.05.
b. Elabore la ecuación de regresión de las tazas con base en el peso del can. ¿Cuánto cambia el peso estimado del perro cada taza adicional de alimento?
c. ¿Come demasiado o come menos uno de los perros?

60. La Waterbury Insurance Company desea estudiar la relación entre la cantidad de daño por fuego, la distancia entre la casa ardiendo y la estación de bomberos más cercana; esta información se empleará en el ajuste de la cobertura del seguro. Mediante una muestra de 30 demandas durante el año previo, el director del departamento de actuarios determinó la distancia de la estación de bomberos (x) y la cantidad de daños, en miles de dólares (y). En la página siguiente se presenta la salida en pantalla de MegaStat.

Siga las indicaciones y responda las preguntas.

a. Elabore la ecuación de regresión. ¿Hay una relación directa o indirecta entre la distancia de la estación de bomberos y la cantidad de daño?
b. ¿Cuánto daño estimaría que provoca un incendio situado a cinco millas de la estación de bomberos más cercana?

Tabla ANOVA

Fuente	SS	gl	MS	F
Regresión	1,864.5782	1	1,864.5782	38.83
Residuo	1,344.4934	28	48.0176	
Total	3,209.0716	29		

Salida de la regresión

Variables	Coeficientes	Error estándar	$t(gl = 28)$
Intersección	12.3601	3.2915	3.755
Distancia–X	4.7956	0.7696	6.231

c. Encuentre e interprete el coeficiente de determinación.

d. Determine el coeficiente de correlación e interprete su valor. ¿Cómo determinó el signo del coeficiente de correlación?

e. Realice una prueba de hipótesis para determinar si hay una relación significativa entre la distancia a la estación de bomberos y la cantidad de daño. Utilice el nivel de significancia 0.01 y una prueba de dos colas.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

61. Un servicio de viajes aéreos muestrea los vuelos de una aerolínea doméstica para explorar la relación entre sus tarifas y las distancias recorridas. Si existe una relación, ¿qué porcentaje de la variación en la tarifa está representado por la distancia? ¿Cuánto añade a la tarifa una milla extra? A continuación se presentan los datos.

Origen	Destino	Distancia	Tarifa (en dólares)
Detroit, MI	Myrtle Beach, SC	636	$109
Baltimore, MD	Sacramento, CA	2 395	252
Las Vegas, NV	Filadelfia, PA	2 176	221
Sacramento, CA	Seattle, WA	605	151
Atlanta, GA	Orlando, FL	403	138
Boston, MA	Miami, FL	1 258	209
Chicago, IL	Covington, KY	264	254
Columbus, OH	Minneapolis, MN	627	259
Fort Lauderdale, FL	Los Ángeles, CA	2 342	215
Chicago, IL	Indianapolis, IN	177	128
Filadelfia, PA	San Francisco, CA	2 521	348
Houston, TX	Raleigh/Durham, NC	1 050	224
Houston, TX	Midland/Odessa, TX	441	175
Cleveland, OH	Dallas/Ft.Worth, TX	1 021	256
Baltimore, MD	Columbus, OH	336	121
Boston, MA	Covington, KY	752	252
Kansas City, MO	San Diego, CA	1 333	206
Milwaukee, WI	Phoenix, AZ	1 460	167
Portland, OR	Washington, DC	2 350	308
Phoenix, AZ	San José, CA	621	152
Baltimore, MD	St. Louis, MO	737	175
Houston, TX	Orlando, FL	853	191
Houston, TX	Seattle, WA	1 894	231
Burbank, CA	Nueva York, NY	2 465	251
Atlanta, GA	San Diego, CA	1 891	291
Minneapolis, MN	Nueva York, NY	1 028	260
Atlanta, GA	West Palm Beach, FL	545	123
Kansas City, MO	Seattle, WA	1 489	211
Baltimore, MD	Portland, ME	452	139
Nueva Orleans, LA	Washington, DC	969	243

a. Trace un diagrama de dispersión con “Distancia” como la variable independiente y “Tarifa” como la variable dependiente. ¿La relación es directa o indirecta?

b. Calcule el coeficiente de correlación. Con el nivel de significancia 0.05, ¿es razonable concluir que el coeficiente de relación es mayor a cero?

c. ¿Qué porcentaje de la variación en “Tarifa” está representado por “Distancia”?

d. Determine la ecuación de regresión. ¿Cuánto añade cada milla adicional a la tarifa? Estime la tarifa para un vuelo de 1 500 millas.

e. Un viajero planea volar de Atlanta a Heathrow, Londres (4 218 millas), y desea usar la ecuación de regresión para estimar la tarifa. Explique por qué no sería buena idea estimar la tarifa para este vuelo internacional con la ecuación de regresión.

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

62. Consulte los datos de sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior.

- **a.** Sea "Precio de venta" la variable dependiente, y "Tamaño de la casa", la variable independiente. Determine la ecuación de regresión y estime el precio de venta de una casa con un área de 2 200 pies cuadrados. Determine el intervalo de confianza de 95% y el intervalo de predicción de 95% del precio de venta de una casa con área de 2 200 pies cuadrados.
- **b.** Sea "Precio de venta" la variable dependiente, y "Distancia desde el centro de la ciudad", la variable independiente. Determine la ecuación de regresión y estime el precio de venta de una casa a 20 millas del centro de la ciudad. Encuentre el intervalo de confianza de 95% y el intervalo de predicción de 95% de las casas a 20 millas del centro de la ciudad.
- **c.** ¿Puede concluir que las variables independientes "Distancia desde el centro de la ciudad" y "Precio de venta" se correlacionan en forma negativa, y que el área de la casa y el precio de venta se correlacionan en forma positiva? Utilice el nivel de significancia 0.05. Reporte el valor p de la prueba y resuma sus resultados en un breve reporte.

63. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Sea "Asistencia" la variable dependiente, y "Salario total del equipo (en millones de dólares)" la variable independiente. Determine la ecuación de regresión y conteste las siguientes preguntas.

- **a.** Trace un diagrama de dispersión; con base en ese diagrama, ¿parece haber una relación directa entre ambas variables?
- **b.** ¿Cuál es la asistencia esperada para un equipo con un salario de 80.0 millones de dólares?
- **c.** Si los propietarios pagan 30 millones de dólares adicionales, ¿cuánta gente más podrían esperar que asistiera?
- **d.** Con el nivel de significancia 0.05, ¿se puede concluir que la pendiente de la recta de regresión es positiva? Realice la prueba de hipótesis correspondiente.
- **e.** ¿Qué porcentaje de la variación en asistencia está representado por el salario?
- **f.** Determine la correlación entre la asistencia y el promedio de bateo por equipo, y entre la asistencia y el promedio de carreras. ¿Cuál es más fuerte? Realice la correspondiente prueba de la hipótesis para cada grupo de variables.

64. Consulte los datos de los autobuses del Distrito Escolar Buena. Desarrolle una ecuación de regresión que exprese la relación entre "Edad del autobús" como variable independiente, y "Mantenimiento" como variable dependiente.

- **a.** Trace un diagrama de dispersión. ¿Qué sugiere el diagrama con respecto a la relación entre las dos variables? ¿Es directa o indirecta, fuerte o débil?
- **b.** Desarrolle una ecuación de regresión. ¿Cuánto añade al mantenimiento un año más de vida? ¿Cuál es el costo estimado de mantenimiento de un camión de diez años de edad?
- **c.** Realice una prueba de hipótesis para determinar si la pendiente de la recta de regresión es mayor a cero. Utilice el nivel de significancia 0.05. Interprete sus resultados de los puntos a, b y c en un breve reporte.

14 Análisis de regresión múltiple

EL BANCO DE NUEVA INGLATERRA estudia datos de créditos recientes. En particular, le interesa saber a qué grado factores tales como el valor de la casa que se desea comprar, el nivel de educación del cabeza de familia, su edad, el pago mensual actual de la hipoteca y el género de dicha persona se relacionan con el ingreso familiar. ¿Las variables propuestas predicen con eficacia la variable dependiente "ingreso familiar"? (vea el ejercicio 12 de la sección "Regresión por pasos" y el OA14-1).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA14-1 Utilizar un análisis de regresión múltiple para describir e interpretar la relación entre diversas variables independientes y una variable dependiente.

OA14-2 Evaluar qué tan bien se ajusta a los datos una ecuación de regresión múltiple.

OA14-3 Probar hipótesis acerca de las relaciones inferidas por un modelo de regresión múltiple.

OA14-4 Evaluar los supuestos de la regresión múltiple.

OA14-5 Utilizar e interpretar una variable ficticia cualitativa en la regresión múltiple.

OA14-6 Incluir e interpretar un efecto de interacción en el análisis de regresión múltiple.

OA14-7 Aplicar la regresión por pasos para desarrollar un modelo de regresión múltiple.

OA14-8 Aplicar las técnicas de regresión múltiple para desarrollar un modelo lineal.

Introducción

En el capítulo 13 se describió la relación entre un par de variables en escala de intervalo o de razón. En este capítulo se comienza con el estudio del coeficiente de correlación, el cual mide la fuerza de una relación. Un coeficiente cercano a 1.00 (por ejemplo, -0.88 o 0.78) indica una relación lineal muy fuerte, en tanto que un valor cercano a 0 (por ejemplo, -0.12 o 0.18) indica que la relación es débil. A continuación se desarrolla un procedimiento que determina una ecuación lineal para expresar la relación entre las dos variables. Este procedimiento se llama *recta de regresión* y describe la relación entre las variables. También describe el patrón general de una variable dependiente (y) de una variable independiente o variable explicativa (x).

En la correlación y regresión lineal múltiple, se emplean variables independientes adicionales (denotadas X_1, X_2,…, y así sucesivamente) que ayudan a explicar o predecir mejor a la variable dependiente (y). Casi todas las ideas estudiadas en la correlación y regresión lineal simple se amplían a esta situación más general; sin embargo, las variables independientes adicionales permiten hacer algunas consideraciones nuevas. El análisis de regresión múltiple sirve como técnica descriptiva o como técnica de inferencia.

Análisis de regresión múltiple

OA14-1
Utilizar un análisis de regresión múltiple para describir e interpretar la relación entre diversas variables independientes y una variable dependiente.

La forma descriptiva general de una ecuación lineal múltiple se muestra en la fórmula [14.1]. Se utiliza k para representar el número de variables independientes; por lo tanto, k puede ser cualquier número entero positivo.

ECUACIÓN GENERAL DE REGRESIÓN MÚLTIPLE	$\hat{y} - a + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_kx_k$	**[14.1]**

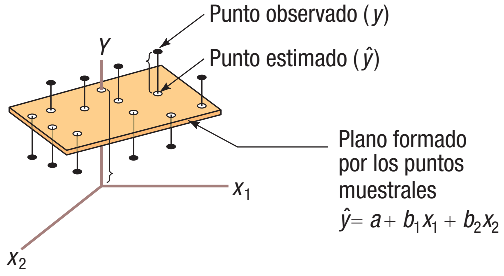
donde:

- a es la intersección, el valor de $\hat{y}$ cuando todas las x son cero;
- b_j es la cantidad en que $\hat{y}$ cambia cuando esa x_j particular aumenta una unidad, mientras los valores de todas las demás variables independientes se mantienen constantes. El subíndice j es solo un identificador de cada variable independiente; no se emplea en los cálculos. En general, el subíndice es un número entero entre 1 y k, el cual representa la cantidad de variables independientes; sin embargo, el subíndice también puede ser un identificador breve o abreviado, por ejemplo, la edad puede servir como subíndice.

En el capítulo 13 (en el análisis de regresión) se describió y comprobó la relación entre una variable dependiente, $\hat{y}$, y una sola variable independiente, x. La relación entre $\hat{y}$ y x se representa en forma esquemática mediante una recta. Cuando hay dos variables independientes, la ecuación de regresión es

$$\hat{y} = a + b_1x_1 + b_2x_2$$

Como hay dos variables independientes, esta relación se representa de forma descriptiva como un plano, y se muestra en la gráfica 14.1; en la cual se presentan los residuos como la diferencia entre la y real y la $\hat{y}$ ajustada en el plano. Cuando un análisis de regresión múltiple incluye más de dos variables independientes, no se pueden emplear gráficas para explicarlo porque están limitadas a tres dimensiones.



GRÁFICA 14.1 Plano de regresión con diez puntos muestrales

La interpretación de la intersección y los dos coeficientes de regresión se ilustran mediante el siguiente ejemplo: suponga que el precio de venta de una casa tiene una relación directa con el número de habitaciones e *inversa* con su edad. Sea x_1 el número de habitaciones, x_2 la edad de la casa en años, y y el precio de venta del inmueble (en miles de dólares).

Suponga que la ecuación de regresión, calculada con software estadístico, es:

$$\hat{y} = 21.2 + 18.7x_1 - 0.25x_2$$

El valor de la intersección de 21.2 indica que la ecuación de regresión (plano) intersecta el eje Y en 21.2. Esto ocurre cuando el número de habitaciones y la edad de la casa son cero; se puede decir que 21 200 dólares es el valor promedio de una propiedad que no está fincada.

El primer coeficiente de relación, 18.7, indica que por cada aumento de una habitación en el tamaño de una casa, el precio de venta se incrementará en 18.7 mil dólares (18 700 dólares), independientemente de la edad del inmueble; el segundo coeficiente de regresión, -0.25, indica que por cada aumento de un año en edad, el precio de venta *disminuirá* en 0.25 mil dólares (250 dólares) independientemente del número de habitaciones. Como ejemplo, se espera que una casa de siete habitaciones y 30 años de antigüedad se venda en 144 600 dólares.

$$\hat{y} = 21.2 + 18.7x_1 - 0.25x_2 = 21.2 + 18.7(7) - 0.25(30) = 144.6$$

Los valores de los coeficientes de la ecuación lineal múltiple se determinan mediante el método de mínimos cuadrados. Recuerde, del capítulo anterior, que en dicho método se suman las diferencias elevadas al cuadrado entre los valores ajustados y reales de y tan pequeña como sea posible, con lo cual el término $\sum(y - \hat{y})^2$ se minimiza. Los cálculos son muy tediosos, por lo que suelen realizarse mediante un paquete de software estadístico.

En el siguiente ejemplo se muestra un análisis de regresión múltiple con tres variables independientes mediante Excel, que produce un grupo estándar de estadísticas y reportes. Otros programas estadísticos, como Minitab y Megastat también incluyen técnicas avanzadas de análisis de regresión.

EJEMPLO

Salsberry Realty vende casas en la costa este de Estados Unidos. Una de las preguntas más frecuentes de los compradores potenciales es: “Si compramos esta casa, ¿cuánto gastaremos en calefacción durante el invierno?”. Al departamento de investigación de Salsberry se le pidió desarrollar algunas directrices respecto de los costos de calefacción de casas unifamiliares. Se considera que tres variables se relacionan con dichos costos: 1) la temperatura externa diaria media, 2) el número de pulgadas de aislamiento en el ático y 3) los años de uso del calentador. Para el estudio, el departamento de investigación de Salsberry seleccionó una muestra aleatoria de 20 casas de venta reciente, y determinó el costo de calefacción de cada casa en enero pasado, así como la temperatura externa en enero en la región, el número de pulgadas de aislamiento del ático y los años de uso del calentador. La información muestral se reporta en la tabla 14.1.

TABLA 14.1 Factores del costo de calefacción en enero de una muestra de 20 casas

Casa	Costo de calefacción (dólares)	Temperatura externa media (°F)	Aislamiento del ático (pulgadas)	Antigüedad del calentador (años)
1	$250	35	3	6
2	360	29	4	10
3	165	36	7	3
4	43	60	6	9
5	92	65	5	6
6	200	30	5	5
7	355	10	6	7
8	290	7	10	10
9	230	21	9	11
10	120	55	2	5
11	73	54	12	4
12	205	48	5	1
13	400	20	5	15
14	320	39	4	7
15	72	60	8	6
16	272	20	5	8
17	94	58	7	3
18	190	40	8	11
19	235	27	9	8
20	139	30	7	5

ESTADÍSTICA EN ACCIÓN

Muchos estudios indican que una mujer ganará cerca de 70% del salario de un hombre en el mismo puesto. Investigadores del University of Michigan Institute for Social Research determinaron que los factores sociales explican alrededor de un tercio de la diferencia, como desigualdades en educación, experiencia e interrupciones en el trabajo. Los dos tercios restantes no se explican por estos factores sociales.

Los datos de la tabla 14.1 se proporcionan en formato de Excel en el sitio web del libro, **www.mhhe.com/uni/lind_ae16e**. Las instrucciones básicas de Excel para utilizar estos datos se encuentran en la sección “Comandos de Software”, en el apéndice C.

Determine la ecuación de regresión múltiple. ¿Cuáles son las variables independientes y cuál es la variable dependiente? Analice los coeficientes de regresión. ¿Qué indica si algunos coeficientes son positivos y otros negativos? ¿Cuál es el valor de la intersección y cuál es el costo de calefacción estimado de una casa si la temperatura externa media es de 30 grados, si el ático tiene 5 pulgadas de aislamiento y el calentador tiene 10 años?

SOLUCIÓN

Inicie el análisis definiendo la variable dependiente y las independientes; la primera es el costo de calefacción en enero, y se representa con y; por otro lado, hay tres variables independientes:

- La temperatura externa media en enero, representada por x_1.
- El número de pulgadas de aislamiento del ático, representado por x_2.
- La antigüedad en años del calentador, representada por x_3.

Con estas definiciones, la forma general de la ecuación de regresión múltiple se muestra a continuación. El valor $\hat{y}$ se emplea para estimar el valor de y.

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$$

Ahora que se definió la ecuación de regresión, calcule con Excel todas las estadísticas necesarias para el análisis. Las capturas de pantalla de Excel se muestran a continuación.

Para predecir el costo de calefacción en enero con la ecuación de regresión es necesario conocer los valores de los coeficientes de regresión, b_j; estos están resaltados en los reportes del software. El software usa los nombres de variables o identificadores asociados con cada variable independiente. La intersección de la ecuación de regresión, a, se identifica “intersección” en la captura de pantalla de Excel.

	A	B	C	D	E	F	G	H	I	J	K
1	Costo	Temp.	Aislam.	Edad		RESUMEN DE SALIDA					
2	250	35	3	6							
3	360	29	4	10		*Estadísticos de regresión*					
4	165	36	7	3		Múltiple *R*	0.897				
5	43	60	6	9		*R* cuadrada	0.804				
6	92	65	5	6		*R* cuadrada ajustada	0.767				
7	200	30	5	5		Error estándar	51.049				
8	355	10	6	7		Observaciones	20				
9	290	7	10	10							
10	230	21	9	11		ANOVA					
11	120	55	2	5			*gl*	*SS*	*MS*	*F*	*Significancia F*
12	73	54	12	4		Regresión	3	171220.473	57073.491	21.901	0.000
13	205	48	5	1		Residual	16	41695.277	2605.955		
14	400	20	5	15		Total	19	212915.750			
15	320	39	4	7							
16	72	60	8	6			*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	272	20	5	8		Intercepción	427.194	59.601	7.168	0.000	
18	94	58	7	3		Temperatura	-4.583	0.772	-5.934	0.000	
19	190	40	8	11		Aislamiento	-14.831	4.754	-3.119	0.007	
20	235	27	9	8		Edad	6.101	4.012	1.521	0.148	
21	139	30	7	5							

En este caso, la ecuación de regresión estimada es:

$$\hat{y} = 427.194 - 4.583x_1 - 14.831x_2 + 6.101x_3$$

Ahora puede estimar o predecir el costo de calefacción en enero de una casa si conoce la temperatura externa media, las pulgadas de aislamiento y la antigüedad del calentador; por ejemplo, para una casa con temperatura externa media mensual de 30 grados (X_1), hay 5 pulgadas de aislamiento en el ático (X_2) y el calentador tiene 10 años (X_3). Al sustituir los valores de las variables independientes:

$$\hat{y} = 427.194 - 4.583(30) - 14.831(5) + 6.101(10) = 276.56$$

El costo estimado de calefacción en enero es de 276.56 dólares.

Los coeficientes de regresión y sus signos algebraicos también proporcionan información acerca de sus relaciones individuales con el costo de calefacción en enero. El coeficiente de regresión de una temperatura externa media es -4.583; observe que este es negativo y presenta una relación inversa entre el costo de calefacción y la temperatura. Eso no es sorprendente. Conforme la temperatura externa aumenta, disminuye el costo para calentar la casa. El valor numérico del coeficiente de regresión proporciona más información: si la temperatura aumenta un grado y las otras dos variables independientes se mantienen constantes, se estima una disminución de 4 583 dólares en el costo de calefacción mensual; por lo tanto, si la temperatura media en Boston es 25 grados y en Filadelfia de 35 grados, y todos los demás elementos son iguales (aislamiento y antigüedad del calentador), se espera que el costo de calefacción sea 45.83 dólares menos en Filadelfia.

La variable “aislamiento del ático” también presenta una relación inversa: mientras más aislamiento tenga este, menor será el costo de calefacción de la casa; por lo tanto, es lógico el signo negativo de este coeficiente. Por cada pulgada adicional de aislamiento, se espera que el costo de calefacción de la casa disminuya 14.83 dólares por mes, si se mantienen constantes la temperatura externa y la antigüedad del calentador.

La variable “antigüedad del calentador” presenta una relación directa. Con un calentador antiguo, aumenta el costo para calentar la casa. Específicamente, por cada año adicional que tenga el calentador, se espera que el costo aumente 6.10 dólares por mes.

AUTOEVALUACIÓN 14-1

En el noreste de Carolina del Sur hay muchos restaurantes que dan servicio a las personas que toman sus vacaciones en la playa en el verano, a golfistas en el otoño y primavera, y a esquiadores en el invierno. Bill y Joyce Tuneall administran varios restaurantes en el área del norte de Jersey y consideran cambiarse a Myrtle Beach, Carolina del Sur, para abrir uno nuevo. Antes de tomar la decisión final desean estudiar algunos restaurantes existentes y las variables que parezcan relacionarse con la rentabilidad; por lo tanto, reúnen información muestral donde las ganancias (reportadas en miles de dólares) forman la variable dependiente, y las variables independientes son:

x_1 el número de cajones de estacionamiento cerca del restaurante;
x_2 el número de horas que está abierto el restaurante por semana;
x_3 la distancia desde el SkyWheel, un sitio emblemático en Myrtle Beach;
x_4 el número de empleados;
x_5 el número de años que el propietario actual ha tenido el restaurante.

La siguiente es parte de la captura de pantalla que se obtuvo con software estadístico.

Factor de predicción	Coeficiente	Error estándar del coeficiente	t
Constante	2.50	1.50	1.667
x_1	3.00	1.50	2.000
x_2	4.00	4.00	1.333
x_3	−3.00	0.20	−15.000
x_4	0.20	0.05	4.000
x_5	1.00	1.50	0.667

(a) ¿Cuál es la ganancia de un restaurante con 40 cajones de estacionamiento, que abre 72 horas a la semana, se encuentra a 10 millas de SkyWheel, tiene 20 empleados y ha sido operado por el dueño actual durante 5 años?
(b) Interprete los valores de b_2 y b_3 en la ecuación de regresión múltiple.

EJERCICIOS

1. El director de marketing de Reeves Wholesales Products está estudiando las ventas mensuales; para tal efecto, seleccionó tres variables independientes para estimar las ventas: población regional, ingreso *per cápita* y la tasa de desempleo regional. La ecuación de regresión se calculó (en dólares):

$$\hat{y} = 64\,100 + 0.094x_1 + 9.6x_2 - 11.600x_3$$

a. ¿Cuál es el nombre completo de la ecuación?
b. Interprete el número 64 100.
c. ¿Cuáles son las ventas mensuales estimadas en una región particular con una población de 796 000, un ingreso *per cápita* de 6 940 dólares y una tasa de desempleo de 6.0%?

2. Thompson Photo Works compró varias máquinas nuevas de procesamiento muy complejas. El departamento de producción necesitó ayuda respecto de las aptitudes necesarias para un operador de estas máquinas. ¿La edad es un factor? ¿Es importante el tiempo de servicio como operador (en años)? Se señalaron cuatro variables a fin de explorar más a fondo los factores necesarios para estimar el desempeño de las nuevas máquinas de procesamiento:

x_1 el tiempo del empleado en la industria;
x_2 la calificación en la prueba de aptitud mecánica;

x_3 las calificaciones anteriores en el trabajo;
x_4 la edad.

El desempeño de la máquina nueva se designa y.

Se seleccionaron 30 empleados al azar, se recopilaron datos de cada uno y se registraron sus desempeños en las máquinas nuevas. Algunos resultados son:

Nombre	Desempeño en la máquina nueva, y	Tiempo en la industria, x_1	Calificación en aptitud mecánica, x_2	Desempeño anterior en el trabajo, x_3	Edad, x_4
Mike Miraglia	112	12	312	121	52
Sue Trythall	113	2	380	123	27

La ecuación es:

$$\hat{y} = 11.6 + 0.4x_1 + 0.286x_2 + 0.112x_3 + 0.002x_4$$

a. ¿Cómo se le denomina a esta ecuación?
b. ¿Cuántas variables dependientes e independientes hay?
c. ¿Cómo se denomina al número 0.286?
d. Conforme aumenta la edad en un año, ¿cuánto aumenta el desempeño estimado en la nueva máquina?
e. Carl Knox solicitó trabajo en Photo Works, él ha estado en el negocio durante seis años, y obtuvo una calificación de 280 en la prueba de aptitud mecánica (la calificación del desempeño anterior en el trabajo de Carl fue 97, y tiene 35 años de edad). Estime su desempeño en la nueva máquina.

3. El Departamento de Recursos Humanos de General Mills, Inc., contrató a un grupo consultor para encuestar a los empleados de la compañía con respecto al grado de satisfacción con su vida actual. Se empleó un índice especial, denominado índice de satisfacción; además, se estudiaron seis factores, a saber, la edad en la que se casaron por primera vez (x_1), el ingreso anual (x_2), el número de hijos vivos (x_3), el valor de todos sus bienes (x_4), el estado de salud en forma de índice (x_5) y el número promedio de actividades sociales por semana, como jugar al boliche y bailar (x_6). Suponga que la ecuación de regresión múltiple es:

$$\hat{y} = 16.24 + 0.017x_1 + 0.0028x_2 + 42x_3 + 0.0012x_4 + 0.19x_5 + 26.8x_6$$

a. ¿Cuál es índice de satisfacción estimado de una persona que se casó por primera vez a los 18 años, con un ingreso anual de 26 500 dólares, tres hijos vivos, bienes por 156 000 dólares, un índice de estado de salud de 141, y tiene un promedio de 2.5 actividades sociales a la semana?
b. ¿Qué daría más satisfacción: un ingreso adicional de 10 000 dólares al año o dos actividades sociales más a la semana?

4. Cellulon, fabricante de aislamiento para casas, desea desarrollar guías para informar a constructores y consumidores acerca de la forma en que el espesor del aislamiento del ático de una casa y la temperatura externa afectan el consumo de gas natural; por lo tanto, modificó el espesor del aislamiento y la temperatura un un laboratorio. Algunos resultados se muestran a la derecha.

Con base en los resultados muestrales, la ecuación de regresión es:

Consumo de gas natural mensual (pies cúbicos), y	Espesor del aislamiento (pulgadas), x_1	Temperatura externa (°F), x_2
30.3	6	40
26.9	12	40
22.1	8	49

$$\hat{y} = 62.65 - 1.86x_1 - 0.52x_2$$

a. ¿Qué cantidad de gas natural esperan consumir por mes los propietarios de las casas si instalan seis pulgadas de aislamiento y la temperatura exterior es de 40 °F?
b. ¿Qué efecto tendría instalar siete pulgadas de aislamiento en lugar de seis en el consumo mensual de gas natural (si la temperatura externa permanece en 40°F)?
c. ¿Por qué son negativos los coeficientes de regresión b_1 y b_2? ¿Es lógico que lo sean?

Evaluación de una ecuación de regresión múltiple

OA14-2
Evaluar qué tan bien se ajusta a los datos una ecuación de regresión múltiple.

Muchas estadísticas y métodos estadísticos se utilizan para evaluar la relación entre una variable dependiente y más de una variable independiente. El primer paso es expresar la relación en términos de una ecuación de regresión múltiple; a continuación se siguen aplicando los conceptos que se presentaron en el capítulo 13, mediante la información que se encuentra en una tabla ANOVA para evaluar con qué nivel de precisión se ajusta la ecuación a los datos.

La tabla ANOVA

Como se hizo en el capítulo 13, el análisis estadístico de una ecuación de regresión múltiple se resume en una tabla ANOVA. Recuerde que la variación total de una variable dependiente, y, se divide en dos componentes: 1) *regresión*, o la variación de y explicada por todas las variables independientes, y 2) *el error* o *residuo*, o variación no explicada de y. En la primera columna de la siguiente tabla ANOVA se identifican ambas categorías. La columna con el encabezado "*gl*" se refiere a los grados de libertad asociados con cada categoría. El número total de *gl* es $n - 1$; su cantidad en la regresión es igual al número de variables independientes en la ecuación de regresión múltiple. Los grados de libertad de la regresión se denominan k. El número de *gl* asociados con el término error es igual al total de estos menos los de la regresión. En una regresión múltiple, los grados de libertad son $n - (k + 1)$.

Fuente	***gl***	**SS**	**MS**	***F***
Regresión	k	SSR	MSR = SSR/k	MSR/MSE
Residuo o error	$n - (k + 1)$	SSE	MSE = SSE/$[n - (k + 1)]$	
Total	$n - 1$	SS total		

El término "SS", que se localiza a la mitad de la tabla ANOVA, se refiere a la suma de los cuadrados; observe que existe una suma de cuadrados en cada fuente de variación. La columna de la suma de los cuadrados muestra la cantidad de variación atribuible a cada fuente. La variación total de la variable independiente, y, está resumido en SS total. Observe que este resultado es simplemente el numerador de la fórmula usual para calcular cualquier variación; en otras palabras, la suma de las desviaciones al cuadrado de la media. Se calcula como:

$$\text{Suma de cuadrados total} = \text{SS total} = \sum(y - \overline{y})^2$$

Como se ha visto, la suma de cuadrados total incluye la de los cuadrados de la regresión y la del residuo. La suma de los cuadrados de la regresión incluye la de las diferencias al cuadrado entre los valores estimados o pronosticados, $\hat{y}$, y la media general de y. La suma de los cuadrados de la regresión se calcula así:

$$\text{Suma de los cuadrados de la regresión} = \text{SSR} = \sum(\hat{y} - \overline{y})^2$$

La suma de los cuadrados del residuo incluye las diferencias al cuadrado entre los valores observados de la variable dependiente y, y sus respectivos estimados o pronosticados, $\hat{y}$. Observe que esta diferencia es el error de estimar o predecir la variable independiente con la ecuación de regresión múltiple; el cual se calcula:

$$\text{Suma de los cuadrados del error o residuo} = \text{SSE} = \sum(y - \hat{y})^2$$

Se utilizará la información de la tabla ANOVA del ejemplo previo para evaluar la ecuación de regresión y estimar los costos de calefacción en enero.

	A	B	C	D	G	H	I	J	K	L	M
1	Costo	Temp.	Aislam.	Edad		RESUMEN DE SALIDA					
2	250	35	3	6							
3	360	29	4	10		*Estadísticos de regresión*					
4	165	36	7	3		Múltiple *R*	0.897				
5	43	60	6	9		*R* cuadrada	0.804				
6	92	65	5	6		*R* cuadrada ajustada	0.767				
7	200	30	5	5		Error estándar	51.049				
8	355	10	6	7		Observaciones	20				
9	290	7	10	10							
10	230	21	9	11		ANOVA					
11	120	55	2	5			*gl*	*SS*	*MS*	*F*	*Significancia F*
12	73	54	12	4		Regresión	3	171220.473	57073.491	21.901	0.000
13	205	48	5	1		Residual	16	41695.277	2605.955		
14	400	20	5	15		Total	19	212915.750			
15	320	39	4	7							
16	72	60	8	6			*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	272	20	5	8		Intercepción	427.194	59.601	7.168	0.000	
18	94	58	7	3		Temperatura	-4.583	0.772	-5.934	0.000	
19	190	40	8	11		Aislamiento	-14.831	4.754	-3.119	0.007	
20	235	27	9	8		Edad	6.101	4.012	1.521	0.148	

Error estándar de estimación múltiple

Comenzamos con el **error estándar de estimación múltiple.** Recuerde que el error estándar de estimación es comparable con la desviación estándar; para explicar los detalles de este, consulte la primera casa muestreada en la segunda fila de la hoja de cálculo de Excel anterior. El costo actual de calefacción de la primera observación, y, es 250 dólares, la temperatura externa, x_1, es 35 grados, el espesor del aislamiento x_2, es 3 pulgadas, y la antigüedad del calentador, x_3, es 6 años. Mediante la ecuación de regresión que se desarrolló en la sección anterior, el costo de calefacción estimado de esta casa es:

$$\begin{aligned}\hat{y} &= 427.194 - 4.583x_1 - 14.831x_2 + 6.101x_3 \\ &= 427.194 - 4.583(35) - 14.831(3) + 6.101(6) \\ &= 258.90\end{aligned}$$

Por lo tanto, se estimaría que la calefacción de una casa con una temperatura externa media en enero de 35 grados, 3 pulgadas de aislamiento y un calentador de 6 años de antigüedad costara 258.90 dólares. El costo de calefacción actual fue 250 dólares, por lo cual el residuo, que es la diferencia entre el valor actual y el estimado, es $y - \hat{y} = 250 - 258.90 = -8.90$. Esta diferencia de 8.90 dólares es el error aleatorio o no explicado del primer elemento muestreado. En el siguiente paso se eleva al cuadrado esta diferencia, es decir; se determina $(y - \hat{y})^2 = (250 - 258.90)^2 = (-8.90)^2 = 79.21$.

Si se repiten estas operaciones con las otras 19 observaciones y se suman las 20 diferencias al cuadrado, el total será la suma de los cuadrados del error o residuo de la tabla ANOVA. Utilizando esta información, el error estándar de estimación múltiple se calcula como:

ERROR ESTÁNDAR DE ESTIMACIÓN MÚLTIPLE

$$s_{Y.123...k} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - (k + 1)}} = \sqrt{\frac{SSE}{n - (k + 1)}} \quad \textbf{[14.2]}$$

donde:

y es la observación actual;
$\hat{y}$ es el valor estimado calculado mediante la ecuación de regresión;
n es el número de observaciones en la muestra;
k es el número de variables independientes;
SSR es la suma de los cuadrados del residuo de la tabla ANOVA.

Todavía se halla más información en la tabla ANOVA que se puede usar para calcular el error estándar de estimación múltiple. En la columna encabezada como MS (*mean squares* que significa *media al cuadrado*) se reportan las medias cuadradas para la regresión y la variación residual. Estos valores se calculan como la suma de los cuadrados divididos entre sus correspondientes grados de libertad. El error estándar de estimación múltiple es igual a la raíz cuadrada de la media cuadrada residual MS, lo que también se llama *error de la media al cuadrado* o MSE.

$$s_{Y.123...K} = \sqrt{MSE} = \sqrt{2\,605.995} = \$51.05$$

¿Cómo se interpreta el error estándar de estimación de 51.05? Es el "error" típico cuando se emplea esta ecuación para predecir el costo. Primero, las unidades son las mismas que en la variable dependiente, por lo cual el error estándar es en dólares ($51.05). Segundo, se espera que los residuos sean aproximados a una distribución más o menos normal, por lo que alrededor de 68% de ellos estará dentro de ±$51.05 y cerca de 95% dentro de ±2(51.05) = ±$102.10. Como ocurrió con similares medidas de dispersión, como el error estándar de estimación del capítulo 13, mediante un error estándar múltiple se indica una mejor o más eficiente ecuación de predicción.

Coeficiente de determinación múltiple

Enseguida se considera el coeficiente de determinación múltiple. Recuerde, del capítulo anterior, que el coeficiente de determinación se define como el porcentaje de la variación de la variable dependiente explicada, o contabilizada, por la variable independiente. En el caso de la regresión múltiple se amplía esta definición, como sigue.

COEFICIENTE DE DETERMINACIÓN MÚLTIPLE Es el porcentaje de variación de la variable dependiente, y, explicada por el conjunto de variables independientes, $x_1, x_2, x_3, \ldots, x_k$.

Las características del coeficiente de determinación múltiple son:

1. **Se representa por una letra *R* (mayúscula) al cuadrado.** En otras palabras, se escribe como R^2 debido a que se comporta como el cuadrado de un coeficiente de correlación.
2. **Puede variar de 0 a 1.** Un valor cercano a 0 indica poca asociación entre el conjunto de variables independientes y la variable dependiente; un valor cercano a 1 significa una asociación fuerte.
3. **No puede adoptar valores negativos.** Ningún número que se eleva al cuadrado (o a la segunda potencia) puede ser negativo.
4. **Es fácil de interpretar.** Como R^2 es un valor entre 0 y 1, es fácil de interpretar, comparar y comprender.

A partir de la información de la tabla ANOVA se puede calcular el coeficiente de determinación. Observe la columna de la suma de los cuadrados, etiquetada con SS en la salida de Excel, y utilice la suma de los cuadrados de la regresión, SSR, y divídala entre la suma total de los cuadrados, SS total.

COEFICIENTE DE DETERMINACIÓN MÚLTIPLE

$$R^2 = \frac{\text{SSR}}{\text{SS total}} \tag{14.3}$$

La regresión y la suma total de los cuadrados de la tabla ANOVA resaltada en la salida de Excel de la sección “La tabla ANOVA” se pueden utilizar para calcular el coeficiente de determinación múltiple.

$$R^2 = \frac{\text{SSR}}{\text{SS total}} = \frac{171\,220.473}{212\,915.750} = 0.804$$

¿Cómo se interpreta este valor? Las variables independientes (la temperatura externa, la cantidad de aislamiento y la antigüedad del calentador) explican, o contabilizan, 80.4% de la variación del costo de calefacción. En otras palabras, 19.6% de la variación se debe a otras fuentes, como el error aleatorio o variables no incluidas en el análisis. Mediante la tabla ANOVA se determina que 19.6% corresponde a la suma de los errores al cuadrado dividida entre la suma total de los cuadrados. Si SSR + SSE = SS total, la siguiente relación se considera válida.

$$1 - R^2 = 1 - \frac{\text{SSR}}{\text{SS total}} = \frac{\text{SSE}}{\text{SS total}} = \frac{41\,695.277}{212\,915.750} = 0.196$$

Coeficiente de determinación ajustado

El número de variables independientes de una ecuación de regresión múltiple aumenta el coeficiente de determinación. Cada nueva variable independiente hace que las predicciones sean más precisas, lo que a su vez reduce el SSE y aumenta el SSR; de aquí, R^2 aumenta solo debido al número total de variables independientes y no porque la variable independiente agregada sea un buen factor de predicción de la variable dependiente. De hecho, si el número de variables, k, y el tamaño de la muestra, n, son iguales, el coeficiente de determinación es 1.0; en la práctica, esta situación es poco frecuente y también sería éticamente cuestionable. Para equilibrar el efecto del número de variables independientes en el coeficiente de determinación múltiple, los paquetes de software estadísticos emplean un coeficiente de determinación *ajustado* múltiple.

COEFICIENTE DE DETERMINACIÓN AJUSTADO

$$R^2_{\text{ajust}} = 1 - \frac{\dfrac{\text{SSR}}{n-(k+1)}}{\dfrac{\text{SS total}}{n-1}} \tag{14.4}$$

Las sumas totales de los cuadrados y del error se dividen entre sus grados de libertad. Observe en especial que los grados de libertad para la suma de los errores al cuadrado incluyen k, el número de variables independientes. En el ejemplo del costo de calefacción, el coeficiente de determinación ajustado es:

$$R^2_{\text{ajust}} = 1 - \frac{\dfrac{41\,695.277}{20 - (3 + 1)}}{\dfrac{212\,915.750}{20 - 1}} = 1 - \frac{2\,605.955}{11\,206.092} = 1 - 0.233 = 0.767$$

Si se compara R^2 (0.80) con R^2 ajustada (0.767), la diferencia en este caso es pequeña.

AUTOEVALUACIÓN 14-2

Consulte la autoevaluación 14.1 respecto de los restaurantes en Myrtle Beach. La parte de la tabla ANOVA de la captura de pantalla de la regresión se muestra a continuación.

Análisis de varianza

Fuente	GL	SS	MS
Regresión	5	100	20
Error residual	20	40	2
Total	25	140	

(a) ¿Cuál fue el tamaño de la muestra?
(b) ¿Cuántas variables independientes hay?
(c) ¿Cuántas variables dependientes hay?
(d) Calcule el error estándar de estimación. ¿Entre qué valores estará aproximadamente 95% de los residuos?
(e) Determine el coeficiente de determinación múltiple e interprete este valor.
(f) Encuentre el coeficiente de determinación múltiple, ajustado según los grados de libertad.

EJERCICIOS

5. Considere la siguiente tabla ANOVA.

Análisis de varianza

Fuente	GL	SS	MS	F	P
Regresión	2	77.907	38.954	4.14	0.021
Error residual	62	583.693	9.414		
Total	64	661.600			

a. Determine el error estándar de estimación. ¿Entre qué valores estará aproximadamente 95% de los residuos?
b. Determine el coeficiente de determinación múltiple e interprete este valor.
c. Encuentre el coeficiente de determinación múltiple, ajustado según los grados de libertad.

6. Considere la siguiente tabla ANOVA.

Análisis de varianza

Fuente	GL	SS	MS	F
Regresión	5	3710.00	742.00	12.89
Error residual	46	2647.38	57.55	
Total	51	6357.38		

a. Determine el error estándar de estimación. ¿Entre qué valores estará aproximadamente 95% de los residuos?
b. Determine el coeficiente de determinación múltiple e interprete este valor.
c. Encuentre el coeficiente de determinación múltiple, ajustado por los grados de libertad.

Inferencias en la regresión lineal múltiple

OA14-3 Probar hipótesis acerca de las relaciones inferidas por un modelo de regresión múltiple.

Hasta este punto, el análisis de regresión múltiple se consideró solo como una forma para describir la relación entre una variable dependiente y varias variables independientes; sin embargo, el método de mínimos cuadrados también permite inferir o generalizar a partir de la relación de una población completa. Recuerde que cuando se crearon intervalos de confianza o se realizaron pruebas de hipótesis como parte de la estadística inferencial, los datos se consideraron una muestra aleatoria tomada de una población.

En el escenario de la regresión múltiple, se supone que hay una ecuación desconocida de regresión múltiple de la población que relaciona la variable dependiente con las k variables independientes. Algunas veces a esto se le denomina **modelo** de la relación; en símbolos se escribe:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Esta ecuación es análoga a la fórmula [14.1], excepto que ahora los coeficientes se denotan con letras griegas; con las cuales se expresan *parámetros poblacionales*. Así, con cierto conjunto de suposiciones, las cuales se analizarán en breve, los valores calculados de a y b_j son estadísticos muestrales; estos son estimadores puntuales de los parámetros poblacionales correspondientes α y β_j; por ejemplo, el coeficiente de regresión de la muestra b_2 es un estimador puntual del parámetro poblacional β_2. La distribución muestral de estos estimadores puntuales sigue la distribución de probabilidad normal; estas distribuciones muestrales se centran en los valores de sus parámetros respectivos. En otras palabras, las medias de las distribuciones muestrales son iguales a los valores de los parámetros que se estimarán; así, con las propiedades de las distribuciones muestrales de estos estadísticos, es posible inferir acerca de los parámetros poblacionales.

Prueba global: prueba del modelo de regresión múltiple

Es posible demostrar la capacidad de las variables independientes $X_1, X_2, \ldots, X_k$ para explicar el comportamiento de la variable dependiente Y. Para expresarlo en forma de pregunta: ¿es posible estimar la variable dependiente sin basarse en las variables independientes? A esta prueba se le denomina **prueba global**; básicamente se investiga si es posible que todas las variables independientes tengan coeficientes de regresión cero.

Para relacionar esta pregunta con el ejemplo del costo de calefacción, se comprobará si las variables independientes (la cantidad de aislamiento del ático, la temperatura externa diaria media y la antigüedad del calentador) sirven para calcular el costo de calefacción de la casa. Para probar una hipótesis, primero se formulan las hipótesis nula y alternativa. En el ejemplo del costo de calefacción hay tres variables independientes; recuerde que b_1, b_2 y b_3 son coeficientes de regresión de la muestra. A los coeficientes correspondientes en la población se les asignan los símbolos β_1, β_2 y β_3. Ahora se comprueba si todos los coeficientes de regresión en la población son cero. La hipótesis nula es:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

La hipótesis alternativa es:

$$H_1: \text{No todas las } \beta_j \text{ son 0.}$$

Si la hipótesis nula es verdadera, todos los coeficientes de regresión son cero y, por lógica, no son útiles para estimar la variable dependiente "costo de calefacción". De ser así, habría que buscar otras variables independientes, o tomar una aproximación distinta, para predecir el costo de calefacción de la casa.

Para comprobar la hipótesis nula (que todos los coeficientes de regresión múltiple son cero) se emplea la distribución F que se presentó en el capítulo 12. Use el nivel de significancia 0.05. Recuerde estas características de la distribución F:

1. **Existe una familia de distribuciones *F*.** Cada vez que los grados de libertad en el numerador o en el denominador cambian, se crea una nueva distribución F.
2. **La distribución *F* no puede ser negativa.** El menor valor posible es 0.
3. **Es una distribución continua**. La distribución puede tomar un número infinito de valores entre 0 y el infinito positivo.
4. **Es sesgada de manera positiva.** La cola de la distribución está a la derecha. Conforme el número de grados de libertad aumenta, tanto en el numerador como en el denominador, la distribución se aproxima a la distribución de probabilidad normal; es decir, la distribución se moverá hacia una distribución simétrica.
5. **Es asintótica.** Conforme aumentan los valores de X, la curva F se aproximará al eje horizontal, pero nunca lo tocará.

A continuación se recurre al estadístico F para probar la hipótesis global; como en el capítulo 12, se trata del rango de dos varianzas. En este caso, el numerador es la suma de los cuadrados de la regresión dividida entre sus grados de libertad, k. El denominador es la suma de los cuadrados del error dividida entre sus grados de libertad, $n - (k + 1)$. La fórmula es:

PRUEBA GLOBAL $$F = \frac{\text{SSR}/k}{\text{SSE}/[n - (k + 1)]} \quad [14.5]$$

Utilizando los valores de la tabla ANOVA que se encuentra en el ejemplo de Salsberry Realty, el estadístico F es:

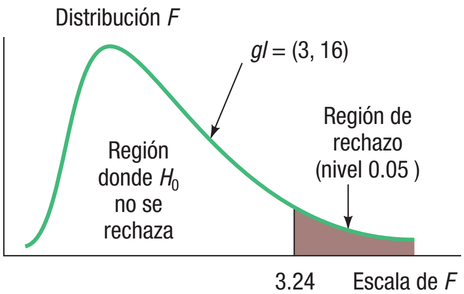
$$F = \frac{\text{SSR}/k}{\text{SSE}/[n - (k + 1)]} = \frac{171\,220.473/3}{41\,695.277/[20 - (3 + 1)]} = 21.90$$

Recuerde que el estadístico F prueba la hipótesis nula básica de que dos varianzas o, en este caso, dos medias cuadradas, son iguales. En esta prueba de hipótesis global de regresión múltiple se rechaza la hipótesis nula, H_0, de que todos los coeficientes de regresión son cero cuando la media cuadrada de la regresión es mayor en comparación con la del residuo. Si esto es cierto, el estadístico F será relativamente grande, y estará en la cola de la extrema derecha de la distribución F; el valor p será pequeño, esto es, menor que el nivel de significancia 0.05; por ello se rechaza la hipótesis nula.

Como con otros métodos de prueba de hipótesis, la regla de decisión puede basarse en cualquiera de dos procedimientos: 1) comparar el estadístico de prueba con un valor crítico o 2) calcular un valor p basado en el estadístico de prueba y comparar el valor p con el nivel de significancia. El método del valor crítico usando el estadístico F requiere tres tipos de información: 1) grados de libertad del numerador, 2) grados de libertad del denominador y 3) nivel de significancia. Los grados de libertad para el numerador y el denominador se determinan en la siguiente tabla ANOVA en Excel. La captura de pantalla de la tabla ANOVA se resalta en color verde claro. El número superior en la columna identificada *"gl"* es 3, para indicar que hay tres grados de libertad en el numerador; este valor corresponde al número de variables independientes. El número a la mitad de la columna "*gl*" (16) indica que hay 16 grados de libertad en el denominador. El número 16 se determina por medio de $(n - (k + 1)) = 20 - (3 + 1) = 16$.

	A	B	C	D	G	H	I	J	K	L	M
1	Costo	Temp.	Aislam.	Edad		RESUMEN DE SALIDA					
2	250	35	3	6							
3	360	29	4	10		*Estadísticos de regresión*					
4	165	36	7	3		Múltiple *R*	0.897				
5	43	60	6	9		*R* cuadrada	0.804				
6	92	65	5	6		*R* cuadrada ajustada	0.767				
7	200	30	5	5		Error estándar	51.049				
8	355	10	6	7		Observaciones	20				
9	290	7	10	10							
10	230	21	9	11		ANOVA					
11	120	55	2	5			*gl*	*SS*	*MS*	*F*	*Significancia F*
12	73	54	12	4		Regresión	3	171220.473	57073.491	21.901	0.000
13	205	48	5	1		Residual	16	41695.277	2605.955		
14	400	20	5	15		Total	19	212915.750			
15	320	39	4	7							
16	72	60	8	6			*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	272	20	5	8		Intercepción	427.194	59.601	7.168	0.000	
18	94	58	7	3		Temperatura	-4.583	0.772	-5.934	0.000	
19	190	40	8	11		Aislamiento	-14.831	4.754	-3.119	0.007	
20	235	27	9	8		Edad	6.101	4.012	1.521	0.148	

El valor crítico de F se encuentra en el apéndice B.6; utilice la tabla con el nivel de significancia 0.05, al moverse por renglones a 3 grados de libertad en el numerador, y luego hacia abajo a 16 grados de libertad en el denominador se lee el valor crítico 3.24. Las regiones de rechazo y aceptación de H_0 se muestran en el diagrama de la derecha.



Al aplicar la prueba global, la regla de decisión es no rechazar la hipótesis nula (que todos los coeficientes de regresión son 0 si el valor calculado de F es menor que o igual que 3.24). Si el valor calculado de F es mayor que 3.24, se rechaza H_0 y se acepta la hipótesis alternativa, H_1.

El valor calculado de F es 21.90, que se encuentra en la región de rechazo; por lo tanto, se descarta la hipótesis nula; esto significa que algunas variables independientes (cantidad de aislamiento, etc.) tienen la capacidad de explicar la variación de la variable dependiente "costo de calefacción". Se esperaba esta decisión, y es lógico que la temperatura externa, la cantidad de aislamiento y la antigüedad del calentador tengan gran peso sobre el costo de calefacción porque la prueba global lo demuestra.

La prueba de la hipótesis nula también puede basarse en un valor p, que se reporta en la salida de software de todas las pruebas de hipótesis. En el caso del estadístico F, el valor p se define como la probabilidad de observar un valor F del mismo tamaño o más grande que el estadístico de prueba F, asumiendo que la hipótesis nula es verdadera. Si el valor p es menor que el nivel de significancia elegido, se decide rechazar la hipótesis nula. En la tabla ANOVA se muestra que el valor p del estadístico F es igual a 0.000, el cual es claramente menor que el nivel de significancia 0.05; por lo tanto, se decide rechazar la hipótesis global nula y se concluye que cuando menos uno de los coeficientes de regresión no es igual a cero.

Evaluación de los coeficientes de regresión individuales

Hasta este punto, al menos uno (no necesariamente todos) de los coeficientes de regresión son distintos de cero, y por ende son útiles para realizar predicciones. En el siguiente paso se prueban las variables independientes de manera *individual* para determinar cuáles coeficientes de regresión pueden ser cero y cuáles no.

¿Por qué es importante saber si algunas de las β_j son iguales a cero? Si una β puede ser igual a cero, esta variable independiente en particular no tiene valor para explicar alguna variación del valor dependiente. Si hay coeficientes con respecto a los cuales H_0 no se puede rechazar, quizá sea prudente eliminarlos de la ecuación de regresión.

Ahora se realizan tres pruebas de hipótesis separadas: para la temperatura, el aislamiento y la antigüedad del calentador.

Para la temperatura:	Para el aislamiento:	Para la antigüedad del calentador:
H_0: $\beta_1 = 0$	H_0: $\beta_2 = 0$	H_0: $\beta_3 = 0$
H_1: $\beta_1 \neq 0$	H_1: $\beta_2 \neq 0$	H_1: $\beta_3 \neq 0$

Se prueba la hipótesis con el nivel de significancia 0.05 (observe que estas son pruebas de dos colas).

El estadístico de prueba sigue la distribución t de Student con $n - (k + 1)$ grados de libertad. El número de observaciones muestrales es n y hay 20 casas en el estudio, por lo cual $n = 20$. El número de variables independientes es k, el cual es 3; así, hay $n - (k + 1) = 20 - (3 + 1) = 16$ grados de libertad.

El valor crítico de t está en el apéndice B.5. En el caso de una prueba de dos colas con 16 grados de libertad y el nivel de significancia 0.05, H_0 se rechaza si t es menor que -2.120 o mayor que 2.120.

Consulte la captura de pantalla de Excel de la sección anterior. La columna que se resalta en color naranja, con encabezado "Coeficientes", muestra los valores de la ecuación de regresión múltiple:

$$\hat{y} = 427.194 - 4.583x_1 - 14.831x_2 + 6.101x_3$$

Al interpretar el término -4.583x_1 en la ecuación, por cada grado de aumento de temperatura, se espera que el costo de calefacción disminuya aproximadamente 4.58 dólares, si las otras dos variables permanecen constantes.

La columna en la salida de Excel identificada como "Error estándar" indica el error estándar del coeficiente de regresión de la muestra. Recuerde que Salsberry Realty seleccionó una muestra de 20 casas a lo largo de la costa este de Estados Unidos. Si la empresa seleccionara una segunda muestra aleatoria y calculara los coeficientes de regresión de esa muestra, los valores no serían exactamente los mismos; sin embargo, si se repitiera el proceso de muestreo muchas veces se podría diseñar una distribución de muestreo de estos coeficientes de regresión. La columna "Error estándar" estima la variabilidad de estos coeficientes de regresión. La distribución de muestreo de los coeficientes sigue la distribución t con $n - (k + 1)$ grados de libertad. De aquí, se pueden probar las variables independientes individualmente para determinar si los coeficientes de regresión difieren de cero. La fórmula es:

PRUEBA DE LOS COEFICIENTES DE REGRESIÓN INDIVIDUALES

$$t = \frac{b_i - 0}{s_{b_i}} \qquad [14.6]$$

El coeficiente b_i se refiere a cualquiera de los coeficientes de regresión, y s_{b_1}, a la desviación estándar de esa distribución del coeficiente de regresión. Se incluye un cero en la ecuación debido a que la hipótesis nula es $\beta_i = 0$.

Mediante la prueba del coeficiente de regresión para la variable independiente "temperatura" se ilustra esta fórmula. Según la salida de la sección "Prueba global", el coeficiente de regresión para la temperatura es -4.583. La desviación estándar de la distribución muestral del coeficiente de regresión de la variable independiente "temperatura" es 0.772. Al sustituir estos valores en la fórmula [14.6]:

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{-4.583 - 0}{0.772} = -5.937$$

El valor t calculado es -5.937 de temperatura (la pequeña diferencia entre el valor calculado y el que se muestra en la salida de Excel se debe al redondeo) y -3.119 del aislamiento; ambos valores t están en la región de rechazo, a la izquierda de -2.120; por ello, se concluye que los coeficientes de regresión de las variables "temperatura" y "aislamiento" *no* son cero. El valor t calculado en el caso de la edad del calentador es 1.521, así que se concluye que podría igualar a 0; la variable independiente "edad del calentador" no es un factor de predicción significativo del costo de la calefacción. Los resultados de estas pruebas de hipótesis indican que el análisis debe enfocarse en la temperatura y el aislamiento como predictores del costo de calefacción.

También es posible utilizar valores p para probar los coeficientes de regresión individual. De nuevo, estos suelen reportarse en una salida de computadora. El valor t calculado de temperatura en la salida de Excel es -5.934 y tiene un valor p de 0.000; como este es menor a 0.05, el coeficiente de regresión de la variable independiente "temperatura" no es igual a cero, y se debe incluir en la ecuación para pronosticar los costos de calefacción. En el caso del aislamiento, el valor t es -3.119 y tiene un valor p de 0.007; como en el caso de la temperatura, este valor es menor a 0.05, así que se concluye que el coeficiente de regresión del aislamiento no es igual a cero y se debe incluir en la ecuación para pronosticar el costo de calefacción. En contraste con ambas variables, el valor p para probar el coeficiente de regresión de la "antigüedad del calefactor" podría igualar a cero; además, como variable independiente no es un factor de predicción significativo del costo de calefacción. De esta forma, la antigüedad del calefactor no se debe incluir en la ecuación para pronosticar los costos de calefacción.

En este punto, es necesario elaborar una estrategia para eliminar variables independientes. En el caso de Salsberry Realty había tres; para la antigüedad del calentador no se puede rechazar la hipótesis nula de que el coeficiente de regresión era cero. Es claro que se debe omitir esa variable y volver a efectuar la ecuación de regresión. En la siguiente pantalla de Excel se muestra el costo de calefacción como la variable dependiente, y la temperatura externa y la cantidad de aislamiento como las variables independientes.

	A	B	C	D	E	F	G	H	I	J	K
1	Costo	Temp.	Aislam.	Edad		RESUMEN DE SALIDA					
2	250	35	3	6							
3	360	29	4	10		*Estadísticos de regresión*					
4	165	36	7	3		Múltiple R	0.881				
5	43	60	6	9		R cuadrada	0.776				
6	92	65	5	6		R cuadrada ajustada	0.749				
7	200	30	5	5		Error estándar	52.982				
8	355	10	6	7		Observaciones	20				
9	290	7	10	10							
10	230	21	9	11		ANOVA					
11	120	55	2	5			*gl*	*SS*	*MS*	*F*	*Significancia F*
12	73	54	12	4		Regresión	2	165194.521	82597.261	29.424	0.000
13	205	48	5	1		Residual	17	47721.229	2807.131		
14	400	20	5	15		Total	19	212915.750			
15	320	39	4	7							
16	72	60	8	6			*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	272	20	5	8		Intercepción	490.286	44.410	11.040	0.000	
18	94	58	7	3		Temperatura	-5.150	0.702	-7.337	0.000	
19	190	40	8	11		Aislamiento	-14.718	4.934	-2.983	0.008	
20	235	27	9	8							

A continuación se resumen los resultados de esta nueva salida:

1. La nueva ecuación de regresión es:

$$\hat{y} = 490.286 - 5.150x_1 - 14.718x_2$$

Observe que los coeficientes de regresión de la temperatura externa (x_1) y la cantidad de aislamiento (x_2) son similares, pero no iguales, cuando se incluyó la variable independiente "antigüedad del calentador". Compare la ecuación anterior con la de la captura de pantalla de Excel de la sección "Prueba global". Los dos coeficientes de regresión son negativos, como en la ecuación previa.

2. Los detalles de la prueba global se muestran a continuación:

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_1: \text{No todas las } \beta_i = 0$$

La distribución F es el estadístico de prueba, y hay $k = 2$ grados de libertad en el numerador y $n - (k + 1) = 20 - (2 + 1) = 17$ grados de libertad en el denominador. Con el nivel de significancia 0.05 y el apéndice B.6, la regla de decisión es rechazar H_0 si F es mayor que 3.59; el valor de F se calcula así:

$$F = \frac{\text{SSR}/k}{\text{SSE}/[n - (k + 1)]} = \frac{165\,194.521/2}{47\,721.229/[20 - (2 + 1)]} = 29.424$$

Como el valor calculado de F (29.42) es mayor que el valor crítico (3.59), se rechaza la hipótesis nula y se acepta la hipótesis alternativa; en conclusión, al menos uno de los coeficientes de regresión es diferente de cero.

Utilizando el valor p, la prueba del estadístico F (29.424) tiene un valor p (0.000) que es claramente menor a 0.05, por lo tanto, se rechaza la hipótesis nula y se acepta la alternativa; en conclusión, cuando menos uno de los coeficientes de regresión es distinto a cero.

3. En el siguiente paso se realiza la prueba de los coeficientes de regresión de manera individual. Se desea saber si uno o ambos coeficientes de regresión son diferentes de cero. Las hipótesis nula y alternativa de cada una de las variables independientes son:

Temperatura externa	Aislamiento
$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
$H_1: \beta_1 \neq 0$	$H_1: \beta_2 \neq 0$

El estadístico de prueba es la distribución t con $n - (k + 1) = 20 - (2 + 1) = 17$ grados de libertad. Con el nivel de significancia 0.05 y el apéndice B.5, la regla de decisión es rechazar H_0 si el valor calculado de t es menor que -2.110 o mayor que 2.110.

Temperatura externa	Aislamiento
$t = \dfrac{b_1 - 0}{s_{b_1}} = \dfrac{-5.150 - 0}{0.702} = -7.337$	$t = \dfrac{b_2 - 0}{s_{b_2}} = \dfrac{-14.718 - 0}{4.934} = -2.983$

En ambas pruebas se rechaza H_0 y se acepta H_1; en conclusión, cada uno de los coeficientes de regresión es diferente de cero. Tanto la temperatura externa como la cantidad de aislamiento son variables útiles para explicar la variación del costo de calefacción.

Utilizando los valores p, el del estadístico t temperatura es 0.000, y el del aislamiento es 0.008; ambos valores p son menores a 0.05, así que en las dos pruebas se rechaza la hipótesis nula y se concluye que cada uno de los coeficientes de regresión es diferente de cero. Tanto la temperatura externa como la cantidad de aislamiento son variables útiles para explicar la variación del costo de calefacción.

En el ejemplo del costo de calefacción fue claro qué variable independiente se debía eliminar; en algunos casos esto no es tan claro. Para explicarlo, suponga que se formula una ecuación de regresión múltiple con base en cinco variables independientes. Se realiza la prueba global y se determina que algunos de los coeficientes de regresión son diferentes de cero; luego, se prueban los coeficientes de regresión de manera individual y se determina que tres son significativos y dos no lo son. El procedimiento preferido es omitir la variable dependiente individual con el *menor valor t absoluto* o *valor p mayor* y volver a formular la ecuación de regresión con las cuatro restantes; después, en la nueva ecuación de regresión con cuatro variables independientes, se realizan las pruebas individuales. Si aún hay coeficientes de regresión que no son significativos, de nuevo se omite la variable con el menor valor t absoluto. Para describir el proceso de otra manera, se debe eliminar una variable a la vez; cada vez que se elimina una, es necesario volver a formular la ecuación de regresión y verificar las restantes.

Este proceso de seleccionar variables para incluirlas en un modelo de regresión se automatiza con Excel, Minitab, MegaStat u otro software estadístico. La mayoría de los sistemas de software incluye métodos para eliminar en secuencia o agregar variables independientes y, al mismo tiempo, proporcionar estimaciones del porcentaje de la variación explicada (el término R cuadrada). Dos de los métodos más comunes son la **regresión por pasos** y la **regresión del mejor subconjunto**. Toma mucho tiempo, pero es posible calcular cada una de las regresiones entre la variable dependiente y cualquier subconjunto posible de variables independientes.

Desafortunadamente, en ocasiones el software puede trabajar "demasiado" para encontrar una ecuación que cumpla con las singularidades de su conjunto de datos particular. La ecuación sugerida quizá no represente la relación en la población; por lo tanto, es necesario discernir para elegir entre las ecuaciones presentadas. Considere si los resultados son lógicos, si tienen una interpretación simple y si son consistentes con su conocimiento de la aplicación en estudio.

AUTOEVALUACIÓN 14-3

La salida de regresión respecto de restaurantes en Myrtle Beach se repite a continuación (vea las autoevaluaciones anteriores).

Factor de predicción	Coeficiente	Error estándar del coeficiente	t	Valor p	
Constante	2.50	1.50	1.667	0.111	
x_1	3.00	1.50	2.000	0.056	
x_2	4.00	3.00	1.333	0.194	
x_3	−3.00	0.20	−15.000	0.000	
x_4	0.20	0.05	4.000	0.000	
x_5	1.00	1.50	0.667	0.511	
Análisis de la varianza					
Fuente	GL	SS	MS	F	Valor p
Regresión	5	100	20	10	0.000
Error residual	20	40	2		
Total	25	140			

(a) Realice una prueba de hipótesis global para verificar si algunos de los coeficientes de regresión son diferentes de cero. ¿Cuál es su decisión? Utilice el nivel de significancia 0.05.
(b) Haga una prueba individual de cada una de las variables independientes. ¿Qué variables consideraría eliminar? Utilice el nivel de significancia 0.05.
(c) Formule un plan para eliminar variables independientes.

EJERCICIOS

7. Con esta salida de regresión:

Factor de predicción	Coeficiente	Error estándar del coeficiente	t	Valor p	
Constante	84.998	1.863	45.62	0.000	
x_1	2.391	1.200	1.99	0.051	
x_2	−0.409	0.172	−2.38	0.021	
Análisis de la varianza					
Fuente	GL	SS	MS	F	Valor p
Regresión	2	77.907	38.954	4.138	0.021
Error residual	62	583.693	9.414		
Total	64	661.600			

Realice lo siguiente:

a. Elabore la ecuación de regresión.
b. Si x_1 es 4 y x_2 es 11, ¿cuál es el valor de la variable dependiente?
c. ¿Cuál es el tamaño de la muestra? ¿Cuántas variables independientes hay?
d. Realice una prueba de hipótesis global para verificar si alguno de los coeficientes de regresión del conjunto es diferente de cero. Utilice el nivel de significancia 0.05. ¿Cuál es su conclusión?
e. Realice una prueba de hipótesis por cada variable independiente. Utilice el nivel de significancia 0.05. ¿Qué variables consideraría eliminar?
f. Formule una estrategia para eliminar variables independientes en este caso.

8. La siguiente salida de regresión se obtuvo de un estudio de empresas de arquitectura. La variable dependiente es la cantidad total de honorarios, en millones de dólares.

Factor de predicción	Coeficiente	Error estándar del coeficiente	t	Valor p	
Constante	7.987	2.967	2.690	0.010	
x_1	0.122	0.031	3.920	0.000	
x_2	−1.220	0.053	−2.270	0.028	
x_3	−0.063	0.039	−1.610	0.114	
x_4	0.523	0.142	3.690	0.001	
x_5	−0.065	0.040	−1.620	0.112	
Análisis de la varianza					
Fuente	GL	SS	MS	F	Valor p
Regresión	5	3710.00	742	12.89	0.000
Error residual	46	2647.38	57.55		
Total	51	6357.38			

x_1 es el número de arquitectos que trabajan en la compañía;
x_2 es el número de ingenieros que trabajan en la compañía;
x_3 es el número de años invertidos en proyectos de cuidado de la salud;
x_4 es el número de estados en los que opera la empresa;
x_5 es el porcentaje del trabajo de la empresa que se relaciona con el cuidado de la salud.

a. Elabore la ecuación de regresión.
b. ¿Cuál es el tamaño de la muestra? ¿Cuántas variables independientes hay?
c. Realice una prueba de hipótesis global para ver si alguno de los coeficientes de regresión del conjunto puede ser diferente de cero. Utilice el nivel de significancia 0.05. ¿Cuál es su conclusión?
d. Realice una prueba de hipótesis por cada variable independiente. Utilice el nivel de significancia 0.05. ¿Qué variables consideraría eliminar?
e. Formule una estrategia para eliminar variables independientes en este caso.

OA14-4
Evaluar los supuestos de la regresión múltiple.

Evaluación de las suposiciones de la regresión múltiple

En la sección anterior se describieron métodos para evaluar, de manera estadística, la ecuación de regresión múltiple. Los resultados de la prueba permitieron saber si al menos uno de los coeficientes no era igual a cero y se describió un proceso de evaluación de cada coeficiente de regresión. También se analizó el proceso de toma de decisiones para incluir y excluir variables independientes en la ecuación de regresión múltiple.

Es importante saber que la validez de las pruebas estadísticas global e individual parte de varias suposiciones; es decir, si estas no son válidas, los resultados pueden estar sesgados o ser confusos, sin embargo, se debe mencionar que en la práctica no siempre es posible apegarse de manera estricta a las suposiciones siguientes. Por fortuna, las técnicas estadísticas que se analizan en este capítulo parecen funcionar muy bien aunque se viole una suposición o más de una. Incluso si los valores de la ecuación de regresión múltiple tienen cierta "desviación", las estimaciones que proporciona estarán más cerca que cualquiera que se pudiera hacer de otra manera; en general, los procedimientos estadísticos son lo bastante robustos para superar las violaciones de algunas suposiciones.

En el capítulo 13 se enumeraron las suposiciones necesarias para la regresión cuando se consideró solo una variable independiente (vea la sección "Estimaciones de intervalo de predicción"). Las suposiciones de la regresión múltiple son similares.

1. **Existe una relación lineal.** Es decir, existe una relación directa entre la variable dependiente y el conjunto de variables independientes.
2. **La variación entre los residuos es la misma tanto en valores grandes como pequeños de $\hat{y}$.** En otras palabras, $(y - \hat{y})$ no está relacionada, ya sea que $\hat{y}$ sea grande o pequeña.
3. **Los residuos siguen la distribución de probabilidad normal.** Recuerde que el residuo es la diferencia entre el valor actual de y y el valor estimado $\hat{y}$; por lo tanto, el término $(y - \hat{y})$ se

calcula para cada observación del conjunto de datos. Estos residuos deberán seguir de manera aproximada una distribución de probabilidad normal. Además, la media de los residuos debe ser cero.

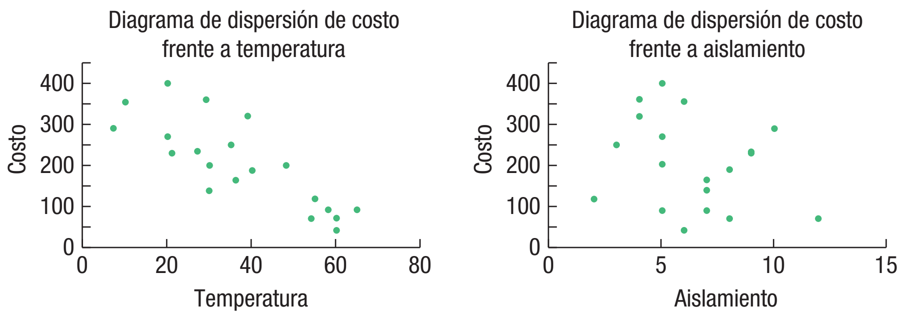
4. **Las variables independientes no deben estar correlacionadas.** Es decir, conviene seleccionar un conjunto de variables independientes que no se relacionen entre sí.
5. **Los residuos son independientes.** Esto significa que las observaciones sucesivas de la variable dependiente no están correlacionadas; esta suposición con frecuencia se viola cuando se comprende el tiempo con los datos muestreados.

En esta sección se presenta un análisis breve de cada una de estas suposiciones; además, se proporcionan métodos para validarlas, y se señalan las consecuencias si estas no se cumplen. Para quienes estén interesados en un análisis adicional, una referencia excelente es Kutner, Nachtscheim y Neter, *Applied Linear Regression Models*, 5a. ed., McGraw-Hill, 2005.

Relación lineal

Comience con la suposición de linealidad; la idea es que la relación entre el conjunto de variables independientes y la variable dependiente es lineal. Si se consideran dos variables independientes, se visualiza esta suposición. Las dos variables independientes y la variable dependiente formarían un espacio tridimensional, por ello, la ecuación de regresión formaría un plano, como se muestra en la gráfica 14.1. Esta suposición se puede evaluar con diagramas de dispersión y gráficas de residuos.

Uso de los diagramas de dispersión Al evaluar una ecuación de regresión múltiple siempre se debe incluir un diagrama de dispersión en el que se trace la variable dependiente contra cada variable independiente. Estos diagramas ayudan a visualizar las relaciones y proporcionan una información inicial respecto de la dirección (positiva o negativa), la linealidad y la fuerza de la relación. Como ejemplo, se analizan a continuación los diagramas de dispersión del caso del costo de calefacción. Las gráficas sugieren una relación muy fuerte, negativa y lineal entre el costo de calefacción y la temperatura, y una relación negativa entre dicho costo y el aislamiento.

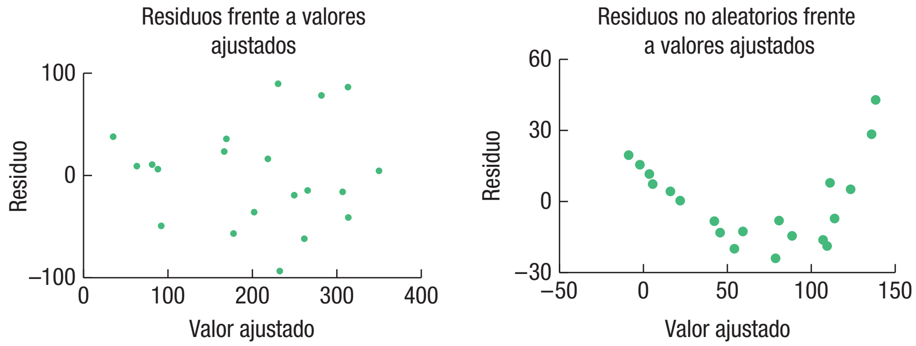


Uso de gráficas de residuos Recuerde que un residuo $(y - \hat{y})$ se calcula mediante la ecuación de regresión múltiple de cada observación en un conjunto de datos. En el capítulo 13 se explicó la idea de que la mejor recta de regresión pasaba por el centro de los datos de un diagrama de dispersión. En este caso, se registra un número grande de observaciones tanto arriba de la recta de regresión (estos residuos tendrían un signo positivo) como debajo de ella (estos residuos tendrían un signo negativo); además, estas estarían dispersas arriba y debajo de la recta, sobre todo el rango de la variable independiente.

El mismo concepto es válido en el caso de la regresión múltiple, pero esta no se puede representar de manera esquemática; sin embargo, las gráficas de los residuos ayudan a evaluar la linealidad de la ecuación de regresión múltiple. Para investigar este tema, los residuos se trazan en el eje vertical frente a la variable del factor de predicción, $\hat{y}$. En la gráfica de la página siguiente a la izquierda se muestran los trazos residuales del ejemplo del costo de calefacción. Observe lo siguiente:

- Los residuos se trazan en el eje vertical y están centrados con respecto a cero. Hay residuos positivos y negativos.
- Los trazos de los residuos muestran una distribución aleatoria de valores positivos y negativos a lo largo de todo el rango de la variable trazada en el eje horizontal.

- Los puntos están dispersos y no hay un patrón obvio, por lo que no hay razón para dudar de la suposición de linealidad; la cual se confirma mediante el siguiente diagrama.



Si hay un patrón en los puntos del diagrama de dispersión, se requiere investigar más; los puntos que se encuentran en la gráfica anterior derecha muestran residuos no aleatorios. Observe que en la gráfica de los residuos *no* se muestra una distribución aleatoria de valores positivos y negativos a lo largo de todo el rango de la variable trazada en el eje horizontal. De hecho, en la gráfica se presenta una curvatura respecto de las gráficas de los residuos; esto indica que la relación quizá no sea lineal. En este caso, tal vez la ecuación sea cuadrática, lo que indica que se necesita el cuadrado de una las variables. Esta probabilidad se analizó en el capítulo 13.

La variación de los residuos es igual en el caso de valores grandes y pequeños de $\hat{y}$

Este requisito indica que la variación respecto de los valores de predicción es constante, sin importar si estos son grandes o pequeños. Para citar un ejemplo específico que puede violar la suposición, suponga que se utiliza la variable independiente individual "antigüedad" para explicar la variación del ingreso. Se sospecha que conforme aumenta la antigüedad también aumenta el salario, pero también parece razonable que a medida que aumenta la antigüedad tal vez haya más variación respecto de la recta de regresión; es decir, es probable que haya más variación del ingreso de una persona de 50 años de edad que de una de 35. El requisito de una variación constante respecto de la recta de regresión se denomina **homoscedasticidad**.

HOMOSCEDASTICIDAD La variación respecto de la ecuación de regresión es igual para todos los valores de las variables independientes.

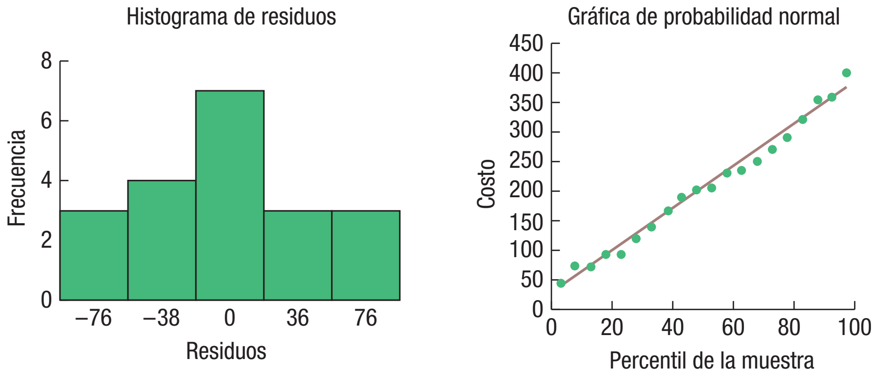
Para verificar la homoscedasticidad, los residuos se trazan contra los valores ajustados de *y*. Esta es la misma gráfica con la cual se evalúa la suposición de linealidad. Con base en el diagrama de dispersión que se muestra en esa salida del software, es razonable concluir que esta suposición no se ha violado.

Distribución de los residuos

Para tener la seguridad de que las inferencias de las pruebas de hipótesis global e individual son válidas, se evalúa la distribución de los residuos. En un caso ideal, los residuos deberán seguir una distribución de probabilidad normal.

Para evaluar esta suposición, los residuos se organizan en una distribución de frecuencias. A continuación se muestra el histograma de los residuos del lado izquierdo para el ejemplo del costo de calefacción de una casa. Aunque es difícil demostrar que los residuos siguen una distribución normal solo con 20 observaciones, parece que la suposición de normalidad es razonable.

En Excel se ofrece otra gráfica que ayuda a evaluar la suposición de residuos con una distribución normal; esta se denomina *Gráfica de probabilidad normal*, y está a la derecha del histograma; dicha gráfica confirma la suposición de residuos normalmente distribuidos si los puntos trazados están muy cerca de la recta que se delineó desde la izquierda inferior hasta la derecha superior; mediante esta se sustenta el supuesto de residuos normalmente distribuidos.



En este caso, mediante ambas gráficas se confirma el supuesto de que los residuos siguen la distribución de probabilidad normal; por lo tanto, las inferencias que se hicieron con base en las hipótesis global e individual se confirman con los resultados de esta evaluación.

Multicolinealidad

La multicolinealidad existe cuando las variables independientes están correlacionadas; lo cual dificulta las inferencias acerca de los coeficientes de regresión individuales y sus efectos individuales sobre la variable dependiente. En la práctica, es casi imposible seleccionar variables que carezcan por completo de alguna relación; en otras palabras, es casi imposible crear un conjunto de variables independientes que no estén correlacionadas hasta cierto punto, sin embargo, comprender de manera general el punto de multicolinealidad es importante.

Primero, la multicolinealidad no afecta la capacidad de una ecuación de regresión múltiple para predecir la variable dependiente; no obstante, cuando se tenga interés en evaluar la relación entre cada variable independiente y la variable dependiente, la multicolinealidad puede presentar resultados inesperados.

Por ejemplo, si se usan dos promedios de calificaciones de preparatoria con multicolinealidad muy alta y la clasificación de un grupo de preparatoria para predecir el promedio de calificaciones de los alumnos de ingreso a la universidad (variable dependiente), se esperaría que las dos variables independientes estén positivamente relacionadas con la variable dependiente; sin embargo, como las variables independientes están muy correlacionadas, una de las variables independientes puede tener un signo negativo inesperado e inexplicable. En esencia, ambas variables independientes son redundantes cuando se trata de explicar el mismo efecto en la variable dependiente.

Una segunda razón para evitar variables independientes correlacionadas es que pueden generar resultados erróneos en las pruebas de hipótesis de las variables independientes individuales; esto se debe a la inestabilidad del error estándar de estimación. Varias pistas mediante las cuales se hallan problemas de multicolinealidad incluyen lo siguiente:

1. Una variable independiente conocida por ser un factor de predicción importante resulta con un coeficiente de regresión que no es significativo.
2. Un coeficiente de regresión que debiera tener un signo positivo resulta negativo, o lo contrario.
3. Cuando se agrega o elimina una variable independiente, hay un cambio drástico de los valores de los coeficientes de regresión restantes.

En la evaluación de una ecuación de regresión múltiple, una aproximación para reducir los efectos de la multicolinealidad es seleccionar con cuidado las variables independientes incluidas en la ecuación de regresión. Una regla general es que, si la correlación entre dos variables independientes está entre -0.70 y 0.70, es probable que no haya problema al emplear las dos variables independientes. Una prueba más precisa es utilizar el **factor de inflación de la varianza**, el cual por lo general se escribe *VIF*. El valor de *VIF* se determina como sigue:

FACTOR DE INFLACIÓN DE LA VARIANZA

$$VIF = \frac{1}{1 - R_j^2} \qquad \textbf{[14.7]}$$

El término R_j^2 se refiere al coeficiente de determinación, donde la *variable independiente* seleccionada sirve como variable dependiente, y las variables independientes restantes, como variables independientes. Un *VIF* mayor que 10 se considera insatisfactorio, e indica que la variable independiente se debe eliminar del análisis. En el siguiente ejemplo se explican los detalles de la determinación del *VIF*.

EJEMPLO

Consulte los datos que se encuentran en la tabla 14.1, donde se relaciona el costo de calefacción con las variables independientes "temperatura externa media", "aislamiento del ático" y "antigüedad del calentador". Elabore una matriz de correlación de las tres variables independientes. ¿Parece que hay un problema de multicolinealidad? Encuentre e interprete el factor de inflación de la varianza de cada una de las variables independientes.

SOLUCIÓN

Comience por determinar la matriz de correlación de la variable dependiente y las cuatro variables independientes. Una parte del resultado se muestra a continuación:

	Costo	*Temperatura*	*Aislamiento*	*Antigüedad*
Costo	1.000			
Temperatura	−0.812	1.000		
Aislamiento	−0.257	−0.103	1.000	
Antigüedad	0.537	−0.486	0.064	1.000

El área resaltada indica la correlación entre las variables independientes; ninguna sobrepasa −0.70 ni 0.70, por lo que no se sospechan problemas de multicolinealidad. La correlación mayor entre las variables independientes es −0.486 entre antigüedad y temperatura.

Para confirmar esta conclusión calcule el *VIF* de cada variable independiente. Primero considere "temperatura"; se emplea Excel para determinar el coeficiente de determinación múltiple con "temperatura" como *variable dependiente,* y "aislamiento" y "antigüedad del calentador" como variables independientes. La salida relevante en pantalla de Excel se muestra a continuación.

RESUMEN					
Estadísticas de regresión					
Múltiple *R*	0.491				
R cuadrada	0.241				
R cuadrada ajustada	0.152				
Error estándar	16.031				
Observaciones	20				
ANOVA					
	GL	*SS*	*MS*	*F*	*Significancia F*
Regresión	2	1 390.291	695.145	2.705	0.096
Residuo	17	4 368.909	256.995		
Total	19	5 759.200			

El coeficiente de determinación es 0.241, por lo que al sustituir este valor en la fórmula del *VIF*:

$$VIF = \frac{1}{1 - R_1^2} = \frac{1}{1 - 0.241} = 1.32$$

El valor del *VIF* de 1.32 es menor que el límite superior de 10, lo que indica que "temperatura" no está muy correlacionada con las demás variables independientes.

Una vez más, para determinar el *VIF* del aislamiento se debe desarrollar una ecuación de regresión con "aislamiento" como *variable dependiente,* y "temperatura" y "antigüedad del calentador" como *variables independientes*. Para esta ecuación, R^2 es 0.011 y, usando la fórmula [14.7], el *VIF* para el asilamiento sería 1.011. Para encontrar el *VIF* para la antigüedad, se desarrolla una ecuación de regresión con "antigüedad" como la variable dependiente, y "temperatura" y "aislamiento" como variables independientes. Para esta ecuación, R^2 es 0.236 y, con la fórmula [14.7], el *VIF* para la antigüedad sería 1.310. Todos los valores *VIF* son menores a 10; por lo tanto, se concluye que no existe un problema de multicolinealidad en este ejemplo.

Observaciones independientes

La quinta suposición respecto del análisis de regresión y correlación es que los residuos sucesivos deberán ser independientes. Esto significa que los residuos no tienen un patrón, que no están muy correlacionados, y que no hay corridas largas de residuos positivos o negativos. A la condición que se presenta cuando los residuos sucesivos están correlacionados se le conoce como **autocorrelación.**

La autocorrelación se presenta con frecuencia cuando los datos se colectan durante un periodo; por ejemplo, se desea predecir las ventas anuales de Ages Software, Inc., con base en el tiempo y la cantidad gastada en publicidad. La variable dependiente es "ventas anuales", y las variables independientes son "tiempo" y "cantidad gastada en publicidad". Es probable que, en un periodo, los puntos actuales estén arriba del plano de regresión (recuerde que hay dos variables independientes), y después, en otro periodo, los puntos estén debajo de este. En la gráfica de la derecha se muestran los residuos graficados en el eje vertical, y los valores ajustados $\hat{y}$, en el horizontal. Observe la corrida de residuos arriba de la media de los residuos, seguida por una corrida debajo de esta. En este diagrama de dispersión se indica una posible autocorrelación.

Existe una prueba para la autocorrelación, denominada Durbin-Watson, que se discute en el capítulo 18.

Variables independientes cualitativas

OA14-5 Utilizar e interpretar una variable ficticia o cualitativa en la regresión múltiple.

En el ejemplo anterior respecto del costo de calefacción, se determinó que las dos variables independientes "temperatura externa" y "aislamiento" fueron cuantitativas; es decir, de naturaleza numérica. Con frecuencia, en el análisis se desea emplear variables de escala nominal; por ejemplo, género, si la casa tiene alberca, o si el equipo fue local o visitante. Estas se denominan **variables cualitativas,** debido a que describen una cualidad particular, como masculino o femenino. Para utilizar una variable cualitativa en el análisis de regresión, se emplea un esquema de **variables ficticias,** en el cual una de las dos condiciones posibles se codifica con un 0 o un 1.

VARIABLE FICTICIA Variable en la que solo existen dos resultados posibles. Para el análisis, uno de los resultados se codifica con un 1, y el otro, con un 0.

Por ejemplo, se quiere estimar el salario de un ejecutivo con base en los años de su experiencia laboral y si él o ella se graduó o no de la universidad. "Graduación de la universidad" solo puede adoptar una de dos condiciones: sí o no; por lo tanto, se considera una variable cualitativa.

Suponga que en el ejemplo de Salsberry Realty se agrega la variable independiente "garaje". Para las casas sin garaje, se utiliza 0; para las que sí tienen se emplea 1. A "garaje" se le designará x_4. Los datos que se encuentran en la tabla 14.2 se ingresan en Excel. Recuerde que la variable "antigüedad del calentador" no se incluye en el análisis porque se determinó que no estaba relacionada significativamente con el costo de calefacción.

TABLA 14.2 Costo de calefacción de las casas, temperatura, aislamiento y garaje de una muestra de 20 casas

Costo, y	Temperatura, x_1	Aislamiento, x_2	Garage, x_4
$250	35	3	0
360	29	4	1
165	36	7	0
43	60	6	0

Costo, y	Temperatura, x_1	Aislamiento, x_2	Garage, x_4
$ 92	65	5	0
200	30	5	0
355	10	6	1

(*continúa*)

ESTADÍSTICA EN ACCIÓN

En años recientes se ha empleado la regresión múltiple en diversos procesos legales. Es particularmente útil en casos contra la discriminación por género o raza; por ejemplo, suponga que una mujer afirma que los salarios de la compañía X son injustos para ellas. Para afirmar su reclamo, la demandante presenta datos para demostrar que, en promedio, las mujeres ganan menos que los hombres; en respuesta, la compañía X argumenta que sus salarios se basan en experiencia, capacitación y aptitudes,

(*continúa*)

(*continuación*)

y que sus empleadas femeninas en promedio son más jóvenes y con menos capacitación que los varones. También, como argumento adicional, la compañía podría afirmar que la situación actual en realidad se debe a sus esfuerzos exitosos para contratar a más mujeres.

(*continuación*)

Costo, y	Temperatura, x_1	Aislamiento, x_2	Garage, x_4
$290	7	10	1
230	21	9	0
120	55	2	0
73	54	12	0
205	48	5	1
400	20	5	1
320	39	4	1

Costo, y	Temperatura, x_1	Aislamiento, x_2	Garage, x_4
$ 72	60	8	0
272	20	5	1
94	58	7	0
190	40	8	1
235	27	9	0
139	30	7	0

La captura de pantalla de Excel es:

	A	B	C	D	E	F	G	H	I	J	K
1	Costo	Temp.	Aislam.	Edad		RESUMEN DE SALIDA					
2	250	35	3	0							
3	360	29	4	1		*Estadísticos de regresión*					
4	165	36	7	0		Múltiple *R*	0.933				
5	43	60	6	0		*R* cuadrada	0.870				
6	92	65	5	0		*R* cuadrada ajustada	0.845				
7	200	30	5	0		Error estándar	41.618				
8	355	10	6	1		Observaciones	20				
9	290	7	10	1							
10	230	21	9	0		ANOVA					
11	120	55	2	0			*gl*		*MS*	*F*	*Significancia F*
12	73	54	12	0		Regresión	3	185202.269	61734.090	35.641	0.000
13	205	48	5	1		Residual	16	27713.481	1732.093		
14	400	20	5	1		Total	19	212915.750			
15	320	39	4	1							
16	72	60	8	0			*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	272	20	5	1		Intercepción	393.666	45.001	8.748	0.000	
18	94	58	7	0		Temperatura	-3.963	0.653	-6.072	0.000	
19	190	40	8	1		Aislamiento	-11.334	4.002	-2.832	0.012	
20	235	27	9	0		Garage	77.432	22.783	3.399	0.004	

¿Cuál es el efecto de la variable "garaje"? ¿Se debe incluir en el análisis? Para mostrar su efecto, suponga que se tienen dos casas exactamente iguales, una al lado de la otra, en Buffalo, Nueva York; una tiene garaje, y la otra no. Las dos casas tienen tres pulgadas de aislamiento y la temperatura media en enero en Buffalo es de 20 grados. Para la casa sin garaje, 0 se sustituye por x_4 en la ecuación de regresión. El costo estimado de la calefacción es de 280.404 dólares, determinado por:

$$\begin{aligned}\hat{y} &= 393.666 - 3.963x_1 - 11.334x_2 + 77.432x_4 \\ &= 393.666 - 3.963(20) - 11.334(3) + 77.432(0) = 280.404\end{aligned}$$

Para la casa con garaje, 1 se sustituye por x_4 en la ecuación de regresión. El costo estimado de la calefacción es de 357.836 dólares, determinado por:

$$\begin{aligned}\hat{y} &= 393.666 - 3.963x_1 - 11.334x_2 + 77.432x_4 \\ &= 393.666 - 3.963(20) - 11.334(3) + 77.432(1) = 357.836\end{aligned}$$

La diferencia entre los dos costos de calefacción estimados es de 77.432 dólares ($358.836 − $280.404); por lo tanto, es de esperar que el costo para calentar la casa con un garaje sea 77.432 dólares más alto que el de una casa equivalente sin garaje.

Se demostró que la diferencia entre los dos tipos de casas es de 77.432 dólares, pero, ¿es significativa la diferencia? Para responder, se realiza la siguiente prueba de hipótesis.

$$H_0\colon \beta_4 = 0$$
$$H_1\colon \beta_4 \neq 0$$

La información necesaria para responder esta pregunta se encuentra en la captura de pantalla de Excel anterior. El coeficiente de regresión de "garaje" es 77.432, y la desviación estándar de la distribución de muestreo es 22.783. Esta se identifica como la cuarta variable independiente, por lo que se emplea un subíndice de 4 (recuerde que la tercera variable independiente, es decir, la antigüedad del calentador, se elimina). Por último, estos valores se sustituyen en la fórmula [14.6].

$$t = \frac{b_4 - 0}{s_{b_4}} = \frac{77.432 - 0}{22.783} = 3.399$$

Hay tres variables independientes en el análisis, por lo cual hay $n - (k + 1) = 20 - (3 + 1) = 16$ grados de libertad. El valor crítico del apéndice B.5 es 2.120. La regla de decisión, con una prueba de dos colas y el nivel de significancia 0.05, es rechazar H_0 si la t calculada se encuentra a la izquierda de -2.120 o a la derecha de 2.120; como el valor calculado de 3.40 está a la derecha, se rechaza la hipótesis nula. Se concluye que el coeficiente de regresión no es cero. La variable independiente "garaje" se debe incluir en el análisis.

Utilizando el método del valor p, el valor t calculado de 3.399 tiene un valor p de 0.004; este es menor que el nivel de significancia 0.05; por lo tanto, se rechaza la hipótesis nula. Se concluye que el coeficiente de regresión no es cero, y que la variable independiente "garaje" se debe incluir en el análisis.

¿Puede emplear una variable cualitativa con más de dos resultados posibles? Sí, pero el esquema de codificación se complica y requiere una serie de variables ficticias; para explicar esto, suponga que una compañía estudia sus ventas, pues se relacionan con el gasto en publicidad trimestral durante los últimos cinco años. Suponga que las ventas son la variable dependiente, y el gasto en publicidad, la primera variable independiente, x_1. Para incluir la información cualitativa respecto del trimestre, se utilizan tres variables independientes adicionales. En el caso de la variable x_2, las cinco observaciones que se refieren al primer trimestre de cada uno de los cinco años se codifican 1, y los otros trimestres, 0. De manera similar, en el de x_3 las cinco primeras observaciones referentes al segundo trimestre se codifican 1, y los otros trimestres, 0; para el de x_4, las cinco referentes al tercer trimestre se codifican 1, y los otros trimestres, 0. Una observación que no aluda a ninguno de los primeros trimestres se debe referir al cuarto trimestre, por lo que no es necesaria una variable independiente distinta concerniente a este trimestre.

AUTOEVALUACIÓN 14-4

En un estudio de la American Realtors Association se investigó la relación entre las comisiones para los agentes de ventas durante el año previo y el número de meses desde que obtuvieron sus licencias para operar en el sector; el género de los agentes también es de interés en el estudio. A continuación se presenta una parte de la salida de la regresión. La variable dependiente es "comisiones", reportadas en miles de dólares, y las variables independientes son los "meses desde que se obtuvo la licencia" y el "género" (mujer = 1 y hombre = 0).

Análisis de regresión

Estadísticas de regresión	
Múltiple R	0.801
R cuadrada	0.642
R cuadrada ajustada	0.600
Error estándar	3.219
Observaciones	20

ANOVA

	gl	*SS*	*MS*	*F*	*Valor p*
Regresión	2	315.9291	157.9645	15.2468	0.0002
Residuo	17	176.1284	10.36049		
Total	19	492.0575			

	Coeficientes	*Error estándar*	*Estadístico t*	*Valor p*
Intercepción	15.7625	3.0782	5.121	.0001
Meses	0.4415	0.0839	5.262	.0001
Género	3.8598	1.4724	2.621	.0179

(a) Escriba la ecuación de regresión. ¿Qué comisión esperaría para una agente que obtuvo su licencia hace 30 meses?
(b) ¿En promedio, las agentes ganan más o menos que sus colegas masculinos? ¿Cuánto más?
(c) Realice una prueba de hipótesis para determinar si se debe incluir la variable independiente "género" en el análisis. Utilice el nivel de significancia 0.05. ¿Cuál es su conclusión?

Modelos de regresión con interacción

OA14-6
Incluir e interpretar un efecto de interacción en el análisis de regresión múltiple.

En el capítulo 12 se analizó la interacción entre variables independientes. Para explicar este tema, suponga que se estudia la pérdida de peso y, además, como se sugiere en la información actual, que la dieta y el ejercicio están relacionados; por lo tanto, la variable dependiente es la cantidad de cambio de peso y las variables independientes son "dieta" (sí o no) y "ejercicio" (nada, moderado, significativo). Nos interesa saber si existe una interacción entre las variables independientes; es decir, si los individuos estudiados son constantes con su dieta y ejercicio, ¿aumentará la cantidad media de pérdida de peso? ¿Es mayor la pérdida de peso total que la suma de la pérdida debida al efecto de la dieta y la pérdida debida al efecto del ejercicio?

Es posible ampliar esta idea; en lugar de tener dos variables en escala nominal ("dieta" y "ejercicio"), se examina el efecto (interacción) de diversas variables en escala de razón. Por ejemplo, suponga que desea estudiar el efecto de la temperatura ambiente (68, 72, 76 u 80 grados Fahrenheit) y el nivel de ruido (60, 70 u 80 decibeles) en el número de unidades producidas; en otras palabras, ¿tiene algún efecto la combinación de nivel de ruido y temperatura en el recinto sobre la productividad de los trabajadores? ¿Producirán más unidades en una habitación en calma y fría que quienes trabajan en una calurosa y ruidosa?

En el análisis de regresión, la interacción se examina como variable independiente separada. Se desarrolla una interacción de la variable de predicción al multiplicar los valores de una variable independiente por los de otra, y por ende, al crear una nueva variable independiente. Un modelo de dos variables que incluye un término de interacción es:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

X_1X_2 es el *término de interacción*; esta variable se creó al multiplicar los valores de X_1 y X_2 para crear la tercera variable independiente. Luego se desarrolló una ecuación de regresión con las tres variables y se probó la significancia de la tercera con la prueba individual para variables independientes que se describió antes en este capítulo. Los detalles se ilustran en el siguiente ejemplo.

EJEMPLO

Consulte el ejemplo del costo de calefacción y los datos que se hallan en la tabla 14.1. ¿Hay alguna interacción entre la temperatura externa y la cantidad de aislamiento? Si ambas variables crecen, ¿será mayor el efecto en el costo de calefacción que la suma de los ahorros derivados por una temperatura más cálida y los de mayor aislamiento, por separado?

SOLUCIÓN

A continuación se repite la información de la tabla 14.1 sobre las variables independientes "temperatura" y "aislamiento". La variable de interacción se crea al multiplicar "temperatura" por "aislamiento". En la primera casa muestreada, el valor de la temperatura es de 35 grados, y el del aislamiento, de 3 pulgadas, por lo que el valor de la variable de interacción es $35 \times 3 = 105$. Los valores de los otros productos de interacción se determinan de manera similar.

	A	B	C	D	E	F	G	H	I	J	K
1	Costo	Temp.	Aislam.	T × A		RESUMEN DE SALIDA					
2	250	35	3	105							
3	360	29	4	116		*Estadísticos de regresión*					
4	165	36	7	252		Múltiple *R*	0.893				
5	43	60	6	360		*R* cuadrada	0.798				
6	92	65	5	325		*R* cuadrada ajustada	0.760				
7	200	30	5	150		Error estándar	51.846				
8	355	10	6	60		Observaciones	20				
9	290	7	10	70							
10	230	21	9	189		ANOVA					
11	120	55	2	110			*gl*	*SS*	*MS*	*F*	*Significancia F*
12	73	54	12	648		Regresión	3	169908.452	56636.151	21.070	0.000
13	205	48	5	240		Residual	16	43007.298	2687.956		
14	400	20	5	100		Total	19	212915.750			
15	320	39	4	156							
16	72	60	8	480			*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	272	20	5	100		Intercepción	598.070	92.265	6.482	0.000	
18	94	58	7	406		Temperatura	-7.811	2.124	-3.678	0.002	
19	190	40	8	320		Aislamiento	-30.161	12.621	-2.390	0.030	
20	235	27	9	243		Temperatura × aislamiento	0.385	0.291	1.324	0.204	
21	139	30	7	210							

La regresión múltiple se encuentra al aplicar la temperatura, el aislamiento y la interacción entre estos como variables independientes. He aquí la ecuación de regresión.

$$\hat{y} = 598.070 - 7.811x_1 - 30.161x_2 + 0.385x_1x_2$$

Lo que se desea saber es si la variable de interacción es significativa. Se utilizará el nivel de significancia 0.05; en términos de una hipótesis:

$$H_0: \beta_3 = 0$$
$$H_1: \beta_3 \neq 0$$

Hay $n - (k + 1) = 20 - (3 + 1) = 16$ grados de libertad. Con el nivel de significancia 0.05 y una prueba de dos colas, los valores críticos de t son -2.120 y 2.120. La hipótesis nula se rechaza si t es menor que -2.120, o si t es mayor que 2.120. De la salida, $b_3 = 0.385$ y $s_{b_3} = 0.291$. Utilice la fórmula [14.6] para determinar el valor de t.

$$t = \frac{b_3 - 0}{s_{b_3}} = \frac{0.385 - 0}{0.291} = 1.324$$

Como el valor calculado de 1.324 es menor que el valor crítico de 2.120, no se rechaza la hipótesis nula; en conclusión, no hay una interacción significativa entre la temperatura y el aislamiento.

Hay otras situaciones factibles cuando se estudia la interacción entre variables independientes.

1. Es posible tener una interacción de tres vías entre las variables independientes. En el ejemplo del costo de la calefacción, podría haber considerado la interacción de tres vías entre la temperatura, el aislamiento y la antigüedad del calentador.
2. Es posible que exista interacción donde una de las variables independientes esté en escala nominal. En el ejemplo del costo de calefacción, podría haber estudiado la interacción entre la temperatura y el garaje.

Estudiar todas las interacciones posibles puede ser muy complejo; sin embargo, con frecuencia una consideración cuidadosa de todas ellas proporciona una visión útil de los modelos de regresión.

Regresión por pasos

OA14-7
Aplicar la regresión por pasos para desarrollar un modelo de regresión múltiple.

En el ejemplo del costo de calefacción (vea la información muestral que se registra en las tablas 14.1 y 14.2) se consideraron cuatro variables independientes: la temperatura externa media, la cantidad de aislamiento en la casa, la antigüedad del calentador, y si había garaje o no. Para elaborar la ecuación, primero se realiza una prueba global o "todo de una vez" para determinar si alguno de los coeficientes de regresión era significativo. Cuando se determina que al menos uno lo era, se prueban los coeficientes de regresión de manera individual para ver cuáles eran importantes. Las variables independientes que no tenían coeficientes de regresión significativos se mantuvieron, y se descartaron las otras. Al retener las variables independientes con coeficientes significativos, se determinó la ecuación de regresión en la que se empleó el número menor de variables independientes; esto facilitó la labor de interpretar la ecuación de regresión. Después se consideró la variable cualitativa "garaje" y se encontró que se relacionaba significativamente con el costo de calefacción; por lo tanto, dicha variable se añadió a la ecuación.

Ahora se describe la técnica denominada **regresión por pasos**; la cual es más eficiente para construir una ecuación que incluya solo las variables independientes que tengan coeficientes significativos de regresión.

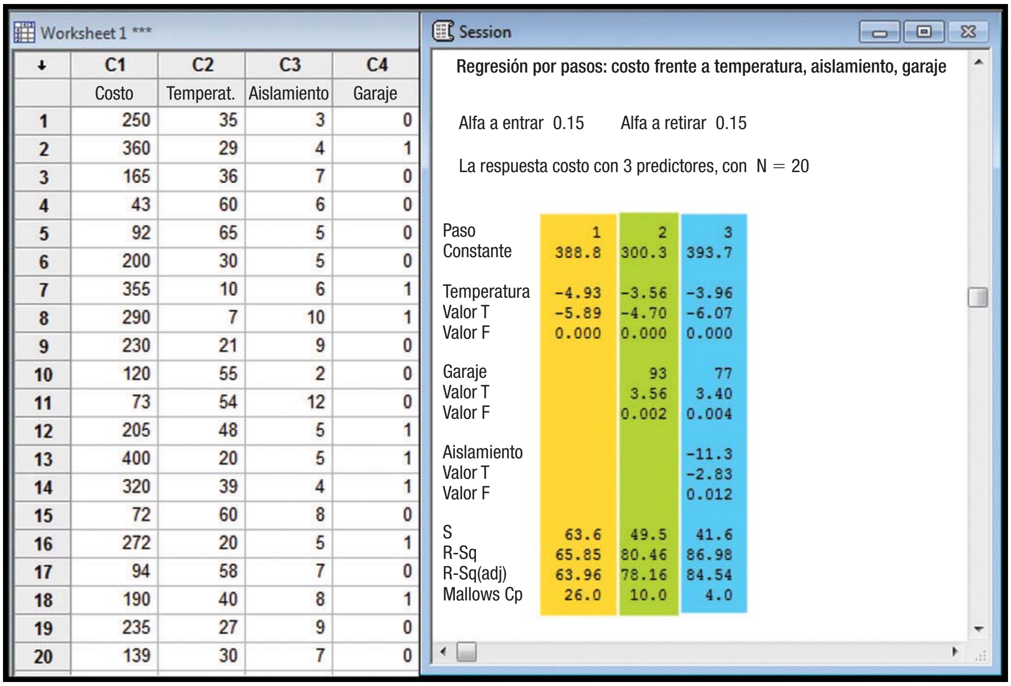
REGRESIÓN POR PASOS Método paso por paso para determinar la ecuación de regresión que se inicia con una sola variable independiente y agrega o elimina otras de manera individual. En la ecuación de regresión solo se incluyen las variables independientes con coeficientes de regresión distintos de cero.

En el método por pasos se desarrolla una secuencia de ecuaciones. La primera de ellas solo contiene una variable independiente; sin embargo, esta proviene del conjunto propuesto de variables independientes que explica la mayoría de las modificaciones en la variable dependiente. En otras palabras, si calcula todas las correlaciones simples entre cada una de las variables independientes y la variable dependiente en el método por pasos, primero se selecciona la variable independiente que tiene la correlación más fuerte con la variable dependiente.

A continuación, mediante este método se analizan las variables independientes y después se selecciona la que explicará el mayor porcentaje de la variación que aún no se aclara. Este proceso se prolonga hasta incluir en la ecuación de regresión todas las variables independientes con coeficientes de regresión significativos. Las ventajas del método por pasos son:

1. Solo se ingresan en la ecuación las variables independientes con coeficientes de regresión significativos.
2. Los pasos comprendidos en el desarrollo de la ecuación de regresión son claros.
3. Es eficaz para determinar la ecuación de regresión solo con coeficientes de regresión significativos.
4. Se muestran los cambios del error estándar de estimación múltiple y el coeficiente de determinación.

La salida de Minitab del método por pasos en el caso del problema del costo de calefacción se muestra a continuación; observe que la ecuación final, la cual se reporta en la tercera columna, incluye las variables independientes “temperatura”, “garaje” y “aislamiento”; las cuales son las mismas que se incluyeron en la ecuación de la prueba global y en la prueba de variables independientes individuales (vea la sección “Variables independientes cualitativas”). No se incluye la variable independiente “antigüedad” (la edad del calentador) debido a que no es un factor de predicción significativo del costo.



Worksheet 1 ***

↓	C1	C2	C3	C4
	Costo	Temperat.	Aislamiento	Garaje
1	250	35	3	0
2	360	29	4	1
3	165	36	7	0
4	43	60	6	0
5	92	65	5	0
6	200	30	5	0
7	355	10	6	1
8	290	7	10	1
9	230	21	9	0
10	120	55	2	0
11	73	54	12	0
12	205	48	5	1
13	400	20	5	1
14	320	39	4	1
15	72	60	8	0
16	272	20	5	1
17	94	58	7	0
18	190	40	8	1
19	235	27	9	0
20	139	30	7	0

Session

Regresión por pasos: costo frente a temperatura, aislamiento, garaje

Alfa a entrar 0.15 Alfa a retirar 0.15

La respuesta costo con 3 predictores, con N = 20

Paso	1	2	3
Constante	388.8	300.3	393.7
Temperatura	-4.93	-3.56	-3.96
Valor T	-5.89	-4.70	-6.07
Valor F	0.000	0.000	0.000
Garaje		93	77
Valor T		3.56	3.40
Valor F		0.002	0.004
Aislamiento			-11.3
Valor T			-2.83
Valor F			0.012
S	63.6	49.5	41.6
R-Sq	65.85	80.46	86.98
R-Sq(adj)	63.96	78.16	84.54
Mallows Cp	26.0	10.0	4.0

He aquí el repaso del método por pasos y la interpretación de la captura de pantalla:

1. En el procedimiento por pasos primero se selecciona la variable independiente, en este caso, “temperatura”; la cual explica más de la variación del costo de calefacción que cualquier otra de las tres variables independientes propuestas (la temperatura explica 65.85% de la variación del costo de calefacción). La ecuación de regresión es:

$$\hat{y} = 388.8 - 4.93x_1$$

   Existe una relación inversa entre el costo de calefacción y la temperatura; por cada grado de aumento de temperatura, el costo de calefacción se reduce 4.93 dólares.

2. La siguiente variable independiente a considerar en la ecuación de regresión es “garaje”; cuando esta se agrega, el coeficiente de determinación aumenta de 65.85% a 80.46%; es decir, al agregarla como variable independiente, el coeficiente de determinación aumenta 14.61%. La ecuación de regresión después del segundo paso es:

$$\hat{y} = 300.3 - 3.56x_1 + 93.0x_2$$

   Usualmente, los coeficientes de regresión cambiarán de un paso al otro. En este caso, el coeficiente de la temperatura retuvo su signo negativo, pero cambió de -4.93 a -3.56; esto se debe a la influencia agregada de la variable independiente “garaje”. ¿Por qué en el método por pasos se seleccionó “garaje” como la variable independiente en lugar de “aislamiento” o “antigüedad?” El aumento en R^2, el coeficiente de determinación, es mayor si se incluye “garaje” en lugar de cualquiera de las otras dos variables.

3. En este punto hay dos variables que no se han usado: “aislamiento” y “antigüedad”. Observe que en el tercer paso se selecciona la primera y después se detiene el procedimiento; lo cual indica que esta explica más de la variación restante del costo de calefacción en comparación con la variable “antigüedad”. Después del tercer paso, la ecuación de regresión es:

$$\hat{y} = 393.7 - 3.96x_1 + 77.0x_2 - 11.3x_3$$

Hasta aquí, 86.98% de la variación del costo de calefacción se explica por las tres variables independientes “temperatura”, “garaje” y “aislamiento”. Este es el mismo valor R^2 y la misma ecuación de regresión determinados en la sección “Variables independientes cualitativas”, excepto por diferencias de redondeo.

4. En esta etapa se detiene el procedimiento por pasos; es decir, la variable independiente “antigüedad” no contribuye de manera significativa al coeficiente de determinación.

El método por pasos desarrolló la misma ecuación de regresión, seleccionó las mismas variables independientes y determinó el mismo coeficiente de determinación que en las pruebas global e individual que se describieron antes en este capítulo. La ventaja del método por pasos es que es más directo que una combinación de los procedimientos global e individual.

También hay otros métodos para seleccionar variables. Al método por pasos también se le denomina **método de selección hacia adelante** debido a que se inicia sin variables independientes y las agrega de manera individual a la ecuación de regresión en cada iteración. Asimismo, existe el **método de eliminación hacia atrás**, que comienza con todo el conjunto de variables y elimina una variable independiente en cada iteración.

En los métodos descritos hasta aquí se considera una variable a la vez, y se decide si esta se incluye o se elimina. Otro enfoque es la **regresión del mejor subconjunto**; en este método se considera el mejor modelo con una variable independiente, el mejor modelo con dos variables independientes, el mejor modelo con tres y así sucesivamente. El criterio es encontrar el modelo con el mayor valor R^2, sin que importe el número de variables independientes; además, no es necesario que cada una tenga un coeficiente de regresión distinto de cero. Como cada variable independiente puede incluirse o no, hay $2^k - 1$ modelos posibles, donde k se refiere al número de variables independientes. En el ejemplo del costo de calefacción hay cuatro variables independientes, por lo que hay 15 modelos de regresión posibles, determinados por $2^4 - 1 = 16 - 1 = 15$. Todos los modelos de regresión se examinarían con una variable independiente, todas las combinaciones con dos variables independientes, todas las combinaciones con tres variables independientes, y la probabilidad de utilizar las cuatro variables independientes. Una ventaja del método del mejor subconjunto es que ayuda a examinar combinaciones de variables independientes no consideradas en el método por pasos (este proceso se encuentra disponible en Minitab y MegaStat).

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

9. El gerente de producción de High Point Sofa and Chair, importante fabricante de muebles ubicado en Carolina del Norte, estudia las calificaciones de desempeño laboral de una muestra de 15 electricistas de mantenimiento que trabajan en la compañía. Para ingresar a la empresa, el Departamento de Recursos Humanos les aplica un examen de aptitud. El gerente de producción obtuvo la calificación de cada electricista incluido en la muestra; además, determinó cuáles electricistas eran miembros de un sindicato (código = 1) y cuáles no lo eran (código = 0). He aquí la información muestral.

Trabajador	Calificación de desempeño laboral	Calificación en el examen de aptitud	Miembro de sindicato
Abbott	58	5	0
Anderson	53	4	0
Bender	33	10	0
Bush	97	10	0
Center	36	2	0
Coombs	83	7	0

(*continúa*)

(*continuación*)

Trabajador	Calificación de desempeño laboral	Calificación en el examen de aptitud	Miembro de sindicato
Eckstine	67	6	0
Gloss	84	9	0
Herd	98	9	1
Householder	45	2	1
Iori	97	8	1
Lindstrom	90	6	1
Mason	96	7	1
Pierse	66	3	1
Rohde	82	6	1

a. Utilice un paquete de software estadístico para desarrollar una ecuación de regresión múltiple con la calificación de desempeño laboral como variable dependiente, y la calificación en el examen de aptitud y la pertenencia a un sindicato como variables independientes.
b. Comente la ecuación de regresión. Incluya el coeficiente de determinación y el efecto de la pertenencia o no a un sindicato. ¿Son eficaces estas dos variables para explicar la variación del desempeño laboral?
c. Realice una prueba de hipótesis para determinar si la pertenencia a un sindicato se debe incluir como variable independiente.
d. Repita el análisis considerando los términos de interacción posibles.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

10. La Cincinnati Paint Company vende marcas de pintura de prestigio en ferreterías de Estados Unidos. La compañía mantiene una fuerza laboral numerosa, cuya tarea es atender a clientes actuales, así como buscar nuevos compradores. El gerente nacional de ventas investiga la relación entre el número de llamadas de ventas y las millas que recorren los agentes de ventas. ¿Ganan más en comisiones por ventas los agentes que recorren más millas y hacen más llamadas? Para investigar esta cuestión, el vicepresidente de ventas seleccionó una muestra de 25 agentes y determinó:

- La cantidad que ganaron por comisiones el mes anterior (y).
- El número de millas que recorrieron durante ese mes (x_1).
- El número de llamadas de ventas del periodo (x_2).

La información se reporta en la siguiente tabla:

Comisiones (en miles de dólares)	Llamadas	Millas recorridas
22	139	2 371
13	132	2 226
33	144	2 731
⋮	⋮	⋮
25	127	2 671
43	154	2 988
34	147	2 829

Formule una ecuación de regresión que incluya un término de interacción. ¿Hay una interacción significativa entre el número de llamadas de ventas y las millas recorridas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

11. Un coleccionista de arte estudia la relación entre el precio de venta de una pintura y dos variables independientes: el número de postores en la subasta particular y la antigüedad de la pintura, en años. Mediante una muestra de 25 pinturas se encontró la siguiente información muestral.

Pintura	Precio en la subasta	Postores	Antigüedad
1	3 470	10	67
2	3 500	8	56
3	3 700	7	73
⋮	⋮		⋮
23	4 660	5	94
24	4 710	3	88
25	4 880	1	84

a. Formule una ecuación de regresión múltiple con el número de variables independientes “postores” y “antigüedad” para estimar el precio en la subasta de la variable dependiente. Analice la ecuación. ¿Le sorprende que haya una relación inversa entre el número de postores y el precio de la pintura?

b. Formule una variable de interacción e inclúyala en la ecuación de regresión. Explique el significado de la interacción. ¿Es significativa esta variable?

c. Utilice el método por pasos y las variables independientes número de postores y antigüedad de la pintura así como la interacción entre ambas. ¿Qué variables seleccionaría?

12. Un constructor inmobiliario desea estudiar la relación entre el tamaño de la casa que compraría un cliente (en pies cuadrados) y otras variables. Las posibles variables independientes son el ingreso familiar, el número de miembros en la familia, si hay un adulto mayor viviendo con la familia (1 para sí, 0 para no), y los años totales de educación adicionales al bachillerato del esposo y la esposa. La información muestral se reporta en la siguiente tabla.

Familia	Pies cuadrados	Ingreso (en miles de dólares)	Miembros en la familia	Adulto mayor	Educación
1	2 240	60.8	2	0	4
2	2 380	68.4	2	1	6
3	3 640	104.5	3	0	7
4	3 360	89.3	4	1	0
5	3 080	72.2	4	0	2
6	2 940	114	3	1	10
7	4 480	125.4	6	0	6
8	2 520	83.6	3	0	8
9	4 200	133	5	0	2
10	2 800	95	3	0	6

Formule una ecuación de regresión múltiple apropiada. ¿Qué variables independientes incluiría en la ecuación de regresión final? Utilice el método por pasos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Repaso de la regresión múltiple

OA14-8 Aplicar las técnicas de regresión múltiple para desarrollar un modelo lineal.

En este capítulo se han descrito varios temas que involucran a la regresión múltiple. Esta sección se enfoca en un solo ejemplo con una solución que repasa el procedimiento y le guiará en su aplicación del análisis de regresión múltiple.

EJEMPLO

El Banco de Nueva Inglaterra es una gran institución financiera que da servicio a los estados de Nueva Inglaterra, así como a Nueva York y Nueva Jersey. El Departamento de Préstamos Hipotecarios del banco está estudiando datos de créditos recientes. En particular, le interesa saber a qué grado factores tales como el valor de la casa que se desea comprar (en miles de dólares), el nivel de educación del cabeza de familia (número de años, comenzando por el primer grado), su edad, el pago mensual actual de la hipoteca (en dólares), y el género de dicha persona (hombre = 1, mujer = 0) se relacionan con el ingreso familiar. Los directivos quieren saber si estas variables predicen con eficacia el ingreso familiar.

SOLUCIÓN

Considere una muestra aleatoria de 25 solicitudes de crédito sometidas al Banco de Nueva Inglaterra el mes anterior. Una parte de dicha información muestral se presenta en la tabla 14.3. El conjunto

TABLA 14.3 Información de la muestra de 25 préstamos del Banco de Nueva Inglaterra

Préstamo	Ingreso (miles de dólares)	Valor (miles de dólares)	Educación	Edad	Hipoteca	Género
1	100.7	190	14	53	230	1
2	99.0	121	15	49	370	1
3	102.0	161	14	44	397	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
23	102.3	163	14	46	142	1
24	100.2	150	15	50	343	0
25	96.3	139	14	45	373	0

completo de datos está disponible en el sitio web (**www.mhhe.com/uni/lind_ae16e**), y se identifica como Banco de Nueva Inglaterra (Bank of New England).

Para comenzar, se calcula la matriz de correlación que aparece a continuación para obtener la relación entre las variables independientes y la variable dependiente; esto ayudará a identificar cuáles de estas se relacionan más con la variable dependiente "ingreso familiar". Mediante la matriz de correlación también se revelan aquellas variables independientes que están altamente relacionadas y quizá sean redundantes.

	Ingreso	Valor	Educación	Edad	Hipoteca	Género
Ingreso	1					
Valor	0.720	1				
Educación	0.188	–0.144	1			
Edad	0.243	0.220	0.621	1		
Hipoteca	0.116	0.358	–0.210	–0.038	1	
Género	0.486	0.184	0.062	0.156	–0.129	1

¿Qué revela esta matriz de correlación?

1. La primera columna muestra las correlaciones entre cada variable independiente y la variable dependiente "ingreso familiar". Observe que cada variable independiente está correlacionada positivamente con el ingreso familiar. El valor de la casa tiene la correlación más fuerte con este; el nivel de educación de la persona que solicita el préstamo y el pago actual de la hipoteca tienen una correlación débil con dicho ingreso. Ambas variables son candidatas a ser eliminadas de la ecuación de regresión.
2. Todas las posibles correlaciones entre las variables independientes están resaltadas en verde. Se sugiere buscar correlaciones que excedan un valor absoluto de 0.700. Ninguna de las variables independientes está fuertemente correlacionada con las demás; esto indica que no hay probabilidad de multicolinealidad.

Después, se calcula la ecuación de la regresión múltiple utilizando todas las variables independientes; a continuación se muestra la salida de software.

	A	B	C	D	E	F
1	RESUMEN DE SALIDA					
2						
3	*Estadísticos de regresión*					
4	Múltiple *R*	0.866				
5	*R* cuadrada	0.750				
6	*R* cuadrada ajustada	0.684				
7	Error estándar	1.478				
8	Observaciones	25				
9						
10	ANOVA					
11		*gl*	*SS*	*MS*	*F*	*Valor p*
12	Regresión	5	124.322	24.864	11.385	0.000
13	Residual	19	41.494	2.184		
14	Total	24	165.815			
15						
16		*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	Intercepción	70.606	7.464	9.459	0.000	
18	Valor ($000)	0.072	0.012	5.769	0.000	
19	Educación	1.624	0.603	2.693	0.014	
20	Edad	-0.122	0.078	-1.566	0.134	
21	Hipoteca	-0.001	0.003	-0.319	0.753	
22	Género	1.807	0.623	2.901	0.009	

Los coeficientes de determinación, esto es, R^2 y R^2 ajustado, se reportan en la parte superior del resumen de la salida y están resaltadas en amarillo. El valor R^2 es de 75.0%, así que las cinco variables independientes representan ese porcentaje en la variación del ingreso familiar. El valor R^2 ajustado mide la fuerza de la relación entre el grupo de variables independientes y el ingreso familiar, y representa el número de variables en la ecuación de regresión. El R^2 ajustado indica que las cinco variables representan 68.4% de la varianza del ingreso familiar. Ambos factores sugieren que las variables independientes propuestas son útiles para pronosticar el ingreso familiar.

La salida incluye también la ecuación de regresión:

$$\hat{y} = 70.606 + 0.072(Valor) + 1.624(Educación) - 0.122(Edad)$$
$$- 0.001(Hipoteca) + 1.807(Género)$$

Tenga cuidado con esta interpretación; tanto el ingreso como el valor de la casa están en miles de dólares. He aquí un resumen:

1. Un aumento de 1 000 dólares del valor de la casa sugiere un incremento de 72 dólares en el ingreso familiar; un aumento de un año de educación eleva el ingreso en 1 624 dólares; un año más de edad reduce el ingreso en 122 dólares y un incremento de 1 000 dólares de la hipoteca reduce el ingreso en un dólar.
2. Si un hombre es el cabeza de familia, el valor del ingreso familiar se eleva en 1 807 dólares; recuerde que "mujer" fue codificado como 0, y "hombre" como 1, así que un hombre como cabeza de familia está relacionado positivamente con el valor de la casa.
3. La edad de la cabeza de familia y el pago mensual de la hipoteca están inversamente relacionados con el ingreso familiar. Esto es cierto porque el signo del coeficiente de regresión es negativo.

A continuación se realiza la prueba de la hipótesis global. Aquí es preciso verificar si cualquiera de los coeficientes de regresión es distinto de cero; por lo tanto, se aplica el nivel de significancia 0.05.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$
$$H_1: \text{No todos los } \beta \text{ son} = 0$$

El valor p de la tabla (celda F12) es 0.000; como el valor p es menor al nivel de significancia, se rechaza la hipótesis nula y se concluye que cuando menos uno de los coeficientes de regresión no es igual a cero.

Enseguida se evalúan los coeficientes individuales de regresión. Consulte los valores p de la salida de software para probar cada coeficiente de regresión; los cuales están reportados en las celdas E18 a E22. Las hipótesis nula y alternativa son:

$$H_0: \beta_i = 0$$
$$H_1: \beta_i \neq 0$$

El subíndice i representa cualquier variable dependiente particular. Si se utiliza otra vez el nivel de significancia 0.05, entonces los valores p de los coeficientes de regresión del valor de la casa, los años de educación y el género son menores a 0.05; por lo tanto, estos coeficientes de regresión no son iguales a 0 y son factores de predicción significativos del ingreso familiar. En el caso de la edad y el monto de la hipoteca, los valores p son mayores al nivel de significancia 0.05, así que no se rechaza la hipótesis nula para estas variables. Los coeficientes de regresión de ambas variables no difieren de cero y no están relacionados con el ingreso familiar.

Con base en los resultados de la prueba de cada coeficiente de regresión, se concluye que las variables "edad" e "hipoteca" no son factores de predicción eficaces del ingreso familiar; por lo tanto, se deben retirar de la ecuación de regresión múltiple. Recuerde que es preciso eliminar una variable independiente a la vez y rehacer el análisis para evaluar el efecto general de dicha eliminación. La estrategia es retirar la variable que tenga el menor estadístico t o el mayor valor p; en este caso, se trata del monto de la hipoteca. A la derecha se presenta el resultado del análisis de regresión sin la variable "hipoteca":

Observe que R^2 y R^2 ajustada cambian muy poco sin "hipoteca", y que el valor p asociado con la edad es mayor que el nivel de significancia 0.05; por lo tanto, se retira "edad" y se rehace el análisis. En la página siguiente se presenta la salida de la regresión sin las variables "edad" e "hipoteca":

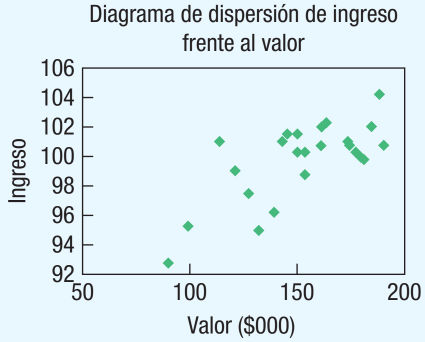
	A	B	C	D	E	F
1	RESUMEN DE SALIDA					
2						
3	*Estadísticos de regresión*					
4	Múltiple *R*	0.865				
5	*R* cuadrada	0.748				
6	*R* cuadrada ajustada	0.698				
7	Error estándar	1.444				
8	Observaciones	25				
9						
10	ANOVA					
11		*gl*	*SS*	*MS*	*F*	*Valor p*
12	Regresión	4	124.099	31.025	14.874	0.000
13	Residual	20	41.716	2.086		
14	Total	24	165.815			
15						
16		*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	Intercepción	70.159	7.165	9.791	0.000	
18	Valor ($000)	0.070	0.011	6.173	0.000	
19	Educación	1.647	0.585	2.813	0.011	
20	Edad	-0.122	0.076	-1.602	0.125	
21	Género	1.846	0.596	3.096	0.006	

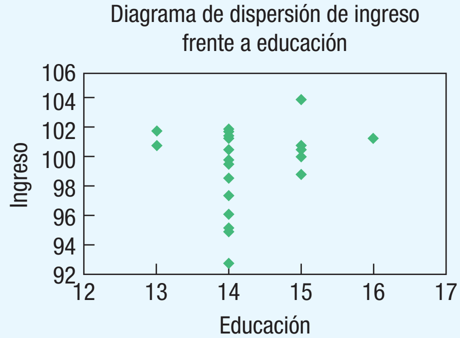
	A	B	C	D	E	F
1	RESUMEN DE SALIDA					
2						
3	*Estadísticos de regresión*					
4	Múltiple *R*	0.846				
5	*R* cuadrada	0.716				
6	*R* cuadrada ajustada	0.676				
7	Error estándar	1.497				
8	Observaciones	25				
9						
10	ANOVA					
11		*gl*	*SS*	*MS*	*F*	*Valor p*
12	Regresión	3	118.743	39.581	17.658	0.000
13	Residual	21	47.072	2.242		
14	Total	24	165.815			
15						
16		*Coeficientes*	*Error estándar*	*Estadístico t*	*Valor p*	
17	Intercepción	74.527	6.870	10.849	0.000	
18	Valor ($000)	0.063	0.011	5.803	0.000	
19	Educación	1.016	0.449	2.262	0.034	
20	Género	1.770	0.616	2.872	0.009	

De esta salida se concluye lo siguiente:

1. Los valores R^2 y R^2 ajustada han disminuido, pero solo ligeramente. Utilizando las cinco variables independientes, el valor R^2 fue de 0.750. Al quitar las dos variables no significativas, los valores R^2 y R^2 ajustada son 0.716 y 0.676, respectivamente. Es preferible tener una ecuación con el menor número de variables independientes porque es más fácil de interpretar.
2. En la tabla ANOVA se observa que el valor *p* es menor que 0.05; por lo tanto, al menos uno de los coeficientes de regresión no es igual a cero.
3. Revisando la significancia de los coeficientes individuales, se comprueba que los valores *p* asociados con cada una de las variables independientes restantes son menores a 0.05; en conclusión, todos los coeficientes de regresión son distintos a cero, y cada variable independiente es un factor de predicción útil del ingreso familiar.

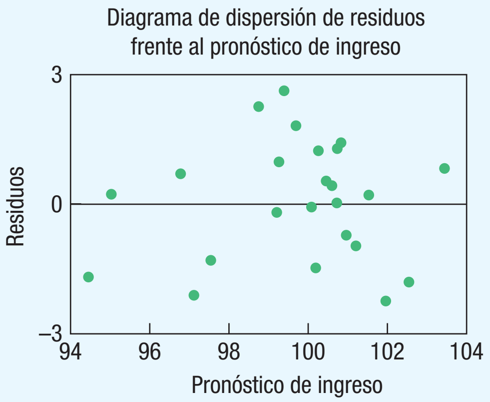
El paso final es examinar las suposiciones de regresión, enumeradas en la sección "Evaluación de las suposiciones de la regresión múltiple", con este modelo de regresión. La primera suposición es que existe una relación lineal entre cada variable independiente y la variable dependiente; no es necesario revisar la variable ficticia "género" porque hay solo dos posibles resultados. A continuación se muestran los diagramas de dispersión del ingreso familiar contra el valor de la casa, y el ingreso familiar frente a los años de educación.





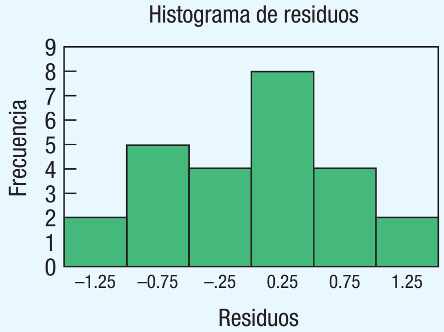
En el diagrama de dispersión del ingreso frente al valor de la casa se muestra una tendencia general ascendente. A medida que se incrementa el valor de la casa, se eleva también el ingreso familiar; los puntos parecen ser lineales. Esto es, no hay un patrón no lineal observable en los datos; en el diagrama de dispersión de la derecha (ingreso frente a los años de educación) se muestra que estos se miden hasta el año inmediatamente anterior. La medida asociada a dicho año es una variable discreta; dado el método de medición, es difícil hacer una observación en el sentido de que la relación es lineal.

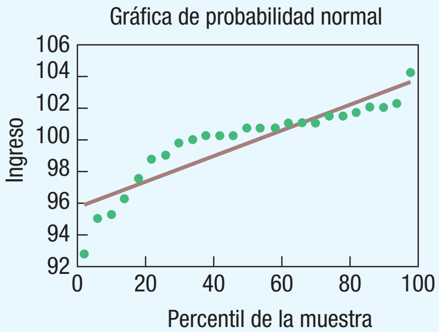
Un trazo de los residuos también permite evaluar la suposición general de linealidad. Recuerde que un residuo es $(y - \hat{y})$ la diferencia entre el valor real de la variable dependiente (y) y el valor pronosticado de la variable independiente $(\hat{y})$. Asumiendo que existe una relación lineal, la distribución de los residuos debería mostrar una proporción aproximadamente igual de los residuos negativos (puntos por encima de la línea) y los positivos (puntos debajo de esta) centrados en cero. No debería haber un patrón observable entre los puntos. El diagrama se muestra a la derecha.



No hay un patrón discernible en el trazo, de modo que se concluye que la suposición de linealidad es razonable.

Si esta suposición de linealidad es válida, entonces la distribución de residuos debe seguir una distribución de probabilidad normal con una media de cero; utilice un histograma y un trazo de probabilidad normal para evaluar esta suposición.





En general, el histograma de la izquierda muestra las mayores características de una distribución normal, esto es, la mayoría de las observaciones están en el medio y centradas en la media de cero, con menores frecuencias en las colas de la distribución. El trazo de probabilidad normal a la derecha se basa en una distribución de probabilidad normal acumulada. La línea muestra la distribución normal acumulada estandarizada. Los puntos verdes muestran la distribución acumulada de los residuos. Para confirmar la distribución normal de los residuos, tales puntos deben estar próximos a la línea. Esto es cierto para la mayoría del trazo; sin embargo, hay alejamientos e incluso quizás se observe un patrón no lineal entre los residuos de la parte baja de la gráfica. Como antes, busque alejamientos importantes de la linealidad que no se indican en estas gráficas.

La suposición final se refiere a la multicolinealidad; esto significa que las variables independientes no deben estar altamente correlacionadas. En una *regla de oro* se sugiere que la multicolinealidad debe generar preocupación si las correlaciones entre las variables independientes están próximas a 0.7 o -0.7; en este caso no hay violaciones a este lineamiento.

Existe una prueba estadística más precisa para evaluar la multicolinealidad, el factor de inflación de la varianza (*VIF*). Para calcular los *VIF* es preciso realizar un análisis de regresión para cada variable independiente como función de las otras variables independientes. En cada uno de estos análisis de regresión se necesita el R^2 para calcular el *VIF* utilizando la fórmula [14.7]. En la tabla siguiente se muestra el R^2 para cada análisis de regresión y el *VIF* calculado. Si los *VIF* son menores a 10, entonces la multicolinealidad no es un problema; en este caso, todos los *VIF* lo son, de manera que la multicolinealidad entre las variables independientes no es relevante.

Variable dependiente	**Variables independientes**	**R cuadrada**	***VIF***
Valor	Educación y género	0.058	1.062
Educación	Género y valor	0.029	1.030
Género	Valor y educación	0.042	1.044

Para resumir, la ecuación de la regresión múltiple es

$$\hat{y} = 74.527 + 0.063(\mathit{Valor}) + 1.016(\mathit{Educación}) + 1.770(\mathit{Género})$$

Esta ecuación explica 71.6% de la variación del ingreso familiar; no hay partidas principales de las suposiciones de linealidad de la regresión múltiple, residuos normalmente distribuidos y multicolinealidad.

RESUMEN DEL CAPÍTULO

I. La fórmula general de una ecuación de regresión múltiple es:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k \tag{14.1}$$

donde a es la intersección con el eje Y cuando todas las x son cero, b_j se refiere a los coeficientes de regresión de la muestra, y x_j, al valor de las diversas variables independientes.

A. Puede haber cualquier cantidad de variables independientes.
B. Se emplea el criterio de mínimos cuadrados para desarrollar la ecuación de regresión.
C. Es necesario un paquete de software estadístico para realizar los cálculos.

II. El análisis de regresión múltiple se resume en una tabla ANOVA.

A. Reporta la cantidad total de la variación de la variable dependiente y divide esta variación entre las que se explican mediante el grupo de variables independientes y las que no.
B. Reporta los grados de libertad asociados con las variables independientes, el error de la variación y la variación total.

III. Hay dos medidas de la eficacia de la ecuación de regresión.

A. El error estándar de estimación múltiple es similar a la desviación estándar.
1. Se mide en las mismas unidades que la variable dependiente.
2. Se basa en desviaciones cuadráticas de la ecuación de regresión.
3. Varía de 0 a más infinito.
4. Se calcula a partir de la siguiente ecuación.

$$s_{Y.123...k} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - (k + 1)}} \tag{14.2}$$

B. El coeficiente de determinación múltiple reporta el porcentaje de la variación de la variable dependiente que explica el conjunto de variables independientes.
1. Puede variar de 0 a 1.
2. También se basa en desviaciones cuadráticas de la ecuación de regresión.
3. Se determina mediante la siguiente ecuación.

$$R^2 = \frac{\text{SSR}}{\text{SS total}} \tag{14.3}$$

4. Cuando el número de variables independientes es grande, se ajusta el coeficiente de determinación de los grados de libertad como sigue.

$$R^2_{\text{ajust}} = 1 - \frac{\dfrac{\text{SSE}}{n - (k + 1)}}{\dfrac{\text{SS total}}{n - 1}} \tag{14.4}$$

IV. Se utiliza una prueba global para investigar si alguna de las variables independientes tiene coeficientes de regresión significativos.

A. La hipótesis nula es: todos los coeficientes de regresión son cero.
B. La hipótesis alternativa es: al menos un coeficiente de regresión no es cero.
C. El estadístico de prueba es la distribución F con k (el número de variables independientes), grados de libertad en el numerador y $n - (k + 1)$, grados de libertad en el denominador, donde n es el tamaño muestral.
D. La fórmula para calcular el valor del estadístico de prueba de la prueba global es:

$$F = \frac{\text{SSR}/k}{\text{SSE}/[n - (k + 1)]} \tag{14.5}$$

V. La prueba de las variables individuales determina cuáles de ellas tienen coeficientes de regresión distintos de cero.

A. En general, las variables con coeficientes de regresión cero se omiten del análisis.
B. El estadístico de prueba es la distribución t con $n - (k + 1)$ grados de libertad.
C. La fórmula para calcular el valor del estadístico de prueba de la prueba individual es:

$$t = \frac{b_i - 0}{s_{b_i}} \tag{14.6}$$

VI. Hay cinco suposiciones para emplear el análisis de regresión.

A. La relación entre la variable dependiente y el conjunto de variables independientes debe ser lineal.
1. Para verificar esta suposición se elabora un diagrama de dispersión, y se trazan los residuos en el eje vertical y los valores ajustados en el eje horizontal.
2. Si las gráficas se perciben como aleatorias, se concluye que la relación es lineal.

B. La variación es la misma tanto para valores grandes como pequeños de $\hat{y}$.
   1. Homoscedasticidad significa que la variación de todos los valores de la variable dependiente es la misma.
   2. Esta condición se verifica cuando se elabora un diagrama de dispersión con los residuos en el eje vertical y los valores ajustados en el eje horizontal.
   3. Si no se halla un patrón en las gráficas, es decir, si se consideran aleatorias, se concluye que los residuos cumplen con el requisito de homoscedasticidad.

C. Los residuos siguen la distribución de probabilidad normal.
   1. Esta condición se verifica al desarrollar un histograma de los residuos para ver si siguen una distribución normal.
   2. La media de la distribución de los residuos es cero.

D. Las variables independientes no están correlacionadas.
   1. Una matriz de correlación muestra todas las correlaciones posibles entre variables independientes. Se considera que hay un problema si las correlaciones son mayores que 0.70 o menores que -0.70.
   2. Entre las señales de variables independientes correlacionadas se encuentran los casos cuando una variable de predicción se determina insignificante, cuando se presenta una inversión obvia de signos en una o más de las variables independientes, o bien cuando, al eliminar una variable de la solución, se produce un gran cambio en los coeficientes de regresión.
   3. El factor de inflación de la varianza se emplea para identificar variables independientes correlacionadas.

$$VIF = \frac{1}{1 - R_j^2} \quad \textbf{[14.7]}$$

E. Cada residuo es independiente de otros residuos.
   1. La autocorrelación ocurre cuando se correlacionan residuos sucesivos.
   2. Cuando existe autocorrelación, el valor del error estándar está sesgado y genera resultados deficientes en las pruebas de hipótesis, sin que importen los coeficientes de regresión.

VII. Hay varias técnicas que ayudan a elaborar un modelo de regresión.

A. Una variable independiente ficticia o cualitativa puede asumir uno de dos resultados posibles.
   1. Se asigna un valor de 1 a uno de los resultados, y al otro, 0.
   2. Se utiliza la fórmula [14.6] para determinar si la variable ficticia debe permanecer en la ecuación.

B. Una interacción se produce cuando una variable independiente (como x_2) afecta la relación con otra variable independiente (x_1) y la variable dependiente (y).

C. La regresión por pasos es un proceso gradual para encontrar la ecuación de regresión.
   1. Solo las variables independientes con coeficientes de regresión distintos de cero entran en la ecuación.
   2. Se agregan variables independientes una a la vez a la ecuación de regresión.

CLAVE DE PRONUNCIACIÓN

Símbolo	**Significado**	**Pronunciación**
b_1	Coeficiente de regresión de la primera variable independiente	*b subíndice 1*
b_k	Coeficiente de regresión de cualquier variable independiente	*b subíndice k*
$s_{Y.123...k}$	Error estándar de estimación múltiple	*s subíndice Y punto 1, 2, 3 . . . k*

EJERCICIOS DEL CAPÍTULO

13. Mediante una ecuación de regresión múltiple se producen los resultados parciales que se muestran abajo a la derecha.
   a. ¿Cuál es el tamaño total de la muestra?
   b. ¿Cuántas variables independientes se consideraron?
   c. Calcule el coeficiente de determinación.
   d. Calcule el error estándar de estimación.
   e. Pruebe la hipótesis de que ninguno de los coeficientes de regresión es igual a cero. Suponga que $\alpha = 0.05$.

Fuente	Suma de cuadrados	*gl*
Regresión	750	4
Error	500	35

14. En una ecuación de regresión múltiple se consideran dos variables independientes, y el tamaño de la muestra es 25. Los coeficientes de regresión y los errores estándar se muestran a continuación.

$$b_1 = 2.676 \qquad s_{b_1} = 0.56$$
$$b_2 = -0.880 \qquad s_{b_2} = 0.71$$

Realice una prueba de hipótesis para determinar si alguna variable independiente tiene un coeficiente igual a cero. ¿Consideraría eliminar alguna variable de la ecuación de regresión? Utilice el nivel de significancia 0.05.

15. Se obtuvo el siguiente resultado.

Análisis de la varianza			
Fuente	*gl*	SS	MS
Regresión	5	100	20
Error residual	20	40	2
Total	25	140	
Factor de predicción	Coeficiente	Error estándar del coeficiente	*t*
Constante	3.00	1.50	2.00
x_1	4.00	3.00	1.33
x_2	3.00	0.20	15.00
x_3	0.20	0.05	4.00
x_4	−2.50	1.00	−2.50
x_5	3.00	4.00	0.75

a. ¿Cuál es el tamaño de la muestra?
b. Calcule el valor de R^2.
c. Calcule el error estándar de estimación múltiple.
d. Realice una prueba global de hipótesis para determinar si algunos de los coeficientes de regresión son significativos. Utilice el nivel de significancia 0.05.
e. Pruebe los coeficientes de regresión de manera individual. ¿Consideraría omitir una o más variables? De ser así, ¿cuál o cuáles? Utilice el nivel de significancia 0.05.

16. En una ecuación de regresión múltiple $k = 5$ y $n = 20$, el valor de MSE es 5.10, y SS total es 519.68. Con el nivel de significancia 0.05, ¿se puede concluir que alguno o varios de los coeficientes de regresión son distintos de cero?

17. La gerente de distrito de Jasons, una cadena grande de productos electrónicos, investiga por qué ciertas tiendas de su región tienen mejor rendimiento que otras; ella considera que tres factores se relacionan con las ventas totales: el número de tiendas de la competencia, la población del área circundante y la cantidad que cada una gasta en publicidad. De su distrito, que consiste en varios cientos de tiendas, selecciona una muestra aleatoria de 30 y reúne la siguiente información de cada tienda.

y = ventas totales del año previo (en miles de dólares)
x_1 = número de tiendas de la competencia en la región.
x_2 = población de la región (en millones de personas).
x_3 = gastos en publicidad (en miles de dólares).

Los datos muestrales se corrieron en Minitab, con los resultados que se muestran a la izquierda.

Análisis de la varianza			
Fuente	*gl*	SS	MS
Regresión	3	3050	1016.67
Error residual	26	2200	84.62
Total	29	5250	
Factor de predicción	Coeficiente	Error estándar del coeficiente	*t*
Constante	14.00	7.00	2.00
x_1	−1.00	0.70	−1.43
x_2	30.00	5.20	5.77
x_3	0.20	0.08	2.50

a. ¿Cuáles son las ventas estimadas de la tienda Byrne, que tiene cuatro competidores, una población regional de 0.4 (400 000) y gastos en publicidad de 30 (30 000 dólares)?
b. Calcule el valor de R^2.
c. Calcule el error estándar de estimación múltiple.
d. Realice una prueba de hipótesis global para determinar si uno o más de los coeficientes de regresión son distintos de cero. Utilice el nivel de significancia 0.05.
e. Realice pruebas de hipótesis para determinar cuál o cuáles de las variables independientes tienen coeficientes de regresión significativos. ¿Cuáles variables consideraría eliminar? Utilice el nivel de significancia 0.05.

18. Suponga que el gerente de ventas de un distribuidor importante de autopartes desea estimar en el mes de abril las ventas totales anuales de una región. Con base en las ventas regionales, también se pueden estimar las ventas totales de la compañía; a partir de la experiencia pasada se determina que las estimaciones de abril de las ventas anuales tienen una precisión razonable y que, en años futuros, esa predicción serviría para revisar los programas de producción y mantener el inventario correcto en las tiendas de descuento minoristas.

Parece que varios factores están relacionados con las ventas, como el número de tiendas de descuento minoristas en la región que ofrecen componentes de la compañía, el número de automóviles en la región registrados desde el 1 de abril, y el ingreso total personal del primer trimestre del año. Al final se seleccionaron cinco variables independientes como las más importantes (según el gerente de ventas). Luego se recopilaron los datos de un año reciente; también se registraron las ventas totales anuales en ese año por cada región. En la siguiente tabla se observa que en la región 1 había 1 739 tiendas de descuento minoristas que vendían los componentes de autos de la compañía y 9 270 000 automóviles registrados en esa región desde el 1 de abril. Las ventas en ese año fueron 37 702 000 dólares.

Ventas anuales (millones de dólares), y	Número de tiendas de descuento, x_1	Número de automóviles registrados (millones), x_2	Ingreso personal (miles de millones de dólares), x_3	Antigüedad promedio de los automóviles (años), x_4	Número de supervisores, x_5
37.702	1 739	9.27	85.4	3.5	9.0
24.196	1 221	5.86	60.7	5.0	5.0
32.055	1 846	8.81	68.1	4.4	7.0
3.611	120	3.81	20.2	4.0	5.0
17.625	1 096	10.31	33.8	3.5	7.0
45.919	2 290	11.62	95.1	4.1	13.0
29.600	1 687	8.96	69.3	4.1	15.0
8.114	241	6.28	16.3	5.9	11.0
20.116	649	7.77	34.9	5.5	16.0
12.994	1 427	10.92	15.1	4.1	10.0

a. Considerando la siguiente matriz de correlación, ¿cuál es la variable individual que tiene la correlación más fuerte con la variable dependiente? Las correlaciones entre las variables independientes “número de tiendas de descuento” e “ingreso personal”, y entre “antigüedad promedio de los automóviles” y “número de tiendas de descuento” son muy fuertes. ¿Esto puede representar un problema? ¿Cómo se denomina esta condición?

	Ventas	Tiendas de descuento	Automóviles	Ingreso	Edad
Tiendas de descuento	0.899				
Automóviles	0.605	0.775			
Ingreso	0.964	0.825	0.409		
Edad	−0.323	−0.489	−0.447	−0.349	
Supervisores	0.286	0.183	0.395	0.155	0.291

b. En la siguiente tabla se presenta el resultado de la ecuación de regresión de las cinco variables. ¿Qué porcentaje de la variación se explica mediante la ecuación de regresión?

```
La ecuación de regresión es
Ventas = −19.7 − 0.00063 tiendas de descuento + 1.74 automóviles
         + 0.410 ingreso + 2.04 edad − 0.034 supervisores

           Predictor                  Coef      SE Coef        T        P
           Constante               −19.672        5.422    −3.63    0.022
           Tiendas de descuento  −0.000629     0.002638    −0.24    0.823
           Automóviles              1.7399       0.5530     3.15    0.035
           Ingreso                 0.40994      0.04385     9.35    0.001
           Edad                     2.0357       0.8779     2.32    0.081
           Supervisores            −0.0344       0.1880    −0.18    0.864

Análisis de varianza
           FUENTE          gl        SS       MS        F        P
           Regresión        5   1593.81   318.76   140.36    0.000
           Error residual   4      9.08     2.27
           Total            9   1602.89
```

c. Realice una prueba global de hipótesis para determinar si uno o más coeficientes de regresión son diferentes de cero. Utilice el nivel de significancia 0.05.

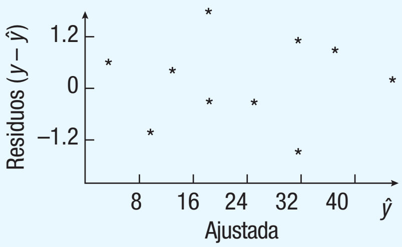
d. Realice una prueba de hipótesis en cada una de las variables independientes. ¿Consideraría eliminar "número de tiendas de descuento" y "número de supervisores"? Utilice el nivel de significancia 0.05.

e. Se vuelve a correr la regresión, pero ahora sin "número de tiendas de descuento" y "número de supervisores", como se muestra a continuación. Calcule el coeficiente de determinación. ¿Cuánto cambió R^2 a partir del análisis anterior?

```
La ecuación de regresión es
Ventas = −18.9 + 1.61 automóviles + 0.400 ingreso + 1.96 edad

          Predictor        Coef      SE Coef         T         P
          Constante     −18.924        3.636     −5.20     0.002
          Automóviles    1.6129       0.1979      8.15     0.000
          Ingreso       0.40031      0.01569     25.52     0.000
          Edad           1.9637       0.5846      3.36     0.015

Análisis de varianza
          FUENTE            GL        SS        MS         F         P
          Regresión          3   1593.66    531.22    345.25     0.000
          Error residual     6      9.23      1.54
          Total              9   1602.89
```

f. Abajo a la izquierda se presenta un histograma de los residuos. ¿Parece razonable la suposición de normalidad?

g. Abajo a la derecha se muestra una gráfica de los valores ajustados de y (es decir, $\hat{y}$) y de los residuos. ¿Observa alguna violación de las suposiciones?



19. El administrador de un nuevo programa para practicantes de leyes en Seagate Technical College desea estimar el promedio de calificaciones en el programa. Considera que el promedio de calificaciones en el bachillerato, la calificación en aptitudes verbales en el Examen de Aptitud Escolar (SAT) y la calificación en matemáticas en el SAT serían buenos factores de predicción de la calificación promedio en el programa. Los datos de nueve estudiantes son:

Estudiante	Promedio de calificaciones en el bachillerato	SAT verbal	SAT matemáticas	Promedio de calificaciones
1	3.25	480	410	3.21
2	1.80	290	270	1.68
3	2.89	420	410	3.58
4	3.81	500	600	3.92
5	3.13	500	490	3.00
6	2.81	430	460	2.82
7	2.20	320	490	1.65
8	2.14	530	480	2.30
9	2.63	469	440	2.33

a. Considere la matriz de correlación que aparece a la derecha. ¿Cuál es la variable que tiene la correlación más fuerte con la variable dependiente? Algunas correlaciones entre las variables independientes son fuertes. ¿Esto representaría un problema?

```
                 legal      SAT   verbal
SAT              0.911
Verbal           0.616    0.609
Matemáticas      0.487    0.636    0.599
```

b. Considere el siguiente resultado y calcule el coeficiente de determinación múltiple.

```
La ecuación de regresión es
Legal = −0.411 + 1.20 GPA + 0.00163 Verbal − 0.00194 Matemáticas

Predictor           Coef        SE Coef         T         P
Constante        −0.4111         0.7823     −0.53     0.622
SAT               1.2014         0.2955      4.07     0.010
Verbal          0.001629       0.002147      0.76     0.482
Matemáticas    −0.001939       0.002074     −0.94     0.393

Análisis de la varianza
FUENTE              GL         SS         MS         F        P
Regresión            3     4.3595     1.4532     10.33    0.014
Error residual       5     0.7036     0.1407
Total                8     5.0631

FUENTE        GL    Seq SS
SAT            1    4.2061
Verbal         1    0.0303
Matemáticas    1    0.1231
```

c. Realice una prueba global de hipótesis a partir del resultado anterior. ¿Alguno de los coeficientes de regresión es diferente de cero?

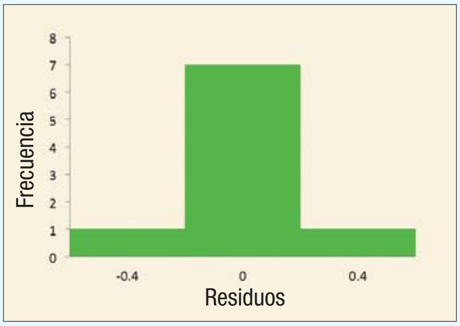
d. Realice una prueba de hipótesis de cada variable independiente. ¿Consideraría eliminar las variables “verbal” y “matemáticas”? Utilice un nivel $\alpha = 0.05$.

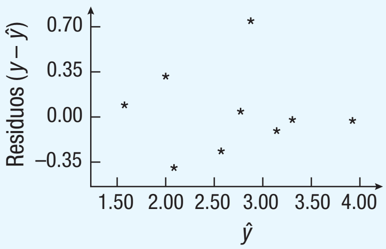
e. El análisis se vuelve a ejecutar, pero ahora sin “verbal” ni “matemáticas”. Observe la siguiente captura de pantalla y calcule el coeficiente de determinación. ¿Cuánto cambió R^2 a partir del análisis anterior?

```
La ecuación de regresión es
Legal = −0.454 + 1.16 SAT

Predictor       Coef     SE Coef        T        P
Constante    −0.4542      0.5542    −0.82    0.439
SAT           1.1589      0.1977     5.86    0.001

Análisis de la varianza
FUENTE            GL       SS        MS        F        P
Regresión          1   4.2061    4.2061    34.35    0.001
Error residual     7   0.8570    0.1224
Total              8   5.0631
```

f. Abajo, en la parte izquierda se presenta un histograma de los residuos. ¿Parece razonable la suposición de normalidad en el caso de los residuos?

g. En la gráfica de la derecha se presentan los valores de los residuos y los de $\hat{y}$. ¿Observa alguna violación de las suposiciones?





20. Mike Wilde es el presidente del sindicato de maestros del Otsego School District. A fin de prepararse para negociaciones próximas, le gustaría investigar la estructura de los salarios de los maestros del distrito. Wilde considera que hay tres factores que influyen en el salario de un profesor: sus años de experiencia, la calificación de su eficiencia como docente por parte del director y si cuenta con un posgrado. Mediante una muestra de 20 maestros se generaron los datos que aparecen en la página siguiente.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

a. Formule una matriz de correlación. ¿Qué variable independiente tiene la correlación más fuerte con la variable dependiente? ¿Habrá problemas respecto de la multicolinealidad?

b. Determine la ecuación de regresión. ¿Qué salario estimaría para un maestro con cinco años de experiencia, una calificación del director de 60 y que no tenga posgrado?

Salario (miles de dólares), y	Años de experiencia, x_1	Calificación del director, x_2	Posgrado,* x_3
31.1	8	35	0
33.6	5	43	0
29.3	2	51	1
⋮	⋮	⋮	⋮
30.7	4	62	0
32.8	2	80	1
42.8	8	72	0

* 1 = sí, 0 = no.

c. Realice una prueba global de la hipótesis para determinar si alguno de los coeficientes de regresión difiere de cero. Utilice el nivel de significancia 0.05.
d. Realice la prueba de hipótesis de los coeficientes de regresión individuales. ¿Consideraría eliminar alguna de las variables independientes? Utilice el nivel de significancia 0.05.
e. Si su conclusión en el punto anterior fue eliminar una o más variables independientes, realice de nuevo el análisis sin ellas.
f. Determine los residuos de la ecuación del punto anterior. Utilice un diagrama de tallo y hojas o un histograma para verificar que la distribución de los residuos sea aproximadamente normal.
g. Trace los residuos calculados en el punto anterior en un diagrama de dispersión con las varianzas residuales en el eje Y y los valores $\hat{y}$ en el eje X. ¿En la gráfica se revela alguna violación de las suposiciones de regresión?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

21. Mediante un análisis de consumidor se recabaron los siguientes datos sobre los tamaños de pantalla de los televisores más populares vendidos recientemente en una gran tienda minorista:

Fabricante	Pantalla	Precio	Fabricante	Pantalla	Precio
Sharp	46	$1 473.00	Sharp	37	$1 314.50
Samsung	52	2 300.00	Sharp	32	853.50
Samsung	46	1 790.00	Sharp	52	2 778.00
Sony	40	1 250.00	Samsung	40	1 749.50
Sharp	42	1 546.50	Sharp	32	1 035.00
Samsung	46	1 922.50	Samsung	52	2 950.00
Samsung	40	1 372.00	Sony	40	1 908.50
Sharp	37	1 149.50	Sony	52	3 103.00
Sharp	46	2 000.00	Sony	46	2 606.00
Sony	40	1 444.50	Sony	46	2 861.00
Sony	52	2 615.00	Sony	52	3 434.00
Samsung	32	747.50			

a. ¿Parece haber una relación lineal entre el tamaño de la pantalla y el precio?
b. ¿Cuál es la variable dependiente?
c. Utilizando software estadístico, determine la ecuación de regresión e interprete el valor de la pendiente en la ecuación de regresión.
d. Incluya al fabricante en un análisis de regresión lineal múltiple empleando una variable "ficticia". ¿Parece que algunos fabricantes pueden establecer un precio especial? *Sugerencia:* utilice un grupo de variables indicadoras.
e. Pruebe cada uno de los coeficientes individuales para ver si son significativos.
f. Haga un trazo de los residuos y comente si parecen seguir una distribución normal.
g. Trace los residuos frente a los valores ajustados. ¿Aparentan la misma cantidad de variación?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

22. Una planeadora regional estudia los datos demográficos en un área de un estado en particular; se ha recabado la información que aparece en la página siguiente en nueve condados.
a. ¿Existe una relación lineal entre el ingreso mediano y la edad mediana?
b. ¿Cuál es la variable "dependiente"?
c. Utilice software estadístico para determinar la ecuación de regresión. Interprete el valor de la pendiente en la ecuación de regresión simple.
d. Incluya el aspecto de que el condado sea "costero" o no en un análisis de regresión lineal múltiple empleando una variable "ficticia". ¿Parece haber una influencia significativa de los ingresos?
e. Pruebe cada uno de los coeficientes individuales para ver si son significativos.

Condado	Ingreso mediano	Edad mediana	Costero
A	$48 157	57.7	1
B	48 568	60.7	1
C	46 816	47.9	1
D	34 876	38.4	0
E	35 478	42.8	0
F	34 465	35.4	0
G	35 026	39.5	0
H	38 599	65.6	0
J	33 315	27.0	0

f. Haga un trazo de los residuos y comente si parecen seguir una distribución normal.
g. Trace los residuos frente a los valores ajustados. ¿Parecen tener la misma cantidad de variación?

23. Great Plains Roofing and Siding Company, Inc., vende productos para techos y recubrimientos de paredes a minoristas en reparación de casas, como Lowe's y Home Depot, y a contratistas comerciales. El propietario desea estudiar los efectos de diversas variables sobre el valor de las tejas locales vendidas (en miles de dólares). El gerente de marketing argumenta que la compañía debe gastar más dinero en publicidad, en tanto que un investigador de mercado sugiere que se deben enfocar más en diferenciar su marca y producto de los de sus competidores.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

La compañía dividió a Estados Unidos en 26 distritos de comercialización; en cada uno reunió información acerca de algunas variables, como el volumen de ventas (en miles de dólares), el dinero gastado en publicidad, la cantidad de cuentas activas, la cantidad de marcas de competidores y una calificación del potencial del distrito.

Ventas (miles de dólares)	Dólares en publicidad (miles)	Número de cuentas	Número de competidores	Potencial de mercado
79.3	5.5	31	10	8
200.1	2.5	55	8	6
163.2	8.0	67	12	9
200.1	3.0	50	7	16
146.0	3.0	38	8	15
177.7	2.9	71	12	17
⋮	⋮	⋮	⋮	⋮
93.5	4.2	26	8	3
259.0	4.5	75	8	19
331.2	5.6	71	4	9

Realice un análisis de regresión múltiple para encontrar los mejores factores de predicción de las ventas.

a. Trace un diagrama de dispersión donde se compare el volumen de ventas con cada una de las variables independientes. Comente los resultados.
b. Formule una matriz de correlación. ¿Hay algún problema? ¿Hay alguna variable independiente redundante?
c. Formule una ecuación de regresión y realice una prueba global. ¿Se puede concluir que algunas de las variables independientes son útiles para explicar las modificaciones de la variable dependiente?
d. Realice una prueba con cada una de las variables independientes. ¿Hay alguna que se deba eliminar?
e. Refine la ecuación de regresión de modo que las variables restantes sean significativas.
f. Elabore un histograma de los residuos y una gráfica de probabilidad normal. ¿Hay algún problema?
g. Determine el factor de inflación de la varianza de cada una de las variables independientes. ¿Hay algún problema?

24. El *Times-Observer* es un periódico de Ciudad Metro. Al igual que muchos periódicos, este pasa por dificultades financieras. La gerente de circulación estudia otros periódicos en ciudades similares en Estados Unidos y Canadá, con interés particular en las variables que se relacionan con el número de suscriptores. Ella reúne la información muestral que aparece en la página siguiente de 25 periódicos de ciudades similares. Se emplea la siguiente notación:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Periódico	Sus	Pob	Pub	Ingreso
1	37.95	588.9	13.2	35.1
2	37.66	585.3	13.2	34.7
3	37.55	566.3	19.8	34.8
⋮	⋮	⋮	⋮	⋮
23	38.83	629.6	22.0	35.3
24	38.33	680.0	24.2	34.7
25	40.24	651.2	33.0	35.8

Sus = Número de suscriptores (en miles de personas).
Pob = Población metropolitana (en miles de personas).
Pub = Presupuesto en publicidad del periódico (miles de dólares).
Ingreso = Ingreso familiar medio en el área metropolitana (miles de dólares).

a. Determine la ecuación de regresión.
b. Realice una prueba global de hipótesis para determinar si algunos de los coeficientes de regresión son distintos de cero.
c. Realice la prueba de los coeficientes individuales. ¿Consideraría eliminar algunos de ellos?
d. Determine los residuos y trácelos contra los valores ajustados. ¿Hay problemas?
e. Elabore un histograma de las varianzas residuales. ¿Hay problemas con la suposición de normalidad?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

25. Fred G. Hire es el gerente de recursos humanos en Crescent Tool and Die, Inc. Como parte de su reporte anual para el presidente, debe presentar un análisis de los empleados asalariados. Como hay más de 1 000 empleados y no tiene personal para reunir información sobre cada uno de ellos, decide seleccionar una muestra aleatoria de 30; por cada empleado registra su salario mensual, los años de servicio en la compañía (en meses), el género (1 = masculino, 0 = femenino), y si ocupa un puesto técnico o administrativo. Los puestos técnicos se codifican 1, y los administrativos, 0.

Empleado muestreado	Salario mensual	Antigüedad en la compañía	Edad	Género	Puesto
1	$1 769	93	42	1	0
2	1 740	104	33	1	0
3	1 941	104	42	1	1
⋮	⋮	⋮	⋮	⋮	⋮
28	1 791	131	56	0	1
29	2 001	95	30	1	1
30	1 874	98	47	1	0

a. Determine la ecuación de regresión; use el salario como variable dependiente y las otras cuatro como variables independientes.
b. ¿Cuál es el valor de R^2? Haga un comentario sobre este valor.
c. Realice una prueba global de hipótesis para determinar si algunas de las variables independientes son diferentes de cero.
d. Realice una prueba individual de hipótesis para determinar si se pueden omitir algunas variables independientes.
e. Determine de nuevo la ecuación de regresión; use solo las variables independientes significativas. ¿Cuánto más gana al mes un hombre que una mujer? ¿Hay alguna diferencia si el empleado ocupa un puesto técnico o uno administrativo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

26. Muchas regiones a lo largo de la costa de Carolina del Norte, de Carolina del Sur y Georgia experimentaron un rápido crecimiento poblacional durante los últimos 10 años, y se espera que el desarrollo continúe durante los próximos 10. Esto ha motivado a muchas de las cadenas importantes de abarrotes a construir nuevas tiendas en la región. La cadena Kelly´s Super Grocery Stores, Inc., no es la excepción, y su director de planeación desea estudiar si es conveniente agregar más tiendas en esta región; él considera que hay dos factores principales que indican la cantidad monetaria que las familias gastan en abarrotes: el primero es su ingreso y el otro es el número de personas que las integran. El director reunió información muestral que aparece a la derecha.

Familia	Alimentos	Ingreso	Tamaño
1	$5.04	$ 73.98	4
2	4.08	54.90	2
3	5.76	94.14	4
⋮	⋮	⋮	⋮
23	4.56	38.16	3
24	5.40	43.74	7
25	4.80	48.42	5

Los alimentos y el ingreso se reportan en miles de dólares por año, y la variable "tamaño" se refiere al número de personas en el hogar.

a. Elabore una matriz de correlación. ¿Detecta algunos problemas de multicolinealidad?
b. Determine la ecuación de regresión. Haga un comentario sobre la ecuación de regresión. ¿Cuánto dinero agrega un miembro familiar adicional a la cantidad que se gasta en alimentos?
c. ¿Cuál es el valor de R^2? ¿Se puede concluir que este valor es mayor que cero?
d. ¿Consideraría eliminar algunas de las variables independientes?
e. Trace los residuos en un histograma. ¿Hay algún problema con la suposición de normalidad?

f. Trace los valores ajustados contra los residuos. ¿En esta gráfica se revelan problemas de homoscedasticidad?

27. Una asesora en inversiones estudia la relación entre un precio accionario común de la razón de ganancias (P/G) y los factores que considera que influirían en él, y para esto cuenta con la siguiente información sobre las ganancias por acción (GPA) y el porcentaje de dividendos (rendimiento) de una muestra de 20 acciones.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Acción	P/G	EPS	Rendimiento
1	20.79	$2.46	1.42
2	3.03	2.69	4.05
3	44.46	−0.28	4.16
⋮	⋮	⋮	⋮
18	30.21	1.71	3.07
19	32.88	0.35	2.21
20	15.19	5.02	3.50

a. Determine una ecuación de regresión lineal múltiple con P/G como variable dependiente.
b. ¿Estas variables independientes son un factor eficaz de predicción de P/G, ya sea de manera conjunta o individual?
c. Interprete los coeficientes de regresión.
d. ¿Alguna de estas acciones parece estar subvalorada de manera particular?
e. Trace los residuos y verifique la suposición de normalidad. Trace los valores ajustados contra los residuos.
f. ¿Parece haber problemas de homoscedasticidad?
g. Determine una matriz de correlación. ¿Alguna de las correlaciones indica multicolinealidad?

28. El Conch Café, ubicado en Gulf Shores, Alabama, ofrece almuerzos casuales con una gran vista al Golfo de México. Para adaptarse al aumento de la clientela durante la temporada vacacional de verano, Fuzzy Conch, el propietario, contrata a un gran número de meseros como ayuda temporal. A Fuzzy le gustaría contar con información sobre la cantidad monetaria en propinas que un mesero puede ganar para proporcionarla durante las entrevistas; él considera que la cantidad de la cuenta y el número de clientes se relacionan con el monto de la propina, y reunió la información que aparece a la derecha.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Cliente	Monto de la propina	Monto de la cuenta	Número de clientes
1	$7.00	$48.97	5
2	4.50	28.23	4
3	1.00	10.65	1
⋮	⋮	⋮	⋮
28	2.50	26.25	2
29	9.25	56.81	5
30	8.25	50.65	5

a. Desarrolle una ecuación de regresión múltiple con la cantidad monetaria en propinas como variable dependiente, y la cantidad monetaria de la cuenta y el número de clientes como variables independientes. Escriba la ecuación de regresión. ¿Cuánto dinero más agrega otro cliente a la cantidad de las propinas?
b. Realice una prueba global de hipótesis para determinar si al menos una de las variables independientes es significativa. ¿Cuál es su conclusión?
c. Realice la prueba individual para cada una de las variables. ¿Se debe eliminar alguna?
d. Utilice la ecuación elaborada en punto anterior para establecer el coeficiente de determinación e interprete su valor.
e. Trace los valores de los residuos. ¿Es razonable suponer que siguen la distribución normal?
f. Trace los valores residuales frente a los ajustados. ¿Es razonable concluir que son aleatorios?

29. El presidente de Blitz Sales Entreprises, una compañía que vende productos de cocina mediante comerciales en televisión, con frecuencia denominados *infomerciales*, reunió datos de las últimas 15 semanas de ventas para determinar la relación entre las ventas y el número de infomerciales.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Infomerciales	Ventas (miles de dólares)	Infomerciales	Ventas (miles de dólares)
20	3.2	22	2.5
15	2.6	15	2.4
25	3.4	25	3.0
10	1.8	16	2.7
18	2.2	12	2.0
18	2.4	20	2.6
15	2.4	25	2.8
12	1.5		

a. Determine la ecuación de regresión. ¿Es posible predecir las ventas a partir del número de comerciales?

b. Determine los residuos y trace un histograma. ¿Parece razonable la suposición de normalidad?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

30. La directora de actos especiales de Sun City consideraba que la cantidad de dinero que se gasta en juegos pirotécnicos el 4 de julio (día de la independencia de Estados Unidos) era un factor de predicción de la asistencia al Festival de Otoño, en octubre, por lo que reunió la siguiente información para probar su hipótesis.

4 de julio (miles de dólares)	Festival de otoño (miles)	4 de julio (miles de dólares)	Festival de otoño (miles)
10.6	8.8	9.0	9.5
8.5	6.4	10.0	9.8
12.5	10.8	7.5	6.6
9.0	10.2	10.0	10.1
5.5	6.0	6.0	6.1
12.0	11.1	12.0	11.3
8.0	7.5	10.5	8.8
7.5	8.4		

Determine la ecuación de regresión. ¿Está relacionada la cantidad que se gasta en juegos pirotécnicos con la asistencia al festival? Evalúe los supuestos de regresión examinando los residuos.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

31. Usted es un empleado nuevo de Laurel Woods Real State, empresa que se especializa en la venta de casas hipotecadas por medio de subastas públicas. Su jefe le pidió aplicar los siguientes datos (saldo de la hipoteca, pagos mensuales, pagos hechos antes de la hipoteca y precio final en la subasta) a una muestra aleatoria de ventas recientes con el fin de estimar el precio real de la subasta.

Préstamo	Pagos mensuales	Pagos hechos	Precio en la subasta
$ 85 600	$ 985.87	1	$16 900
115 300	902.56	33	75 800
103 100	736.28	6	43 900
⋮	⋮	⋮	⋮
119 400	1 021.23	58	69 000
90 600	836.46	3	35 600
104 500	1 056.37	22	63 000

a. Realice la prueba global de hipótesis para verificar si algunos de los coeficientes de regresión son diferentes de cero.

b. Realice la prueba individual de las variables independientes. ¿Eliminaría alguna?

c. Si parece que una o más de las variables independientes no son necesarias, elimínela y resuelva la nueva ecuación de regresión.

32. Considere las cifras del ejercicio anterior y agregue una variable nueva que describa la interacción potencial entre la cantidad del préstamo y el número de pagos hechos; después, haga una prueba de hipótesis para verificar si la interacción es significativa.

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

33. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en Goodyear, Arizona, el año anterior. Utilice el precio de venta de la casa como variable dependiente y determine la ecuación de regresión con el número de recámaras, tamaño de la casa, si tiene alberca, si tiene garaje, distancia desde el centro de la ciudad, y el número de baños como variables independientes.

a. Utilice un software estadístico para determinar la ecuación de regresión y analice cada una de las variables; por ejemplo, ¿le sorprende que el coeficiente de regresión de la distancia desde el centro de la ciudad sea negativo? ¿Cuánto agrega un garaje o una alberca al precio de una casa?

b. Determine el valor de la intersección.

c. Desarrolle una matriz de correlación. ¿Cuáles variables independientes tienen correlaciones fuertes o débiles con la variable dependiente? ¿Detecta algunos problemas de multicolinealidad?
d. Realice la prueba global en el conjunto de variables independientes e interprétela.
e. Realice la prueba de hipótesis de cada una de las variables independientes. ¿Consideraría eliminar algunas de ellas? Si es así, ¿cuáles?
f. Efectúe de nuevo el análisis hasta que solo tenga coeficientes de regresión significativos e identifique las variables.
g. Elabore un histograma o un diagrama de tallo y hojas de los residuos a partir de la ecuación de regresión final desarrollada en el punto anterior. ¿Es razonable concluir que se cumplió la suposición de normalidad?
h. Trace los residuos contra los valores ajustados a partir de la ecuación de regresión final desarrollada en el punto f contra los valores ajustados de *y*. Trace los residuos en el eje vertical, y los valores ajustados, en el eje horizontal.

34. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Sea "número de juegos ganados" la variable dependiente, y "promedio de bateo del equipo", "número de bases robadas", "número de errores cometidos", "promedio de carreras del equipo", "número de cuadrangulares" y "el equipo juega en la Liga Nacional o en la Americana" las variables independientes.
a. Utilice un software estadístico para determinar la ecuación de regresión. Comente sobre cada una de las variables; por ejemplo, ¿le sorprende que el coeficiente de regresión del promedio de carreras sea negativo? ¿El número de victorias se ve afectado si el equipo juega en la Liga Nacional o en la Americana?
b. Encuentre el coeficiente de determinación de este grupo de variables independientes.
c. Formule una matriz de correlación. ¿Cuáles son las variables independientes que tienen correlaciones fuertes o débiles con la variable dependiente? ¿Detecta algunos problemas de multicolinealidad?
d. Realice una prueba global en el conjunto de variables independientes e interprétela.
e. Realice una prueba de hipótesis en cada una de las variables independientes. ¿Consideraría eliminar algunas de ellas? Si es así, ¿cuáles?
f. Vuelva a efectuar el análisis hasta que solo permanezcan coeficientes de regresión netos significativos e identifique las variables.
g. Elabore un histograma o un diagrama de tallo y hojas de los residuos a partir de la ecuación de regresión final desarrollada en el punto f. ¿Es razonable concluir que se cumplió la suposición de normalidad?
h. Trace los residuos contra los valores ajustados a partir de la ecuación de regresión final desarrollada en el punto f. Trace los residuos en el eje vertical, y los valores ajustados, en el eje horizontal.

35. Consulte los datos de los autobuses del Distrito Escolar Buena. Primero, añada una variable para cambiar el tipo de autobús (diesel o gasolina) a una variable cualitativa. Si el tipo de autobús es diesel, establezca la variable cualitativa 0; si es de gasolina, 1. Desarrolle una ecuación de regresión mediante un software estadístico, con "mantenimiento" como variable dependiente y "edad", "millas" y "tipo de autobús" como variables independientes.
a. Escriba la ecuación del análisis de regresión múltiple y comente cada variable.
b. Determine e interprete el valor R^2.
c. Elabore una matriz de correlación. ¿Qué variables independientes tienen correlaciones fuertes o débiles con la variable dependiente? ¿Detecta algunos problemas de multicolinealidad?
d. Realice una prueba global de hipótesis en el conjunto de variables independientes e interprete sus resultados.
e. Realice una prueba de hipótesis con cada una de las variables independientes. ¿Consideraría eliminar algunas de ellas? Si es así, ¿cuáles?
f. Realice de nuevo el análisis hasta que solo queden los coeficientes de regresión significativos e identifique estas variables.
g. Elabore un histograma o un diagrama de tallo y hojas de los residuos a partir de la ecuación de regresión final desarrollada en el punto f. ¿Es razonable concluir que se cumplió la suposición de normalidad?
h. Trace los residuos contra los valores ajustados a partir de la ecuación de regresión final. Trace los residuos en el eje vertical, y los valores ajustados, en el eje horizontal.

REPASO DE LOS CAPÍTULOS 13 y 14

En esta sección se repasan los conceptos y términos más importantes que se presentaron en los capítulos 13 y 14. En el capítulo 13 se indicó que la fuerza de la relación entre la variable independiente y la dependiente se mide con el *coeficiente de correlación*; el cual se designa con la letra r, y adopta cualquier valor entre -1.00 y $+1.00$ inclusive. Los coeficientes de -1.00 y $+1.00$ indican una relación perfecta, y un 0 indica que no hay relación. Un valor cercano a 0, como -0.14 o 0.14, indica una relación débil. Una valor cercano a -1 o $+1$, como -0.90 o $+0.90$, indica una relación fuerte. Al elevar r al cuadrado se obtiene el *coeficiente de determinación*, designado r^2, el cual indica la proporción de la variación total en la variable dependiente explicada por la variable independiente.

De igual forma, la fuerza de la relación entre diversas variables independientes y una variable dependiente se mide por el *coeficiente de determinación múltiple*, R^2, que mide la proporción de la variación en y explicada por dos o más variables independientes.

La relación lineal en el caso simple que implica una variable independiente y una variable dependiente se describe por la ecuación $\hat{y} = a + bx$. En el caso de k variables independientes, x_1, x_2 y x_3, la misma ecuación de regresión múltiple es la siguiente:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

Despejar $b_1, b_2, b_3, \ldots, b_n$ implicaría cálculos muy tediosos; por fortuna, este problema se resuelve pronto con uno de los muchos paquetes de software estadístico y paquetes de hojas de cálculo. En los resultados de la mayoría de los programas de software se reportan varias mediciones, como el coeficiente de determinación, el error estándar de estimación múltiple, los resultados de la prueba global y la prueba de las variables individuales.

PROBLEMAS

1. El Departamento de Contabilidad de Crate and Barrel desea estimar la ganancia de cada una de las muchas tiendas de la cadena con base en el número de empleados de estas, los costos generales, los márgenes de ganancia promedio y las pérdidas por robo. Algunos estadísticos de las tiendas son:

Tienda	Ganancias netas (miles de dólares)	Número de empleados	Costo general (miles de dólares)	Margen de ganancia promedio (porcentaje)	Promedio por robo (miles de dólares)
1	$846	143	$79	69%	$52
2	513	110	64	50	45

a. La variable dependiente es _________.
b. La ecuación general de este problema es _________.
c. La ecuación de regresión múltiple se calculó $\hat{y} = 67 + 8x_1 - 10x_2 + 0.004x_3 - 3x_4$. ¿Cuáles son las ventas anticipadas de una tienda con 112 empleados, un costo general de 65 000 dólares, una tasa del margen de ganancia de 50% y pérdidas por robo de 50 000 dólares?
d. Suponga que R^2 se calculó en 0.86. Explique este valor.
e. Suponga que el error estándar de estimación múltiple fue 3 (en miles de dólares). Explique qué significa este valor en este problema.

2. Las compañías de impresión rápida en un área comercial importante del centro gastan la mayoría de su dinero en publicidad en anuncios en las bancas de espera del autobús. Un proyecto de investigación implica predecir las ventas mensuales con base en la cantidad anual que gastan en dichos anuncios. Una muestra de compañías de impresión rápida reveló que se tienen los siguientes gastos en publicidad y ventas:

Compañía	Publicidad anual en bancas de autobuses (miles de dólares)	Ventas mensuales (miles de dólares)
A	2	10
B	4	40
C	5	30
D	7	50
E	3	20

a. Trace el diagrama de dispersión.
b. Determine el coeficiente de correlación.
c. ¿Cuál es el coeficiente de determinación?

d. Calcule la ecuación de regresión.
e. Estime las ventas mensuales de una compañía de impresión rápida que gasta 4 500 dólares en publicidad en bancas de autobuses.
f. Resuma sus resultados.

3. Se proporciona en la segunda columna la siguiente salida de una tabla ANOVA:
a. Calcule el coeficiente de determinación.
b. Calcule el error estándar de estimación múltiple.
c. Realice una prueba de hipótesis para determinar si algunos de los coeficientes de regresión son diferentes de cero.
d. Realice una prueba de hipótesis de los coeficientes de regresión individuales. ¿Se puede eliminar alguna de las variables?

Análisis de la varianza			
Fuente	GL	SS	MS
Regresión	4	1050.8	262.70
Error residual	20	83.8	4.19
Total	24	1134.6	
Factor de predicción	Coeficiente	Error estándar del coeficiente	t
Constante	70.05	2.13	32.00
x_1	0.42	0.17	2.47
x_2	0.27	0.21	1.29
x_3	0.75	0.30	2.50
x_4	0.42	0.07	6.00

CASOS

A. El Century National Bank

Consulte los datos del Century National Bank. Utilice el saldo de cuentas de cheques como variable dependiente y, como variables independientes, el número de transacciones en cajeros automáticos, el número de otros servicios empleados, si el individuo tiene tarjeta de crédito y si se paga interés en la cuenta en particular; indique en un reporte cuáles son las variables que parecen relacionarse con los saldos de las cuentas y si explican bien la variación de estos. ¿Se deben emplear todas las variables propuestas en el análisis, o se pueden eliminar algunas?

B. Terry and Associates: tiempo para entregar equipos médicos

Terry and Associates es un centro especializado en pruebas médicas de Denver, Colorado. Una de las fuentes principales de ingresos de la compañía es un equipo para detectar cantidades elevadas de plomo en la sangre. Los trabajadores en talleres de hojalatería de autos, en la industria de jardinería y los pintores comerciales de casas están expuestos a grandes cantidades de esta sustancia y, por lo tanto, se deben someter a una prueba de forma aleatoria; pero es muy costoso realizarla, por lo que los equipos se suministran por pedido a diversos lugares del área de Denver.

Kathleen Terry, la propietaria, tiene interés en determinar los costos adecuados por entrega. Para investigar esto, ella reunió información sobre una muestra aleatoria de 50 entregas recientes. Los factores que se consideran relacionados con el costo de entrega de un equipo son:

Preparación — El tiempo en minutos desde la recepción del pedido por teléfono y cuando el equipo está listo para su entrega.
Entrega — El tiempo de recorrido real en minutos desde la planta de Terry hasta el cliente.
Millas — La distancia en millas desde la planta de Terry hasta el cliente.

Número de muestra	Costo	Preparación	Entrega	Millas	Número de muestra	Costo	Preparación	Entrega	Millas
1	$32.60	10	51	20	20	25.22	6	41	14
2	23.37	11	33	12	21	24.29	3	28	13
3	31.49	6	47	19	22	22.76	4	26	10
4	19.31	9	18	8	23	28.17	9	54	16
5	28.35	8	88	17	24	19.68	7	18	8
6	$22.63	9	20	11	25	25.15	6	50	13
7	22.63	9	39	11	26	20.36	9	19	7
8	21.53	10	23	10	27	21.16	3	19	8
9	21.16	13	20	8	28	25.95	10	45	14
10	21.53	10	32	10	29	18.76	12	12	5
11	$28.17	5	35	16	30	18.76	8	16	5
12	20.42	7	23	9	31	$24.29	7	35	13
13	21.53	9	21	10	32	19.56	2	12	6
14	27.55	7	37	16	33	22.63	8	30	11
15	23.37	9	25	12	34	21.16	5	13	8
16	17.10	15	15	6	35	21.16	11	20	8
17	27.06	13	34	15	36	19.68	5	19	8
18	15.99	8	13	4	37	18.76	5	14	7
19	17.96	12	12	4	38	17.96	5	11	4

(*continúa*)

(*continuación*)

Número de muestra	Costo	Preparación	Entrega	Millas
39	23.37	10	25	12
40	25.22	6	32	14
41	27.06	8	44	16
42	21.96	9	28	9
43	22.63	8	31	11
44	19.68	7	19	8
45	22.76	8	28	10
46	21.96	13	18	9
47	25.95	10	32	14
48	26.14	8	44	15
49	24.29	8	34	13
50	24.35	3	33	12

1. Formule la ecuación de regresión lineal múltiple que describa la relación entre "costo de entrega" y las demás variables. ¿Estas tres variables explican a grado razonable los cambios en la variable dependiente? Estime el costo de entrega de un equipo cuya preparación tarda 10 minutos; su entrega, 30 minutos, y debe recorrer una distancia de 14 millas.
2. Haga una prueba para determinar que al menos un coeficiente de regresión neto difiere de cero. Asimismo, pruebe si algunas variables se pueden omitir del análisis; en tal caso, efectúe de nuevo la ecuación de regresión hasta que solo se incluyan variables significativas. Interprete en un reporte breve la ecuación de regresión final.
3. Escriba un breve reporte en el cual interprete la ecuación de regresión final.

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. En un diagrama de dispersión, ¿en qué eje se registra siempre la variable dependiente? 1. ________________
2. ¿Qué nivel de medición se requiere para calcular el coeficiente de correlación? 2. ________________
3. Si no existe correlación entre dos variables, ¿cuál es el valor del coeficiente de correlación? 3. ________________
4. ¿En cuál de los siguientes valores se indica la correlación más fuerte entre dos variables (0.65, −0.77, 0, −45)? 4. ________________
5. ¿Bajo qué condiciones asumirá el coeficiente de determinación un valor mayor a 1? 5. ________________

Considere la siguiente ecuación de regresión $\hat{Y} = 7 - 0.5X$, y que el coeficiente de determinación es 0.81 para contestar las preguntas 7, 8 y 9.

6. ¿En qué punto cruza la ecuación de regresión el eje Y? 6. ________________
7. ¿Un aumento de una unidad en la variable independiente resultará en qué cantidad de incremento o disminución de la variable independiente? 7. ________________
8. ¿Cuál es el coeficiente de correlación (cuidado con el signo)? 8. ________________
9. Si todos los puntos de un diagrama de dispersión se hallan en la recta de regresión, ¿cuál es el valor del error estándar de estimación? 9. ________________
10. En una ecuación de regresión múltiple, ¿cuál es el máximo número permitido de variables independientes (2, 10, 30, ilimitado)? 10. ________________
11. En un análisis de regresión múltiple, ¿qué tipo de relación supuesta existe entre la variable independiente y el grupo de variables independientes (lineal, múltiple, curva, ninguna de las anteriores)? 11. ________________
12. La diferencia entre Y y $\hat{Y}$ se denomina __________. 12. ________________
13. ¿Cuántos resultados diferentes son posibles para una variable ficticia en particular, como el género? 13. ________________
14. ¿Cuál es el nombre que se da a una tabla en la que se muestran todos los posibles coeficientes de relación entre la variable dependiente y todas las variables independientes, y entre todas estas? 14. ________________
15. Si existe una relación lineal entre la variable dependiente y el grupo de variables independientes, ¿qué tipo de gráfica de residuos se mostrará en el tipo de distribución? 15. ________________

Parte 2: Problemas

1. Dado el siguiente resultado:

Análisis de regresión

Tabla ANOVA

Fuente	SS	*gl*	MS	*F*	Valor *p*
Regresión	129.7275	1	129.7275	14.50	0.0007
Residuo	250.4391	28	8.9443		
Total	380.1667	29			

(*continúa*)

(*continuación*)

Salida de regresión			
Variables	**Coeficientes**	**Error estándar**	**t (gl = 28)**
Intersección	90.6190	1.5322	59.141
Pendiente	−0.9401	0.2468	−3.808

a. ¿De qué tamaño es la muestra?
b. Escriba la ecuación de regresión e interprete los valores de la pendiente y de la intersección.
c. Si el valor de la variable independiente es 10, ¿cuál es el valor de la variable dependiente?
d. Calcule el coeficiente de determinación e interprete su valor.
e. Calcule el coeficiente de correlación y realice una prueba de hipótesis para determinar si existe una asociación negativa significativa entre las variables.

2. Dado el siguiente resultado:

Análisis de regresión

Tabla ANOVA					
Fuente	**SS**	***gl***	**MS**	***F***	**Valor *p***
Regresión	227.0928	4	56.7732	9.27	0.000
Residual	153.0739	25	6.1230		
Total	380.1667	29			

Salida de regresión				
Variables	**Coeficientes**	**Error estándar**	**t (gl = 25)**	**Valor *p***
Intersección	68.3366	8.9752	7.614	0.000
x_1	0.8595	0.3087	2.784	0.010
x_2	−0.3380	0.8381	−0.403	0.690
x_3	−0.8179	0.2749	−2.975	0.006
x_4	−0.5824	0.2541	−2.292	0.030

a. ¿De qué tamaño es la muestra?
b. ¿Cuántas variables independientes hay en el estudio?
c. Determine el coeficiente de determinación.
d. Realice una prueba global de la hipótesis. ¿Puede concluir que al menos una de las variables independientes es distinta de cero? Utilice el nivel de significancia 0.01.
e. Realice la prueba de hipótesis individual a cada una de las variables independientes. ¿Consideraría retirar alguna de ellas? Si es así, ¿cuál variable o cuáles variables eliminaría? Utilice el nivel de significancia 0.01.

15 Métodos no paramétricos:

PRUEBAS DE NIVEL NOMINAL

DURANTE MUCHOS AÑOS, los ejecutivos de televisión dieron crédito a la pauta de que 30% de la audiencia veía cada una de las cadenas televisivas de mayor audiencia, y 10%, canales de televisión por cable durante una noche a la semana. Una muestra aleatoria de 500 televidentes del área de Tampa-St. Petersburg, Florida, el pasado lunes por la noche, reveló que 165 hogares sintonizaron ABC; 140, CBS; 125, NBC, y el resto vio un canal de televisión por cable. Con el nivel de significancia 0.05, ¿es posible concluir que la pauta aún es razonable? (vea el ejercicio 24 y el OA15-3).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA15-1 Probar una hipótesis acerca de una proporción poblacional.

OA15-2 Probar una hipótesis acerca de dos proporciones poblacionales.

OA15-3 Realizar una prueba de hipótesis para comparar un grupo de frecuencias observadas con una distribución esperada de frecuencias.

OA15-4 Explicar las limitaciones de utilizar el estadístico *ji* cuadrada en pruebas de bondad de ajuste.

OA15-5 Probar una hipótesis de que una distribución observada de frecuencias tiene una distribución normal.

OA15-6 Realizar una prueba de *ji* cuadrada de la independencia en una tabla de contingencia.

Introducción

En los capítulos 9 a 12 se describieron las pruebas de hipótesis para datos con escala de intervalo o de razón; en los ejemplos respectivos se incluyen las puntuaciones de un examen de estadística, los ingresos de funcionarios ejecutivos corporativos en las compañías tecnológicas, o los años de empleo de los trabajadores de producción en la planta de BMW en Greer, Carolina del Sur.

Se realizaron pruebas de hipótesis con respecto a una sola media poblacional (capítulo 10), dos medias poblacionales (capítulo 11) y tres o más (capítulo 12). Para estas pruebas se utilizaron datos de intervalo o de razón y se asumió que las poblaciones seguían la distribución de probabilidad normal; sin embargo, hay pruebas de hipótesis que no requieren ningún supuesto con respecto a la forma de la población, por lo tanto, no es necesario suponer una población normal. Estas se denominan *pruebas de hipótesis no paramétricas*.

En este capítulo se comienza con pruebas de hipótesis para datos de escala nominal; recuerde que estos se clasifican simplemente en categorías mutuamente excluyentes. En las primeras dos secciones de este capítulo se analizan las pruebas de proporciones, en las cuales se clasifica a los individuos o a los objetos en uno o dos grupos mutuamente excluyentes; por ejemplo, el género (hombre o mujer), la calidad (aceptable o inaceptable), la presencia de diabetes (sí o no) y las llegadas de las líneas aéreas (a tiempo o demoradas).

También se amplían las pruebas de escala nominal para incluir situaciones en donde los datos se clasifican en varias categorías mutuamente excluyentes aunque la escala siga siendo de ese tipo; por ejemplo, los colores de los dulces M&M (rojo, verde, azul, amarillo, anaranjado y café), la marca comprada de mantequilla de maní (Peter Pan, Jif, Skippy y otras), o los días de la semana laboral (lunes, martes, miércoles, jueves y viernes). Se introduce la distribución *ji* cuadrada como un nuevo estadístico de prueba, la cual se utiliza principalmente cuando hay más de dos categorías de escala nominal.

Probar una hipótesis de una proporción de una población

OA15-1
Probar una hipótesis acerca de una proporción poblacional.

Los intervalos de confianza para proporciones se comenzaron a analizar a partir de la sección "Intervalo de confianza de una proporción" del capítulo 9. También se puede llevar a cabo una prueba de hipótesis de una proporción; recuerde que una proporción es la razón entre el número de éxitos y el de observaciones. Si X se refiere al número de éxitos y n, al de observaciones, la proporción de éxitos en una cantidad fija de ensayos es X/n; por consiguiente, la fórmula para calcular una proporción muestral, p, es $p = X/n$. Considere los siguientes casos de posibles pruebas de hipótesis.

- Históricamente, General Motors informa que 70% de los vehículos arrendados se devuelve con menos de 36 000 millas recorridas. Una muestra reciente de 200 vehículos devueltos al final de su periodo de arrendamiento mostró que 158 tenían menos de dicha cantidad de millas. ¿Se incrementó la proporción?
- La American Association of Retired Persons (AARP) informa que 60% de los retirados menores de 65 años de edad regresaría a trabajar de tiempo completo si hubiera disponible un trabajo adecuado. Una muestra de 500 retirados menores de dicha edad reveló que 315 volverían a trabajar. ¿Puede concluir que más de 60% lo haría?
- Able Moving and Storage, Inc., anuncia a sus clientes que el traslado a largas distancias de los bienes familiares se entregarán de tres a cinco días a partir del momento de recogerlos. Los registros de Able muestran que han tenido éxito 90% de las veces. Una auditoría reciente mostró que de 200 veces, 190 tuvieron éxito. ¿Puede concluir la compañía que aumentó este registro de éxitos?

Se deben hacer algunas suposiciones antes de probar una proporción de población. Para probar una hipótesis relativa a una proporción de población, se elige una muestra aleatoria de esta. Se supone que se satisfacen los supuestos binomiales del capítulo 6: 1) los datos de la muestra que se recogen son resultado de conteos, 2) el resultado de un experimento se clasifica en una de dos categorías mutuamente excluyentes ("éxito" o "fracaso"), 3) la probabilidad de un éxito es la misma para cada ensayo y 4) los ensayos son independientes, lo cual significa que el resultado de uno no influye en el resultado de los demás. Esta prueba es adecuada cuando $n\pi$ y $n(1 - \pi)$ son de al menos 5. El tamaño de la muestra es n, y la proporción poblacional es π. Se tiene la ventaja de que una distribución binomial puede aproximarse por medio de la distribución normal.

EJEMPLO

Un gobernador republicano de un estado del oeste de Estados Unidos piensa postularse para la reelección. Históricamente, para que un candidato republicano sea reelegido necesita ganar cuando menos 80% de los votos en la sección norte del estado. El gobernador contrata una organización de encuestas para sondear a los votantes de esa sección y determinar qué porcentaje de ellos votará por él. La compañía de encuestas entrevistará 2 000 votantes. Utilice un procedimiento para probar hipótesis y evaluar las probabilidades de que el gobernador se reelija.

SOLUCIÓN

Este caso, la reelección del gobernador satisface las condiciones binomiales.

- Solo hay dos posibles resultados; es decir, un votante entrevistado votará o no por él.
- La probabilidad de un éxito es la misma para cada ensayo. En este caso, la probabilidad de que cualquier votante entrevistado apoye la reelección es de 0.80.
- Los ensayos son independientes; esto significa, por ejemplo, que la probabilidad de que en la entrevista número 23 el elector apoye la reelección no resulta afectada por lo que se responda en las entrevistas número 24 o 52.
- Los datos de la muestra son el resultado de conteos; en este caso, es preciso contar a los votantes que apoyan la reelección en la muestra de 2 000.

Se puede utilizar la aproximación normal de la distribución binomial que se analizó en el capítulo 7, pues $n\pi$ y $n(1 - \pi)$ son superiores a 5. En este caso, $n = 2\,000$ y $\pi = 0.80$ (π es la proporción de votos en la parte norte del estado, u 80%, necesarios); por lo tanto, $n\pi = 2\,000(0.80) = 1\,600$ y $n(1 - \pi) = 2\,000(1 - 0.80) = 400$. Ambos, 1 600 y 400, son mayores que 5.

Paso 1: Se establecen las hipótesis nula y alternativa. La hipótesis nula, H_0, consiste en que la proporción de la población π es 0.80 o mayor. La hipótesis alternativa, H_1, es que la proporción es menor a 0.80. Desde un punto de vista práctico, al gobernador solo le interesa cuando la proporción es menor de 0.80. Si es igual o mayor que 0.80, no pondrá objeciones; es decir, los datos de la muestra indicarían que probablemente se le reelija. Estas hipótesis se escriben simbólicamente de la siguiente manera:

$$H_0\colon \pi \geq 0.80$$
$$H_1\colon \pi < 0.80$$

H_1 establece una dirección; por consiguiente, como se observó antes, la prueba es de una cola, en la que el signo de desigualdad apunta a la cola de la distribución que contiene la región de rechazo.

Paso 2: Se selecciona el nivel de significancia. El nivel de significancia es de 0.05. Esta es la probabilidad de rechazar una hipótesis verdadera.

Paso 3: Seleccione el estadístico de prueba. El estadístico adecuado es z, que se determina de la siguiente manera:

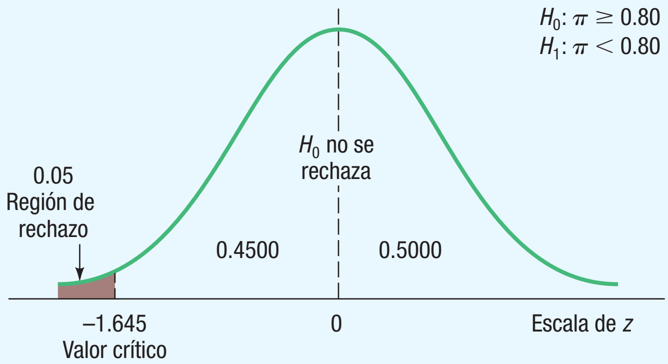
PRUEBA DE HIPÓTESIS DE UNA PROPORCIÓN

$$z = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \qquad \textbf{[15.1]}$$

donde:

π es la proporción poblacional;
p es la proporción muestral;
n es el tamaño de la muestra.

Paso 4: Se formula la regla de decisión. El valor o los valores críticos de z forman el punto o los puntos de división entre las regiones en las que se rechaza H_0 y en las que no. Como la hipótesis alternativa indica una dirección, se trata de una prueba de una cola. El signo de la desigualdad apunta hacia la izquierda, así que solo se utiliza el lado izquierdo de la curva (vea la gráfica 10.8). El nivel de significancia es 0.05. Esta probabilidad se encuentra en la cola izquierda y determina la región de rechazo. El área entre cero y el valor crítico es de 0.4500, que se calcula mediante 0.5000 − 0.0500. Refiérase al apéndice B.3, vaya a la columna que indica el valor de significancia 0.05 para una prueba de una cola, encuentre la fila con grados infinitos de libertad, y vea el valor crítico de z de 1.65; por lo tanto, la regla de decisión es: se rechaza la hipótesis nula y se acepta la hipótesis alternativa si el valor calculado de z cae a la izquierda de −1.645; de otra forma, H_0 no se rechaza.



GRÁFICA 15.1 Región de rechazo del nivel de significancia 0.05, prueba de una cola

Paso 5: **Se toma una decisión.** Seleccione una muestra y tome una decisión respecto de H_0. Un sondeo de muestra de 2 000 posibles electores en la parte norte del estado reveló que 1 550 pensaban votar por la reelección del gobernador. ¿Se encuentra la proporción muestral de 0.775 (calculada con la operación 1 550/2 000) lo bastante cerca de 0.80 para concluir que la diferencia se debe al error de muestreo? En este caso:

p tiene un valor de 0.775 y representa la proporción en la muestra que planea votar por el gobernador;
n tiene un valor de 2 000 y representa el número de votantes entrevistados;
π tiene un valor de 0.80 y representa la proporción de población hipotética;
z es un estadístico de prueba con una distribución normal cuando la hipótesis es verdadera y los demás supuestos son verdaderos.

Con la fórmula [15.1] se calcula el valor de z:

$$z = \frac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}} = \frac{\dfrac{1\,550}{2\,000} - 0.80}{\sqrt{\dfrac{0.80(1-0.80)}{2\,000}}} = \frac{0.775 - 0.80)}{\sqrt{0.00008}} = -2.80$$

El valor calculado de z (−2.80) es menor que el valor crítico, por lo que la hipótesis nula se rechaza en el nivel 0.05. La diferencia de 2.5 puntos porcentuales entre el porcentaje de la muestra (77.5%) y el de la población hipotética en la parte norte del estado que se requiere para ganar las elecciones estatales (80%) resulta estadísticamente significativa. Según el apéndice B-3, la probabilidad de hallar un valor z entre cero y −2.80 es de 0.4974. Así que el valor p es 0.0026, que se determina con el cálculo de 0.5000 − 0.4974. Como el valor p es inferior al nivel de significancia, se rechaza la hipótesis nula.

Paso 6: **Se interpreta el resultado.** El gobernador puede concluir que no cuenta con el apoyo necesario en la sección norte del país para ganar la reelección; dicho de otra forma, en este punto la evidencia no sustenta la afirmación de que el gobernador en funciones regresará a la mansión de gobierno por otros cuatro años.

AUTOEVALUACIÓN 15-1

Un informe reciente de la industria de seguros indicó que 40% de las personas implicadas en accidentes de tránsito menores había tenido por lo menos un accidente los últimos cinco años. Un grupo de asesoría decidió investigar dicha afirmación, pues creía que la cantidad era muy grande. Una muestra de 200 accidentes de tránsito de este año mostró que 74 personas también estuvieron involucradas en otro accidente los últimos cinco años. Utilice el nivel de significancia 0.01.

(a) ¿Se puede emplear z como estadístico de prueba? Indique la razón.
(b) Formule la hipótesis nula y la hipótesis alternativa.
(c) Muestre gráficamente la regla de decisión.
(d) Calcule el valor z y plantee su decisión respecto de la hipótesis nula.
(e) Determine e interprete el valor p.

EJERCICIOS

1. Considere las siguientes hipótesis:

$$H_0: \pi \leq 0.70$$
$$H_1: \pi > 0.70$$

Una muestra de 100 observaciones reveló que $p = 0.75$. ¿Puede rechazar la hipótesis nula con el nivel de significancia 0.05?
   a. Formule la regla de decisión.
   b. Calcule el valor del estadístico de prueba.
   c. ¿Cuál es su decisión respecto de la hipótesis nula?

2. Considere las siguientes hipótesis:

$$H_0: \pi = 0.40$$
$$H_1: \pi \neq 0.40$$

Una muestra de 120 observaciones reveló que $p = 0.30$. ¿Puede rechazar la hipótesis nula con el nivel de significancia 0.05?
   a. Formule la regla de decisión.
   b. Calcule el valor del estadístico de prueba.
   c. ¿Cuál es su decisión respecto de la hipótesis nula?

Nota: Se recomienda utilizar el procedimiento de los seis pasos de la prueba de hipótesis y resolver los siguientes problemas.

3. El National Safety Council informó que 52% de los conductores estadounidenses que viajan por autopista de cuota es de género masculino; una muestra de 300 automóviles que circularon el día de ayer por la autopista de Nueva Jersey reveló que a 170 los manejaban hombres. Con el nivel de significancia 0.01, ¿puede concluir que por la autopista de cuota de Nueva Jersey manejaba una proporción mayor de hombres que lo indicado por las estadísticas nacionales?
4. Un artículo reciente informó que solo hay un trabajo disponible por cada tres nuevos graduados de universidad; las principales razones fueron una sobrepoblación de graduados universitarios y una economía débil. Una encuesta de 200 recién graduados reveló que 80 estudiantes tenían trabajo. Con el nivel de significancia 0.01, ¿puede concluir que una proporción mayor de estudiantes de su escuela tiene empleo?
5. Chicken Delight afirma que 90% de sus pedidos se entrega en 10 minutos desde que se hace el pedido; una muestra de 100 pedidos mostró que 82 se entregaron en el tiempo prometido. Con el nivel de significancia 0.10, ¿puede concluir que menos de 90% de los pedidos se entregó en un lapso inferior a 10 minutos?
6. Una investigación de la Universidad de Toledo indica que 50% de los estudiantes cambia de área de estudios después del primer año; una muestra aleatoria de 100 estudiantes de la Facultad de Administración reveló que 48 habían cambiado de área de estudio después del primer año del programa de estudios. ¿Hubo una reducción significativa en la proporción de estudiantes que cambian de área el primer año en este programa? Realice una prueba con el nivel de significancia 0.05.

OA15-2
Probar una hipótesis acerca de dos proporciones poblacionales.

Prueba de proporciones de dos muestras

En la sección anterior se consideró una prueba de medias poblacionales; sin embargo, con frecuencia también se tiene interés en saber si dos proporciones de muestras provienen de poblaciones iguales. A continuación se presentan algunos ejemplos.

- El vicepresidente de recursos humanos desea saber si hay alguna diferencia entre la proporción de empleados asalariados por hora que faltan más de 5 días de trabajo por año en las plantas de Atlanta y Houston.
- General Motors considera un diseño nuevo para el Chevy Malibú. El diseño se muestra a un grupo de compradores potenciales menores de 30 años de edad y a otro grupo de mayores de

60 años. La compañía quiere saber si hay alguna diferencia entre la proporción de los dos grupos a quienes les gusta el diseño nuevo.

- Un asesor de la industria de aerolíneas está investigando el miedo a volar entre los adultos; en específico, la compañía desea saber si hay alguna diferencia entre la proporción de hombres con respecto mujeres que temen viajar en avión.

En los casos anteriores, cada elemento o individuo muestreado se clasifica como "éxito" o "fracaso"; es decir, en el ejemplo del Chevy Malibú, cada comprador potencial se clasifica como "le gusta el diseño nuevo" o "no le gusta el diseño nuevo". Después, se compara la proporción del grupo de menores de 30 años de edad con la proporción del grupo de mayores de 60 años que indique el gusto por el diseño nuevo. ¿Las diferencias se deben a la casualidad? En este estudio no se obtiene ninguna medida, solo se clasifican los individuos u objetos.

Para realizar la prueba, suponga que la muestra es lo bastante grande para que la distribución normal sirva como una buena aproximación a la distribución binomial. El estadístico de prueba sigue la distribución normal estándar. El valor de z se calcula a partir de la fórmula siguiente:

PRUEBA DE PROPORCIONES DE DOS MUESTRAS

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_c(1-p_c)}{n_1} + \dfrac{p_c(1-p_c)}{n_2}}} \qquad \textbf{[15.2]}$$

donde:

n_1 es el número de observaciones en la primera muestra;
n_2 es el número de observaciones en la segunda muestra;
p_1 es la proporción en la primera muestra que posee la característica;
p_2 es la proporción en la segunda muestra que posee la característica;
p_c es la proporción conjunta que posee la característica en las muestras combinadas. Se denomina estimación conjunta de la proporción poblacional y se calcula a partir de la fórmula siguiente.

PROPORCIÓN CONJUNTA

$$p_c = \frac{x_1 + x_2}{n_1 + n_2} \qquad \textbf{[15.3]}$$

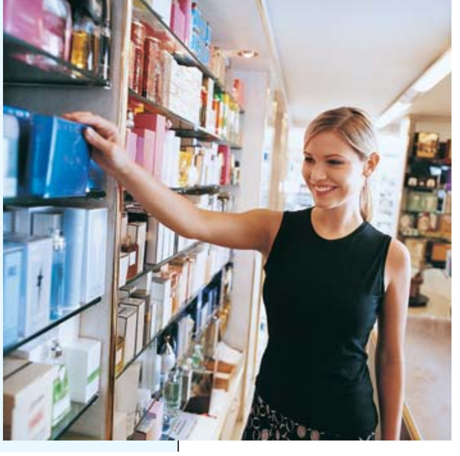
donde:

x_1 es el número que posee la característica en la primera muestra;
x_2 es el número que posee la característica en la segunda muestra.

En el ejemplo siguiente se ilustra la prueba de proporciones de dos muestras.

EJEMPLO

La compañía de perfumes Manelli desarrolló una fragancia nueva que planea comercializar con el nombre de Heavenly. Varios estudios de mercado indican que Heavenly tiene buen potencial de mercado; al departamento de ventas de Manelli le interesa saber si hay alguna diferencia entre las proporciones de mujeres jóvenes y mayores que comprarían el perfume si saliera al mercado. Hay dos poblaciones independientes, una de mujeres jóvenes y la otra, de mujeres mayores. A cada una de las mujeres muestreadas se le pedirá oler el perfume e indicar si le gusta lo suficiente para comprar un frasco.



SOLUCIÓN

Se utilizará el procedimiento usual de prueba de hipótesis de seis pasos.

Paso 1: **Formule H_0 y H_1.** En este caso, la hipótesis nula es: "No hay diferencia en la proporción de mujeres jóvenes y mayores que prefieren Heavenly". Designe a π_1 como la proporción de mujeres jóvenes que comprarían

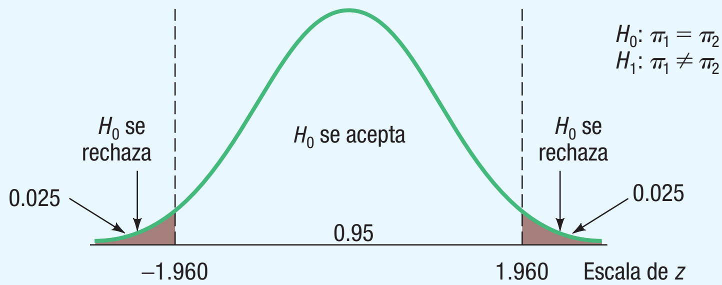
Heavenly y π_2 como la proporción de mujeres mayores que lo comprarían. La hipótesis alternativa es que las dos proporciones no son iguales.

$$H_0\colon \pi_1 = \pi_2$$
$$H_1\colon \pi_1 \neq \pi_2$$

Paso 2: **Seleccione el nivel de significancia.** En este ejemplo se elige el nivel de significancia 0.05.

Paso 3: **Determine el estadístico de prueba.** El estadístico de prueba sigue la distribución normal estándar; su valor se calcula a partir de la fórmula [15.2].

Paso 4: **Formule la regla de decisión.** Recuerde que la hipótesis alternativa del primer paso no indica una dirección, de modo que esta es una prueba de dos colas. Para determinar el valor crítico, vaya a la distribución *t* de Student (apéndice B.5). En los encabezados, encuentre la fila etiquetada **“Nivel de significancia para la prueba de dos colas”** y seleccione la columna para un alfa de 0.05. Vaya a la fila inferior con grados infinitos de libertad. El valor crítico es 1.960, así que los valores críticos son -1.960 y $+1.960$; como antes, si el valor calculado de z es menor a -1.960 o mayor a 1.960, se rechaza la hipótesis nula. Esta información se muestra en la gráfica 15.2.



GRÁFICA 15.2 Reglas de decisión de la prueba de la fragancia Heavenly, nivel de significancia 0.05

Paso 5: **Seleccione una muestra y tome una decisión.** Una muestra aleatoria de 100 mujeres jóvenes reveló que a 19 les gustó la fragancia Heavenly lo suficiente para comprarla; de manera similar, una muestra de 200 mujeres mayores reveló que a 62 les gustó la fragancia lo suficiente para comprarla. Se designa p_1 como el número de mujeres jóvenes y p_2 como el de mujeres mayores.

$$p_1 = \frac{x_1}{n_1} = \frac{19}{100} = 0.19 \qquad p_2 = \frac{x_2}{n_2} = \frac{62}{200} = 0.31$$

La pregunta de investigación es si la diferencia de 0.12 en las dos proporciones de ambas muestras se debe a la casualidad o si hay alguna diferencia entre las proporciones de mujeres jóvenes y mayores a quienes les gusta la fragancia Heavenly.

Después, se combinan o se conjuntan las proporciones de las muestras mediante la fórmula [15.3].

$$p_c = \frac{x_1 + x_2}{n_1 + n_2} = \frac{19 + 62}{100 + 200} = \frac{81}{300} = 0.27$$

Observe que la proporción conjunta se aproxima más a 0.31 que a 0.19 debido a que se muestrearon más mujeres mayores que jóvenes.

Con la fórmula [15.2] se determina el valor del estadístico de prueba.

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_c(1 - p_c)}{n_1} + \dfrac{p_c(1 - p_c)}{n_2}}} = \frac{0.19 - 0.31}{\sqrt{\dfrac{0.27(1 - 0.27)}{100} + \dfrac{0.27(1 - 0.27)}{100}}} = -0.207$$

El valor calculado de -2.207 se encuentra en el área de rechazo; es decir, está a la izquierda de -1.960; por lo tanto, rechace la hipótesis nula en el nivel de significancia 0.05. En otras palabras, se rechaza la hipótesis nula de que la proporción de mujeres jóvenes que comprarían la fragancia es igual a la de mujeres mayores que también la comprarían. Es improbable que la diferencia entre las proporciones de las muestras se deba a la casualidad.

Para determinar el valor p, se debe redondear el estadístico de prueba z de 2.207 a 2.21 para que se puedan usar las áreas de la tabla bajo la "Curva normal" en el apéndice B.3. En la tabla, encuentre la probabilidad de un valor z menor que -2.21 o mayor que 2.21. El valor z que corresponde a 2.21 es 0.4864; por ello, la probabilidad de determinar que el valor del estadístico de prueba sea menor que -2.21 o mayor que 2.21 es:

$$\text{valor } p = 2(0.5000 - 0.4864) = 2(0.0136) = 0.0272$$

El valor p de 0.0272 es menor que el nivel de significancia 0.05, por lo cual se debe rechazar la hipótesis nula.

Paso 6: Interprete el resultado. Los resultados de la prueba de hipótesis indican que las mujeres mayores y las jóvenes compran Heavenly en distintos rangos o proporciones.

El complemento de Excel, MegaStat, tiene un procedimiento para determinar el valor del estadístico de prueba y calcular el valor p. Observe que la salida de MegaStat incluye las dos proporciones muestrales, el valor z y el valor p. Los resultados se indican a continuación.

Prueba de hipótesis para dos proporciones independientes

p_1	p_2	p_c	
0.19	0.31	0.27	p (como decimal)
19/100	62/200	81/300	p (como fracción)
19.	62.	81.	X
100	200	300	n
	−0.12	diferencia	
	0.	diferencia hipotética	
	0.0544	error estándar	
	−2.21	z	
	0.273	valor p (dos colas)	

AUTOEVALUACIÓN 15-2

De 150 adultos que probaron un nuevo pastel sabor durazno, 87 lo calificaron como excelente; de 200 niños muestreados, 123 le asignaron esa misma calificación. Con el nivel de significancia 0.10, ¿puede concluir que existe una diferencia significativa entre la proporción de adultos y la de niños que calificaron al nuevo sabor como excelente?

(a) Formule las hipótesis nula y alternativa.
(b) ¿Cuál es la probabilidad de un error tipo I?
(c) ¿Se trata de una prueba de una o dos colas?
(d) ¿Cuál es la regla de decisión?
(e) ¿Cuál es el valor del estadístico de prueba?
(f) ¿Cuál es su decisión respecto de la hipótesis nula?
(g) ¿Cuál es el valor p? Explique qué significa en términos de este problema.

EJERCICIOS

7. Las hipótesis nula y alternativa son:

$$H_0: \pi_1 \leq \pi_2$$
$$H_1: \pi_1 > \pi_2$$

Una muestra de 100 observaciones de la primera población indicó que X_1 es 70; una muestra de 150 observaciones de la segunda población reveló que X_2 es 90. Utilice el nivel de significancia 0.05 para probar la hipótesis.

a. Formule la regla de decisión.
b. Calcule la proporción conjunta.
c. Calcule el valor del estadístico de prueba.
d. ¿Cuál es su decisión respecto de la hipótesis nula?

8. Las hipótesis nula y alternativa son:

$$H_0: \pi_1 = \pi_2$$
$$H_1: \pi_1 \neq \pi_2$$

Una muestra de 200 observaciones de la primera población indicó que X_1 es 170; otra, de 150 observaciones de la segunda población, reveló que X_2 es 110. Utilice el nivel de significancia 0.05 para probar la hipótesis.

a. Formule la regla de decisión.
b. Calcule la proporción conjunta.
c. Estime el valor del estadístico de prueba.
d. ¿Cuál es su decisión respecto de la hipótesis nula?

Nota: Para resolver los ejercicios siguientes se debe utilizar el procedimiento de prueba de hipótesis de seis pasos.

9. La familia Damon posee un viñedo grande en el oeste de Nueva York, a orillas de lago Erie; los viñedos deben fumigarse al inicio de la temporada de cultivo para protegerlos contra diversos insectos y enfermedades. Dos nuevos insecticidas acaban de salir al mercado: Pernod 5 y Action; para probar su eficacia, se seleccionaron tres hileras y se fumigaron con Pernod 5, y otras tres se fumigaron con Action. Cuando las uvas maduraron, se revisaron 400 vides tratadas con Pernod 5 para saber si no estaban infectadas. De igual forma, se revisó una muestra de 400 vides fumigadas con Action. Los resultados son:

Insecticida	Número de vides revisadas (tamaño de la muestra)	Número de vides infectadas
Pernod 5	400	24
Action	400	40

Con el nivel de significancia 0.05, ¿se puede concluir que existe una diferencia entre la proporción de vides infectadas empleando Pernod 5 en comparación con las fumigadas con Action?

10. GfK Custom Research North America realizó encuestas idénticas en un intervalo de cinco años. Una pregunta para mujeres fue: “¿La mayoría de los hombres son amables, gentiles y considerados?”. La primera encuesta reveló que, de las 3 000 mujeres entrevistadas, 2 010 dijeron que sí; la última encuesta reveló que 1 530 de las 3 000 mujeres a las cuales se les formuló la pregunta pensaban que los hombres eran amables, gentiles y considerados. Con el nivel de significancia 0.05, ¿se puede concluir que las mujeres consideran que los hombres son menos amables, gentiles y considerados en la última encuesta en comparación con la primera?

11. A una muestra nacional de republicanos y demócratas influyentes se les preguntó, como parte de una encuesta muy amplia, si estaban en favor de relajar las normas ambientales para que se pudiera quemar carbón con alto contenido de azufre en las plantas eléctricas. Los resultados fueron:

	Republicanos	Demócratas
Número en la muestra	1 000	800
Número a favor	200	168

Con el nivel de significancia 0.02, ¿puede concluir que hay una proporción mayor de demócratas a favor de relajar las normas? Determine el valor p.

12. El departamento de investigación de la oficina matriz de New Hampshire Insurance realiza investigaciones continuas sobre las causas de accidentes automovilísticos, las características de los conductores, etcétera. Mediante una muestra aleatoria de 400 pólizas de personas solteras se reveló que 120 habían estado en al menos un accidente en el periodo anterior de tres años. De forma similar, a través de una muestra de 600 pólizas de personas casadas se supo que 150 habían estado involucradas en al menos un accidente. Con el nivel de significancia 0.05, ¿hay una diferencia significativa entre las proporciones de personas solteras y casadas involucradas en un accidente durante un periodo de tres años? Determine el valor p.

OA15-3
Realizar una prueba de hipótesis para comparar un grupo de frecuencias observadas con una distribución esperada de frecuencias.

Prueba de bondad de ajuste: comparación de las distribuciones de frecuencias observada y esperada

Enseguida, se comparan las dos distribuciones de datos a escala nominal que se han resumido en varias categorías; por ejemplo, una compañía de seguros de vida clasificó sus pólizas en cuatro rubros: vida completa, término medio, término decreciente y otros. En la siguiente tabla se muestra la distribución histórica relativa de frecuencias de los tipos de póliza.

Tipo de póliza	Porcentaje
Vida completa	40
Término medio	25
Término decreciente	15
Otros	20

La compañía de seguros desea comparar esta distribución con la de los tipos de póliza que una muestra de 2 000 adultos jóvenes casados compraron. A la situación que se presenta cuando los datos están en escala nominal y se quieren cotejar dos distribuciones se le llama prueba de bondad de ajuste; la cual es una de las pruebas estadísticas de uso más común. Es particularmente útil porque requiere solo un nivel nominal de medición; por ello es posible llevar a cabo una prueba de hipótesis con datos que se han clasificado en grupos.

Prueba de hipótesis de frecuencias iguales esperadas

En la primera ilustración de la prueba de bondad de ajuste se supone el caso en que las frecuencias esperadas de las celdas son iguales. Como su nombre lo indica, el propósito de la prueba de bondad de ajuste es comparar una distribución observada con una distribución esperada. En el siguiente ejemplo se describe la situación de una prueba de hipótesis.

EJEMPLO

Bubba's Fish and Pasta es una cadena de restaurantes ubicados a lo largo de la costa del Golfo de Florida. Bubba, el propietario, desea añadir filete a su menú, pero antes de hacerlo contrata a Magnolia Research, LLC, para que lleve a cabo una encuesta entre personas adultas para saber cuál es su platillo favorito cuando comen fuera de casa. Magnolia seleccionó una muestra de 120 adultos y les pidió que indicaran su comida favorita cuando salen a cenar; los resultados se reportan en la tabla de la derecha.

¿Es razonable concluir que no hay preferencia entre los cuatro platillos?

TABLA 15.1 Plato fuerte seleccionado por una muestra de 120 adultos

Plato favorito	Frecuencia
Pollo	32
Pescado	24
Carne	35
Pasta	29
Total	120

SOLUCIÓN

Si no existe diferencia entre la popularidad de los cuatro platillos, se podría esperar que las frecuencias observadas fueran iguales, o casi iguales; para decirlo de otro modo, se esperaría que el mismo número de adultos indicara que prefiere pollo o pescado. Así, cualquier discrepancia entre las frecuencias observadas y esperadas se atribuye al azar, o a un error de muestreo.

¿Cuál es el nivel de medición en este problema? Observe que cuando se selecciona a una persona, solo se le puede clasificar en una de las categorías de platillos preferidos; no se obtiene ningún tipo de lectura o medición. La "medida" o "clasificación" se basa en el platillo seleccionado; además, no existe un orden natural entre los platillos. No se supone que alguno sea mejor que otro; por lo pronto, la escala nominal es apropiada.

Si los platillos tienen la misma popularidad, se esperaría que 30 adultos eligieran cada uno de ellos. ¿Por qué? Si hay 120 adultos en la muestra, y cuatro categorías, lo esperado sería que una cuarta parte de los encuestados eligiera cada platillo; por lo tanto, la frecuencia esperada de cada categoría (o celda) sería 30, calculada mediante 120/4, asumiendo que no existiera preferencia por ningún platillo. Esta información se resume en la tabla 15.2. Un examen de los datos indica que la carne es el platillo seleccionado con más frecuencia (35 de 120), y que el pescado es el que cuenta con menos preferencia (24 de 120). ¿La diferencia entre los números de veces que cada platillo es seleccionado se debe al azar, o se concluye que los platillos no tienen el mismo grado de popularidad?

Para dilucidar este problema, se utiliza el procedimiento de la prueba de hipótesis en seis pasos.

TABLA 15.2 Frecuencias observadas y esperadas de la encuesta entre 120 personas adultas

Plato favorito	Frecuencia observada, f_o	Frecuencia esperada, f_e
Pollo	32	30
Pescado	24	30
Carne	35	30
Pasta	29	30
Total	120	120

Paso 1: Formule las hipótesis nula y alternativa. La hipótesis nula, H_0, es que no hay diferencia entre los conjuntos de frecuencias observadas y esperadas; en otras palabras, cualquier diferencia entre am-

bos conjuntos de frecuencias se puede atribuir al error de muestreo. La hipótesis alternativa, H_1, es que sí hay una diferencia entre ellos. Si se rechaza la hipótesis nula y se acepta la hipótesis alternativa, se concluye que las preferencias no se distribuyen de igual forma entre las cuatro categorías (celdas).

H_0: No hay diferencia entre las proporciones de adultos que eligen cada platillo.
H_1: Existe diferencia entre las proporciones de adultos que eligen cada platillo.

Paso 2: **Seleccione el nivel de significancia.** Si selecciona el nivel de significancia 0.05, la probabilidad de que rechace la hipótesis nula verdadera es 0.05.

Paso 3: **Seleccione el estadístico de prueba.** El estadístico de prueba sigue la distribución *ji* cuadrada, designada como χ^2.

ESTADÍSTICO DE PRUEBA *ji* CUADRADA

$$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] \qquad \textbf{[15.4]}$$

con $k - 1$ grados de libertad, donde:

k es el número de categorías;
f_o es la frecuencia observada en una categoría particular;
f_e es la frecuencia esperada en una categoría particular.

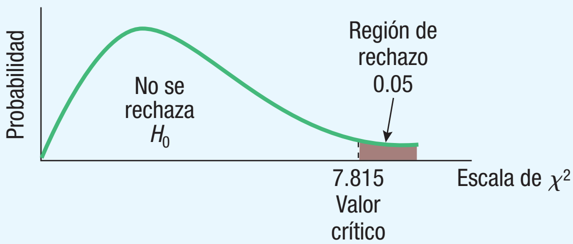
En breve se estudiarán las características de la distribución *ji* cuadrada con más detalle.

Paso 4: **Formule la regla de decisión.** Recuerde que, en las pruebas de hipótesis, la regla de decisión requiere determinar un número que separe la región donde no se rechaza H_0 de la región de rechazo; este número se denomina *valor crítico*. Como verá, la distribución *ji* cuadrada en realidad es una familia de distribuciones donde cada una tiene una forma un poco diferente, según el número de grados de libertad. El número de grados de libertad en este tipo de problema se encuentra mediante $k - 1$, donde k es el número de categorías. En este problema en particular hay cuatro; por lo tanto, hay $k - 1 = 4 - 1 = 3$ grados de libertad. Como se observó, una categoría se denomina *celda*, por lo que hay cuatro celdas. El valor crítico para 3 grados de libertad y el nivel de significancia 0.05 se encuentran en el apéndice B.7. Una parte de esa tabla se muestra en la tabla 15.3. El valor crítico es 7.815, determinado al ubicar 3 grados de libertad en el margen izquierdo, y luego, por la horizontal (a la derecha), y leyendo el valor crítico en la columna 0.05.

TABLA 15.3 Parte de la tabla de *ji* cuadrada

	Área de la cola derecha			
gl	**0.10**	**0.05**	**0.02**	**0.01**
1	2.706	3.841	5.412	6.635
2	4.605	5.991	7.824	9.210
3	6.251	7.815	9.837	11.345
4	7.779	9.488	11.668	13.277
5	9.236	11.070	13.388	15.086

La regla de decisión es rechazar H_0 si el valor calculado de *ji* cuadrada es mayor que 7.815; si es menor o igual a 7.815, no se rechaza H_0. En la gráfica 15.3 se muestra la regla de decisión.



GRÁFICA 15.3 Distribución de probabilidad *ji* cuadrada para 3 grados de libertad, con la región de rechazo y el nivel de significancia 0.05

ESTADÍSTICA EN ACCIÓN

Durante muchos años, investigadores y estadísticos creyeron que todas las variables se distribuían normalmente. De hecho, en general se suponía una ley universal; sin embargo, Karl Pearson observó que los datos experimentales no siempre se ajustaban a este supuesto, pero no había forma de demostrar que estos eran correctos. Para resolver el problema, Pearson descubrió el estadístico *ji* cuadrada, que en esencia compara una distribución de la frecuencia observada con una supuesta distribución normal. Su descubrimiento demostró que no todas las variables tenían una distribución normal.

La regla de decisión indica que si hay diferencias grandes entre las frecuencias observada y esperada, lo que genera una χ^2 calculada mayor que 7.815, se debe rechazar la hipótesis nula; sin embargo, si las diferencias entre f_o y f_e son pequeñas, el valor χ^2 calculado será 7.815 o menor, por lo que la hipótesis nula se debe aceptar. El razonamiento es que es probable que esas pequeñas diferencias entre las frecuencias observada y esperada se deban a la casualidad. Recuerde que las 120 observaciones son una muestra de la población.

Paso 5: Calcule el valor de *ji* cuadrada y tome una decisión. De los 120 adultos que integraban la muestra, 32 indicaron que su platillo favorito era el pollo. Los conteos se registraron en la tabla 15.1; a continuación se muestran los cálculos de la *ji* cuadrada (observe una vez más que las frecuencias esperadas son las mismas para cada celda).

Columna D: Determine las diferencias entre cada f_o y f_e; es decir, $(f_o - f_e)$. La suma de estas diferencias es cero.

Columna E: Eleve al cuadrado la diferencia entre cada frecuencia observada y esperada, es decir, $(f_o - f_e)^2$.

Columna F: Divida el resultado de cada observación entre la frecuencia esperada; es decir, $(f_o - f_e)^2/f_e$ Finalmente, sume estos valores.
El resultado es el valor de χ^2, que es 2.20.

	A	B	C	D	E	F	G
1	Plato favorito	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$	
2	Pollo	32	30	2	4	0.133	
3	Pescado	24	30	-6	36	1.200	
4	Carne	35	30	5	25	0.833	valor χ^2
5	Pasta	29	30	-1	1	0.033	
6	Total	120	120			2.200	

La $\chi 2$ calculada (2.20) no está en la región de rechazo, es menor que el valor crítico de 7.815; por lo tanto, la decisión es no rechazar la hipótesis nula.

Paso 6: Interprete el resultado. Se concluye que las diferencias entre las frecuencias observada y esperada podrían deberse al azar; los datos no sugieren que las preferencias entre los cuatro platillos sean diferentes.

Se puede emplear MegaStat para calcular la prueba de bondad de ajuste como sigue. Los pasos se muestran en la sección "Comandos de software" en el apéndice C. El valor calculado de *ji* cuadrada es 2.20, el mismo valor que se obtuvo en los cálculos anteriores; también observe que el valor *p* es 0.5319, mucho mayor que 0.05.

Prueba de bondad de ajuste

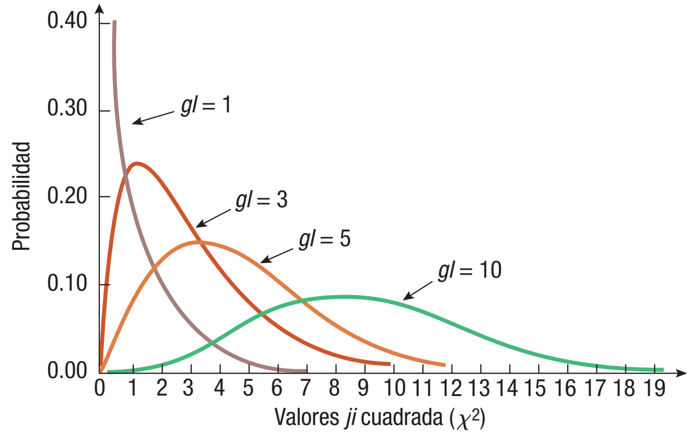
Observado	Esperado	O − E	(O − E)^2/E	% de ji cuad.
32	30.000	2.000	0.133	6.06
24	30.000	−6.000	1.200	54.55
35	30.000	5.000	0.833	37.88
29	30.000	−1.000	0.033	1.52
120	120.000	0.000	2.200	100.00

2.20	ji cuadrada
3	gl
0.5319	valor p

La distribución del estadístico *ji* cuadrada se caracteriza por lo siguiente:

1. **Los valores de *ji* cuadrada nunca son negativos.** Esto se debe a que la diferencia entre f_o y f_e se eleva al cuadrado, es decir, $(f_o - f_e)^2$.
2. **Existe una familia de distribuciones de *ji* cuadrada.** Hay una distribución de *ji* cuadrada para 1 grado de libertad, otra para 2, otra para 3 grados de libertad, etcétera. En este tipo de problema, el número de grados de libertad se determina mediante $k - 1$, donde k es el número de categorías; por lo tanto, la forma de la distribución *ji* cuadrada *no* depende del tamaño de la muestra, sino del número de categorías, por ejemplo, si clasifica a 200 empleados de una aerolínea en una de tres categorías: personal de vuelo, apoyo terrestre y personal administrativo, tendría $k - 1 = 3 - 1 = 2$ grados de libertad.

3. **La distribución *ji* cuadrada tiene un sesgo positivo.** Sin embargo, a medida que aumenta el número de grados de libertad, la distribución comienza a aproximarse a la distribución normal. En la gráfica 15.4 se muestran las distribuciones de grados de libertad seleccionados; observe que, para los 10 grados de libertad, la curva se aproxima a una distribución normal.



GRÁFICA 15.4 Distribuciones *ji* cuadrada de grados de libertad seleccionados

AUTOEVALUACIÓN 15-3

La directora de recursos humanos de Georgetown Paper, Inc., está preocupada por el absentismo entre los trabajadores por hora, por lo que decide tomar una muestra de los registros de la compañía y determinar si el absentismo está distribuido de manera uniforme en toda la semana de seis días. Las hipótesis son:

H_0: El absentismo está distribuido de manera uniforme en toda la semana de trabajo.
H_1: El absentismo *no* está distribuido de manera uniforme en toda la semana de trabajo.

Los resultados de la muestra son:

	Número de ausencias		Número de ausencias
Lunes	12	Jueves	10
Martes	9	Viernes	9
Miercoles	11	Sábado	9

(a) ¿Cómo se denominan los números 12, 9, 11, 10, 9 y 9?
(b) ¿Cuántas categorías (celdas) hay?
(c) ¿Cuál es la frecuencia *esperada* de cada día?
(d) ¿Cuántos grados de libertad hay?
(e) ¿Cuál es el valor crítico de *ji* cuadrada con un nivel de significancia de 1%?
(f) Calcule el estadístico de prueba χ^2.
(g) ¿Cuál es su regla de decisión respecto de la hipótesis nula?
(h) Específicamente, ¿qué le indica lo anterior a la directora de recursos humanos?

EJERCICIOS

13. En una prueba de bondad de ajuste de *ji* cuadrada hay cuatro categorías y 200 observaciones. Utilice el nivel de significancia 0.05.
a. ¿Cuántos grados de libertad hay?
b. ¿Cuál es el valor crítico de *ji* cuadrada?

14. En una prueba de bondad de ajuste de *ji* cuadrada hay seis categorías y 500 observaciones. Utilice el nivel de significancia 0.01.
a. ¿Cuántos grados de libertad hay?
b. ¿Cuál es el valor crítico de *ji* cuadrada?

15. Las hipótesis nula y alternativa son:

H_0: Las frecuencias son iguales.
H_1: Las frecuencias no son iguales.

a. Formule la regla de decisión, con el nivel de significancia 0.05.
b. Calcule el valor de *ji* cuadrada.
c. ¿Cuál es su decisión respecto de H_0?

Categoría	f_o
A	10
B	20
C	30

16. Las hipótesis nula y alternativa son:

H_0: Las frecuencias son iguales.
H_1: Las frecuencia no son iguales.

a. Formule la regla de decisión, con el nivel de significancia 0.05.
b. Calcule el valor de *ji* cuadrada.
c. ¿Cuál es su decisión respecto de H_0?

Categoría	f_o
A	10
B	20
C	30
D	20

17. Un dado se lanza 30 veces y los números 1 a 6 aparecen como se muestra en la siguiente distribución de frecuencia. Con el nivel de significancia 0.10, ¿es posible concluir que el dado no está cargado?

Resultado	Frecuencia	Resultado	Frecuencia
1	3	4	3
2	6	5	9
3	2	6	7

18. Classic Golf, Inc., administra cinco campos de golf en el área de Jacksonville, Florida. El director quiere estudiar el número de rondas de golf que se juegan por día en los cinco campos, por lo que reunió la siguiente información de una muestra. Con el nivel de significancia 0.05, ¿hay una diferencia entre el número de rondas jugadas por día de la semana?

Día	Rondas
Lunes	124
Martes	74
Miércoles	104
Jueves	98
Viernes	120

19. Un grupo de compradoras en tiendas departamentales vio una línea nueva de vestidos y opinó al respecto. Los resultaron fueron:

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Opinión	Número de compradoras	Opinión	Número de compradoras
Sobresaliente	47	Bueno	39
Excelente	45	Regular	35
Muy bueno	40	Indeseable	34

Como el número mayor (47) indicó que la línea nueva es extraordinaria, el jefe de diseño piensa que esta es una razón para iniciar la producción masiva de los vestidos. El jefe de mantenimiento (que de alguna manera participó en el estudio) considera que no hay una razón clara y afirma que las opiniones están distribuidas de manera uniforme entre las seis categorías. Además, dice que las pequeñas diferencias entre los diversos conteos quizá se deban a la casualidad. Demuestre que en la hipótesis nula no hay una diferencia relevante entre las opiniones de las compradoras; pruebe con el nivel de riesgo 0.01. Siga un enfoque formal, es decir, formule la hipótesis nula, la hipótesis alternativa, etcétera.

20. El director de seguridad de una empresa siderúrgica tomó muestras aleatorias de los registros de la compañía sobre accidentes menores relacionados con el trabajo, y los clasificó de acuerdo con la hora en que ocurrieron.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Hora	Número de accidentes	Hora	Número de accidentes
8 hasta 9 a.m.	6	1 hasta 2 p.m.	7
9 hasta 10 a.m.	6	2 hasta 3 p.m.	8
10 hasta 11 a.m.	20	3 hasta 4 p.m.	19
11 hasta 12 p.m.	8	4 hasta 5 p.m.	6

Utilice la prueba de bondad de ajuste y el nivel de significancia 0.01, y determine si los accidentes están distribuidos de manera uniforme durante el día. Explique brevemente su conclusión.

Prueba de hipótesis de frecuencias esperadas desiguales

Las frecuencias esperadas (f_e) del ejemplo acerca de los platillos preferidos eran iguales; de acuerdo con la hipótesis nula, se esperaba que de los 120 adultos que participaron en el estudio, un número igual seleccionara cada uno de los cuatro platillos. Así que se esperaba que 30 eligieran pollo, 30 eligieran pescados y así sucesivamente. La prueba *ji* cuadrada también es útil si las frecuencias esperadas no son iguales.

En el siguiente ejemplo se ilustra el caso de frecuencias desiguales y también se presenta un uso práctico de la prueba de bondad de ajuste de *ji* cuadrada para determinar si una experiencia local difiere de una experiencia más amplia, la nacional, por ejemplo.

EJEMPLO

La American Hospital Administrators Association (AHAA) reporta la siguiente información respecto del número de veces que los adultos mayores son admitidos en un hospital durante un periodo de un año: 40% no es admitido, 30% es admitido una vez, 20% son admitidos dos veces y el 10% restante es admitido tres o más veces.

Una encuesta que abarcó a 150 residentes de Bartow Estates, comunidad con una población predominante de adultos mayores activos en el centro de Florida, reveló que 55 residentes no ingresaron durante el año anterior, 50 fueron admitidos en un hospital una vez, 32 fueron admitidos dos veces, y 13 fueron admitidos tres o más veces. ¿Es posible concluir que la encuesta en Bartow Estates es consistente con la información sugerida por la AHAA? Utilice el nivel de significancia 0.05.

ESTADÍSTICA EN ACCIÓN

Muchos gobiernos estatales organizan loterías a fin de recaudar fondos para la educación; en muchas de ellas se mezclan pelotas numeradas que se seleccionan mediante una máquina. En el juego Select Three, las pelotas de tres grupos de pelotas numeradas del cero al nueve se seleccionan al azar. La selección aleatoria pronostica que la frecuencia de cada número sea igual. ¿Cómo probaría que la máquina asegure un proceso de selección aleatoria? Puede usar la prueba de bondad de ajuste de *ji* cuadrada para investigar esta cuestión.

SOLUCIÓN

Primero, la información anterior se debe organizar como en la tabla 15.4. Es evidente que los porcentajes del estudio del Hospital Administrators no se pueden comparar con las frecuencias reportadas por Bartow Estates; sin embargo, estos porcentajes se pueden convertir en frecuencias esperadas, f_e. De acuerdo con Hospital Administrators, 40% de los residentes de Bartow en la encuesta no requirió hospitalización; por lo tanto, si no hay una diferencia entre la experiencia nacional y la de Bartow Estates, 40% de los 150 adultos mayores encuestados (60 residentes) no habrían sido hospitalizados. Además, 30% de los encuestados fue admitido una vez (45 residentes), etcétera. Las frecuencias observadas en Bartow y las frecuencias esperadas con base en los porcentajes del estudio nacional se dan en la tabla 15.4.

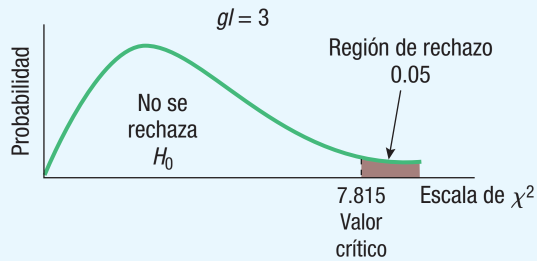
TABLA 15.4 Resumen del estudio de la AHAA y de una encuesta de los residentes en Bartow Estates

Número de admisiones	Porcentaje AHAA del total	Número de residentes en Bartow (f_o)	Número esperado de residentes (f_e)
0	40	55	60
1	30	50	45
2	20	32	30
3 o más	10	13	15
Total	100	150	

Las hipótesis nula y alternativa son:

H_0: No hay diferencias entre la experiencia local y la nacional respecto de las admisiones en un hospital

H_1: Hay diferencias entre la experiencia local y la nacional respecto de las admisiones en un hospital.



GRÁFICA 15.5 Criterios de decisión del estudio de investigación de Bartow Estates

Para determinar la regla de decisión, utilice el apéndice B.7 y el nivel de significancia 0.05. Hay cuatro categorías de admisión, por lo cual los grados de libertad son $gl = 4 - 1 = 3$.

El valor crítico es 7.815; así, la regla de decisión es rechazar la hipótesis nula si $\chi^2 > 7.815$. En la gráfica 17.3 se representa la regla de decisión.

Ahora calcule el estadístico de prueba *ji* cuadrada:

Número de admisiones	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
0	55	60	−5	25	0.4167
1	50	45	5	25	0.5556
2	32	30	2	4	0.1333
3 o más	13	15	−2	4	0.2667
Total	150				1.3723

valor de χ^2

El valor calculado de χ^2 (1.3723) aparece a la izquierda de 7.815; por lo tanto, no se rechaza la hipótesis nula. En conclusión, no hay evidencia de una diferencia entre las experiencias local y la nacional respecto de las admisiones en hospitales.

Limitaciones de *ji* cuadrada

OA15-4 Explicar las limitaciones de utilizar el estadístico *ji* cuadrada en pruebas de bondad de ajuste.

Si en una celda existe una frecuencia esperada pequeña inusual, aplicar *ji* cuadrada puede generar una conclusión errónea. Esto sucede debido a que f_e aparece en el denominador y, al dividirlo entre un número muy pequeño, hace el cociente muy grande. En general, dos pautas aceptadas respecto de las frecuencias de celdas pequeñas son:

1. Si solo hay dos celdas, la frecuencia *esperada* en cada una deberá ser al menos 5. El cálculo de *ji* cuadrada sería permisible en el siguiente problema para determinar el mínimo de f_e de 6.

Persona	f_o	f_e
Alfabetizada	641	642
Analfabeta	7	6

2. En caso de más de dos celdas, *no* se deberá utilizar *ji* cuadrada si más de 20% de las celdas f_e tiene frecuencias esperadas menores que 5. De acuerdo con esta pauta, lo adecuado es que se utilice la prueba de bondad de ajuste en los siguientes datos. Tres de las siete celdas, o 43%, tienen frecuencias esperadas (f_e) menores que 5.

Nivel de administración	f_o	f_e
Capataz	30	32
Supervisor	110	113
Gerente	86	87
Gerencia de nivel medio	23	24
Asistente del vicepresidente	5	2
Vicepresidente	5	4
Vicepresidente ejecutivo	4	1
Total	263	263

Para demostrar la razón de la pauta de 20%, realice la prueba de bondad de ajuste de los datos anteriores en los niveles de administración. La salida que se registra en MegaStat es la siguiente.

Prueba de bondad de ajuste

Observada	Esperada	O-E	(O-E)^2/E	Porcentaje de *ji* cuadrada
30	32.000	−2.000	0.125	0.89
110	113.000	−3.000	0.080	0.57
86	87.000	−1.000	0.011	0.08
23	24.000	−1.000	0.042	0.30
5	2.000	3.000	4.500	32.12
5	4.000	1.000	0.250	1.78
4	1.000	3.000	9.000	64.25
263	263.000	0.000	14.008	100.00

14.01 *ji* cuadrada
6 *gl*
0.0295 valor *p*

En el caso de esta prueba, con el nivel de significancia 0.05, rechace H_0 si el valor calculado de *ji* cuadrada es mayor que 12.592. El valor calculado es 14.008, por lo que se rechaza la hipótesis nula de que las frecuencias observadas representan una muestra aleatoria de la población de los valores esperados. Examine la salida de MegaStat, en la cual, más de 98% del valor calculado de *ji* cuadrada se explica por las tres categorías de vicepresidentes [(4.500 + 0.250 + 9.000)/14.008 = 0.9815], lo cual es lógico, pues a estas tres categorías se les dio mucha ponderación.

El dilema se resuelve mediante la combinación de categorías si es lógico hacerlo; en el ejemplo anterior se combinaron tres categorías de vicepresidentes, lo que satisface la pauta de 20%.

Nivel de administración	f_o	f_e
Capataz	30	32
Supervisor	110	113
Gerente	86	87
Gerencia de nivel medio	23	24
Vicepresidente	14	7
Total	263	263

El valor calculado de *ji* cuadrada con las categorías revisadas es 7.258 (vea la siguiente salida de MegaStat que se muestra a continuación); el cual es menor que el valor crítico de 9.488 para el nivel de significancia 0.05; por lo tanto, la hipótesis nula no se rechaza con el nivel de significancia 0.05. Esto indica que no hay una diferencia relevante entre las distribuciones observada y esperada.

Prueba de bondad de ajuste

Observada	Esperada	O-E	$(O-E)^2/E$	Porcentaje de *ji* cuadrada
30	32.000	−2.000	0.125	1.72
110	113.000	−3.000	0.080	1.10
86	87.000	−1.000	0.011	0.16
23	24.000	−1.000	0.042	0.57
14	7.000	7.000	7.000	96.45
263	263.000	0.000	7.258	100.00

7.26 *ji* cuadrada
4 *gl*
0.1229 valor *p*

AUTOEVALUACIÓN 15-4

La American Accounting Association clasifica las cuentas por cobrar como “actuales”, “atrasadas” e “irrecuperables”. Las cifras de la industria muestran que 60% de las cuentas por cobrar son actuales, 30% atrasadas y 10% irrecuperables. Massa and Barr, despacho de abogados de Greenville, Ohio, tiene 500 cuentas por cobrar: 320 son actuales, 120 están atrasadas y 60 son irrecuperables. ¿Concuerdan estas cifras con la distribución de la industria? Utilice el nivel de significancia 0.05.

EJERCICIOS

21. Con las siguientes hipótesis:

H_0: 40% de las observaciones se encuentra en la categoría A, 40% en la categoría B y 20% en la C.
H_1: La distribución de las observaciones no es como se describe en H_0.

En una muestra de 60 se dieron los resultadosque aparecen a la derecha:

Categoría	f_o
A	30
B	20
C	10

a. Formule la regla de decisión con el nivel de significancia 0.01.
b. Calcule el valor de *ji* cuadrada.
c. ¿Cuál es su decisión respecto de H_0?

22. Al jefe de seguridad de Mall of the Dakotas se le pidió estudiar el problema de la pérdida de mercancía. Seleccionó una muestra de 100 cajas que se manipularon de forma indebida y averiguó que, en 60 de ellas, los pantalones, zapatos y demás mercancía faltante se debían a hurtos en las tiendas. En otras 30 cajas, los empleados sustrajeron las mercancías, y en las restantes 10, lo atribuyó a un control de inventario deficiente. En su reporte a la gerencia del centro comercial, ¿es posible que concluyera que tal vez el hurto sea el *doble* de la causa de la pérdida en comparación con el robo por parte de los empleados o un control de inventario deficiente, y que el robo por parte de los empleados y el control de inventario deficiente quizá sean iguales? Utilice el nivel de significancia 0.02.

23. El departamento de tarjetas de crédito del Carolina Bank sabe por experiencia que 5% de sus tarjetahabientes estudió algunos años de la preparatoria; 15%, la terminó; 25%, fue varios años a la universidad, y 55% terminó una carrera. De los 500 tarjetahabientes a quienes se les llamó por no pagar sus cargos del mes, 50 estudiaron algunos años de preparatoria; 100, la terminaron; 190, cursaron ciertos años de la universidad, y 160 se graduaron de esta. ¿Es posible concluir que la distribución de los tarjetahabientes que no pagan sus cargos es diferente a los demás? Utilice el nivel de significancia 0.01.

24. Durante muchos años, los ejecutivos de televisión dieron crédito a la pauta de que 30% de la audiencia veía cada una de las cadenas televisivas de mayor audiencia, y 10%, canales de televisión por cable durante una noche a la semana. Una muestra aleatoria de 500 televidentes del área de Tampa-St. Petersburg, Florida, el pasado lunes por la noche, reveló que 165 hogares sintonizaron ABC; 140, CBS; 125, NBC, y el resto vio un canal de televisión por cable. Con el nivel de significancia 0.05, ¿es posible concluir que la pauta aún es razonable?

Prueba de hipótesis de que la distribución es normal

OA15-5 Probar una hipótesis de que una distribución observada de frecuencias tiene una distribución normal.

La prueba de bondad de ajuste se utiliza para comparar una distribución observada de frecuencias con una distribución esperada de frecuencias. En el ejemplo sobre Bubba's Fish and Pasta, las frecuencias observadas son el conteo de cada platillo seleccionado por una muestra de 120 adultos. Las frecuencias esperadas se determinan asumiendo que no existe preferencia por ninguno de los cuatro platillos, así que se espera que una cuarta parte, o 30 adultos, elijan cada platillo. En esta sección se quiere probar la hipótesis de que la distribución es normal, utilizando la prueba de bondad de ajuste para comparar una distribución observada de frecuencias con una distribución esperada de frecuencias que es normal. ¿Por qué es importante esta prueba? En el capítulo 11, al hacer la prueba para encontrar diferencias entre dos medias poblacionales, se asume que ambas poblaciones seguían la distribución normal. Se parte de la misma suposición en el capítulo 12, cuando se probó si varias medias poblacionales eran iguales. En la sección "Estimaciones de intervalo de predicción" del capítulo 13, se asumió que la distribución de los residuos en un análisis de regresión de mínimos cuadrados seguía una distribución de probabilidad normal.

En el siguiente ejemplo se muestran los detalles de esta prueba de bondad de ajuste para investigar la razonabilidad del supuesto de normalidad.

EJEMPLO

En la sección "Construcción de distribuciones de frecuencias: datos cuantitativos" del capítulo 2 se utilizó una distribución de frecuencia para organizar las ganancias de la venta de 180 vehículos en Applewood Auto Group; esta se repite en la tabla 15.5.

TABLA 15.5 Distribución de frecuencia de las ganancias por vehículos vendidos el mes anterior por Applewood Auto Group

Ganancia	Frecuencia
\$ 200 hasta \$ 600	8
600 hasta 1 000	11
1 000 hasta 1 400	23
1 400 hasta 1 800	38
1 800 hasta 2 200	45
2 200 hasta 2 600	32
2 600 hasta 3 000	19
3 000 hasta 3 400	4
Total	180

Utilizando un software estadístico se determinó en la sección "Solución con software" del capítulo 3, que la ganancia media sobre un vehículo del Applewood Auto Group era de 1 843.17 dólares, y que la desviación estándar era de 643.63 dólares.

¿Es razonable concluir que los datos acerca de las ganancias son una muestra obtenida de una población normal? En otras palabras, ¿los datos de ganancia siguen una distribución normal? Utilice el nivel de significancia 0.05.

SOLUCIÓN

Para probar una distribución normal, es preciso encontrar las frecuencias esperadas de cada clase de dicha distribución, asumiendo que la distribución esperada sigue una distribución de probabilidad normal. Inicie con la distribución normal calculando las probabilidades de cada clase; después, utilice estas probabilidades para calcular las frecuencias esperadas de cada una.

Para comenzar, se debe encontrar el área, o probabilidad, de cada una de las ocho clases en la tabla 15.5, asumiendo una población normal con una media de 1 843.17 dólares y una desviación estándar de 643.63 dólares. Para hallar esta probabilidad se debe adaptar la fórmula [7.1] del capítulo 7, reemplazando μ con $\overline{x}$ y σ con s. Así que se utiliza la siguiente fórmula para determinar los diversos valores de z

$$z = \frac{x - \overline{x}}{s}$$

En este caso, z es el valor de la distribución normal estándar; $\overline{x}$ (1 843.17 dólares) es la media simple; y s (643.63 dólares) es la desviación estándar de la muestra. Para ilustrar estos cálculos, se selecciona la clase $200 hasta $600 de la tabla 15.5. La meta es determinar la frecuencia esperada de esta clase, bajo el supuesto de que la distribución de ganancias sigue una distribución normal. Primero, se calcula el valor z correspondiente a 200 dólares.

$$z = \frac{x - \overline{x}}{s} = \frac{\$200 - \$1\,843.17}{\$643.63} = 2.55$$

Este resultado indica que el límite inferior de esta clase está a 2.55 desviaciones estándar por debajo de la media. Según el apéndice B.3, la probabilidad de encontrar un valor z menor a -2.55 es $0.5000 - 0.4946 = 0.0054$.

En el caso del límite superior de la clase $200 hasta $600:

$$z = \frac{x - \overline{x}}{s} = \frac{\$600 - \$1\,843.17}{\$643.63} = -1.93$$

El área a la izquierda de $600 es la probabilidad de un valor z menor a -1.93; para encontrar este valor, se utiliza de nuevo el apéndice B.3 y se calcula que $0.5000 - 0.4732 = 0.0268$.

Finalmente, para encontrar el área entre $200 y $600:

$$P(\$200 < x < \$600) = P(-2.55 < z < -1.93) = 0.268 - 0.0054 = 0.0214$$

Esto es, alrededor de 2.14% de los vehículos vendidos generará una ganancia de entre 200 y 600 dólares.

Existe una probabilidad de que la ganancia obtenida sea menor a 200 dólares; para encontrarla:

$$P(x < \$200) = P(z < -2.55) = 0.5000 - 0.4946 = 0.0054$$

Ambas probabilidades se registran en la segunda y tercera filas de la tercera columna de la tabla 15.6.

TABLA 15.6 Ganancias en Applewood Auto Group, valores *z*, áreas bajo la distribución normal y frecuencias esperadas

Ganancia	Valores *z*	Área	Calculada por	Frecuencia esperada
Menor a $200	Menor a −2.55	0.0054	0.5000 − 0.4946	0.97
$ 200 hasta $ 600	−2.55 hasta −1.93	0.0214	0.4946 − 0.4732	3.85
600 hasta 1 000	−1.93 hasta −1.31	0.0683	0.4732 − 0.4049	12.29
1 000 hasta 1 400	−1.31 hasta −0.69	0.1500	0.4049 − 0.2549	27.00
1 400 hasta 1 800	−0.69 hasta −0.07	0.2270	0.2549 − 0.0279	40.86
1 800 hasta 2 200	−0.07 hasta 0.55	0.2367	0.0279 + 0.2088	42.61
2 200 hasta 2 600	0.55 hasta 1.18	0.1722	0.3810 − 0.2088	31.00
2 600 hasta 3 000	1.18 hasta 1.80	0.0831	0.4641 − 0.3810	14.96
3 000 hasta 3 400	1.80 hasta 2.42	0.0281	0.4922 − 0.4641	5.06
3 400 o más	2.42 o más	0.0078	0.5000 − 0.4922	1.40
Total		1.0000		180.00

Lógicamente, si se vendieron 180 vehículos, se espera obtener una ganancia de entre 200 y 600 dólares en 3.852 de ellos, calculado por 0.0214 (180), y se esperaría vender 0.972 vehículos con una ganancia menor a 200 dólares, calculada por 180(0.0054). El proceso continúa con las clases restantes; esta información se resume en la tabla 15.7. No se preocupe de que se estén reportando fracciones de vehículos.

TABLA 15.7 Cálculo del estadístico *ji* cuadrada

Ganancia	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
Menor a \$600	8	4.82	3.18	10.1124	2.098
\$ 600 hasta \$1 000	11	12.29	−1.29	1.6641	.135
1 000 hasta 1 400	23	27.00	−4.00	16.0000	.593
1 400 hasta 1 800	38	40.86	−2.86	8.1796	.200
1 800 hasta 2 200	45	42.61	2.39	5.7121	.134
2 200 hasta 2 600	32	31.00	1.00	1.0000	.032
2 600 hasta 3 000	19	14.96	4.04	16.3216	1.091
3 000 y más	4	6.46	−2.46	6.0516	.937
Total	180	180.00	0		5.220

Antes de seguir, es preciso destacar una de las limitaciones de las pruebas que utilizan *ji* cuadrada como estadístico de prueba. La segunda limitación, que se encuentra en la sección "Limitaciones de *ji* cuadrada", indica que si más de 20% de las celdas tienen *frecuencias esperadas* menores a 5, algunas de las categorías deben combinarse. En la tabla 15.6 se encuentran tres clases en donde las frecuencias esperadas son menores a 5; por lo tanto, se combina la clase "Menor a \$200" con la clase "\$200 hasta \$600", y la clase "\$3 400 o más" con la clase "\$3 000 hasta \$3 400"; por ello, la frecuencia esperada en la clase "Menor a \$600" es ahora 4.82, calculada por 0.97 más 3.85. Se hace lo mismo con la clase "\$3 000 o más": 5.06 + 1.40 = 6.46; los resultados se muestran en la tabla 15.7. El valor calculado de *ji* cuadrada es 5.220.

Ahora ponga esta información en un formato formal de prueba de hipótesis; las hipótesis nula y alternativa son:

H_0: La población de ganancias sigue la distribución normal.
H_1: La población de ganancias no sigue la distribución normal.

Para determinar el valor crítico de *ji* cuadrada es necesario saber los grados de libertad; en este caso, hay ocho categorías (o clases), así que los grados de libertad son $k - 1 = 8 - 1 = 7$. Además, los valores \$1 843.17 (la ganancia media) y \$643.63 (la desviación estándar de las ganancias de Applewood Auto Group) se calcularon a partir de una muestra. Cuando se estiman parámetros poblacionales a partir de datos muestrales, se pierde un grado de libertad por cada estimación, de modo que se pierden dos grados más de libertad por estimar la media poblacional y la desviación estándar de la población. Así, el número de grados de libertad en este problema es 5, calculados por $k - 2 - 1 = 5$.

De acuerdo con el apéndice B.7, utilizando el nivel de significancia 0.05, el valor crítico de *ji* cuadrada es 11.070. La regla de decisión es rechazar la hipótesis nula si el valor calculado de *ji* cuadrada es mayor a 11.070.

Para calcular el valor de *ji* cuadrada se utiliza la fórmula [15.4]:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(8 - 4.82)^2}{4.82} + \cdots + \frac{(4 - 6.46)^2}{6.46} = 5.220$$

Los valores de cada clase se muestran en la columna de la derecha de la tabla 15.7, así como en la columna total, que es 5.220; debido a que el valor calculado de 5.220 es menor que el valor crítico, no se rechaza la hipótesis nula. Se concluye que la evidencia no sugiere que la distribución de ganancias sea distinta de la normal.

Para expandir el cálculo del número de grados de libertad, si se conoce la media y la desviación estándar de una población, y se desea determinar si algunos de los datos muestrales se conforman a una población normal, los grados de libertad son $k - 2 - 1$. En general, cuando se utilizan estadísticas de muestras para estimar parámetros poblacionales, se pierde un grado de libertad por cada parámetro estimado. Esto es paralelo a la situación que se planteó en la sección "Prueba global: prueba del modelo de regresión múltiple” del capítulo 14 sobre la regresión múltiple, donde se perdió un grado de libertad en el denominador del estadístico *F* por cada variable independiente considerada.

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

25. El IRS estaba interesado en el número de declaraciones de impuestos individuales preparadas por pequeñas firmas contables; por lo tanto, seleccionó al azar una muestra de 50 despachos contables que tuvieran 10 empleados o menos en el área de Dallas-Fort-Worth. En la siguiente tabla de frecuencias se reportan los resultados del estudio. Suponga que la media muestral es 44.8 clientes y que la desviación estándar de la muestra es 9.37 clientes. ¿Es razonable concluir que los datos muestrales provienen de una población que sigue una distribución de probabilidad normal? Utilice el nivel de significancia 0.05.

Número de clientes	Frecuencia
20 hasta 30	1
30 hasta 40	15
40 hasta 50	22
50 hasta 60	8
60 hasta 70	4

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

26. Los gastos publicitarios son un componente significativo del costo de venta de los bienes. Abajo se presenta una distribución de frecuencia que muestra los gastos publicitarios de 60 compañías manufactureras ubicadas en el sudoeste de Estados Unidos. El gasto medio es de 52.0 millones de dólares, y la desviación estándar, 11.32 millones de dólares. ¿Es razonable concluir que los datos muestrales provienen de una población que sigue una distribución de probabilidad normal? Utilice el nivel de significancia 0.05.

Gastos publicitarios (millones de dólares)	Número de compañías
25 hasta 35	5
35 hasta 45	10
45 hasta 55	21
55 hasta 65	16
65 hasta 75	8
Total	60

OA15-6
Realizar una prueba de *ji* cuadrada de la independencia de una tabla de contingencia

Análisis de tablas de contingencia

En el capítulo 4 se analizaron datos bivariados, y se estudió la relación entre dos variables; además, se describió una tabla de contingencia, que resume de manera simultánea dos variables de interés de escala nominal; por ejemplo, una muestra de estudiantes inscritos en la Escuela de Negocios por género (masculino o femenino) y especialidad (contabilidad, administración, finanzas, marketing o métodos cuantitativos). Esta clasificación tiene como base la escala nominal debido a que las clasificaciones no siguen un orden natural.

En el capítulo 5 se estudiaron las tablas de contingencia; en el ejemplo correspondiente se ilustró la relación entre el número de películas vistas por mes y la edad de quienes las vieron. Se puede usar la distribución *ji* cuadrada para probar si hay o no una relación entre dos variables con escala nominal; en otras palabras, ¿es *independiente* una variable de la otra?

He aquí algunos ejemplos interesantes para probar si dos variables están relacionadas.

- Ford Motor Company opera una planta de ensamble en Dearborn, Michigan. La planta opera tres turnos por día, cinco días a la semana. El gerente de control de calidad quiere comparar el nivel de calidad en cada horario. Los vehículos se clasifican por sus niveles de calidad (aceptable, inaceptable) y por turno (matutino, vespertino, nocturno). ¿Hay alguna diferencia en el nivel de calidad en los tres horarios? Es decir, ¿está relacionada la calidad del producto con la jornada en que se fabricó? ¿O es independiente la calidad del producto de este?
- Una muestra de 100 conductores detenidos por rebasar los límites de velocidad se clasificó por género y el uso del cinturón de seguridad. En esta muestra, ¿el uso del cinturón de seguridad se relaciona con el género?

- ¿Un hombre liberado de una prisión federal tiene una adaptación diferente a la vida civil si regresa a su ciudad natal o si va a vivir a otra parte? Las dos variables son: “adaptación a la vida civil” y “lugar de residencia”. Observe que ambas se miden en una escala nominal.

En el siguiente ejemplo se proporcionan los detalles del análisis y las posibles conclusiones.

EJEMPLO

Rainbow Chemical, Inc., tiene empleados por hora y asalariados. El vicepresidente de recursos humanos encuestó a una muestra de 380 trabajadores con respecto a su nivel de satisfacción con el programa vigente de beneficios de salud. Después clasificó a los empleados de acuerdo con su forma de pago, es decir, por hora o por salario; los resultados se muestran en la tabla 15.8

TABLA 15.8 Nivel de satisfacción con los servicios de salud de los empleados de Rainbow Chemical

Tipo de pago	Satisfecho	Neutral	Insatisfecho	Total
Salario	30	17	8	55
Por hora	140	127	58	325
Total	170	144	66	380

Con el nivel de significancia 0.05, ¿es razonable concluir que el tipo de pago y el nivel de satisfacción con los beneficios de salud están relacionados?

SOLUCIÓN

El primer paso es formular las hipótesis nulas y alternativa.

H_0: No hay relación entre el nivel de satisfacción y el tipo de pago.
H_1: Hay relación entre el nivel de satisfacción y el tipo de pago.

El nivel de satisfacción, según lo solicitó el vicepresidente de RH, es 0.05; el nivel de medición para el tipo de pago es la escala nominal; el nivel de satisfacción con los beneficios de salud es en realidad la escala ordinal, pero se utiliza como una variable con escala nominal. Cada empleado de la muestra se clasificó de acuerdo con dos criterios: el nivel de satisfacción con los beneficios y el tipo de pago. La información se tabuló en la tabla 15.8, que se denomina tabla de contingencia.

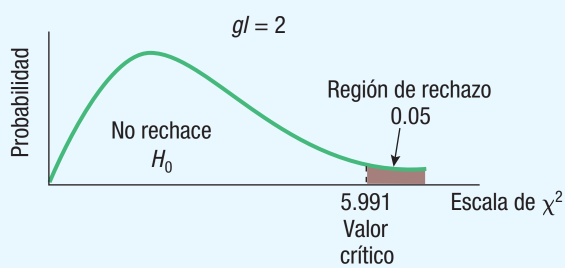
La distribución *ji* cuadrada se utiliza como el estadístico de prueba. Para determinar el valor crítico de *ji* cuadrada debe saber cuántos grados de libertad (*gl*) se obtienen por medio de:

$$gl = (\text{número de filas} - 1)(\text{número de columnas} - 1) = (r - 1)(c - 1)$$

En este ejemplo, hay dos filas y tres columnas, así que hay 2 grados de libertad.

$$gl = (r - 1)(c - 1) = (2 - 1)(3 - 1) = 2$$

Para encontrar el valor crítico de 2 grados de libertad y el nivel de 0.05, consulte el apéndice B.7, descienda por la columna de grados de libertad en el margen izquierdo, a la fila con 2; desplácese a lo largo de esta fila hasta la columna de 0.05. En la intersección, el valor crítico de *ji* cuadrada es 5.991. La regla de decisión es: rechazar la hipótesis nula si el valor calculado de χ^2 es mayor que 5.991 (vea la gráfica 15.6).



GRÁFICA 15.6 Distribución *ji* cuadrada para 2 grados de libertad

A continuación se determina el valor calculado de χ^2 usando la fórmula [15.4]; las frecuencias observadas, f_o, se muestran en la tabla 15.9. ¿Cómo se determinan las frecuencias esperadas correspondientes, f_e? Para comenzar, en la tabla 15.8 se observa que 55 de los 380 empleados de Rainbow Chemical son asalariados. Así que la fracción de empleados asalariados en la muestra es 55/380 = 0.14474. *Si no hubiera relación* entre el tipo de pago y el nivel de satisfacción con el programa de beneficios de salud, se esperaría que aproximadamente la misma fracción de los empleados que están satisfechos con el programa de salud fueran asalariados. Hay 170 empleados satisfechos con el programa, así que el número esperado de empleados satisfechos que son asalariados es 24.61, calculado por (0.14474)(170); por lo tanto, la frecuencia esperada de la celda superior izquierda es 24.61. De igual forma, si no hubiera relación entre el tipo de pago y el nivel de satisfacción, se esperaría que 0.14474 de los 144 empleados (20.84) que fueron neutrales con respecto al programa sean asalariados. Se continúa este proceso, llenando el resto de las celdas; no es necesario calcular cada uno de estos valores porque algunos pueden determinarse por sustracción.

La frecuencia esperada para cualquier celda se determina mediante:

FRECUENCIA ESPERADA $$f_e = \frac{(\text{Total de filas})(\text{Total de columnas})}{(\text{Gran total})} \qquad \textbf{[15.5]}$$

A partir de esta fórmula, la frecuencia que se espera en la celda superior izquierda en la tabla 15.8 es:

$$f_e = \frac{(\text{Total de filas})(\text{Total de columnas})}{(\text{Gran total})} = \frac{(55)(170)}{380} = 24.61$$

Las frecuencias observadas, f_o, y las frecuencias esperadas, f_e, de todas las celdas de la tabla de contingencia se presentan en la tabla 15.9.

TABLA 15.9 Frecuencias observadas y esperadas

	Nivel de satisfacción con los servicios de salud						
	Satisfecho		Neutral		Insatisfecho		
Tipo de pago	f_o	f_e	f_o	f_e	f_o	f_e	Total
Salario	30	24.61	17	20.84	8	9.55	55
Por hora	140	145.39	127	123.16	58	56.45	325
Total	170	170.00	144	144.00	66	66.00	380

Se utiliza la fórmula [15.4] para determinar el valor de *ji* cuadrada. Comenzando por la celda superior izquierda:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(30 - 24.61)^2}{24.61} + \frac{(17 - 20.84)^2}{20.84} + \cdots + \frac{(58 - 56.45)^2}{56.45}$$
$$= 1.181 + 0.708 + \cdots + 0.043 = 2.506$$

Como el valor calculado de *ji* cuadrada (2.506) aparece en la región a la izquierda de 5.991, no se rechaza la hipótesis nula con el nivel de significancia 0.05; en conclusión, los datos simples no proporcionan evidencia de una relación entre el tipo de pago y el nivel de satisfacción con los beneficios de salud.

La siguiente es una salida de MegaStat Excel:

Prueba de *ji* cuadrada de la independencia de una tabla de contingencia

		Nivel de satisfacción con los servicios de salud			
Tipo de pago		Satisfecho	Neutral	Insatisfecho	Total
Salario	Observada	**30**	**17**	**8**	55
	Esperada	24.61	20.84	9.55	55.00
Por hora	Observada	**140**	**127**	**58**	325
	Esperada	145.39	123.16	56.45	325.00
Total	Observada	170	144	66	380
	Esperada	170.00	144.00	66.00	380.00
		2.506 *ji* cuadrada 2 *gl* 0.286 valor *p*			

Observe que el valor de *ji* cuadrada es el mismo que el que se calculó antes, 2.506; además, el valor p reportado es 0.286, por lo tanto, la probabilidad de encontrar un valor del estadístico de prueba igual o mayor, asumiendo que la hipótesis nula es verdadera es 0.286. El valor p también da por resultado la misma decisión: no se rechaza la hipótesis nula.

AUTOEVALUACIÓN 15-5

Un científico social tomó una muestra de 140 personas y las clasificó de acuerdo con su nivel de ingresos, y si jugaron o no en la lotería estatal el mes anterior. La información de la muestra se presenta a continuación. ¿Es posible concluir que jugar a la lotería se relaciona con el nivel de ingresos? Utilice el nivel de significancia 0.05.

	Ingreso			
	Bajo	**Medio**	**Alto**	**Total**
Jugaron	46	28	21	95
No jugaron	14	12	19	45
Total	60	40	40	140

(a) ¿Cómo se denomina esta tabla?
(b) Formule las hipótesis nula y alternativa.
(c) ¿Cuál es su regla de decisión?
(d) Determine el valor de *ji* cuadrada.
(e) Tome una decisión respecto de la hipótesis nula e interprete el resultado.

EJERCICIOS

27. La directora de publicidad del *Carolina Sun Times*, el periódico más importante de Carolina del Norte y del Sur, estudia la relación entre el tipo de comunidad en que residen sus suscriptores y la sección del periódico que leen primero; por lo tanto, recopiló la siguiente información de una muestra de lectores.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Noticias nacionales	**Deportes**	**Tiras cómicas**
Ciudad	170	124	90
Suburbios	120	112	100
Rural	130	90	88

Con el nivel de significancia 0.05, ¿se puede concluir que existe relación entre el tipo de comunidad donde reside la persona y la sección del periódico que lee primero?

28. Se considera usar cuatro marcas de lámparas en el área de ensamblado final de la planta Saturn de Spring Hill, Tennessee; así que el director de compras pidió muestras de 100 lámparas de cada fabricante. Los números de lámparas aceptables e inaceptables de cada uno se muestran en la siguiente tabla. Con el nivel de significancia 0.05, ¿hay una diferencia entre las calidades de las lámparas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Fabricante			
	A	**B**	**C**	**D**
Inaceptable	12	8	5	11
Aceptable	88	92	95	89
Total	100	100	100	100

29. El departamento de control de calidad de Food Town, Inc., cadena de abarrotes del norte de Nueva York, mensualmente compara los precios registrados con los precios anunciados. En la siguiente gráfica se resumen los resultados de una muestra de 500 artículos del mes anterior. La gerencia de la compañía quiere saber si existe relación entre las tasas de error de los artículos con precios normales y los artículos con precios especiales. Utilice el nivel de significancia 0.01.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Precio regular	**Precio especial anunciado**
Precio bajo	20	10
Precio mayor	15	30
Precio correcto	200	225

30. El uso de teléfonos celulares en automóviles aumentó de forma impresionante en los últimos años. El efecto en los índices de accidentes es de interés para los expertos de tránsito, así como para los fabricantes de teléfonos celulares. ¿Es más probable que quien usa un teléfono celular se vea involucrado en un accidente de tránsito? ¿Cuál es su conclusión a partir de la siguiente información? Utilice el nivel de significancia 0.05.

	Tuvo un accidente el año previo	No tuvo un accidente el año previo
Usa teléfono celular	25	300
No usa teléfono celular	50	400

RESUMEN DEL CAPÍTULO

I. En este capítulo se consideraron las pruebas de hipótesis para los datos con nivel nominal.

II. Cuando se toma una muestra de una sola población, y la variable de interés tiene solo dos posibles resultados, se llama prueba de proporción.

A. Las condiciones binomiales deben estar presentes.

B. Tanto $n\pi$ como $n(1 - \pi)$ deben ser cuando menos 5.

C. El estadístico de prueba es

$$z = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \qquad \textbf{[15.1]}$$

III. También se puede comprobar si dos muestras provienen de poblaciones con la misma proporción de éxitos.

A. Las dos proporciones muestrales se agrupan con la fórmula siguiente:

$$p_c = \frac{x_1 + x_2}{n_1 + n_2} \qquad \textbf{[15.3]}$$

B. Se calcula el valor del estadístico de prueba a partir de la fórmula siguiente:

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_c(1 - p_c)}{n_1} + \dfrac{p_c(1 - p_c)}{n_2}}} \qquad \textbf{[15.2]}$$

IV. Las características de la distribución *ji* cuadrada son:

A. El valor de *ji* cuadrada nunca es negativo.

B. La distribución *ji* cuadrada tiene sesgo positivo.

C. Hay una familia de distribuciones *ji* cuadrada.

1. Cada vez que cambian los grados de libertad se forma una nueva distribución.

2. A medida que aumentan los grados de libertad, la distribución se aproxima a una distribución normal.

V. Una prueba de bondad de ajuste indica si un conjunto de frecuencias observadas puede provenir de una distribución normal.

A. Los grados de libertad son $k - 1$, donde k es el número de categorías.

B. La fórmula para calcular el valor de *ji* cuadrada es

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \qquad \textbf{[15.4]}$$

VI. La prueba de bondad de ajuste también puede utilizarse para determinar si una muestra proviene de una población normal; he aquí sus pasos:

A. Primero se determina la media y la desviación estándar de los datos muestrales.

B. Se agrupan en una distribución de frecuencia.

C. Los límites de clase se convierten en valores z y se encuentra la distribución estándar de probabilidad normal de cada clase.

D. Se halla la frecuencia esperada de distribución normal de cada clase al multiplicar la distribución estándar de probabilidad normal por la frecuencia de clase.

E. Se calcula el estadístico de bondad de ajuste *ji* cuadrada, basándose en la frecuencia de clase observada y esperada.

F. Se encuentra la frecuencia esperada de cada celda determinando el producto de la probabilidad de encontrar un valor en cada celda por el número total de celdas.
G. Si se utiliza la información de la media muestral y la desviación estándar de la muestra de los datos muestrales, los grados de libertad son $k - 3$.

VII. Una tabla de contingencia sirve para probar si hay relación entre dos rasgos o características.
A. Cada observación se clasifica de acuerdo con dos rasgos.
B. La frecuencia esperada se determina de la siguiente manera:

$$f_e = \frac{\text{(Total de filas)(Total de columnas)}}{\text{(Gran total)}} \qquad \textbf{[15.5]}$$

C. Los grados de libertad se determinan mediante:

$$gl = (\text{Filas} - 1)(\text{Columnas} - 1)$$

D. Se emplea el procedimiento de prueba de hipótesis habitual.

CLAVE DE PRONUNCIACIÓN

Símbolo	**Significado**	**Pronunciación**
p_c	Proporción conjunta	*p sub c*
χ^2	Distribución de probabilidad	*ji cuadrada*
f_o	Frecuencia observada	*f subíndice o*
f_e	Frecuencia esperada	*f subíndice e*

EJERCICIOS DEL CAPÍTULO

31. En la mayoría de los deportes se acostumbra lanzar una moneda para decidir qué equipo obtiene la pelota primero; esto requiere poco esfuerzo y se cree que concede la misma oportunidad a ambos equipos. En los 47 juegos del Súper Tazón, la Conferencia Nacional ha ganado estos “volados” 31 veces, mientras que la Conferencia Americana solo ha ganado 16 veces. Utilice el procedimiento de seis pasos de prueba de la hipótesis y el nivel de significancia 0.01 para probar si estos datos sugieren que es justo lanzar la moneda.
a. ¿Por qué es posible emplear z como el estadístico de prueba?
b. Establezca las hipótesis nula y alternativa.
c. Elabore un diagrama de la regla de decisión.
e. ¿Cuál es el valor p y qué es lo que implica?

32. De acuerdo con un estudio de la American Pet Food Dealers Association, 63% de las familias estadounidenses tiene mascotas. Se prepara un informe para una editorial del *San Francisco Chronicle* y, como parte del editorial, una muestra aleatoria de 300 familias mostró que poseía mascotas. ¿Estos datos contradicen los de la Pet Food Dealers Association? Aplique el nivel de significancia 0.05.

33. Tina Dennis es contralora de Meek Industries y cree que el problema actual de flujo de efectivo en la empresa es consecuencia de la tardanza en el cobro de cuentas; ella cree que la liquidación de más de 60% de las cuentas tarda más de tres meses. Una muestra aleatoria de 200 cuentas reveló que 140 tenían más de tres meses de antigüedad. Con el nivel de significancia 0.01 ¿Puede concluir que más de 60% de las cuentas permanece sin cobrarse tres meses?

34. La política de la Suburban Transit Authority (STA) consiste en añadir una ruta de autobús cuando más de 55% de los pasajeros potenciales indiquen que la utilizarán. Una muestra de 70 pasajeros reveló que 42 utilizarían una ruta propuesta que va de Bowman Park al área del centro de la ciudad. ¿La ruta de Bowman al centro cumple con el criterio de la STA? Aplique el nivel de significancia 0.05.

35. La experiencia en Crowder Travel Agency indicó que 44% de las personas que le solicitaron planear sus vacaciones deseaba ir a Europa. Durante la temporada de vacaciones reciente, se eligió una muestra aleatoria de 1 000 planes vacacionales archivados y se descubrió que 480 personas querían ir a Europa de vacaciones. ¿Hubo un incremento significativo en el porcentaje de personas que quieren ir a ese continente? Lleve a cabo la prueba con el nivel de significancia 0.05.

36. Una investigación en la industria del juego reveló que 10% de las máquinas tragamonedas en Estados Unidos deja de funcionar cada año. Short’s Game Arcade tiene 60 de estas máquinas y solo tres fallaron el año anterior. Utilice el procedimiento de cinco pasos de la prueba de hipótesis con el nivel 0.05 para probar si estos datos contradicen el reporte de la investigación.
a. Por qué es posible emplear z como el estadístico de prueba?
b. Establezca las hipótesis nula y alternativa.

c. Evalúe el estadístico de prueba y tome la decisión.
d. ¿Cuál es el valor *p* y qué es lo que implica?

37. Un planeador urbano afirma que, en todo el país, 20% de las familias que rentan condominios se muda en el lapso de un año. Una muestra de 200 familias que rentan condominios en Dallas Metroplex reveló que 56 se mudaron el año anterior. Con el nivel de significancia 0.01, ¿sugieren estas evidencias que una proporción mayor de propietarios de condominios se mudaron en el área de Dallas? Determine el valor *p.*

38. Después de perder una temporada, hay un gran clamor para que se despida al director técnico. En una muestra aleatoria de 200 alumnos universitarios, 80 están de acuerdo en conservarlo. Utilice el nivel de significancia 0.05 para probar si la proporción de alumnos que apoyan al director técnico es menor a 50%.

39. En la década de 1990, el índice de mortalidad por cáncer de pulmón era de 80 por cada 100 000 personas. A la vuelta del siglo y el establecimiento de nuevos tratamientos y ajustes en la publicidad de salud pública, una muestra aleatoria de 10 000 personas exhibe solo seis muertes debidas al cáncer de pulmón. Con el nivel 0.05, pruebe si los datos comprueban una reducción del índice de mortalidad de ese tipo de cáncer.

40. Cada mes, la Asociación Nacional de Gerentes de Compras (NAPM) de Estados Unidos entrevista a gerentes de compras y publica el índice NAPM. Una de las preguntas que se hacen en la encuesta es: "¿Piensa que la economía se está contrayendo?". El mes anterior, de los 300 gerentes encuestados, 160 contestaron afirmativamente; este mes, 170 de los 290 gerentes indicaron que sentían que la economía se estaba contrayendo. Con el nivel de significancia 0.05, ¿es posible concluir que una proporción mayor de los gerentes de compras piensa que la economía se está contrayendo este mes?

41. Como parte de una encuesta reciente entre parejas en que ambos cónyuges trabajan, un psicólogo industrial determinó que 990 hombres de 1 500 encuestados creen que es justa la división de tareas domésticas. Una muestra de 1 600 mujeres reveló que 970 creen que la división es justa. Con el nivel de significancia 0.01, ¿es razonable concluir que es más alta la proporción de hombres que creen que es justa la división de tareas domésticas? ¿Cuál es el valor *p*?

42. En el área de Colorado Springs, Colorado, hay dos proveedores de teléfonos celulares: HTC y Mountain Communications. Se desea investigar si hay alguna diferencia en la proporción de veces que un cliente puede conectarse a internet. Durante un periodo de una semana, se hicieron 500 intentos mediante HTC en diversas horas del día y la noche; se logró una conexión a internet en 450 ocasiones. Un estudio similar durante una semana con Mountain Communications reveló que la conexión se logró en 352 de 400 intentos. Con el nivel de significancia 0.01, ¿hay alguna diferencia en el porcentaje de veces que se logró la conexión a internet?

43. La Consumer Confidence Survey es una revisión mensual que mide la confianza del consumidor en la economía estadounidense. Se basa en una muestra típica de 5 000 hogares. El mes anterior, 9.1% de los consumidores dijo que las condiciones eran "buenas"; el mes anterior, solo 8.1% sostuvo los mismo. Utilice el método de prueba de hipótesis de seis pasos con el nivel de significancia 0.05 para ver si se puede determinar que hubo un incremento de la proporción que consideraba que las condiciones eran "buenas". Encuentre el valor *p* y explique lo que significa.

44. Se realizó un estudio para determinar si había una diferencia entre el contenido humorístico de los anuncios en revistas inglesas y estadounidenses. En una muestra aleatoria independiente de 270 anuncios en revistas estadounidenses, 56 tenían contenido humorístico; una muestra aleatoria independiente de 203 revistas inglesas encontró 52 anuncios humorísticos. ¿Estos datos proporcionan evidencia, con el nivel de significancia 0.05, de que hay una diferencia entre las proporciones de anuncios humorísticos en las revistas inglesas en comparación con las estadounidenses?

45. La encuesta de AP-Petside.com contactó a personas casadas que tuvieran mascotas: 300 mujeres y 200 hombres; de estos, 100 mujeres y 36 hombres contestaron que sus mascotas sabían escuchar mejor que sus cónyuges. Con el nivel de significancia 0.05, ¿existe una diferencia entre las respuestas de hombres y mujeres?

46. La Asociación Nacional de Basquetbol (NBA, por sus siglas en inglés) tiene 39 altos ejecutivos de color (presidentes o vicepresidentes) entre sus 388 directivos; por su parte, las Ligas Mayores de Béisbol tienen solo 11 miembros de color entre sus 307 altos administradores. Con el nivel de significancia 0.05, prueba si estos datos revelan que la NBA tiene una participación significativamente mayor de directivos de color en los altos niveles de administración.

47. Los vehículos que se dirigen hacia el oeste sobre Front Street pueden dar vuelta a la derecha, a la izquierda o seguir de frente hacia Elm Street. El ingeniero de tráfico de la ciudad considera que la mitad de los vehículos continuará de frente cruzando la intersección; de la mitad restante, proporciones iguales darán vuelta a la derecha e izquierda. Se observaron 200 vehículos, con los resultados que aparecen en la página siguiente. ¿Es posible concluir que el ingeniero de tráfico tiene razón? Utilice el nivel de significancia 0.10.

	De frente	Vuelta a la dercha	Vuelta a la izquierda
Frecuencia	112	48	40

48. El editor de una revista deportiva piensa ofrecer a los nuevos suscriptores uno de tres regalos: una sudadera, una taza o un par de aretes, todos ellos con el logotipo de su equipo favorito. En una muestra de 500 suscriptores nuevos, el número seleccionado de regalos se registra en la siguiente tabla. Con el nivel de significancia 0.05, ¿existe una preferencia por los regalos o es posible concluir que esta preferencia es igual?

Regalo	Frecuencia
Sudadera	183
Taza	175
Aretes	142

49. En un área particular de la ciudad hay tres estaciones de televisión comercial, cada una con su propio noticiero de 6:00 a 6:30 p.m. De acuerdo con el reporte de un periódico local matutino, una muestra aleatoria de 150 televidentes reveló que anoche 53 vieron las noticias en WNAE (canal 5), 64 en WRRN (canal 11) y 33 en WSPD (canal 13). Con el nivel de significancia 0.05, ¿hay una diferencia entre las proporciones de televidentes que ven los tres canales?

50. Hay cuatro entradas en el Government Center Building, en el centro de Filadelfia, y al supervisor de mantenimiento del edificio le gustaría saber si estas se utilizan por igual; para investigar esto, observó a 400 personas que entraron al edificio. El número de personas por cada entrada se registra en la siguiente tabla. Con el nivel de significancia 0.01, ¿hay una diferencia entre el empleo de las cuatro entradas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Entrada	Frecuencia
Main Street	140
Broad Street	120
Cherry Street	90
Walnut Street	50
Total	400

51. El propietario de un negocio de ventas por catálogo quiere comparar sus ventas con la distribución geográfica de la población. De acuerdo con el United States Bureau of the Census, 21% de la población vive en el noreste, 24% en el medio oeste, 35% en el sur, y 20% en el oeste. El desglose de una muestra de 400 pedidos seleccionados de manera aleatoria de los envíos del mes previo se registra en la siguiente tabla. Con el nivel de significancia 0.01, ¿la población refleja la distribución de los pedidos?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Región	Frecuencia
Noreste	68
Medio oeste	104
Sur	155
Oeste	73
Total	400

52. Banner Mattress and Furniture Company quiere estudiar el número de solicitudes de crédito que recibió a diario durante los últimos 300 días. La información se muestra a continuación.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Número de solicitudes de crédito	Frecuencia (número de días)
0	50
1	77
2	81
3	48
4	31
5 o más	13

Para interpretar los datos anteriores, hubo 50 días en los que no se recibieron solicitudes de crédito, 77 en los que solo se recibió una solicitud, etcétera. ¿Es razonable concluir que la distribución de población tiene una distribución de Poisson con una media de 2.0? Utilice el nivel de significancia 0.05. *Sugerencia:* para determinar las frecuencias esperadas utilice la distribución de Poisson con una media de 2.0, encuentre la probabilidad exacta de un éxito dada esta distribución y multiplique esta probabilidad por 300 para encontrar la frecuencia esperada del número de días en que hubo exactamente una solicitud. De manera similar, determine la frecuencia esperada de los demás días.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

53. Se piensa que cada uno de los dígitos de una rifa tiene la misma probabilidad de salir. En la siguiente tabla se muestra la frecuencia de cada dígito al ser elegido al azar y consecutivamente en la lotería de California. Realice la prueba de *ji* cuadrada para ver si se rechaza la hipótesis de que los dígitos provienen de una población uniforme, con el nivel de significancia 0.05.

Dígito	Frecuencia	Dígito	Frecuencia
0	44	5	24
1	32	6	31
2	23	7	27
3	27	8	28
4	23	9	21

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

54. John Isaac, Inc., un diseñador e instalador de señalamientos industriales, tiene 60 empleados. La compañía registró el tipo de la visita al médico más reciente de cada empleado. Una evaluación nacional que se realizó en Estados Unidos en 2004 reveló que 53% de todas las visitas al médico eran a profesionales de atención primaria, 19% a especialistas, 17% a cirujanos y 11% a atención de emergencia. Con el nivel de significancia 0.01, pruebe si los empleados de Isaac difieren significativamente de la distribución derivada de la encuesta. He aquí los resultados:

Tipo de visita	Número de visitas
Atención primaria	29
Especialista	11
Cirujano	16
Emergencia	4

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

55. Eckel Manufacturing Company piensa que sus salarios por hora siguen una distribución de probabilidad normal; para confirmarlo se eligió una muestra de 300 empleados, organizados en la siguiente distribución de frecuencia. Utilice los métodos que se incluyen en la sección 3.15, capítulo 3, para encontrar la media y la desviación estándar de estos datos agrupados en una distribución de frecuencia. Con el nivel de significancia 0.10, ¿es razonable concluir que la distribución de los salarios mensuales sigue una distribución normal?

Salario por hora	Frecuencia
\$5.50 hasta \$ 6.50	20
6.50 hasta 7.50	24
7.50 hasta 8.50	130
8.50 hasta 9.50	68
9.50 hasta 10.50	28
Total	270

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

56. La Asociación Nacional de Cable y Telecomunicaciones reportó que el número medio de televisores de alta definición (HDTV) por hogar en Estados Unidos es 2.30, con una desviación estándar de 1.474 televisores. Mediante una muestra de 100 hogares en Boise, Idaho, se reveló la siguiente información muestral.

Número de HDTV	Número de hogares
0	7
1	27
2	28
3	18
4	10
5 o más	10
Total	100

Con el nivel de significancia 0.05, ¿es razonable concluir que el número de HDTV por hogar sigue una distribución normal? (*Sugerencia:* Utilice límites como 0.5 hasta 1.5, 1.5 hasta 2.5, y así sucesivamente.)

57. Una encuesta investigó la actitud pública hacia la deuda federal. Cada ciudadano encuestado opinó si el gobierno debería reducir el déficit, aumentarlo o se abstuvo de opinar. Los resultados de la muestra del estudio por género se reportan enseguida.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Género	Reducir el déficit	Aumentar el déficit	Se obstuvo
Masculino	244	194	68
Femenino	305	114	25

Con el nivel de significancia 0.05, ¿es razonable concluir que el género es independiente de la posición de una persona respecto al déficit?

58. Mediante un estudio acerca de la relación entre la edad y la cantidad de presión que siente el personal de ventas en su trabajo se reveló la siguiente información de una muestra. Con el nivel de significancia 0.01, ¿hay alguna relación entre la presión en el trabajo y la edad?

	Grado de presión en el trabajo		
Edad (años)	**Bajo**	**Medio**	**Alto**
Menores de 25	20	18	22
25 hasta 40	50	46	44
40 hasta 60	58	63	59
60 y mayores	34	43	43

59. El departamento de reclamaciones de Wise Insurance Company cree que los conductores jóvenes tienen más accidentes, por lo cual se les deben cobrar primas mayores. Mediante una muestra de 1 200 de sus asegurados se reveló el siguiente análisis acerca de las reclamaciones en los últimos tres años y la edad del cliente. ¿Es razonable concluir que hay una relación entre la edad del asegurado y si hizo una reclamación o no? Utilice el nivel de significancia 0.05.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Grupo de edad	Sin reclamación	Reclamación
16 hasta 25	170	74
25 hasta 40	240	58
40 hasta 55	400	44
55 y mayores	190	24
Total	1 000	200

60. A una muestra de empleados de una gran planta química se le pidió indicar su preferencia por uno de tres planes de pensión. Los resultados se registran en la siguiente tabla. ¿Parece haber una relación entre el plan seleccionado y la clasificación del trabajo de los empleados? Utilice el nivel de significancia 0.01.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Plan de pensión		
Clase de trabajo	**Plan A**	**Plan B**	**Plan C**
Supervisor	10	13	29
De oficina	19	80	19
Obrero	81	57	22

61. ¿Alguna vez compró una bolsa de chocolates M&M's y se preguntó acerca de la distribución de los colores? Visite el sitio **web www.baking.m-ms.com** y en el mapa haga clic en United States, luego en **About M&M's,** después en **History of M&M's Brand, Product Information,** y **Peanut,** y encuentre el análisis del porcentaje de acuerdo con el fabricante, así como una historia breve del producto. ¿Sabía que al inicio todos los chocolates eran color marrón? En el caso de los M&M's de cacahuate, 12% es color marrón, 15% amarillo, 12% rojo, 23% azul, 23% naranja y 15% verde. Una bolsa de 6 onzas comprada en la Book Store en Coastal Carolina University el 1 de noviembre de 2005 tenía 12 chocolates color azul, 14 marrón, 13 amarillo, 14 rojo, 7 naranja y 12 verde. ¿Es razonable concluir

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

que la distribución actual concuerda con la distribución esperada? Utilice el nivel de significancia 0.05. Realice su propio ensayo e informe al maestro sus resultados.

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

62. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en el área de Goodyear, Arizona, el año anterior.

a. Determine la proporción de casas que tienen garaje. Con el nivel de significancia 0.05, ¿es posible concluir que más de 60% de las casas vendidas en esa zona tenía garaje? ¿Cuál es el valor *p?*

b. Determine la proporción de casas con alberca. Con el nivel de significancia 0.05, ¿es posible concluir que 60% de las casas vendidas en ese sitio tenía alberca? ¿Cuál es el valor *p*?

c. Elabore una tabla de contingencia que muestre si una casa tiene alberca y el poblado de su ubicación. ¿Hay alguna asociación entre las variables "alberca" y "poblado"? Utilice el nivel de significancia 0.05.

d. Elabore una tabla de contingencia que muestre si una casa tiene garaje y el poblado de su ubicación. ¿Hay alguna asociación entre las variables "garaje" y "poblado"? Utilice el nivel de significancia 0.05.

63. Consulte los datos sobre Baseball 2012 que contienen información de los 30 equipos de las Ligas Mayores de Béisbol durante la temporada 2012. Establezca una variable que divida los equipos en dos grupos: los que tuvieron una temporada ganadora y los que no. La temporada se compone de 162 juegos; por lo tanto, defina una temporada ganadora con 81 juegos o más; luego, divida los equipos en dos grupos de salarios. Coloque los 15 equipos con los salarios mayores en un grupo y los otros 15 equipos con los salarios menores en el otro. Con el nivel de significancia 0.05, ¿hay una relación entre los salarios y los juegos ganados?

64. Consulte los datos de los autobuses del Distrito Escolar Buena.

a. Suponga que un autobús es "viejo" si ha estado en servicio más de ocho años. Con el nivel de significancia 0.01, ¿es posible concluir que menos de 40% de los autobuses del distrito son viejos? Reporte el valor *p*.

b. Encuentre el costo mediano de mantenimiento y la edad mediana de los autobuses. Organice los datos en una tabla de contingencia dos por dos, con los autobuses por encima y por debajo de la mediana de cada variable. Determine si la edad del autobús se relaciona con el costo de mantenimiento. Utilice el nivel de significancia 0.05.

c. ¿Existe una relación entre el costo de mantenimiento y el fabricante del autobús? Utilice el desglose del punto anterior (los autobuses por encima y por debajo del costo mediano de mantenimiento) y los fabricantes de los autobuses para crear una tabla de contingencia, con el nivel de significancia 0.05.

Métodos no paramétricos:

ANÁLISIS DE DATOS ORDINALES

16

LOS TRABAJADORES DE COASTAL Computer, Inc. (CC), ensamblan uno o dos montajes parciales y los insertan en un chasis. Los ejecutivos de CC consideran que los empleados estarían más orgullos de su trabajo si ensamblaran todos los componentes y probaran la computadora terminada. Se seleccionó una muestra de 25 empleados para probar la idea; a 20 les gustó ensamblar la unidad completa y probarla. Con el nivel de significancia 0.05, ¿es posible concluir que los empleados prefieren ensamblar toda la unidad y probarla? (vea el ejercicio 8 y el OA16-1).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA16-1 Utilizar la prueba de signos para comparar dos poblaciones independientes.

OA16-2 Realizar una prueba de hipótesis sobre la mediana utilizando la prueba de signos.

OA16-3 Realizar una prueba de hipótesis de poblaciones dependientes utilizando la prueba de los rangos con signo de Wilcoxon.

OA16-4 Realizar una prueba de hipótesis de poblaciones independientes utilizando la prueba de la suma de los rangos de Wilcoxon.

OA16-5 Realizar una prueba de hipótesis de varias poblaciones independientes utilizando la prueba de Kruskal-Wallis.

OA16-6 Realizar e interpretar una prueba de hipótesis no paramétrica de la correlación.

Introducción

En el capítulo 15 se introdujeron las pruebas de hipótesis de variables en *escala nominal*. Del capítulo 1, recuerde que un nivel de medición nominal implica que los datos solo se clasifican en categorías, y estas no reconocen un orden particular. El propósito de estas pruebas es determinar si un conjunto de frecuencias observadas, f_o, tiene una diferencia significativa con un conjunto correspondiente de frecuencias esperadas, f_e; de igual forma, si le interesa la relación entre dos características, como la edad de un individuo o su preferencia musical, será preciso ordenar los datos en una tabla de contingencia y utilizar *ji* cuadrada como el estadístico de prueba. En ambos tipos de problemas no es necesario hacer suposiciones acerca de la forma de la población; por ejemplo, no necesita suponer que la población de interés sigue la distribución normal, como lo hizo con las pruebas de hipótesis en los capítulos 10 a 12.

En este capítulo se continúa con la prueba de hipótesis especial para datos no paramétricos. Para realizar estas pruebas no necesita hacer ninguna suposición acerca de la distribución de la población; en ocasiones, se usa el término *pruebas libres de distribución*, las cuales no requieren que las respuestas estén clasificadas u ordenadas, así que deben medirse con una escala ordinal, de intervalo o de razón. Un ejemplo de clasificación es el título de los ejecutivos corporativos, quienes se clasifican como asistente de la vicepresidencia, vicepresidente, vicepresidente senior y presidente. Un vicepresidente se clasifica más alto que su asistente, un vicepresidente senior se clasifica más alto que un vicepresidente y así sucesivamente.

En este capítulo se consideran cinco pruebas sin distribución y coeficiente de correlación de los rangos de Spearman. Las pruebas son: de los signos, de la mediana, de los rangos con signo de Wilcoxon, de la suma de los rangos de Wilcoxon y el análisis de la varianza por rangos de Kruskal-Wallis.

OA16-1
Utilizar la prueba de signos para comparar dos poblaciones independientes.

Prueba de los signos

La **prueba de los signos** se basa en el signo de una diferencia entre dos observaciones relacionadas; en general, se designa con un signo de adición (+) una diferencia positiva, y con un signo de resta (−), una negativa; por ejemplo, una dietista quiere ver si disminuirá el nivel de colesterol de una persona si la dieta se complementa con cierto mineral. Ella selecciona una muestra de 20 obreros mayores de 40 años de edad y mide su nivel de colesterol; después que los 20 sujetos toman el mineral durante seis semanas, vuelve a medir su nivel de colesterol; si este disminuye, se registra un signo "+"; si aumentó, se registra un signo "−"; si no hay cambio, se registra cero (y esa persona sale del estudio). En el caso de una prueba de los signos, no interesa la magnitud de la diferencia, solo la dirección de ella.

La prueba de los signos tiene muchas aplicaciones; una es en los experimentos de "antes/después". Para ilustrar este punto, suponga la evaluación de un nuevo programa de afinación de automóviles. Se registra el número de millas recorridas por galón de gasolina antes y después de la afinación; si la afinación no es eficaz, es decir, si no tuvo efecto en el desempeño, casi la mitad de los automóviles probados presentará un aumento en las millas por galón, y la otra mitad, una disminución. Se asigna "+" a un aumento y "−" a una disminución.

Mediante un experimento sobre la preferencia de un producto se ilustra otro uso de la prueba de los signos. Taster´s Choice vende dos tipos de café: descafeinado y regular. Su departamento de investigación de mercado quiere determinar si los bebedores de café lo prefieren descafeinado o regular; para saberlo les dan dos tazas con ambos tipos de esa bebida y sin ninguna marca que distinga una de la otra y a cada uno se le pregunta cuál prefiere. La preferencia por café descafeinado se codifica con el signo "+", y la preferencia por el regular, con "−". En cierto sentido, los datos están en un nivel ordinal debido a que los bebedores de café le dan a su bebida preferida un rango más alto, mientras que el otro tipo de café queda en un rango menor. Una vez más, si la población de consumidores de café no tiene una preferencia, se debe esperar que la mitad de la muestra de consumidores seleccione café descafeinado, y la otra mitad, regular.

Un ejemplo ayudará a mostrar mejor la aplicación de la prueba de los signos. A continuación se presenta un experimento de "antes/después".

EJEMPLO

El director de Sistemas de Información de Samuelson Chemicals recomendó implementar un programa de capacitación para gerentes en la planta; el objetivo es aumentar los conocimientos de computación en los departamentos de Nómina, Contabilidad y Producción.

Se seleccionó de forma aleatoria una muestra de 15 gerentes de los tres departamentos; un panel de expertos clasificó a cada uno de acuerdo con sus conocimientos en computación. Se calificaron como sobresalientes, excelentes, buenos, regulares o deficientes (consulte la tabla 16.1); después del programa de capacitación de tres meses, el mismo panel de expertos en sistemas de información calificó a cada gerente una vez más. Ambas calificaciones (antes y después) aparecen con el signo de la diferencia; un signo "+" indica una mejora, y un signo "−", que la competencia del gerente con las bases de datos declinó después del programa de capacitación.

TABLA 16.1 Nivel de competencia antes y después del programa de capacitación

	Nombre	Antes	Después	Signo de la diferencia
	T. J. Bowers	Buena	Extraordinaria	+
	Sue Jenkins	Regular	Excelente	+
	James Brown	Excelente	Buena	−
	Tad Jackson	Deficiente	Buena	+
Eliminado del análisis	~~Andy Love~~	~~Excelente~~	~~Excelente~~	0
	Sarah Truett	Buena	Extraordinaria	+
	Antonia Aillo	Deficiente	Regular	+
	Jean Unger	Excelente	Extraordinaria	+
	Coy Farmer	Buena	Deficiente	−
	Troy Archer	Deficiente	Buena	+
	V. A. Jones	Buena	Extraordinaria	+
	Juan Guillen	Regular	Excelente	+
	Candy Fry	Buena	Regular	−
	Arthur Seiple	Buena	Extraordinaria	+
	Sandy Gumpp	Deficiente	Buena	+

Lo que interesa saber es si el programa de capacitación en la planta aumentó la eficacia de los gerentes en el uso de la base de datos de la compañía; es decir, ¿los gerentes son más competentes después la de capacitación que antes?

SOLUCIÓN

Utilice el procedimiento de prueba de hipótesis de seis pasos.

Paso 1: Formule las hipótesis nula y alternativa.

H_0: $\pi \leq 0.50$ No hay aumento del conocimiento en el uso de las bases de datos como resultado del programa de capacitación en la planta.

H_1: $\pi > 0.50$ Aumenta el conocimiento de los gerente en el uso de las bases de datos después del programa de capacitación.

El símbolo π es la proporción de la población con una característica particular; si *no se rechaza* la hipótesis nula, se indica que el programa de capacitación no produjo ningún cambio en el nivel de competencia o que la competencia en realidad disminuyó. Si se *rechaza*, se indica que la competencia de los gerentes aumentó como resultado de a capacitación.

El estadístico de prueba sigue la distribución de probabilidad binomial; es apropiado debido a que la prueba de los signos cumple con todas las suposiciones binomiales, las cuales se mencionan a continuación:

1. Solo hay dos resultados: "éxito" o "fracaso". Un gerente o aumentó sus conocimientos (éxito) o no.
2. Por cada intento, se supone que la probabilidad de éxito es 0.50; así, la probabilidad de un éxito es la misma en todos los intentos (en este caso, los gerentes).
3. El número total de intentos es fijo (15 en este experimento).
4. Cada intento es independiente; eso significa, por ejemplo, que el desempeño de Arthur Seiple en el curso de tres meses no se relaciona con el desempeño de Sandy Gumpp.

ESTADÍSTICA EN ACCIÓN

Una investigación reciente aplicada a estudiantes universitarios de la University of Michigan reveló que los alumnos con los peores registros de asistencia suelen obtener las calificaciones más bajas. ¿Le sorprende? Los estudiantes que se ausentan menos de 10% del tiempo suelen obtener una calificación de 9 o mejor. El mismo estudió determinó que los estudiantes que se sientan al frente de la clase obtienen calificaciones mayores que quienes se sientan en la parte posterior.

Paso 2: Seleccione el nivel de significancia. Elija 0.10.

Paso 3: Decida sobre el estadístico de prueba. Es el *número de signos más* que resulta del experimento.

Paso 4: Formule una regla de decisión. En el curso de capacitación se inscribieron 15 gerentes, pero el nivel de conocimientos de Andy Love no mostró aumento ni reducción (consulte la tabla 16.1); por lo tanto, se eliminó del estudio debido a que no se pudo incluir en ningún grupo, entonces $n = 14$. A partir de la tabla de distribución de probabilidad binomial del apéndice B.1, para una n de 14 y una probabilidad de 0.50, se presenta la distribución de probabilidad binomial en la tabla 16.2. El número de éxitos aparece en la primera columna, las probabilidades de éxito en la segunda, y las probabilidades acumuladas en la tercera; para llegar a las probabilidades acumuladas, *sume* las probabilidades de éxito de la segunda columna desde la parte inferior. Con fines de ilustración, para obtener la probabilidad acumulada de 11 o más éxitos, sume $0.000 + 0.001 + 0.006 + 0.022 = 0.029$.

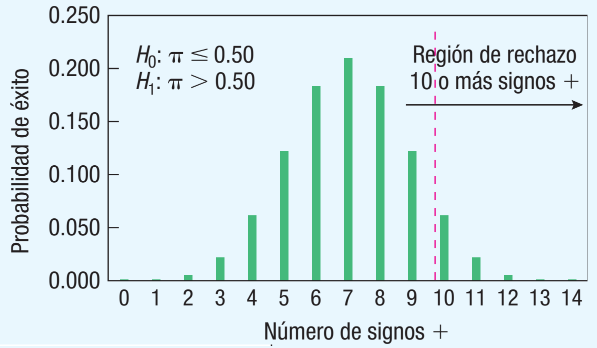
Esta es una prueba de una cola debido a que la hipótesis alternativa proporciona una dirección. La desigualdad ($>$) apunta hacia la derecha; por lo tanto, la región de rechazo está en la cola superior o derecha. Si el signo de desigualdad apuntara hacia la cola izquierda ($<$), la región de rechazo estaría en la cola inferior o izquierda. En tal caso, sumaría las probabilidades de la segunda columna *hacia abajo* para obtener las probabilidades acumuladas en la tercera columna.

Recuerde que se seleccionó el nivel de significancia 0.10; para llegar a la regla de decisión para este problema, recurra a las probabilidades acumuladas que se registran en la tercera columna de la tabla 16.2. Se lee de abajo hacia arriba hasta llegar a la *probabilidad acumulada más cercana, pero sin exceder el nivel de significancia* (0.10); esa probabilidad acumulada es 0.090. El número de éxitos (signos más) que corresponde a 0.090 en la columna 1 es 10; por lo tanto, la regla de decisión es: si el número de signos más en la muestra es 10 o mayor, se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

TABLA 16.2 Distribución de probabilidad binomial para $n = 14$, $\pi = 0.50$

Número de éxitos	Probabilidad de éxito		Probabilidad acumulada
0	0.000		1.000
1	0.001		0.999
2	0.006		0.998
3	0.022		0.992
4	0.061		0.970
5	0.122		0.909
6	0.183		0.787
7	0.209	**Suma hacia arriba**	0.604
8	0.183		0.395
9	0.122		0.212
10	0.061		0.090
11	0.022		0.029
12	0.006		0.007
13	0.001		0.001
14	0.000		0.000

$.000 + .001 + .006 + .022 + .061$



GRÁFICA 16.1 Distribución binomial, $n = 14$, $\pi = 0.50$

Repaso: se suman las probabilidades de abajo hacia arriba porque la dirección de la desigualdad ($>$) es hacia la derecha, lo que indica que la región de rechazo está en la cola superior. Si el número de signos más en la muestra es 10 o mayor, se rechaza la hipótesis nula; de lo contrario, no se rechaza H_0. La representación de la región de rechazo se indica en la gráfica 16.1.

¿Qué procedimiento se sigue en el caso de una prueba de dos colas? Se combinan (suman) las probabilidades de éxito en ambas colas hasta estar lo más cerca posible del nivel de significancia deseado (α) sin sobrepasarlo. En este ejemplo, α es 0.10; la probabilidad de 3 o menos éxitos es 0.029, determinada mediante $0.000 + 0.001 + 0.006 + 0.022$. La probabilidad de 11 o más éxitos también es 0.029; si se suman las probabilidades, $0.029 + 0.029$, se obtiene 0.058. Esto es lo más cercano que se puede estar de 0.10 sin sobrepasarlo. Si hubiera incluido las probabilidades de 4 y 10 éxitos, 0.090 + 0.090, el total sería 0.180, que excede 0.10; por lo tanto, la regla de decisión en el caso de una prueba de dos colas sería rechazar la hipótesis nula si hay 3 o menos signos más (+), u 11 o más signos más.

Paso 5: **Tome una decisión respecto de la hipótesis nula.** De los 14 gerentes en el curso de capacitación, 11 aumentaron su competencia para las bases de datos; el número 11 está en la región de rechazo, que inicia en 10, por tanto, se rechaza H_0.

Paso 6: **Interprete los resultados.** Se concluye que el curso de capacitación de tres meses fue eficaz porque el conocimiento de los gerentes sobre computación se incrementó.

Si la hipótesis nula no ofrece una dirección, por ejemplo, H_0: $\pi = 0.50$ y H_1: $\pi \neq 0.50$, la prueba de hipótesis es de *dos colas*; en esos casos hay dos regiones de rechazo, una en la cola inferior y la otra en la cola superior. Si $\alpha = 0.10$ y la prueba es de dos colas, el área en cada una es 0.05 ($\alpha/2 = 0.10/2 = 0.05$). En la autoevaluación 16.1 se ilustra lo anterior.

AUTOEVALUACIÓN 16-1

Recuerde el ejemplo de Taster´s Choice descrito en la sección "Prueba de los signos", con la cual se buscaba determinar la preferencia de los consumidores por el café descafeinado en comparación con el normal. Las hipótesis nula y alternativa son:

$$H_0: \pi = 0.50 \qquad n = 12$$
$$H_1: \pi \neq 0.50$$

(a) ¿Se trata de una hipótesis de prueba de una o dos colas?
(b) Ilustre la regla de decisión en una gráfica.
(c) Al designar la preferencia del consumidor por café descafeinado como "+" y por café normal como "−", se determinó que dos consumidores prefirieron café descafeinado. ¿Cuál es su decisión? Explique su respuesta.

EJERCICIOS

1. Considere la siguiente situación de prueba de hipótesis: H_0: $\pi \leq 0.50$ y H_1: $\pi > 50$. El nivel de significancia es 0.10, y el tamaño de la muestra es 12.
   a. ¿Cuál es su regla de decisión?
   b. Hubo nueve éxitos. ¿Cuál es su decisión respecto de la hipótesis nula? Explique su respuesta.
2. Considere la siguiente situación de prueba de hipótesis: H_0: $\pi = 0.50$ y H_1: $\pi \neq 0.50$. El nivel de significancia es 0.05, y el tamaño de la muestra es 9.
   a. ¿Cuál es su regla de decisión?
   b. Hubo cinco éxitos. ¿Cuál es su decisión respecto de la hipótesis nula?
3. Calorie Watchers tiene desayunos, comidas y cenas bajas en calorías. Si usted se une al club, recibe dos alimentos empacados al día; la empresa afirma que usted puede comer todo lo que quiera en su tercera comida y aun así perderá al menos cinco libras el primer mes. Los miembros del club se pesan antes de comenzar el programa y de nuevo al cabo del primer mes. Las experiencias de una muestra aleatoria de 11 miembros son:

Nombre	Cambio de peso	Nombre	Cambio de peso
Foster	Bajó	Hercher	Bajó
Taoka	Bajó	Camder	Bajó
Lange	Subió	Hinckle	Bajó
Rousos	Bajó	Hinkley	Bajó
Stephens	Sin cambio	Justin	Bajó
Cantrell	Bajó		

   Lo que interesa saber es si los miembros perdieron peso como resultado del programa de Calorie Watchers.
   a. Formule H_0 y H_1.
   b. Con el nivel de significancia 0.05, ¿cuál es su regla de decisión?
   c. ¿Cuál es su conclusión respecto del programa de Calorie Watchers?

4. Muchos corredores de bolsa nuevos no se atreven a realizar presentaciones frente a banqueros y otros grupos. Al detectar esta falta de autoestima, la gerencia organizó un seminario de motivación para una muestra de corredores de bolsa nuevos y contrató a Career Boosters para que diera un curso de tres semanas. Antes de la primera sesión, Career Boosters midió el nivel de autoestima de cada participante, y volvió a hacerlo después del seminario de tres semanas. Los niveles de autoestima antes y después de los 14 participantes en el curso se muestran en la siguiente tabla (la autoestima se clasificó como negativa, baja, alta o muy alta).

Corredor de bolsa	Antes del seminario	Después del seminario	Corredor de bolsa	Antes del seminario	Después del seminario
J. M. Martin	Negativa	Baja	F. M. Orphey	Baja	Muy alta
T. D. Jagger	Negativa	Negativa	C. C. Ford	Baja	Alta
A. D. Hammer	Baja	Alta	A. R. Utz	Negativa	Baja
T. A. Jones Jr.	Muy alta	Baja	M. R. Murphy	Baja	Alta
B. G. Dingh	Baja	Alta	P. A. Lopez	Negativa	Baja
D. A. Skeen	Baja	Alta	B. K. Pierre	Baja	Alta
C. B. Simmer	Negativa	Alta	N. S. Walker	Baja	Muy alta

El propósito del estudio es determinar si Career Boosters fue capaz de aumentar la autoestima de los corredores de bolsa nuevos; es decir, ¿el nivel de autoestima fue más alto después del seminario que antes? Utilice el nivel de significancia 0.05.

a. Formule las hipótesis nula y alternativa.

b. Con el nivel de significancia 0.05 indique la regla de decisión, ya sea en palabras o en forma gráfica.

c. Apunte sus conclusiones acerca del seminario ofrecido por Career Boosters.

Uso de la aproximación normal a la binomial

Si el número de observaciones en la muestra es mayor que 10, se puede utilizar la distribución normal para aproximar la binomial. En la sección "Tablas de probabilidad binomial", del capítulo 6, se calculó la media de la distribución normal a partir de $\mu = n\pi$, y la desviación estándar de $\sigma = \sqrt{n\pi(1 - \pi)}$. En este caso, $\pi = 0.50$, por lo que puede reducir las ecuaciones a $\mu = 0.50n$ y $\sigma = 0.50\sqrt{n}$, respectivamente.

El estadístico de prueba z es:

PRUEBA DE LOS SIGNOS, $n > 10$

$$z = \frac{(x \pm 0.50) - \mu}{\sigma} \quad [16.1]$$

Si el número de signos más (+) o menos (−) es *mayor que* $n/2$, emplee la siguiente fórmula como estadístico de prueba:

PRUEBA DE LOS SIGNOS, $n > 10$, SIGNOS + MAYORES QUE $n/2$

$$z = \frac{(x - 0.50) - \mu}{\sigma} = \frac{(x - 0.50) - 0.50n}{0.50\sqrt{n}} \quad [16.2]$$

Si el número de signos más (+) o menos (−) es *menor que* $n/2$, el estadístico de prueba z es:

PRUEBA DE LOS SIGNOS, $n > 10$, SIGNOS + MENORES QUE $n/2$

$$z = \frac{(x + 0.50) - \mu}{\sigma} = \frac{(x + 0.50) - 0.50n}{0.50\sqrt{n}} \quad [16.3]$$

En las fórmulas anteriores, el número de signos (+ o −) se representa mediante X. El valor +0.50 o −0.50 es el *factor de corrección de continuidad*, que se estudió en la sección "Factor de corrección de continuidad" del capítulo 7. En resumen, se aplica cuando una distribución continua como la normal (que se está utilizando) sirve para aproximar una distribución discreta (la binomial).

En el siguiente ejemplo se ilustran los detalles de la prueba de los signos cuando n es mayor que 10.

EJEMPLO

El departamento de investigación de mercado de Cola, Inc., tiene la tarea de probar una nueva bebida. Se consideran dos versiones: un refresco más bien dulce y uno un tanto amargo. La prueba de preferencia que se realizará consiste en una muestra de 64 consumidores; cada uno degustará ambas bebidas, la dulce (con la etiqueta A) y la amarga (con la etiqueta B), e indicará su preferencia. Realice una prueba de hipótesis para determinar si hay una diferencia entre las preferencias por el refresco dulce o por el amargo; utilice el nivel de significancia 0.05.

SOLUCIÓN

Paso 1: Formule las hipótesis nula y alternativa.

H_0: $\pi = 0.50$ No hay preferencia.
H_1: $\pi \neq 0.50$ Sí hay preferencia.

Paso 2: Seleccione el nivel de significancia. En el problema se indica que es 0.05.

Paso 3: Seleccione el estadístico de prueba. Es z, dado en la fórmula [16.1].

$$z = \frac{(x \pm 0.50) - \mu}{\sigma}$$

donde $\mu = 0.50n$ y $\sigma = 0.50\sqrt{n}$.

Paso 4: Formule la regla de decisión. Use la *distribución de Student,* en el apéndice B.5, con grados infinitos de libertad, para una prueba de dos colas (debido a que H_1 estipula que $\pi \neq 0.50$) y el nivel de significancia 0.05, los valores críticos son $+1.960$ y -1.960; por lo tanto, no rechace H_0 si el valor z calculado se encuentra entre $+1.960$ y -1.960. De lo contrario, rechace H_0 y acepte H_1.

Paso 5: Calcule *z*, compárelo con el valor crítico y tome una decisión respecto de H_0. A la preferencia por el refresco *A* se le asignó un signo "+", y a la preferencia por el *B*, un signo "−". De las 64 personas de la muestra, 42 prefirieron el sabor dulce, que es el refresco A; por lo tanto, hay 42 signos más. Como 42 es *mayor que* $n/2 = 64/2 = 32$, emplee la fórmula [16.2] de z:

$$z = \frac{(x - 0.50) - 0.50n}{0.50\sqrt{n}} = \frac{(42 - 0.50) - 0.50(64)}{0.50\sqrt{64}} = 2.38$$

El valor z calculado de 2.38 es mayor que el valor crítico de 1.96; en consecuencia, se debe rechazar la hipótesis nula de que no hay diferencia con el nivel de significancia 0.05.

Paso 6: Interprete los resultados. Hay evidencia de una diferencia en la preferencia de los consumidores; esto es, se concluye que los consumidores prefieren el refresco de cola dulce al otro.

El valor p es la probabilidad de encontrar un valor z mayor que 2.38 o menor que -2.38; para calcularlo, se usa primero el apéndice B.3, Áreas bajo la curva normal, donde se menciona que la probabilidad de un valor z mayor que 2.38 es $0.5000 - 0.4913 = 0.0087$. Para una prueba de dos colas, esta probabilidad se multiplica por dos; el valor p es 0.0174; por lo tanto, la probabilidad de obtener un estadístico de la muestra tan extremo cuando la hipótesis nula es verdadera es menor que 2%.

AUTOEVALUACIÓN 16-2

El departamento de recursos humanos de Ford Motor Company comenzó un programa piloto de educación para la salud al comienzo del año; el director del estudio seleccionó aleatoriamente a 100 empleados para que participaran. Para evaluar la efectividad del programa, se registró la presión arterial de esos empleados en julio; 80 de ellos mostraron una reducción en su presión arterial. ¿Es posible concluir que el programa fue eficaz para reducir la presión arterial?

(a) Formule las hipótesis nula y alternativa.
(a) ¿Cuál es la regla de decisión con el nivel de significancia 0.05?
(a) Calcule el valor del estadístico de prueba.
(a) ¿Cuál es su decisión respecto de la hipótesis nula?
(a) Interprete su decisión.

EJERCICIOS

5. Una muestra de 45 hombres con sobrepeso participó en un programa de ejercicio. Al término del programa, el peso de 32 de ellos se redujo. Con el nivel de significancia 0.05, ¿es posible concluir que el programa es eficaz?

a. Formule las hipótesis nula y alternativa.
b. Formule la regla de decisión.
c. Calcule el valor del estadístico de prueba.
d. ¿Cuál es su decisión respecto de la hipótesis nula?

6. Una muestra de 60 estudiantes universitarios participó en un programa de capacitación especial para mejorar su administración del tiempo. Un mes después de terminar el curso se contactó a los estudiantes y se les preguntó si las habilidades adquiridas en el programa fueron eficaces; en total, 42 respondieron que sí. Con el nivel de significancia 0.05, ¿es posible concluir que el programa es eficaz?
a. Formule las hipótesis nula y alternativa.
b. Formule su regla de decisión.
c. Calcule el valor del estadístico de prueba.
d. ¿Cuál es su decisión respecto de la hipótesis nula?

7. Pierre's Restaurant anunció que la noche del jueves el menú consistirá en platillos gourmet poco comunes, como calamar, conejo, caracoles de Escocia y hojas de diente de león. Como parte de un estudio más extenso, a una muestra de 81 comensales frecuentes se le preguntó si prefieren el menú normal o el gourmet; de ellos, 43 prefirieron la segunda opción. Con el nivel de significancia 0.02, ¿es posible concluir que los comensales prefieren el menú gourmet?

8. Los trabajadores de Costal Computers, Inc. (CC), ensamblan uno o dos montajes parciales y los insertan en un chasis. Los ejecutivos de CC consideran que los empleados estarían más orgullosos de su trabajo si ensamblaran todos los componentes y probaran la computadora terminada. Se seleccionó una muestra de 25 empleados para probar la idea; después de un programa de capacitación, se les preguntó su preferencia; a 20 les gustó ensamblar la unidad completa y probarla. Con el nivel de significancia 0.05, ¿es posible concluir que los empleados prefieren ensamblar toda la unidad? Explique los pasos que siguió para llegar a su decisión.

OA16-2
Realizar una prueba de hipótesis sobre la mediana utilizando la prueba de signos.

Prueba de hipótesis acerca de una mediana

La mayoría de las pruebas de hipótesis que se realizaron hasta este punto comprendieron la media de la población o una proporción. La prueba de los signos es una de las pocas pruebas con que se demuestra el valor de una mediana. Recuerde, del capítulo 3, que la mediana es el valor sobre el cual están la mitad de las observaciones y debajo del cual se encuentra la otra mitad. Para los honorarios por hora de 7, 9, 11 y 18 dólares, la mediana es 10 dólares; la mitad de los honorarios están arriba de esa cantidad, y la otra mitad, por debajo.

Para realizar una prueba de hipótesis, a un valor por arriba de la mediana se le da un signo más, y a un valor debajo de la mediana, un signo menos; si un valor es el mismo que la mediana, en el análisis posterior se le elimina.

EJEMPLO

En 2012, el Buró de Estadísticas Laborales de Estados Unidos reportó que la cantidad mediana que las familias estadounidenses gastan anualmente comiendo fuera de casa es de 2 505 dólares. La editora de la sección de alimentos del *Portland (Oregon) Tribune* desea saber si los ciudadanos de Portland difieren de este valor nacional del año previo en términos de comer fuera. Seleccionó una muestra aleatoria de 102 parejas y descubrió que 60 gastaban más de 2 505 dólares anuales por este concepto, 40 gastaban menos de esa cantidad y, sorprendentemente, dos gastaban exactamente ese dinero. Con el nivel de significancia 0.10, ¿es razonable concluir que la cantidad mediana gastada este año en Portland, Oregon, no es igual a 2 505 dólares?

SOLUCIÓN

Si la mediana de la población es 2 505 dólares, se espera que casi la mitad de las parejas muestreadas haya gastado más de esa cantidad el último año, y que casi toda la otra mitad haya gastado menos. Después de eliminar a las dos parejas que gastaron exactamente 2 505 dólares, se esperaría que 50 estén arriba de la mediana y 50 por debajo de ella. ¿Es posible que la diferencia entre las 60 parejas que gastaron más de esa cifra, y el número esperado de parejas que gastó más sea atribuible a la casualidad? ¿Es la mediana algún valor distinto a 2 505 dólares? La prueba estadística de la mediana ayudará a responder esta pregunta.

Las hipótesis nula y alternativa son:

$$H_0\text{: Mediana} = \$2\ 505$$
$$H_1\text{: Mediana} \neq \$2\ 505$$

Esta es una prueba de dos colas debido a que la hipótesis alternativa no indica una dirección; es decir, no interesa si la mediana es menor o mayor que 2 505 dólares, solo que es diferente. El estadístico de prueba cumple con las suposiciones binomiales; esto es:

1. Una observación es mayor o es menor que la mediana propuesta, por lo que solo hay dos resultados posibles.
2. La probabilidad de éxito permanece constante en 0.50; es decir, $\pi = 0.50$.
3. Los matrimonios seleccionados como parte de la muestra representan intentos independientes.
4. El número de éxitos se cuenta en un número fijo de intentos; en este caso, se consideran 100 parejas y se cuenta el número de los que gastan más de 2 505 dólares en comida fuera de casa.

El tamaño útil de la muestra es 100, y π es 0.50, por lo que $n\pi = 100(0.50) = 50$ y $n(1 - \pi) = 100(1 - 0.50) = 50$, que son mayores que 5, por lo que se utiliza la distribución normal para aproximar la binomial; es decir, en realidad se emplea la distribución normal estándar como el estadístico de prueba. El nivel de significancia es 0.10; por lo tanto, $\alpha/2 = 0.10/2 = 0.05$ del área se encuentra en cada cola de una distribución normal. Del apéndice B.5, en la fila con grados infinitos de libertad, los valores críticos son -1.645 y 1.645. La regla de decisión es rechazar H_0 si z es menor que -1.645 o mayor que 1.645.

Utilice la fórmula [16.2] para calcular z, porque 60 es mayor que $n/2$ o $(100/2 = 50)$.

$$z = \frac{(x - 0.5) - 0.5n}{0.50\sqrt{n}} = \frac{(60 - 0.5) - 50(100)}{0.50\sqrt{100}} = 1.90$$

Se rechaza la hipótesis nula debido a que el valor calculado (1.90) es mayor que el valor crítico (1.645). La evidencia de la muestra indica que la cantidad mediana gastada anualmente *no* es de 2 505 dólares. La editora de la sección de alimentos de Portland debe concluir que existe una diferencia en la cantidad mediana gastada anualmente en Portland, comparada con la reportada por el Buró de Estadísticas Laborales en 2012. El valor p es 0.0574, determinado mediante $2(0.5000 - 0.4713)$; el cual es menor que el nivel de significancia (0.10) para esta prueba. Así que, con base en el valor p y el nivel de significancia 0.10, la hipótesis nula también se rechaza y se concluye que las parejas de Portland gastan una cantidad diferente al valor mediano nacional.

AUTOEVALUACIÓN 16-3

Tras leer los resultados del estudio de Portland, Oregon, la editora de la sección de alimentos del *Tampa Times* decidió conducir un estudio similar; pero ella decide alterar ligeramente el estudio, investigando si las familias de su región gastan *más* que la cantidad mediana (2 505 dólares). Una muestra de 64 parejas de Tampa reveló que 42 gastaron más de esa cantidad por año comiendo fuera. Utilice el nivel de significancia 0.05.

EJERCICIOS

9. De acuerdo con el U.S. Department of Labor, en Estados Unidos el salario mediano de un quiropráctico es de 81 500 dólares al año. Un grupo de graduados recientes considera que esta cantidad es muy baja porque en una muestra de 205 quiroprácticos recién graduados, 170 iniciaron con un salario superior a 81 500 dólares, y cinco ganaban un salario de exactamente esa cantidad.
  a. Formule las hipótesis nula y alternativa.
  b. Formule la regla de decisión; utilice el nivel de significancia 0.15.
  c. Realice los cálculos necesarios e interprete los resultados.

10. Central Airlines afirma que la mediana del precio de un boleto de ida y vuelta de Chicago a Jackson Hole, Wyoming, es de 503 dólares; pero la Association of Travel Agents duda de esta afirmación y sostiene que la mediana del precio es menor que esa cantidad. Una muestra aleatoria de 400 boletos de ida y vuelta de Chicago a Jackson Hole reveló que 160 boletos costaban menos de 503 dólares y ninguno costaba exactamente esa cantidad. Sea $\alpha = 0.05$.
  a. Formule las hipótesis nula y alternativa.
  b. ¿Cuál es su decisión respecto de H_0? Haga un comentario sobre su decisión.

OA16-3
Realizar una prueba de hipótesis de poblaciones dependientes utilizando la prueba de los rangos con signo de Wilcoxon.

Prueba de rangos con signo de Wilcoxon para muestras dependientes

La prueba *t* por pares (o apareada) , que se describe en la sección "Pruebas de hipótesis de dos muestras: muestras dependientes" del capítulo 11, tiene dos requisitos. Primero, las muestras deben ser dependientes. Recuerde que estas se caracterizan por una medición, algún tipo de intervención y luego otra medición; por ejemplo, una compañía inició un programa de "bienestar" al inicio del año al cual se inscribieron 20 personas en la parte de reducción de peso del programa; para comenzar, se pesaron todos los participantes, luego se pusieron a dieta, hicieron ejercicio, etcétera, para reducir de peso. Al final del programa, que duró seis meses, todos los participantes se pesaron de nuevo; la diferencia entre sus pesos al inicio y al final del programa es la variable de interés. Observe que hay una medición, una intervención y luego otra medición.

El segundo requisito de la prueba *t* por pares es que la distribución de las diferencias siga la distribución normal de probabilidad. En el ejemplo sobre el bienestar de la compañía, esto se aplica a las diferencias entre los pesos de los 20 participantes; en ese caso, dicha suposición es razonable, sin embargo, hay casos en que interesarán las diferencias entre observaciones independientes y no se podrá suponer que la distribución de las diferencias se aproxima a una distribución normal. Con frecuencia, encontrará problemas con la suposición de normalidad cuando el nivel de medición en las muestras sea ordinal, en lugar de ser de intervalo o de razón; por ejemplo, suponga que hoy, en la Clínica 3, hay 10 pacientes en cirugía. La supervisora de enfermería pide a las enfermeras Benner y Jurris que califiquen a cada uno de los pacientes en una escala de 1 a 10 de acuerdo con la dificultad de los cuidados que deben recibir; como la distribución de las diferencias entre las calificaciones quizá no se aproxime a la distribución normal, no sería adecuado utilizar la prueba *t* por pares.

En 1945, Frank Wilcoxon desarrolló una prueba no paramétrica, con base en las diferencias entre muestras dependientes, que no requiere la suposición de normalidad; esta se denomina **prueba de rangos con signo de Wilcoxon**. En el siguiente ejemplo se dan los detalles de su aplicación.

EJEMPLO

Fricker's es una cadena de restaurantes familiares cuya mayoría de establecimientos está en el sureste de Estados Unidos. Ofrece un menú muy completo, pero su especialidad es el pollo; y hace poco, Bernie Frick, propietario y fundador, elaboró un nuevo sabor con especias para la salsa en la que se cocina el pollo. Antes de reemplazar el sabor actual, quiere realizar algunas pruebas para estar seguro de que a los comensales les gusta más este nuevo sabor.

Para iniciar, Bernie selecciona una muestra aleatoria de 15 clientes; y a cada uno le da una pieza de pollo actual y le pide que califique su sabor en una escala de 1 a 20. Un valor cercano a 20 indica

Participante	Calificación del sabor a especias	Calificación del sabor actual
Arquette	14	12
Jones	8	16
Fish	6	2
Wagner	18	4
Badenhop	20	12
Hall	16	16
Fowler	14	5
Virost	6	16
Garcia	19	10
Sundar	18	10
Miller	16	13
Peterson	18	2
Boggart	4	13
Hein	7	14
Whitten	16	4

que al participante le gustó el sabor, en tanto que una calificación cerca de 1 indica que no le gustó; luego, a los mismos participantes les da una muestra del pollo con el nuevo sabor a especias y vuelve a pedirles que califiquen su sabor en una escala de 1 a 20. Los resultados se muestran en la tabla de la página anterior. Con el nivel de significancia 0.05, ¿es razonable concluir que el sabor a especias es el preferido?

SOLUCIÓN

Se pidió a cada participante calificar los dos sabores del pollo. Así que las puntuaciones son dependientes (o están relacionadas) y, por cada participante, se calcula la diferencia entre la calificación del sabor a especias y la del sabor actual; el valor resultante muestra que la cantidad de participantes favorecen un sabor en comparación con el otro. Si elige restar la calificación del sabor actual a la calificación del sabor a especias, un resultado positivo es la "cantidad" con que los participantes prefieren el sabor a especias; las diferencias negativas de las calificaciones indican que el participante prefirió el sabor actual. Debido a la naturaleza subjetiva de las calificaciones, no es seguro que la distribución de las diferencias siga la distribución normal, por lo que es conveniente utilizar la prueba de rangos con signo de Wilcoxon no paramétrica.

Como es habitual, se emplea el procedimiento de prueba de hipótesis de seis pasos. La hipótesis nula es que no hay diferencias entre las calificaciones de los participantes de los sabores del pollo; la hipótesis alternativa es que las calificaciones son más altas para el sabor a especias. De manera más formal:

H_0: No hay diferencia entre las calificaciones de ambos sabores.
H_1: Las calificaciones son más altas para el sabor a especias.

Se trata de una prueba de una cola porque Bernie Frick, propietario de Fricker's, cambiará el sabor del pollo solo si los participantes en la muestra indican que a la población de clientes le gusta más el nuevo sabor. El nivel de significancia de la prueba es 0.05, como se indicó antes.

A continuación se indican los pasos para realizar la prueba de rangos con signo de Wilcoxon:

1. Calcule la diferencia entre la calificación del sabor a especias y la del sabor actual de cada participante; por ejemplo, la calificación del sabor a especias de Arquette fue de 14, y del sabor actual, de 12, por lo que la diferencia es 2. Para Jones, la diferencia es −8, determinada mediante 8 − 16, y para Fish es 4, determinada por 6 − 2. Las diferencias de todos los participantes se registran en la columna D de la tabla 16.3.

TABLA 16.3 Calificación de los sabores actual y de especias

A	B	C	D	E	F	G	H	I
Participante	Calificación del sabor a especias	Calificación del sabor actual	Diferencia entre calificaciones	Diferencia absoluta	Rango	Rango con signo R^+	R^-	
Arquette	14	12	2	2	1	1		
Jones	8	16	−8	8	6		6	
Fish	6	2	4	4	3	3		
Wagner	18	4	14	14	13	13		
Badenhop	20	12	8	8	6	6		
Hall	16	16	0	*	*			
Fowler	14	5	9	9	9	9		
Virost	6	16	−10	10	11		11	
Garcia	19	10	9	9	9	9		
Sundar	18	10	8	8	6	6		
Miller	16	13	3	3	2	2		
Peterson	18	2	16	16	14	14		
Boggart	4	13	−9	9	9		9	
Hein	7	14	−7	7	4		4	
Whitten	16	4	12	12	12	12		
					Suma	75	30	Suma del menor rango

2. En el análisis posterior solo se consideran las diferencias positivas y negativas; es decir, si la diferencia entre las calificaciones del sabor es 0, ese participante se elimina de un análisis pos-

terior y se reduce el número de integrantes de la muestra. De la tabla 16.3, Hall, el sexto participante, calificó al sabor a especias y al actual con 16; por lo tanto, se elimina del estudio y se reduce el tamaño útil de la muestra de 15 a 14.

3. Determine las diferencias absolutas de los valores calculados en la columna D; recuerde que la diferencia absoluta ignora el signo de la diferencia y se enfoca en la magnitud de las diferencias entre calificaciones. Las diferencias absolutas se muestran en la columna E.
4. Luego, ordene las diferencias absolutas de menor a mayor. Arquette, el primer participante, calificó al pollo con especias con 14 y al actual con 12; la diferencia de 2 en las dos calificaciones del sabor es la diferencia absoluta menor, por lo cual se le asigna un rango de 1. La siguiente diferencia mayor es 3, de Miller, por lo que se le asigna un rango de 2. Las otras diferencias se ordenan de manera similar. Hay tres participantes que calificaron la diferencia entre los sabores con 8; es decir, Jones, Badenhop y Sundar tuvieron esa diferencia entre la calificación del sabor a especias y la del sabor actual. Para resolver este problema, promedie sus clasificaciones y registre la clasificación promedio de cada uno; esta situación comprende las clasificaciones de 5, 6 y 7, de modo que a los tres participantes se les asigna la clasificación de 6; es la misma situación de los participantes con una diferencia de 9; donde las clasificaciones comprendidas son 8, 9 y 10, de manera que a estos participantes se les asigna una clasificación de 9.
5. A cada clasificación asignada en la columna F se le da el mismo signo que tenía en la diferencia original, y los resultados se reportan en las columnas G o H; por ejemplo, el segundo participante tiene una diferencia de -8 y un rango de 6. Este valor se coloca en la sección R^- de la columna H.
6. Finalmente, se obtienen los totales de las columnas R^+ y R^-. La suma de los rangos positivos es 75, y la suma de los rangos negativos es 30; la menor de las dos sumas de los rangos se utiliza como el estadístico de prueba y se conoce como T.

En el apéndice B.8 aparecen los valores críticos de la prueba de rangos con signo de Wilcoxon; una parte de esa tabla se muestra a continuación (la fila α se utiliza para pruebas de una cola, y la fila 2α, para pruebas de dos colas). En este caso se desea demostrar que a los clientes les gusta más el sabor a especias, que es una prueba de una cola, por lo que seleccione la fila α. Elija el nivel de significancia 0.05 y vaya hasta la columna con el encabezado 0.05; baje por la columna hasta la fila donde n es 14 (recuerde que una persona calificó igual a ambos sabores y fue eliminada del estudio; entonces, el tamaño útil de la muestra es 14). El valor en la intersección (el valor crítico) es 25; la regla de decisión es rechazar la hipótesis nula si el *menor* de los totales de los rangos es 25 o menos. El valor que se obtuvo del apéndice B.8 es el *valor mayor en la región de rechazo*; en otras palabras, la regla de decisión es rechazar H_0 si la menor de las dos sumas de los rangos es 25 o menor. En este caso, la suma menor del rango es 30; en consecuencia, la decisión es no rechazar la hipótesis nula. No es posible concluir que hay una diferencia entre las calificaciones del sabor actual y el sabor a especias; en el estudio no se pudo demostrar que los clientes prefieran el nuevo sabor. El señor. Frick debería continuar con el sabor actual del pollo.

	2α 0.15	**0.10**	**0.05**	**0.04**	**0.03**	**0.02**	**0.01**
n	**α 0.075**	**0.05**	**0.025**	**0.02**	**0.015**	**0.01**	**0.005**
4	0						
5	1	0					
6	2	2	0	0			
7	4	3	2	1	0	0	
8	7	5	3	3	2	1	0
9	9	8	5	5	4	3	1
10	12	10	8	7	6	5	3
11	16	13	10	9	8	7	5
12	19	17	13	12	11	9	7
13	24	21	17	16	14	12	9
14	28	25	21	19	18	15	12
15	33	30	25	23	21	19	15

AUTOEVALUACIÓN 16-4

El área de ensamblado de Gotrac Products se rediseñó hace poco; la instalación de un nuevo sistema de iluminación y la compra de nuevas mesas de trabajo son dos características de las modificaciones. El supervisor de producción quiere saber si los cambios generaron un aumento de la productividad de los empleados y, con el fin de investigar esta cuestión, seleccionó una muestra de 11 empleados para determinar las tasas de producción antes y después de los cambios; en la página siguiente se indica la información de la muestra.

(a) ¿Cuántos pares útiles hay? Es decir, ¿cuál es el valor de *n*?
(b) Mediante la prueba de rangos con signo de Wilcoxon, determine si en realidad los nuevos procedimientos incrementaron la producción; utilice el nivel de significancia 0.05 y una prueba de una cola.
(c) ¿Qué suposición debe hacer acerca de la distribución de las diferencias entre las producciones antes y después del rediseño?

Operador	Producción antes	Producción después	Operador	Producción antes	Producción después
S. M.	17	18	U. Z.	10	22
D. J.	21	23	Y. U.	20	19
M. D.	25	22	U. T.	17	20
B. B.	15	25	Y. H.	24	30
M. F.	10	28	Y. Y.	23	26
A. A.	16	16			

EJERCICIOS

11. Un psicólogo industrial seleccionó una muestra aleatoria de siete parejas de profesionales urbanos jóvenes que viven en casa propia, y comparó el tamaño de su casa (en pies cuadrados) con la de sus padres. Con el nivel de significancia 0.05, ¿es posible concluir que las parejas de profesionales viven en casas más grandes que las de sus padres?

Apellido de la pareja	Profesionales	Padres	Apellido de la pareja	Profesionales	Padres
Gordon	1 725	1 175	Kuhlman	1 290	1 360
Sharkey	1 310	1 120	Welch	1 880	1 750
Uselding	1 670	1 420	Anderson	1 530	1 440
Bell	1 520	1 640			

12. La Toyota Motor Company estudia el efecto de la gasolina regular en comparación con la de alto octanaje sobre el ahorro de combustible de su nuevo motor V6 de alto desempeño de 3.5 litros; se seleccionan 10 ejecutivos y se les pide registrar el número de millas que recorren por galón. Los resultados son:

	Millas por galón			Millas por galón	
Ejecutivo	Regular	Alto octanaje	Ejecutivo	Regular	Alto octanaje
Bowers	25	28	Rau	38	40
Demars	33	31	Greolke	29	29
Grasser	31	35	Burns	42	37
DeToto	45	44	Snow	41	44
Kleg	42	47	Lawless	30	44

Con el nivel de significancia 0.05, ¿hay alguna diferencia entre las millas que recorren por galón con gasolina regular y con la de alto octanaje?

13. El señor Mump sugiere un nuevo procedimiento para incrementar la producción de la línea de ensamblado; para probar si este es mejor que el anterior, selecciona una muestra aleatoria de 15 trabajadores de la línea. Se determina el número de unidades que se producen en una hora con el procedimiento anterior y luego se aplica el nuevo procedimiento de Mump y, después de un periodo prudente para que los operarios se familiarizaran con el nuevo procedimiento, se midió de nuevo su producción. Los resultados son:

	Producción			Producción	
Empleado	Sistema anterior	Sistema de Mump	Empleado	Sistema anterior	Sistema de Mump
A	60	64	I	87	84
B	40	52	J	80	80
C	59	58	K	56	57
D	30	37	L	21	21
E	70	71	M	99	108
F	78	83	N	50	56
G	43	46	O	56	62
H	40	52			

Con el nivel de significancia 0.05, ¿es posible concluir que la producción aumenta con el sistema de Mump?

a. Formule las hipótesis nula y alternativa.
b. Formule la regla de decisión.
c. Llegue a una decisión respecto de la hipótesis nula.

14. Se sugirió que la producción diaria de una parte de subensamblado aumentaría si se instalara una mejor iluminación, se tocara música de fondo y se ofreciera café y rosquillas gratis durante el día; por lo tanto, la gerencia acordó probar el esquema durante cierto tiempo. El número de subensamblados que una muestra de empleados producen por día es el siguiente.

Empleado	Registro de la producción anterior	Producción después de los cambios	Empleado	Registro de la producción anterior	Producción después de los cambios
JD	23	33	WWJ	21	25
SB	26	26	OP	25	22
MD	24	30	CD	21	23
RCF	17	25	PA	16	17
MF	20	19	RRT	20	15
UHH	24	22	AT	17	9
IB	30	29	QQ	23	30

Aplique la prueba de rangos con signo de Wilcoxon y determine si los cambios sugeridos valen la pena.
a. Formule la hipótesis nula.
b. Decida sobre la hipótesis alternativa.
c. Elija el nivel de significancia.
d. Formule la regla de decisión.
e. Calcule T y tome una decisión.
f. ¿Qué supuso acerca de la distribución de las diferencias?

OA16-4
Realizar una prueba de hipótesis de poblaciones independientes utilizando la prueba de la suma de los rangos de Wilcoxon.

Prueba de Wilcoxon de la suma de rangos de muestras independientes

Un procedimiento diseñado para determinar si dos muestras *independientes* provienen de poblaciones equivalentes es la **prueba de Wilcoxon de la suma de rangos**; la cual es una alternativa a la prueba t de dos muestras que se describe en la sección “Prueba de dos muestras agrupadas” del capítulo 11. Recuerde que la prueba t requiere que ambas poblaciones sigan la distribución normal y tengan varianzas poblacionales iguales. La prueba de Wilcoxon de la suma de rangos no requiere estas condiciones.

Esta prueba se basa en la suma de los rangos; por lo tanto, los datos se clasifican como si fueran parte de una sola muestra. Si la hipótesis nula es verdadera, los rangos tendrán una distribución casi uniforme entre las dos muestras, y la suma de los rangos de las dos muestras será casi igual; es decir, los rangos bajo, medio y alto deberán dividirse en forma equitativa entre las dos muestras. Si la hipótesis alternativa es verdadera, una de las muestras tendrá mayor cantidad de rangos bajos y, por lo tanto, una suma de rangos menor; además, la otra muestra tendrá mayor cantidad de rangos altos, por lo que la suma de rangos será mayor. Si cada una de las muestras contiene *al menos ocho observaciones*, se utiliza la distribución normal estándar como estadístico de prueba. El valor del estadístico de prueba se estima mediante la siguiente fórmula.

PRUEBA DE WILCOXON DE LA SUMA DE RANGOS

$$z = \frac{W - \dfrac{n_1(n_1 - n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}} \qquad \textbf{[16.4]}$$

donde:

n_1 es el número de observaciones de la primera muestra;
n_2 es el número de observaciones de la segunda muestra;
W es la suma de los rangos de la primera población.

EJEMPLO

Dan Thompson, presidente de CEO Airlines, observó recientemente un aumento del número de personas que no llegan a tomar los vuelos que salen de Atlanta; por lo tanto, su interés principal es determinar si hay más personas que no se presentan a tomar los vuelos que salen de Atlanta en comparación con vuelos que salen de Chicago. En la tabla 16.4 se registra una muestra de nueve vuelos de Atlanta y ocho de Chicago. Con el nivel de significancia 0.05, ¿es posible concluir que hay más personas que no se presentan a tomar los vuelos que salen de Atlanta?

TABLA 16.4 Número de personas que no se presentan a los vuelos programados

Atlanta	Chicago
11	13
15	14
10	10
18	8
11	16
20	9
24	17
22	21
25	

SOLUCIÓN

Si las poblaciones de personas que no se presentan a tomar los vuelos siguen la distribución normal de probabilidad y tienen varianzas iguales, es adecuada la prueba *t* de dos muestras que se estudió en la sección "Pruebas de medias de dos muestras" del capítulo 11. En este caso, Thompson considera que ambas condiciones no se pueden cumplir; por lo tanto, la prueba adecuada es la no paramétrica de Wilcoxon de la suma de rangos.

Si el número de personas que no se presentan a tomar los vuelos es el mismo en Atlanta que en Chicago, ambas poblaciones serán casi iguales; en otras palabras, el rango promedio de ambos grupos será aproximadamente el mismo. Si el número de personas que no se presentan no es el mismo, se espera que el promedio de los rangos sea muy diferente.

Thompson considera que más personas pierden su vuelo en Atlanta; por ello, es adecuada una prueba de una cola. Las hipótesis nula y alternativa son:

H_0: El número de personas que no se presentan es la misma o menor en Atlanta que en Chicago.
H_1: El número de personas que no se presentan en Atlanta es mayor que en Chicago.

El estadístico de prueba sigue la distribución normal estándar. Con el nivel de significancia 0.05 se determina, del apéndice B.5, que el valor crítico de *z* es 1.645; la hipótesis nula se rechaza si el valor calculado de *z* es mayor que 1.645.

La hipótesis alternativa es que hay más personas que no se presentan en Atlanta, lo que significa que la distribución se ubica a la derecha de la distribución de Chicago. Los detalles de la asignación del rango se muestran en la tabla 16.5; además, las observaciones de *ambas* muestras se clasificaron como si fueran un solo grupo. El vuelo de Chicago con solo 8 personas que no se presentaron tuvo la menor cantidad, por lo que se le asignó un rango de 1, al vuelo de Chicago con 9 personas ausentes, un rango de 2, y así en lo sucesivo. El vuelo de Atlanta con 25 personas que no se presentaron es el mayor, por lo que se le asigna el mayor rango (17). También hay dos casos de rangos iguales. Hay un vuelo de Atlanta y otro de Chicago a los que no se presentaron 10 personas, y dos vuelos de Atlanta con 11 asientos vacíos. ¿Cómo manejar estos empates? La solución es promediar los rangos y asignar el rango promedio a los dos vuelos; en el caso que comprende 10 personas que no se presentaron, los rangos comprendidos son 3 y 4. La media de estos es 3.5, por lo que se asigna ese rango a los dos vuelos de Atlanta y de Chicago con 10 personas ausentes.

TABLA 16.5 Números de rango de las personas que no se presentaron a los vuelos programados

Atlanta		Chicago	
No se presentaron	Rango	No se presentaron	Rango
11	5.5	13	7
15	9	14	8
10	3.5	10	3.5
18	12	8	1
11	5.5	16	10
20	13	9	2
24	16	17	11
22	15	21	14
25	17		
	96.5		56.5

Suma de rangos de Atlanta

La suma de rangos de los vuelos de Atlanta es 96.5 (este es el valor de *W* en la fórmula [16.4]); en la tabla 16.5 se observan nueve vuelos que salen de Atlanta y ocho de Chicago, por lo que $n_1 = 9$ y $n_2 = 8$. Al calcular *z* a partir de la fórmula [16.4]:

$$z = \frac{W - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}} = \frac{96.5 - \dfrac{9(9 + 8 + 1)}{2}}{\sqrt{\dfrac{9(8)(9 + 8 + 1)}{12}}} = 1.49$$

Como el valor *z* calculado (1.49) es menor que 1.645, no se rechaza la hipótesis nula. La evidencia no muestra una diferencia entre las distribuciones de los números de personas que no se presentaron; es decir, parece que el número de personas que pierden el vuelo es el mismo en Atlanta que en Chicago. El valor *p* es 0.0681, que se encontró al determinar el área a la derecha de 1.49 (0.5000 − 0.4319), indica el mismo resultado.

El software de MegaStat produce los mismos resultados: el valor *p* de MegaStat es 0.0679, que se aproxima al valor anterior; la diferencia es por el redondeo del sistema y la corrección de los empates.

Wilcoxon – Prueba de U Mann-Whitney

n	suma de rangos	
9	96.5	Atlanta
8	56.5	Chicago
17	153	Total
	81.000	valor esperado
	10.392	desviación estándar
	1.491	z
	0.0679	valor p (una cola, superior)

Al emplear la prueba de Wilcoxon de la suma de los rangos, puede numerar las dos poblaciones en cualquier orden; sin embargo, una vez que haga una elección, W debe ser la suma de los rangos identificados como la población 1. Si, en el ejemplo de las personas que no se presentaron a los vuelos, la población de Chicago se identificara como el número 1, la dirección de la hipótesis alternativa cambiaría, pero el *valor absoluto de z* aún sería el mismo.

H_0: La distribución de la población de personas que no se presentaron en Chicago es la misma o mayor que en Atlanta.

H_1: La distribución de la población de personas que no se presentaron en Chicago es menor que en Atlanta.

El valor calculado de z es -1.49, determinado por:

$$z = \frac{W - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}} = \frac{56.5 - \dfrac{8(8 + 9 + 1)}{2}}{\sqrt{\dfrac{8(9)(8 + 9 + 1)}{12}}} = -1.49$$

La conclusión es la misma que antes: no hay una diferencia entre los números habituales de personas que no se presentaron en Chicago y Atlanta.

AUTOEVALUACIÓN 16-5

El director de investigación de Top Flite quiere saber si hay una diferencia entre las distribuciones de las distancias recorridas por dos pelotas de golf de la compañía. Se lanzaron ocho pelotas de su modelo XL-5000 y ocho D2 con un dispositivo automático; las distancias (en yardas) son las siguientes:

XL-5000:	252, 263, 279, 273, 271, 265, 257, 280
D2:	262, 242, 256, 260, 258, 243, 239, 265

No suponga que las distribuciones de las distancias recorridas siguen la distribución normal de probabilidad. Con el nivel de significancia 0.05, ¿hay alguna diferencia entre ambas distribuciones?

EJERCICIOS

15. Se seleccionaron ocho observaciones de manera aleatoria de dos poblaciones (A y B), que no necesariamente tenían una distribución normal. Utilice el nivel de significancia 0.05, una prueba de dos colas y la prueba de Wilcoxon de la suma de los rangos para determinar si hay una diferencia entre ambas poblaciones.

Población A	38, 45, 56, 57, 61, 69, 70, 79
Población B	26, 31, 35, 42, 51, 52, 57, 62

16. Se seleccionaron aleatoriamente nueve observaciones de la población A y ocho de la población B, que no necesariamente tienen una distribución normal. Utilice el nivel de significancia 0.05, una prueba de dos colas y la prueba de Wilcoxon de la suma de los rangos para determinar si hay una diferencia entre ambas poblaciones.

Población A	12, 14, 15, 19, 23, 29, 33, 40, 51
Población B	13, 16, 19, 21, 22, 33, 35, 43

17. La Tucson State University ofrece dos programas de maestría en administración de empresas; en el primero, los estudiantes se reúnen dos noches por semana en el campus principal, en el centro de Tucson; en el segundo, solo se comunican por internet con el profesor. El director de la maestría de Tucson quiere comparar el número de horas que estudiaron la semana pasada ambos grupos; mediante una muestra de 10 estudiantes en el campus y otra de 12 estudiantes por internet se obtuvo la siguiente información.

Campus	28, 16, 42, 29, 31, 22, 50, 42, 23, 25
Por internet	26, 42, 65, 38, 29, 32, 59, 42, 27, 41, 46, 18

No suponga que las dos distribuciones del tiempo de estudio, que se reportan en horas, siguen una distribución normal. Con el nivel de significancia 0.05, ¿es posible concluir que los estudiantes por internet estudian más?

18. En fechas recientes, debido a los bajos niveles de las tasas hipotecarias, las instituciones financieras han comenzado a ofrecer mayores beneficios a los clientes; en ese sentido, una innovación de Coastal National Bank and Trust es la presentación de solicitudes por internet. En la siguiente tabla se muestra el tiempo (en minutos) necesario para completar el proceso de solicitud para ocho clientes que piden un préstamo hipotecario de tasa fija a 15 años y nueve clientes que solicitan un préstamo a tasa fija a 30 años.

Tasa fija a 15 años	41, 36, 42, 39, 36, 48, 49, 38
Tasa fija a 30 años	21, 27, 36, 20, 19, 21, 39, 24, 22

Con el nivel de significancia 0.05, ¿es posible concluir que el proceso que deben cubrir los clientes que solicitan un préstamo hipotecario a tasa fija a 30 años tarda menos? No suponga que la distribución del tiempo sigue una distribución normal para ningún grupo.

Prueba de Kruskal-Wallis: análisis de la varianza por rangos

OA16-5
Realizar una prueba de hipótesis de varias poblaciones independientes utilizando la prueba de Kruskal-Wallis.

El procedimiento del análisis de la varianza (ANOVA) que se estudió en el capítulo 12 se relaciona con la igualdad de las medias de diversas poblaciones. Los datos estaban en un nivel de intervalo o de razón; asimismo, se supuso que las poblaciones seguían la distribución normal de probabilidad y que sus desviaciones estándar eran iguales. ¿Qué sucede si los datos están en escala ordinal o las poblaciones no siguen una distribución normal? En 1952, W.H. Kruskal y W.A. Wallis reportaron una prueba no paramétrica que solo requería datos de nivel ordinal (clasificados), pero no precisaba suposiciones acerca de la forma de las poblaciones. A la prueba se le conoce como **análisis en una dirección de la varianza por rangos de Kruskal-Wallis**.

Para aplicar la prueba de Kruskall-Wallis, las muestras seleccionadas de la población deben ser *independientes*; por ejemplo, si selecciona y entrevista muestras de tres grupos —ejecutivos, personal y supervisores—, las respuestas de un grupo (ejecutivos) no deben por ningún motivo influir en las respuestas de los demás.

Para calcular el estadístico de prueba de Kruskal-Wallis, 1) se combinan todas las muestras, 2) se ordenen los valores combinados de bajo a alto y 3) los valores ordenados se *reemplazan por rangos, a partir de 1 para el valor menor*. En el siguiente ejemplo se aclaran los detalles del procedimiento.

EJEMPLO

El Hospital System of the Carolinas opera tres hospitales en el área de Great Charlotte: St. Luke's Memorial, en el lado poniente de la ciudad, Swedish Medical Center, al sur, y el Piedmont Hospital en el lado este. Al director de administración le preocupa el tiempo de espera de los pacientes con lesiones de tipo deportivo, que no ponen en peligro la vida, y que llegan durante las tardes entre semana a los tres hospitales; específicamente, ¿existe una diferencia en los tiempos de espera en los tres hospitales?

SOLUCIÓN

Para averiguarlo, el director seleccionó una muestra aleatoria de pacientes en los tres hospitales y determinó el tiempo (en minutos) en que se llega al hospital y el momento en que termina el tratamiento; los tiempos se reportan en la tabla 16.6.

TABLA 16.6 Tiempos de espera de los pacientes en la sala de urgencias en el Hospital System of the Carolinas

St. Luke's Memorial	Swedish Medical Center	Piedmont Hospital
56	103	42
39	87	38
48	51	89
38	95	75
73	68	35
60	42	61
62	107	
	89	

En la tabla 16.6 se observa que el tiempo de espera más corto (35 minutos) es del quinto paciente muestreado en el Piedmont Hospital; el tiempo más largo (107 minutos) se registró con el séptimo paciente muestreado en el Swedish Medical Center.

Probablemente, el primer enfoque para comparar los periodos de espera es determinar si existe una diferencia entre los tiempos medios en los tres hospitales, esto es, utilizar la ANOVA de una vía que se describe en la sección "ANOVA: Análisis de la varianza" del capítulo 12; sin embargo, esta prueba tiene tres requisitos:

1. Las muestras deben ser de poblaciones independientes.
2. Las varianzas de la población deben ser iguales.
3. Las muestras deben ser de poblaciones normales.

En este caso, las muestras provienen de poblaciones independientes (los tres hospitales); pero suponga que no quiere asumir que hay una varianza igual en los tiempos de espera en los tres hospitales o que estos siguen una distribución de probabilidad normal; la falta de ambos criterios significa que no se cubren los requisitos de ANOVA, así que no se puede utilizar esta técnica; en vez de eso, recurra a la prueba de Kruskal-Wallis, donde no se requieren estas suposiciones.

El primer paso en la prueba de hipótesis es formular las hipótesis nula y alternativa.

H_0: Las distribuciones de las poblaciones de los tiempos de espera son iguales para los tres hospitales.

H_1: No todas las distribuciones de las poblaciones son iguales.

El director de administración seleccionó el nivel de significancia 0.05.

El estadístico de prueba de la prueba de Kruskal-Wallis se designa como H, y su fórmula es:

PRUEBA DE KRUSKAL-WALLIS

$$H = \frac{12}{n(n+1)}\left[\frac{(\Sigma R_1)^2}{n_1} + \frac{(\Sigma R_2)^2}{n_2} + \cdots + \frac{(\Sigma R_k)^2}{n_k}\right] - 3(n+1) \qquad \textbf{[16.5]}$$

con $k - 1$ grados de libertad (k es el número de poblaciones), donde:

$\Sigma R_1, \Sigma R_2, \ldots, \Sigma R_k$ son las sumas de los rangos de las muestras 1, 2, . . . , k, respectivamente;
$n_1, n_2, \ldots, n_k$ son los tamaños de las muestras 1, 2, . . . , k, respectivamente;
n es el número combinado de observaciones de todas las muestras.

La distribución del estadístico de prueba H es muy similar a la distribución *ji* cuadrada con $k - 1$ grados de libertad; es preferible que cada muestra incluya al menos cinco observaciones, y utilizar *ji* cuadrada para formular la regla de decisión. En este ejemplo hay tres poblaciones: los pacientes de St. Luke's Memorial, los del Swedish Medical Center, y los del Piedmont Hospital; por lo tanto, hay $k - 1$, es decir, $3 - 1 = 2$ grados de libertad. Consulte la tabla de *ji* cuadrada de los valores críticos en el apéndice B.7; el valor crítico de 2 grados de libertad y el nivel de significancia 0.05 es 5.991. No rechace H_0 si el valor calculado del estadístico de prueba H es menor o igual a 5.991; rechace H_0 si el valor calculado de H es mayor que 5.991 y acepte H_1.

A continuación se determina el valor del estadístico de prueba; reemplace los tiempos de espera en los tres hospitales por sus rangos correspondientes. Considerando los tiempos de espera como una sola población, el paciente de Piedmont con un tiempo de espera de 35 minutos aguardó el tiempo más corto y, por lo tanto, se le otorga el rango más bajo (1). Hay dos pacientes que esperaron 38 minutos, uno en St. Luke's y el otro en Piedmont; para resolver este empate, se otorga a cada paciente un rango de 2.5, calculado mediante $(2 + 3)/2$. El proceso continúa con todos los tiempos de espera; el más largo es de 107 minutos, y ese paciente del Swedish Medical Center recibe un rango de 21. En la tabla 16.7 se muestran las calificaciones, los rangos y la suma de los rangos en cada uno de los tres hospitales.

Al despejar H, se obtiene

$$H = \frac{12}{n(n+1)}\left[\frac{(\Sigma R_1)^2}{n_1} + \frac{(\Sigma R_2)^2}{n_2} + \frac{(\Sigma R_3)^2}{n_k}\right] - 3(n+1)$$

$$= \frac{12}{21(21+1)}\left[\frac{58.5^2}{7} + \frac{120^2}{8} + \frac{52.5^2}{6}\right] - 3(21+1) = 5.38$$

Como el valor calculado de H (5.38) es menor que el valor crítico de 5.991, no se rechaza la hipótesis nula; por lo tanto, no hay evidencia suficiente para concluir que existe una diferencia entre los tiempos de espera en los tres hospitales.

TABLA 16.7 Tiempos de espera, rangos y suma de rangos en el Hospital System of the Carolinas

St. Luke's Memorial		Swedish Medical Center		Piedmont Hospital	
Tiempo de espera	**Rango del tiempo de espera**	**Tiempo de espera**	**Rango del tiempo de espera**	**Tiempo de espera**	**Rango del tiempo de espera**
56	9	103	20	42	5.5
39	4	87	16	38	2.5
48	7	51	8	89	17.5
38	2.5	95	19	75	15
73	14	68	13	35	1
60	10	42	5.5	61	11
62	12	107	21		
		89	17.5		
	$\Sigma R_1 = 58.5$		$\Sigma R_2 = 120$		$\Sigma R_3 = 52.5$

Sumas de los rangos de tiempos de espera

También es posible realizar el procedimiento de Kruskal-Wallis con el software complementario de Excel, MegaStat. A la derecha se muestra la salida en pantalla del ejemplo respecto del conocimiento de los principios de administración de ejecutivos de varias industrias. El valor calculado de H es 5.39, y el valor p que se reporta en la salida es 0.067; observe que los valores son ligeramente distintos debido al redondeo. Con base en estos resultados, se determina que la decisión y la conclusión son iguales.

Prueba de Kruskal-Wallis				
	Mediana	*N*	*Rango promedio*	
	86.00	7	8.36	St. Luke's
	88.00	8	15.00	Swedish
	51.50	8	8.75	Piedmont
	61.00	21		Total
			5.39	*H* (corregido por empates)
			2	*gl*
			0.067	valor *p*

Recuerde, del capítulo 12, que los supuestos para la aplicación de la técnica del análisis de la varianza son: 1) las poblaciones están normalmente distribuidas, 2) estas tienen desviaciones estándar iguales y 3) las muestras se seleccionan de manera independiente. Si se cumplen en el ejemplo de los tiempos de espera en los hospitales, utilice la distribución F como estadístico de prueba; si no es así, aplique la prueba de Kruskal-Wallis sin distribución. Para resaltar las similitudes entre ambos enfoques, se resuelve el ejemplo respecto del conocimiento de los principios de administración de ejecutivos mediante la técnica ANOVA.

Para iniciar, formule las hipótesis nula y alternativa de los tres grupos.

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: No todas las medias de tratamiento son iguales.

Con el nivel de significancia 0.05, con $k - 1 = 3 - 1 = 2$ grados de libertad en el numerador y $n - k = 21 - 3 = 18$ grados de libertad en el denominador, el valor crítico de F es 3.55. La regla de decisión es rechazar la hipótesis nula si el valor calculado de F es mayor que 3.55. A continuación se muestra la salida en pantalla con Excel.

	A	B	C	D	E	F	G	H	I	J	K
1	St Luke's	Swedisch	Piedmont		ANOVA de una vía						
2	56	103	42								
3	39	87	38		Resumen						
4	48	51	89		*Grupos*	*Cuenta*	*Suma*	*Promedio*	*Varianza*		
5	38	95	75		St Luke's	7	376	53.714	163.571		
6	73	68	35		Swedisch	8	642	80.250	577.357		
7	60	42	61		Piedmont	6	340	56.667	486.667		
8	62	107									
9		89									
10					ANOVA						
11					*Fuente de variación*	*SS*	*gl*	*MS*	*F*	*Valor P*	*F crítica*
12					Entre grupos	3166.4	2	1583.202	3.822	0.041	3.555
13					En los grupos	7456.26	18	414.24			
14											
15					TOTAL	10622.7	20				

En la salida anterior, el valor calculado de F es 3.822, y el valor p, 0.41; la decisión es rechazar la hipótesis nula y aceptar la hipótesis alternativa. Utilizando la prueba ANOVA de una vía, se concluye que los tiempos de espera medios en los tres hospitales del Hospital System of the Carolinas son distintos.

¿Por qué hay conclusiones contradictorias sobre los mismos datos? Si compara los resultados con el empleo de valores p, las respuestas son similares. En el caso de la prueba de Kruskal-Wallis el valor p fue 0.067, que solo es un poco mayor que el nivel de significancia 0.05, pero la regla de decisión fue no rechazar H_0. El valor p mediante ANOVA es 0.041, que no es mucho menor que el valor crítico en la región de rechazo; entonces, para resumir, apenas falló en rechazar H_0 con la prueba de Kruskal-Wallis y apenas estuvo en la región de rechazo mediante ANOVA; la diferencia entre los valores p es 0.026; por lo tanto, en realidad los resultados están muy cercanos en términos de los valores p.

AUTOEVALUACIÓN 16-6

Al gerente del banco regional Statewide Financial Bank le interesa el índice de movimientos de dinero de las cuentas de cheques personales en cuatro sucursales (el índice de movimientos es la velocidad a la que el dinero en una cuenta se deposita y se retira; una cuenta extremadamente activa puede tener un índice de 300; si solo se emiten uno o dos cheques, el índice puede ser de 30, aproximadamente). Los índices de rotación de las muestras seleccionadas de las cuatro sucursales se registran en la siguiente tabla; con el nivel de significancia 0.01 y la prueba de Kruskal-Wallis, determine si hay una diferencia entre los índices de rotación de las cuentas de cheques personales de las cuatro sucursales.

Sucursal Englewood	Sucursal West Side	Sucursal Great Northern	Sucursal Sylvania
208	91	302	99
307	62	103	116
199	86	319	189
142	91	340	103
91	80	180	100
296			131

EJERCICIOS

19. ¿En qué condiciones debe utilizar la prueba de Kruskal-Wallis en lugar del análisis de la varianza?

20. ¿En qué condiciones debe utilizar la prueba de Kruskal-Wallis en lugar de la prueba de Wilcoxon de la suma de los rangos?

21. Los siguientes datos de la muestra se obtuvieron de tres poblaciones que no siguen una distribución normal.

Muestra 1	Muestra 2	Muestra 3
50	48	39
54	49	41
59	49	44
59	52	47
65	56	51
	57	

a. Formule la hipótesis nula.
b. Con el nivel de significancia 0.05, formule la regla de decisión.
c. Calcule el valor del estadístico de prueba.
d. ¿Cuál es su decisión respecto de la hipótesis nula?

22. Los siguientes datos de una muestra se obtuvieron de tres poblaciones donde las varianzas no son iguales, pero usted quiere compararlas.

Muestra 1	Muestra 2	Muestra 3
21	15	38
29	17	40
35	22	44
45	27	51
56	31	53
71		

a. Formule la hipótesis nula.
b. Con el nivel de significancia 0.01, formule la regla de decisión.
c. Calcule el valor del estadístico de prueba.
d. ¿Cuál es su decisión respecto de la hipótesis nula?

23. Hace poco, Davis Outboard Motors, Inc., desarrolló un proceso de pintura epóxica para proteger los componentes de sus sistemas de escape contra la oxidación. Bill Davis, el propietario, quiere determinar si la duración de la vida útil de la pintura es igual en tres condiciones diferentes: agua salada, agua dulce sin algas y agua dulce con una alta concentración de algas; por lo tanto, realizó pruebas aceleradas de la duración en el laboratorio y registró el número de horas que duró la pintura sin caerse. Probó cinco botes para cada condición.

Agua salada	Agua dulce	Agua dulce con algas
167.3	160.6	182.7
189.6	177.6	165.4
177.2	185.3	172.9
169.4	168.6	169.2
180.3	176.6	174.7

Utilice la prueba de Kruskal-Wallis y el nivel de significancia 0.01 para determinar si la calidad de duración de la pintura es la misma en las tres condiciones.

24. La National Turkey Association quiere experimentar con tres mezclas diferentes de alimentos para pavos muy jóvenes; pero no existen registros respecto de las tres mezclas, así que es imposible hacer suposiciones acerca de la distribución de los pesos. Para estudiar los efectos de las tres mezclas, cinco pavos reciben el alimento A, seis el B y otros cinco el C. Con el nivel de significancia 0.05, evalúe la hipótesis de que la mezcla de alimento no tiene efecto en el peso.

Peso (en libras)		
Mezcla de alimento A	**Mezcla de alimento B**	**Mezcla de alimento C**
11.2	12.6	11.3
12.1	10.8	11.9
10.9	11.3	12.4
11.3	11.0	10.6
12.0	12.0	12.0
	10.7	

OA16-5
Realizar e interpretar una prueba de hipótesis no paramétrica de la correlación.

Correlación por orden de rango

En el capítulo 13 se analizó el coeficiente de correlación. Recuerde que este mide la asociación entre dos variables en escala de intervalo o de razón; por ejemplo, el coeficiente de correlación reporta el vínculo entre el salario de ejecutivos y sus años de experiencia, o entre el número de millas que un embarque tiene que recorrer y el número de días que tarda en llegar a su destino. El coeficiente de correlación es una medida muy versátil de asociación, sin embargo, existen varias circunstancias en las que no es apropiado o puede ser engañoso; estas condiciones incluyen:

1. Cuando la escala de medición de una de las variables es ordinal (ordenada).
2. Cuando la relación entre estas no es lineal.
3. Cuando uno o más de los puntos de datos son muy diferentes de los otros.

Charles Spearman, estadístico británico, introdujo una medida para correlacionar datos de nivel ordinal la cual permite describir la relación entre conjuntos de datos clasificados; por ejemplo, a dos miembros del personal en la Office of Research de la University of the Valley se les pide clasificar 10 propuestas de investigación de la facultad con la finalidad de recolectar fondos. Aquí interesa estudiar la relación entre las calificaciones de los dos miembros del personal; es decir, ¿los empleados califican las mismas propuestas como las más valiosas y las menos valiosas para los fondos?

Este coeficiente de correlación de rangos también se aplica a las circunstancias 2 y 3 ya mencionadas. En vez de utilizar los valores actuales en el grupo, ordene los datos muestrales y calcule la correlación entre los ordenados. Al igual que el coeficiente de correlación de Pearson que se describió en el capítulo 13, en el coeficiente de correlación por rangos se adopta cualquier valor en el intervalo de -1.00 hasta 1.00. Los valores de -1.00 o 1.00 indican una asociación perfecta entre los rangos, mientras que un valor de 0 señala que no hay asociación entre los rangos de las dos variables, y los valores de -0.84 y 0.84 marcan el mismo grado de asociación, pero -0.84 muestra una relación inversa, y 0.084, una relación directa. El coeficiente de correlación de rangos de Spearman se denota como r_s.

Es posible utilizar la fórmula [13.1] para encontrar el coeficiente de correlación por rangos; sin embargo, utilice los valores ordenados en vez de los actuales; estos se ordenan siempre de manera ascendente. Un método más simple es utilizar la siguiente fórmula, que utiliza los rangos.

ESTADÍSTICA EN ACCIÓN

Los manatíes son mamíferos grandes que suelen flotar a pocos centímetros por debajo de la superficie del agua; debido a esto, están en peligro de ser alcanzados por las hélices del motor de las embarcaciones. Un estudio de la correlación entre el número de embarcaciones registradas en los condados de la costa de Florida y el número de muertes accidentales de manatíes reveló una fuerte correlación positiva; como resultado, en Florida se establecieron regiones donde se prohíben las embarcaciones de motor, a fin de proteger estos mamíferos.

COEFICIENTE DE CORRELACIÓN POR RANGOS DE SPEARMAN

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 + 1)} \qquad [16.6]$$

donde:

d es la diferencia entre los rangos por cada par;
n es el número de observaciones por pares.

En el siguiente ejemplo se proporcionan los detalles para calcular el coeficiente de correlación por rangos.

EJEMPLO

Estudios recientes se han enfocado en la relación entre la edad de los compradores en línea y el número de minutos que pasan navegando en internet. En la tabla 16.8 se registra una muestra de 15 compradores en línea que hicieron una compra la semana pasada; se incluyen la edad y el tiempo (en minutos) que pasaron en internet la semana pasada.

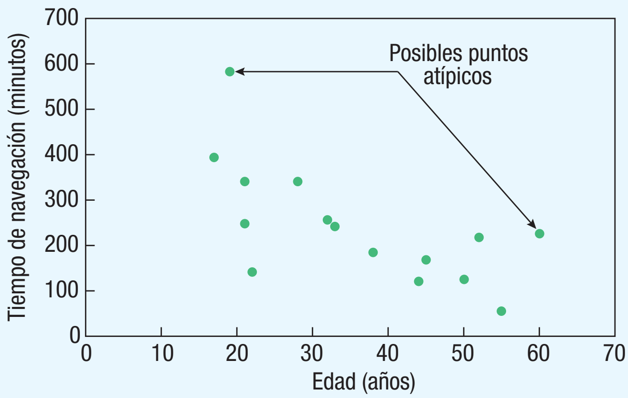
1. Trace un diagrama de dispersión; coloque la edad en el eje horizontal y los minutos de navegación en el eje vertical.
2. ¿Qué tipo de asociación sugieren los datos de la muestra? ¿Fuerte o débil, directa o inversa?
3. ¿Puede ver algún problema con la relación entre variables?
4. Encuentre el coeficiente de correlación por rangos.
5. Realice una prueba de hipótesis para determinar si existe una asociación negativa entre los rangos.

TABLA 16.8 Edad y minutos de navegación de una muestra de compradores por internet

Comprador	Edad	Tiempo de navegación (minutos)
Spina, Sal	28	342
Gordon, Ray	50	125
Schnur, Roberta	44	121
Alvear, José	32	257
Myers, Tom	55	56
Lyons, George	60	225
Harbin, Joe	38	185
Bobko, Jack	22	141
Koppel, Marty	21	342
Rowatti, Marty	45	169
Monahan, Joyce	52	218
Lanoue, Bernie	33	241
Roll, Judy	19	583
Goodall, Jody	17	394
Broderick, Ron	21	249

SOLUCIÓN

El primer paso es crear un diagrama de dispersión (se muestra en la gráfica 16.2) para entender mejor la relación entre ambas variables.



GRÁFICA 16.2 Diagrama de dispersión de edad y minutos de navegación en internet

Parece existir una relación inversa bastante fuerte entre la edad de los compradores y la cantidad de tiempo que pasan navegando en línea; esto es, los compradores más jóvenes parecen pasar más tiempo en internet, lo cual no es de sorprender. Hay un par de puntos de datos que parecen ser muy distintos a los otros; los cuales se identifican en la tabla 16.2; es debido a estos que se decide usar el coeficiente de correlación por rangos.

Para calcular dicho coeficiente, primero se ordena la edad de los compradores. El más joven es Jody Goodall, de 17, así que recibe el número 1; el siguiente es Judy Roll, que recibe el rango 2; hay dos compradores, Marty Kippel y Ron Broderick, que tienen 21 años (este empate se resuelve asignando a cada uno un rango de 3.5, que es el promedio de los rangos 3 y 4); el comprador de más edad de la muestra es George Lyons, quien recibe un rango de 15. Utilice la misma estrategia de otorgamiento de rangos para los minutos de navegación. Tom Myers pasó el menor tiempo en internet, así que tiene el rango 1; sigue Roberta Schnur, con 141 minutos, por lo tanto tiene un rango de 2. Todos los datos de rangos se muestran en la tabla 16.9.

El coeficiente de correlación por rangos es -0.724, el cual se determina mediante la fórmula [16.6].

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} = 1 - \frac{6(965.5)}{15(15^2 - 1)} = 1 - 1.724 = -0.724$$

TABLA 16.9 Edad, tiempo de navegación, rangos, diferencias y diferencias al cuadrado

Comprador	Edad	Rango de edad	Tiempo de navegación (minutos)	Rango de navegación	d	d^2
Spina, Sal	28	6.0	342	12.5	−6.50	42.25
Gordon, Ray	50	12.0	125	3.0	9.00	81.00
Schnur, Roberta	44	10.0	121	2.0	8.00	64.00
Alvear, José	32	7.0	257	11.0	−4.00	16.00
Myers, Tom	55	14.0	56	1.0	13.00	169.00
Lyons, George	60	15.0	225	8.0	7.00	49.00
Harbin, Joe	38	9.0	185	6.0	3.00	9.00
Bobko, Jack	22	5.0	141	4.0	1.00	1.00
Koppel, Marty	21	3.5	342	12.5	−9.00	81.00
Rowatti, Marty	45	11.0	169	5.0	6.00	36.00
Monahan, Joyce	52	13.0	218	7.0	6.00	36.00
Lanoue, Bernie	33	8.0	241	9.0	−1.00	1.00
Roll, Judy	19	2.0	583	15.0	−13.00	169.00
Goodall, Jody	17	1.0	394	14.0	−13.00	169.00
Broderick, Ron	21	3.5	249	10.0	−6.50	42.25
					Suma d^2	965.50

El valor −0.724 indica una asociación negativa bastante fuerte entre la edad del comprador por internet y los minutos que pasa navegando. Los compradores más jóvenes tendieron a pasar más tiempo en internet.

Prueba de significancia de r_s

¿Cómo saber si el grado de asociación entre las variables ordenadas ocurrió por casualidad? Recuerde que en el capítulo 13 se investigó este mismo problema; en aquel caso, se utilizaron datos de intervalo o de razón; ahora se emplearán los rangos. En el ejemplo anterior se encontró una correlación de rangos de −0.724, ¿es posible que esta sea casual, y que la correlación entre los rangos en la población en realidad sea 0? Mediante la siguiente prueba de significancia se despejará esa duda.

En el caso de una muestra de 10 o más, la significancia de r_s se determina al calcular t con la siguiente fórmula. La distribución de muestreo de r_s sigue la distribución t con $n - 2$ grados de libertad.

PRUEBA DE HIPÓTESIS, CORRELACIÓN POR RANGOS

$$t = r_s\sqrt{\frac{n-2}{1-r_s^2}} \qquad \textbf{[16.7]}$$

Las hipótesis nula y alternativa son:

H_0: La correlación por rangos entre la población es cero.
H_1: Hay una asociación positiva entre los rangos.

Utilice el nivel de significancia 0.05; además, la regla de decisión es rechazar la hipótesis nula si el valor calculado de t es menor a −1.771. Vaya al apéndice B.5 y utilice la prueba de una cola y $15 - 2 = 13$ grados de libertad.

$$t = r_s\sqrt{\frac{n-2}{1-r_s^2}} = -0.724\sqrt{\frac{15-2}{1-(0.724)^2}} = -3.784$$

Se rechaza H_0 debido a que el valor calculado de t es −3.784; así, se rechaza la hipótesis nula y se acepta la hipótesis alternativa. El valor calculado de t de −3.784 es menor que −1.771. Existe evidencia de una asociación negativa entre la edad del comprador por internet y el tiempo que pasa navegando.

AUTOEVALUACIÓN 16-7

Mediante una muestra de personas que solicitan empleo en una fábrica de Davis Enterprises se encontraron las siguientes calificaciones sobre una prueba de percepción ocular (*X*) y una prueba de aptitudes para la mecánica (*Y*):

Sujeto	Percepción ocular	Aptitud para la mecánica	Sujeto	Percepción ocular	Aptitud para la mecánica
001	805	23	006	810	28
002	777	62	007	805	30
003	820	60	008	840	42
004	682	40	009	777	55
005	777	70	010	820	51

(a) Calcule el coeficiente de correlación por rangos entre la percepción ocular y la aptitud para la mecánica.
(b) Con el nivel de significancia 0.05, ¿es posible concluir que la correlación entre la población es diferente de 0?

EJERCICIOS

25. ¿A los esposos y las esposas les gustan los mismos programas de televisión? En un estudio reciente, Nielsen Media Research les pidió a matrimonios jóvenes que calificaran programas de mejor a peor. Una calificación de 1 indica el programa más agradable, y una calificación de 14, el menos agradable. Los resultados de una pareja casada son:

Programa	Calificación de los hombres	Calificación de las mujeres
60 Minutes	4	5
CSI—New York	6	4
Bones	7	8
SportsCenter	2	7
Late Show with David Letterman	12	11
NBC Nightly News	8	6
Law and Order: Los Angeles	5	3
Miami Medical	3	9
Survivor	13	2
Office	14	10
American Idol	1	1
Grey's Anatomy	9	13
House	10	12
Criminal Minds	11	14

a. Elabore un diagrama de dispersión y coloque las calificaciones de los hombres en el eje horizontal y las de las mujeres en el eje vertical;
b. Calcule el coeficiente de correlación por rangos entre las calificaciones de los hombres y las mujeres.
c. Con el nivel de significancia 0.05, ¿es posible concluir que hay una asociación positiva entre ambas calificaciones?

26. Far West University ofrece clases diurnas y nocturnas en administración. Mediante una encuesta a estudiantes se intenta saber cómo perciben el prestigio asociado con ciertas carreras; a un alumno diurno se le pidió calificar las carreras de 1 a 8, con 1 como la calificación de mayor prestigio y 8 la de menor prestigio. A un estudiante nocturno se le pidió hacer lo mismo.

Carrera	Calificación de los estudiantes diurnos	Calificación de los estudiantes nocturnos	Carrera	Calificación de los estudiantes diurnos	Calificación de los estudiantes nocturnos
Contador	6	3	Estadístico	1	7
Programador de computadoras	7	2	Investigador de marketing	4	8
Gerente bancario	2	6	Analista bursátil	3	5
Administrador de hospital	5	4	Gerente de producción	8	1

Encuentre el coeficiente de correlación por rangos de Spearman.

27. Se solicitó a 10 nuevos representantes de Clark Sprocket and Chain, Inc., que asistieran a un programa de capacitación antes de asignárseles a una oficina regional de ventas; al final del programa, los representantes tomaron una serie de pruebas y se asignó un rango a las calificaciones; por ejemplo, Arden tuvo la puntuación más baja y quedó en el rango 1; Arbuckle tuvo la puntuación más alta, y el rango 10. Al término del primer año de ventas, se compararon los rangos de los representantes basados en las puntuaciones de las pruebas con sus ventas en ese periodo:

Representante	Ventas anuales (miles de dólares)	Calificación en el programa de capacitación	Representante	Ventas anuales (miles de dólares)	Calificación en el programa de capacitación
Kitchen	319	8	Arden	300	1
Bond	150	2	Crane	280	6
Gross	175	5	Arthur	200	9
Arbuckle	460	10	Keene	190	4
Greene	348	7	Knopf	300	3

a. Calcule e interprete el coeficiente de correlación por rangos entre las ventas en el primer año y la calificación después del programa de capacitación.
b. Con el nivel de significancia 0.05, ¿es posible concluir que hay una asociación positiva entre las ventas del primer año y la calificación en el programa de capacitación?

28. Suponga que la Texas A & M University—Commerce tiene cinco becas disponibles para el equipo de basquetbol femenil. El entrenador dio a sus dos asistentes los nombres de 10 jugadoras de preparatoria con potencial para jugar en la universidad; cada asistente fue a tres juegos y calificó a las 10 jugadoras respecto de su potencial. Para explicar lo anterior, el primer asistente calificó a Norma Tidwell como la mejor jugadora entre las 10 observadas, y a Jeannie Black, como la peor.

	Rango, calificación del asistente			Rango, calificación del asistente	
Jugadora	Jean Cann	John Cannelli	Jugadora	Jean Cann	John Cannelli
Cora Jean Seiple	7	5	Candy Jenkins	3	1
Bette Jones	2	4	Rita Rosinski	5	7
Jeannie Black	10	10	Anita Lockes	4	2
Norma Tidwell	1	3	Brenda Towne	8	9
Kathy Marchal	6	6	Denise Ober	9	8

a. Determine el coeficiente de correlación por rangos de Spearman.
b. Con el nivel de significancia 0.05, ¿es posible concluir que hay una asociación positiva entre los rangos?

RESUMEN DEL CAPÍTULO

I. La prueba de los signos se basa en la diferencia de signos entre dos observaciones relacionadas.
A. No es necesario hacer suposiciones acerca de la forma de ambas poblaciones.
B. Se basa en muestras por pares o dependientes.
C. En el caso de muestras pequeñas, se encuentra el número de signos más (+) o menos (−) y se consulta la distribución binomial para el valor crítico.
D. En el caso de una muestra de 10 signos más se utiliza la distribución normal estándar y la siguiente fórmula.

$$z = \frac{(x \pm 0.50) - 0.50n}{0.50\sqrt{n}} \qquad \textbf{[16.2] [16.3]}$$

II. La prueba de la mediana valida una hipótesis acerca de la mediana de una población.
A. Se encuentra μ y σ de una distribución normal.
B. Se utiliza la distribución z como el estadístico de prueba.
C. El valor de z se calcula a partir de la siguiente fórmula, donde X es el número de observaciones arriba y debajo de la media.

$$z = \frac{(x \pm 0.50) - \mu}{\sigma} \qquad \textbf{[16.1]}$$

III. La prueba de Wilcoxon de los rangos con signo es una prueba no paramétrica donde no se requiere la suposición de normalidad.

A. Los datos deben estar al menos en una escala ordinal, y las muestras deben ser dependientes.
B. Los pasos para realizar la prueba son:
  1. Clasificar las diferencias absolutas entre las observaciones relacionadas.
  2. Aplicar el signo de las diferencias a los rangos.
  3. Sumar los rangos negativos y los positivos.
  4. La menor de las dos sumas es el valor T calculado.
  5. Hallar en el apéndice B.8 el valor crítico y tomar una decisión respecto de H_0.

IV. La prueba de Wilcoxon de la suma de rangos se usa para probar si dos muestras independientes provienen de poblaciones iguales.
A. No se requiere una suposición acerca de la forma de la población.
B. Los datos deben estar al menos en escala ordinal.
C. Cada muestra debe contener al menos ocho observaciones.
D. Para determinar el valor del estadístico de prueba W, las observaciones de las muestras se clasifican de bajo a alto como si fueran de una sola población.
E. Se determina la suma de los rangos de cada muestra.
F. W se utiliza para calcular z, donde W es la suma de los rangos de la primera población.

$$z = \frac{W - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2(n_1 + n_2 + 1)}{12}}} \qquad \textbf{[16.4]}$$

G. La distribución normal estándar, del apéndice B.1, es el estadístico de prueba.

V. El análisis de Kruskal-Wallis de la varianza por rangos se usa para probar si varias poblaciones son iguales.
A. No se requieren suposiciones respecto de que las poblaciones tengan una distribución normal.
B. Las muestras deben ser independientes y, al menos, de escala ordinal.
C. Las observaciones de las muestras se clasifican de menor a mayor, como si fueran un solo grupo.
D. El estadístico de prueba sigue la distribución *ji* cuadrada, con la condición de que haya al menos cinco observaciones en cada muestra.
E. El valor del estadístico de prueba se calcula a partir de la siguiente fórmula:

$$H = \frac{12}{n(n+1)}\left[\frac{(\Sigma R_1)^2}{n_1} + \frac{(\Sigma R_2)^2}{n_2} + \cdots + \frac{(\Sigma R_k)^2}{n_k}\right] - 3(n+1) \qquad \textbf{[16.5]}$$

VI. El coeficiente de correlación por rangos de Spearman es una medida de la asociación entre dos variables en escala ordinal.
A. Puede variar de -1 a 1.
  1. Un valor de 0 indica que no hay asociación entre las variables.
  2. Un valor de -1 indica una correlación negativa perfecta, y un valor de 1, una correlación positiva perfecta.
B. El valor de r_s se calcula a partir de la siguiente fórmula:

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 + 1)} \qquad \textbf{[16.6]}$$

C. Con la condición de que el tamaño de la muestra sea de al menos 10, se puede realizar una prueba de hipótesis mediante la siguiente fórmula:

$$t = r_s\sqrt{\frac{n-2}{1-r_s^2}} \qquad \textbf{[16.7]}$$

  1. El estadístico de prueba sigue la distribución t.
  2. Hay $n - 2$ grados de libertad.

CLAVE DE PRONUNCIACIÓN

Símbolo	Significado	Pronunciación
$(\Sigma R_1)^2$	Cuadrado del total de los rangos de la primera *columna*	*Sigma R subíndice 1 al cuadrado*

r_s	Coeficiente de correlación por rangos de Spearman	*r subíndice s*

EJERCICIOS DEL CAPÍTULO

29. La vicepresidente de programación de NBC terminó la programación del horario estelar para el otoño, y decidió incluir un drama que se desarrolla en un hospital, pero no está segura sobre cuál elegir entre dos alternativas. Tiene dos programas piloto, uno se llama "The Surgeon" y, el otro, "Critical Care". Para ayudarla a tomar una decisión, a una muestra de 20 televidentes de Estados Unidos se les pidió ver los dos programas e indicar sus preferencias. Los resultados fueron que a 12 les gustó "The Surgeon", a 7 les gustó "Critical Care" y 1 no tuvo preferencia. ¿Hay alguna preferencia por uno de los dos programas? Utilice el nivel de significancia 0.10.

30. IBM, Inc., quiere otorgar un contrato para suministrar bolígrafos de punto fino que se van a utilizar en sus oficinas en todo el país; dos proveedores, Bic y Pilot, presentaron ofertas. Para determinar la preferencia de los empleados, corredores y otros interesados, se realiza una prueba de preferencia personal con una muestra de 20 empleados seleccionada al azar. Se utilizará el nivel de significancia 0.05.

a. Si la hipótesis alternativa establece que Bic tiene preferencia en comparación con Pilot, ¿la prueba de los signos que se va a realizar es de una o dos colas? Explique su respuesta.

b. Conforme cada uno de los miembros de la muestra indicó a los investigadores su preferencia, se registró un signo "+" para Bic y un "−" para Pilot. Un conteo de los signos más (+) reveló que 12 empleados preferían Bic, 5 preferían Pilot y 3 no se decidieron. ¿Cuál es el valor de *n*?

c. ¿Cuál es su regla de decisión expresada en palabras?

d. ¿A qué conclusión llegó respecto de la preferencia por los bolígrafos? Explique su respuesta.

31. Cornwall and Hudson, importante tienda departamental al menudeo, quiere manejar una sola marca de reproductores de CD de alta calidad. La lista se redujo a dos marcas: Sony y Panasonic. Para ayudar a tomar una decisión, se reunió un panel de 16 expertos en audio; se tocó una pieza musical con componentes Sony (identificados como A), y luego se tocó la misma pieza, ahora con componentes Panasonic (identificados B). En la siguiente tabla se indica la preferencia de una persona por los componentes Sony ("+"), la preferencia por Panasonic ("−") y cuando no hay preferencia (0).

Experto															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
+	−	+	−	+	+	−	0	−	+	−	+	+	−	+	−

Realice una prueba de hipótesis con el nivel de significancia 0.10 para determinar si hay una diferencia entre las preferencias por las dos marcas.

32. La Greater Jacksonville, Florida, Real Estate Association afirma que la mediana de la renta de condominios de tres recámaras es mayor a 1 200 dólares por mes. Una muestra de 149 unidades reveló que cinco se rentaban exactamente por esa cantidad y 75 se por más de esa cifra; con el nivel de significancia 0.05, ¿es posible concluir que la mediana de la renta es mayor a 1 200 dólares por mes?

a. Formule H_0 y H_1.

b. Establezca la regla de decisión.

c. Haga los cálculos necesarios y tome una decisión.

33. El Citrus Council of America quiere determinar si los consumidores prefieren jugo de naranja sin pulpa o con ella; por lo tanto, se seleccionó una muestra aleatoria de 212 consumidores, y cada miembro de la muestra probó un vaso pequeño, sin identificación, de una clase de jugo y luego la otra. Doce clientes dijeron que no tenían preferencia, 40 preferían el jugo sin pulpa y al resto le gustó el jugo con pulpa. Pruebe con el nivel de significancia 0.05 que las preferencias por jugo sin pulpa y para jugo con pulpa son iguales.

34. El objetivo de un proyecto de investigación comunitario es determinar si las mujeres tienen más conciencia respecto de la comunidad antes de casarse o después de cinco años de matrimonio; por lo tanto, se aplicó una prueba diseñada para medir la conciencia comunitaria a una muestra de mujeres solteras, y se les aplicó la misma prueba después de cinco años de matrimonio. Las calificaciones de la prueba se muestran a la izquierda.

Pruebe con el nivel de significancia 0.05. H_0: no hay diferencia en la conciencia comunitaria antes ni después del matrimonio; H_1 : hay una diferencia.

Nombre	Antes de casarse	Después de casarse	Nombre	Antes de casarse	Después de casarse
Beth	110	114	Carol	186	196
Jean	157	159	Lisa	116	116
Sue	121	120	Sandy	160	140
Cathy	96	103	Petra	149	142
Mary	130	139			

35. ¿Hay alguna diferencia entre las tasas de divorcios anuales en condados predominantemente rurales de tres regiones geográficas, suroeste, sureste y noroeste? Pruébelo con el nivel de significancia 0.05. Las tasas de divorcio anuales por 1 000 habitantes de cinco condados seleccionados al azar en el sudoeste, seis condados en el sudeste y cinco en el noroeste son:

Suroeste:	5.9, 6.2, 7.9, 8.6, 4.6
Sureste:	5.0, 6.4, 7.3, 6.2, 8.1, 5.1
Noroeste:	6.7, 6.2, 4.9, 8.0, 5.5

36. El gerente de producción de MPS Audio Systems, Inc., tiene interés en el tiempo de inactividad de los trabajadores; en particular, le gustaría saber si hay una diferencia entre los minutos inactivos de los trabajadores del turno diurno y del turno nocturno. En la siguiente tabla se indica el número de minutos de inactividad del día de ayer de los trabajadores en cinco días a la semana y de los trabajadores en seis noches a la semana. Utilice el nivel de significancia 0.05.

Turno diurno	Turno nocturno
92	96
103	114
116	80
81	82
89	88
	91

37. Los doctores Trythall y Kerns estudian la movilidad de los ejecutivos en ciertas industrias; su investigación mide la movilidad a partir de una calificación basada en el número de veces que un ejecutivo se ha mudado, cambiado de compañía o de trabajo durante los últimos 10 años. El mayor número de puntos se otorga a los que se mudan y cambian de compañías, y el menor número de puntos, a los que cambian de trabajo en la misma compañía sin mudarse. Seleccionaron aleatoriamente a cinco ejecutivos de la industria química, cinco de la venta al detalle, cinco en la industria de internet y cinco en la industria aeroespacial. La distribución de las calificaciones no sigue la distribución normal de probabilidad; utilizando el nivel de significancia 0.05, desarrolle una prueba adecuada para determinar si hay una diferencia entre las calificaciones de movilidad en las cuatro industrias.

Química	Detallista	Internet	Aeroespacial
4	3	62	30
17	12	40	38
8	40	81	46
20	17	96	40
16	31	76	21
	19		

38. Se formuló una serie de preguntas sobre deportes y sucesos mundiales a un grupo seleccionado al azar de 14 ciudadanos naturalizados; los resultados se convirtieron en las siguientes calificaciones de "conocimiento".

Ciudadano	Deportes	Sucesos mundiales	Ciudadano	Deportes	Sucesos mundiales
J. C. McCarthy	47	49	L. M. Zaugg	87	75
A. N. Baker	12	10	J. B. Simon	59	86
B. B. Beebe	62	76	J. Goulden	40	61
L. D. Gaucet	81	92	A. A. Fernandez	87	18
C. A. Jones	90	86	A. M. Carbo	16	75
J. N. Lopez	35	42	A. O. Smithy	50	51
A. F. Nissen	61	61	J. J. Pascal	60	61

a. Determine el grado de asociación entre cómo calificaron los ciudadanos respecto del conocimiento sobre deportes y cómo calificaron en relación con los sucesos mundiales.

b. Con el nivel de significancia 0.05, ¿es mayor que cero la correlación de rangos en la población?

39. A principios de la temporada de basquetbol, 12 equipos parecen sobresalir. A un panel de comentaristas deportivos y a otro de entrenadores de basquetbol colegial se les pidió calificar a esos equipos; a continuación se indican sus calificaciones compuestas:

Equipo	Entrenadores	Comentaristas deportivos	Equipo	Entrenadores	Comentaristas deportivos
Duke	1	1	Syracuse	7	10
UNLV	2	5	Georgetown	8	11
Indiana	3	4	Villanova	9	7
North Carolina	4	6	LSU	10	12
Louisville	5	3	St. Johns	11	8
Ohio State	6	2	Michigan	12	9

Determine la correlación entre las calificaciones de los entrenadores y las de los comentaristas deportivos. Con el nivel de significancia 0.05, ¿es posible concluir que hay una correlación positiva entre las calificaciones?

40. El profesor Bert Forman considera que los estudiantes que terminan sus exámenes en el menor tiempo posible reciben las calificaciones más altas, y los que tardan más en terminarlos, las más bajas; para verificar su sospecha, asigna una calificación al orden en que terminan los alumnos y luego califica los exámenes. A continuación se muestran los resultados:

Estudiante	Orden en que terminó	Calificación (50 puntos posibles)	Estudiante	Orden en que terminó	Calificación (50 puntos posibles)
Gromney	1	48	Smythe	7	39
Bates	2	48	Arquette	8	30
MacDonald	3	43	Govito	9	37
Sosa	4	49	Gankowski	10	35
Harris	5	50	Bonfigilo	11	36
Cribb	6	47	Matsui	12	33

Convierta las calificaciones de los exámenes en un rango y determine el coeficiente de correlación por rangos. Con el nivel de significancia 0.05, ¿es posible que el profesor Forman concluya que hay una asociación positiva entre el orden en que terminaron los alumnos los exámenes y las calificaciones que obtuvieron?

EJERCICIOS DE LA BASE DE DATOS

(Los datos para estos ejercicios están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**)

41. Consulte los datos sobre Real State, que contienen información acerca de casas que se vendieron en el área de Goodyear, Arizona, el año anterior.

a. Utilice una prueba no paramétrica apropiada para determinar si hay una diferencia entre los precios de venta habituales de las casas en varias colonias; suponga que los precios de venta no están normalmente distribuidos. Utilice el nivel de significancia 0.05.

b. Clasifique las casas con seis o más recámaras en un grupo y determine si hay una diferencia de acuerdo con el número de recámaras entre los precios de venta habituales de las casas. Utilice el nivel de significancia 0.05 y suponga que la distribución de los precios de venta no está normalmente distribuida.

c. Suponga que la distribución de la distancia desde el centro de la ciudad tiene un sesgo positivo; es decir, no es razonable la suposición de normalidad. Compare la distribución de la distancia desde el centro de la ciudad de las casas que tienen una alberca con las no que la tienen. ¿Es posible concluir que hay una diferencia entre las distribuciones? Utilice el nivel de significancia 0.05.

42. Consulte los datos sobre Baseball 2012 que contienen información de la temporada 2012 de las Ligas Mayores de Béisbol.

a. Clasifique los equipos por el número de partidos ganados y el salario total del equipo; además, calcule el coeficiente de correlación por rangos entre las dos variables. Con el nivel de significancia 0.01, ¿es posible concluir que es mayor que cero?

b. Suponga que las distribuciones de los salarios de los equipos de la Liga Americana y de la Liga Nacional no siguen la distribución normal y realice una prueba de hipótesis para ver si hay una diferencia entre ambas distribuciones.

c. Clasifique los 30 equipos por asistencia y salario del equipo y determine el coeficiente de correlación por rangos entre estas dos variables. Con el nivel de significancia 0.05, ¿es razonable concluir que están relacionados los rangos de ambas variables?

43. Consulte los datos de los autobuses del Distrito Escolar Buena.

a. Suponga que la distribución del costo de mantenimiento de tres fabricantes de autobuses no sigue una distribución normal y realice una prueba de hipótesis con el nivel de significancia 0.05 para determinar si las distribuciones son diferentes.

b. Asuma que la distribución del costo de mantenimiento de la flotilla de autobuses no sigue una distribución normal y realice una prueba de hipótesis con el nivel de significancia 0.05 para determinar si las distribuciones son diferentes.

c. Suponga que la distribución del costo de mantenimiento de los tipos de autobús, de diesel o de gasolina, no sigue una distribución normal y realice una prueba de hipótesis con el nivel de significancia 0.05 para determinar si las distribuciones son diferentes.

REPASO DE LOS CAPÍTULOS 15 y 16

En los capítulos 15 y 16 se describieron métodos estadísticos para estudiar datos en escala nominal u ordinal de medición; estos son estadísticos *no paramétricos* o *sin distribución* (no requieren suposiciones respecto de la forma de la población). Recuerde, por ejemplo, del capítulo 12, que al investigar las medias de varias poblaciones se supuso que las poblaciones seguían la distribución de probabilidad normal.

En el capítulo 15 se describieron las pruebas para los datos en el nivel nominal. Primero se estudiaron las pruebas de proporciones de una y dos muestras. Una proporción es la fracción de individuos u objetos que posee una cierta característica; en la cual, el objeto o individuo muestreado tiene o no tiene dicha característica, por ejemplo, en una prueba de proporciones de una muestra, se estudiaron 100 compras de gasolina en la estación local de Kwick Fill. El cliente compró o no la gasolina regular (hay solo dos posibles resultados). En una prueba de proporciones de dos muestras se compara la proporción de personas que compraron gasolina regular en la estación Corry con la de quienes compraron gasolina regular en Tyrone. El estadístico de prueba es la distribución normal estándar en ambas pruebas.

También se describió la distribución *ji* cuadrada; la cual se utiliza para comparar el conjunto observado de frecuencias en una muestra aleatoria con el conjunto correspondiente de frecuencias esperadas en la población. El nivel de medición también es de escala nominal. En el ejemplo previo existían solo dos posibles resultados: el comprador adquirió o no gasolina regular. Mediante la distribución *ji* cuadrada se investiga una opción donde existen diversos resultados posibles en la escala nominal; por ejemplo, un individuo puede comprar gasolina regular, de mediano octanaje o Premium. Recuerde que cuando los datos se miden en un nivel nominal, estos solo se clasifican de acuerdo con alguna identificación, nombre o característica.

En el capítulo 15 también se estudió la relación entre dos variables en una tabla de contingencia; es decir, se observaron dos características de cada individuo u objeto muestreado, por ejemplo, ¿hay alguna relación entre la calidad del producto (aceptable o inaceptable) y el turno en que se fabricó (diurno, vespertino o nocturno)? La distribución *ji* cuadrada es el estadístico de prueba.

En el capítulo 16 se describieron cinco pruebas no paramétricas de hipótesis y el coeficiente de correlación por rangos; cada una de estas pruebas requiere la escala de medición ordinal, es decir, la capacidad de clasificar u ordenar las variables de interés.

La *prueba de los signos* para muestras dependientes se basa en el signo de la diferencia entre observaciones relacionadas. La distribución binomial es el estadístico de prueba; pero en los casos donde la muestra es mayor que 10, la aproximación normal a la distribución de probabilidad binomial sirve como el estadístico de prueba.

Cuando se utiliza la *prueba de la mediana*, el primer paso es contar el número de observaciones arriba (o debajo) de la mediana propuesta. Luego se emplea la distribución normal estándar para determinar si este número es razonable o demasiado grande para haber ocurrido por azar.

La *prueba de Wilcoxon de los rangos con signo* requiere muestras dependientes; es una extensión de la prueba de los signos pues emplea tanto la dirección como la magnitud de la diferencia entre los valores relacionados; además, tiene su propia distribución muestral (se reporta en el apéndice B.8).

La *prueba de Wilcoxon de la suma de los rangos* supone poblaciones independientes, pero no requiere que las poblaciones sigan la distribución de probabilidad normal; una alternativa es la prueba *t* para muestras independientes, la cual se describió en el capítulo 11. Cuando hay al menos ocho observaciones en cada muestra, el estadístico de prueba es la distribución normal estándar.

La *prueba de Kruskal-Wallis* es una extensión de la prueba de Wilcoxon de la suma de los rangos, en el sentido de que maneja más de dos poblaciones; es una alternativa al método de la ANOVA en una dirección, el cual se describió en el capítulo 12; pero no requiere que las poblaciones sigan la distribución de probabilidad normal, o que las poblaciones tengan desviaciones estándar iguales.

El estadístico *coeficiente de correlación por rangos de Spearman* es un caso especial del coeficiente de correlación de Pearson, el cual se describió en el capítulo 13; este se basa en la correlación entre los *rangos* de observaciones relacionadas. Puede variar de -1.00 a 1.00, en donde 0 indica que no hay asociación entre los rangos.

PROBLEMAS

1. El propietario de Beach Front Snow Cones, Inc., considera que la mediana del número de conos de nieve que vende por día entre el Memorial Day y el Labor Day es 60. A continuación se registra una muestra de 20 días. ¿Es razonable concluir que la mediana en realidad es mayor que 60? Utilice el nivel de significancia 0.05.

65	70	65	64	66	54	68	61	62	67
65	50	64	55	74	57	67	72	66	65

2. Un fabricante de impermeables para niños quiere saber si ellos tienen preferencia por un color específico. En la siguiente información se registra la preferencia del color de una muestra de 50 niños de 6 a 10 años de edad; para investigar esta cuestión utilice el nivel de significancia 0.05.

Color	Frecuencia
Azul	17
Rojo	8
Verde	12
Amarillo	13

3. ¿Hay alguna diferencia (en pies) entre las longitudes de los puentes colgantes de las zonas del noreste, sureste y oeste de Estados Unidos? En la siguiente tabla se muestran las longitudes de siete puentes en el noreste, nueve en el sureste y ocho en el lejano oeste. Realice una prueba de hipótesis adecuada con base en los siguientes datos; pero no suponga que las longitudes de los puentes siguen una distribución de probabilidad normal. Utilice el nivel de significancia 0.05.

Noreste	Sureste	Lejano oeste
3 645	3 502	3 547
3 727	3 645	3 636
3 772	3 718	3 659
3 837	3 746	3 673
3 873	3 758	3 728
3 882	3 845	3 736
3 894	3 940	3 788
	4 070	3 802
	4 081	

4. Una investigación del First Bank of Illinois reveló que 8% de sus clientes esperan más de cinco minutos en el lobby de un banco para ser atendidos en ventanilla. La administración del banco considera que ese tiempo es razonable y no incrementará el número de cajas a menos que la proporción de espera de cinco minutos aumente en más de 8%. El gerente de la sucursal de Litchfield considera que la espera en su sucursal es mayor que el estándar y ha solicitado cajeros adicionales de medio tiempo. Para apoyar su solicitud, encontró que, en una muestra de 100 clientes, 10 de ellos esperan más de cinco minutos. En un nivel de significancia de 0.01, ¿qué tan razonable es concluir que 8% de los clientes esperan más de cinco minutos?

CASOS

A. Century National Bank

¿Hay alguna relación entre la ubicación de la sucursal bancaria y el hecho de que un cliente tenga una tarjeta de débito? Con base en la información disponible, elabore una tabla en la que se indique la relación entre ambas variables. Con el nivel de significancia 0.05, ¿es posible concluir que hay una relación entre la ubicación de la sucursal y un cliente con tarjeta de débito?

B. Thomas Testing Labs

John Thomas, propietario de Thomas Testing, ha trabajado durante cierto tiempo como contratista para compañías de seguros en lo que concierne a los conductores en estado de ebriedad; para mejorar sus capacidades de investigación, hace poco compró el Rupple Driving Simulator. Este dispositivo permite que un sujeto haga una "prueba del camino" y proporciona una calificación que indica el número de errores de conducción cometidos durante la prueba. Las calificaciones más altas indican más errores, por ejemplo, no detenerse por completo en una señal de alto, no utilizar las señales de vuelta, no tener precaución en el pavimento húmedo o con nieve, etcétera. Durante la prueba del camino, los problemas aparecen al azar, y no se presentan todos en cada prueba del camino. Estas son ventajas importantes para el Rupple Driving Simulator debido a que los sujetos no tienen ventaja al realizar la prueba varias veces.

Con el nuevo simulador de conducción, Thomas quiere estudiar con detalle el problema de la conducción en estado de ebriedad; por lo tanto, selecciona una muestra aleatoria de 25 conductores, y pide a cada individuo realizar la prueba de conducción en el simulador. En la tabla de la derecha se registra el número de errores de cada uno; luego, pide a cada integrante del grupo que beba tres latas de 16 onzas de cerveza en un periodo de 60 minutos y regrese al simulador para hacer otra prueba de conducción. En la tabla también se muestra el número de errores después de beber la cerveza. La pregunta de la investigación es: ¿afecta el consumo de alcohol la habilidad del conductor y, por lo tanto, aumenta el número de errores de conducción?

Thomas considera que la distribución de las calificaciones en la prueba de manejo no sigue una distribución normal, y, en consecuencia, deberá utilizar una prueba no paramétrica; como las observaciones son apareadas, decide emplear las pruebas de los signos y de Wilcoxon por rangos con signo.

a. Compare los resultados que se obtuvieron con los dos procedimientos. Realice una prueba apropiada de hipótesis para determinar si el alcohol se relaciona con errores al conducir.
b. Redacte un reporte acerca de sus resultados.

	Errores de conducción			Errores de conducción	
Sujeto	**Sin alcohol**	**Con alcohol**	**Sujeto**	**Sin alcohol**	**Con alcohol**
1	75	89	14	72	106
2	78	83	15	83	89
3	89	80	16	99	89
4	100	90	17	75	77
5	85	84	18	58	78
6	70	68	19	93	108
7	64	84	20	69	69
8	79	104	21	86	84
9	83	81	22	97	86
10	82	88	23	65	92
11	83	93	24	96	97
12	84	92	25	85	94
13	80	103			

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. Se requiere un ______ nivel de medición para la prueba de bondad de ajuste *ji* cuadrada. **1.** ____________
2. ¿Cuál de las siguientes características *no* corresponde a la distribución *ji* cuadrada? Sesgo positivo, basada en grados de libertad, no puede ser negativa, cuando menos 30 observaciones. **2.** ____________
3. En una tabla de contingencia, ¿cuántas colas de cada variable se consideran? **3.** ____________
4. En una tabla de contingencia, hay cuatro filas y tres columnas; por lo tanto, hay ________ grados de libertad. **4.** ____________
5. En una prueba de bondad de ajuste, el valor crítico de *ji* cuadrada se basa en _________. Tamaño de la muestra, número de categorías, número de variables o ninguna de las anteriores. **5.** ____________
6. En una prueba de signos, ¿las muestras son dependientes o independientes? **6.** ____________
7. En una prueba de signos de ocho observaciones apareadas, el estadístico de prueba es la distribución ________. Binomial, *z*, *t* o *ji* cuadrada. **7.** ____________
8. ¿Cuál es la diferencia principal entre la prueba de Kruskal-Wallis y la de suma de los rangos de Wilcoxon? Una se basa en muestras dependientes y la otra en muestras independientes, uno sirve para comparar dos muestras independientes y la otra para comparar dos o más muestras independientes. **8.** ____________
9. ¿Bajo qué condiciones puede el coeficiente de rango de correlación ser menor a -1.00? **9.** ____________
10. La prueba de Kruskal-Wallis se utiliza en vez de la ANOVA cuando no se cumple uno de los siguientes criterios: población normal, desviaciones estándar iguales, más de 12 elementos en la muestra, las poblaciones son independientes. **10.** ____________

Parte 2: Problemas

Utilice el procedimiento estándar de la prueba de hipótesis de seis pasos.

1. Un reciente reporte de censo indicó que 65% de las familias tienen a ambos padres presentes, 20% solo a la madre, 10% solo al padre, y 5% no tienen padres presentes. Mediante una muestra aleatoria de 200 niños de un gran distrito escolar rural se determinaron las siguientes frecuencias de niños con dos padres, solo la madre, solo el padre, sin padres y el total (200).

Ambos padres	Solo la madre	Solo el padre	Sin padres	Total
120	40	30	10	200

¿Hay suficiente evidencia para concluir que la proporción de familias por tipo de padre presente en este distrito escolar en particular difieren de las reportadas en el censo?

2. Un editor de libros quiere investigar el tipo de obras que eligen los hombres y las mujeres como lectura de entretenimiento. Mediante una muestra aleatoria de 540 hombres y 500 mujeres se halló la siguiente información con respecto a sus preferencias por novelas de misterio, romance o libros de autoayuda. Con el nivel de significancia 0.05, ¿puede concluir que el género está o no relacionado con el tipo de libro elegido?

	Misterio	Romance	Autoayuda	Total
Hombres	250	100	190	540
Mujeres	130	170	200	500

3. Un instructor tiene tres secciones de estadística básica: 8:00 a.m., 10:00 a.m. y 1:30 p.m. A continuación se presentan las calificaciones del primer examen de cada sección. Asuma que las distribuciones no siguen una distribución de probabilidad normal. Con el nivel de significancia 0.05, ¿hay diferencia entre las distribuciones de calificaciones?

8 a.m.	10 a.m.	1:30 p.m.
68	59	67
84	59	69
75	63	75
78	62	76
70	78	79
77	76	83
88	80	86
71		86
		87

4. De acuerdo con un estudio en *Health Magazine*, uno de cada tres niños en Estados Unidos es obeso o tiene sobrepeso. Un médico de Louisiana muestreó a 500 niños y encontró que 210 eran obesos o tenían sobrepeso. ¿Sugiere la evidencia de que la proporción actual de niños obesos o con sobrepeso es de más de uno en tres? Utilice el nivel de significancia 0.01.

Números índices

17

EN EL EJERCICIO 27 se proporciona información sobre precios y cantidades de margarina, manteca, leche y papas fritas de los años 2000 y 2013. Calcule un índice de precios simple de cada uno de los cuatro artículos; considere el año 2000 como periodo base (vea el ejercicio 27 y el OA17-1).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA17-1 Calcular e interpretar un índice simple y no ponderado.

OA17-2 Calcular e interpretar un índice agregado y no ponderado.

OA17-3 Calcular e interpretar un índice agregado y ponderado.

OA17-4 Enumerar y describir índices que tienen un propósito especial.

OA17-5 Aplicar el Índice de Precios al Consumidor.

Introducción

En este capítulo se analiza una útil herramienta descriptiva denominada índice; el cual expresa el cambio relativo de un valor de un periodo a otro; sin duda, conoce índices como el Índice de Precios al Consumidor (IPC), que en Estados Unidos se publica cada mes mediante el U.S. Department of Labor. Hay muchos índices, como el **Dow Jones Industrial Average** (DJIA), **NASDAQ, NIKKEI 225 y Standard & Poor's 500 Stock Average.** El gobierno federal estadounidense publica índices de manera periódica en revistas de negocios como *Bloomberg BusinessWeek* y *Forbes*, en la mayoría de los periódicos y en internet.

¿Qué importancia tiene un índice? ¿Por qué es tan importante y popular el IPC? Como su nombre lo indica, mide el cambio de precios de un grupo grande de artículos que compran los consumidores. El Departamento de la Reserva Federal de Estados Unidos, los grupos de consumidores, los sindicatos, los gerentes, las organizaciones de personas de la tercera edad, y otras organizaciones de negocios y de la economía se preocupan por los cambios de precios; todos ellos vigilan muy de cerca el IPC, así como el Índice de Precios al Productor (IPP), que mide las fluctuaciones de los precios en todas las etapas de la producción. Con el fin de combatir aumentos de precios repentinos, con frecuencia la Reserva Federal estadounidense aumenta la tasa de interés para "enfriar" la economía. De igual forma, el Promedio Industrial Dow Jones, que se actualiza de manera continua, describe el cambio general de precios de las acciones comunes de 30 grandes compañías.

Algunos índices del mercado accionario aparecen a diario en la sección financiera de la mayoría de los periódicos; muchos se reportan en tiempo real. Abajo se muestran el Promedio Industrial Dow Jones, el NASDAQ y otros índices obtenidos del sitio de Yahoo.com (**http://financeyahoo.com**).

OA17-1
Calcular e interpretar un índice simple y no ponderado.

Números índices simples

¿Qué es un número índice? Un índice, o número índice, mide el cambio que se produce en un artículo en particular (un producto o servicio) entre dos periodos.

> **NÚMERO ÍNDICE** Número que expresa el cambio relativo de precio, cantidad o valor comparado con un periodo base.

Si el número índice se utiliza para medir el cambio relativo de una sola variable, como los salarios por hora en la manufactura, es un índice simple porque es la razón de dos variables, y dicha razón se convierte en un porcentaje. En los siguientes cuatro ejemplos se ilustra el uso de los números índices; como se observa en la definición, su uso principal en los negocios es mostrar el cambio de uno o más aspectos de un periodo a otro.

EJEMPLO

De acuerdo con el Bureau of Labor Statistics, en 2000 el salario promedio por hora de los obreros en Estados Unidos era de 14.02 dólares; en 2013 fue de 24.92 dólares. ¿Cuál es el índice de salarios por hora de los obreros en 2013 con base en los datos de 2000?

SOLUCIÓN

Es 177.75, determinado por:

$$P = \frac{\text{Salario por hora promedio en 2013}}{\text{Salario por hora promedio en 2000}}(100) = \frac{\$24.92}{\$14.02}(100) = 177.75$$

Por lo tanto, el salario por hora de 2013 comparado con el de 2000 fue de 177.75%; esto significa que el salario aumentó en 77.75% por hora durante el periodo, determinado por $177.75 - 100.0 = 77.75$.

Puede revisar la información más reciente sobre salarios, los IPC y otros valores relacionados con los negocios en el sitio de internet del Bureau of Labor Statistics (BLS), http://www.bls.gov. En la siguiente tabla se muestran algunos estadísticos del BLS.

NÚMEROS RECIENTES

Índice de precios al consumidor:
−0.4% en abr 2013

Tasa de desempleo:
7.5% en abr 2013

Empleo asalariado:
+165 000(p) en abr 2013

Ingreso promedio por hora:
+$0.04(p) en abr 2013

Índice de precios al productor (IPP):
−0.7% en abr 2013

Índice de costos del empleo (ICE):
+0.3% en 1er trimestre de 2013

Productividad:
+0.7% en el 1er trimestre de 2013

Índice de precios a las importaciones (EE.UU.):
−0.5% en abr 2013

Índice de precios a las exportaciones (EE.UU.):
−0.7% en abr 2013

EJEMPLO

Un índice también compara un artículo con otro: la población de la provincia canadiense de Columbia Británica en 2012 fue de 4 606 451, y en Ontario, 13 472 428. ¿Cuál es el índice de población de Columbia Británica comparado con el de Ontario?

SOLUCIÓN

El índice de población de Columbia Británica es 34.2, determinado por:

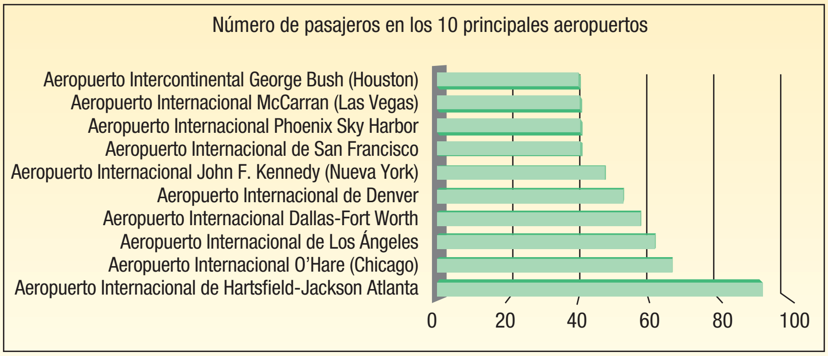
$$P = \frac{\text{Población de Columbia Británica}}{\text{Población de Ontario}}(100) = \frac{4\,606\,451}{13\,472\,438}(100) = 34.2$$

Esto indica que la población de Columbia Británica suma 34.2% (cerca de un tercio) de la población de Ontario, o que la población de Columbia Británica es 65.8% menor que la población de Ontario (100 − 34.2 = 65.8).

EJEMPLO

En la siguiente tabla se muestra el número de pasajeros (en millones) de los diez aeropuertos más grandes de Estados Unidos en 2011. Si se utiliza el Aeropuerto Intercontinental George Busch en Houston como base, ¿cuál es el índice de los otros aeropuertos comparados con Houston?

Rango	Aeropuerto	Pasajeros (millones)	Índice
1	Aeropuerto Internacional de Hartsfield-Jackson Atlanta	92.3	230.2
2	Aeropuerto Internacional O'Hare (Chicago)	66.7	166.3
3	Aeropuerto Internacional de Los Ángeles	61.9	154.4
4	Aeropuerto Internacional Dallas-Fort Worth	57.7	143.9
5	Aeropuerto Internacional de Denver	52.9	131.9
6	Aeropuerto Internacional John F. Kennedy (Nueva York)	47.7	119.0
7	Aeropuerto Internacional de San Francisco	40.8	101.7
8	Aeropuerto Internacional Phoenix Sky Harbor	40.6	101.2
9	Aeropuerto Internacional McCarran (Las Vegas)	40.6	101.2
10	Aeropuerto Intercontinental George Bush (Houston)	40.1	100.0



SOLUCIÓN

Para determinar los diez índices divida los pasajeros de Houston entre los de los otros nueve aeropuertos; así que el índice de Atlanta es 230.2, calculado por (92.3/40.1) × (100), y el del aeropuerto de San Francisco es 101.7, calculado por (40.8/40.1) × (100). De manera que Atlanta tiene 130.2% más pasajeros que Houston, y San Francisco tiene 1.7% más que Houston. En la gráfica anterior se resumen los índices y se revela que Houston, Las Vegas, Phoenix y San Francisco atendieron aproximadamente el mismo número de pasajeros en 2011 (estos cuatro índices están próximos a 100) comparados con el Aeropuerto Intercontinental George Bush; por otra parte, Atlanta atendió a casi el doble de pasajeros el mismo año.

En el análisis anterior se observa que:

1. El índice de salarios por hora promedio de los obreros (177.75) es un porcentaje, pero el símbolo casi siempre se omite.
2. Cada índice tiene ya sea una **base** o un **periodo base.** En el ejemplo sobre el salario por hora promedio de los obreros, se utilizó 2 000 como periodo base. El periodo base del IPC es 1982-84; en contraste, en el ejemplo sobre los pasajeros de los aeropuertos, el aeropuerto de Houston se usó como la base de comparación.

3. En negocios y economía, la mayoría de los índices se calculan hasta el número entero más cercano, como 214 o 96, o hasta el décimo más cercano de un porcentaje, como 83.4% o 118.7%.

¿Por qué convertir datos en índices?

La recopilación de números índices no es una innovación reciente; se acredita a un italiano, G.R. Carli, la creación de los números índices en 1764, cuando los incorporó en un reporte que hizo respecto de las fluctuaciones de precios en Europa de 1500 a 1750. En Estados Unidos no hubo un enfoque sistemático evidente para recopilar y reportar datos en forma de índice hasta alrededor de 1900. El índice del costo de la vida (que en la actualidad se denomina Índice de Precios al Consumidor, o IPC) se introdujo en 1913, y desde entonces se compila una larga lista de índices.

¿Por qué convertir los datos en índices? Un índice es una forma conveniente para expresar un cambio en un grupo diverso de artículos; el IPC, por ejemplo, abarca cerca de 200 categorías de artículos, clasificados en ocho grupos: 1) alimentos y bebidas, 2) vivienda, 3) vestido, 4) transporte, 5) cuidados médicos, 6) recreación, 7) educación y comunicación y 8) otros bienes y servicios. Los precios se expresan en muchas unidades diferentes, como dólares por libra o docenas de huevos. Solo mediante la conversión de los precios de estos diversos bienes y servicios en un número índice, el gobierno federal estadounidense y otros organismos preocupados por la inflación se mantienen informados del movimiento general de los precios al consumidor.

La conversión de datos en índices también facilita la evaluación de la tendencia en una serie compuesta de números muy grandes; por ejemplo, la estimación de las ventas al menudeo por internet (e-commerce) del primer trimestre de 2010, ajustado a la variación estacional, fue de 61 200 000 dólares. Las ventas de e-commerce del cuarto trimestre de 2009 sumaron 30 700 000 dólares; esta cifra representa un incremento de 30 500 000 dólares. Si las ventas de e-commerce del primer trimestre de 2013 se expresan como un índice basado en el cuarto trimestre de 2009, el índice es 199.3, y el incremento es de 99.3%.

$$\text{Índice} = \frac{\text{Ventas de e-commerce, 1er. trim. 2013}}{\text{Ventas de e-commerce, 4to. trim. 2009}}(100) = \frac{\$61\,200\,000}{\$30\,700\,000}(100) = 199.3$$

Elaboración de números índices

Ya se vio cómo elaborar un índice de precios simple: el precio de un año seleccionado (como 2014) se divide entre el precio del año base. El precio en el periodo base se designa p_0, y un precio que no sea el periodo base se conoce como *periodo dado* o *seleccionado,* y se designa p_t. Para calcular este índice de precios simple P con 100 como valor base de un periodo dado, utilice la fórmula:

ÍNDICE SIMPLE $$P = \frac{P_t}{P_0} \times 100 \quad \textbf{[17.1]}$$

Suponga que el precio de un paquete de vacaciones de fin de semana durante el otoño (con alojamiento y todos los alimentos) en Tryon Mountain Lodge, en el oeste de Carolina del Norte en 2000, fue de 450 dólares. El precio aumentó a 795 dólares en 2013. ¿Cuál es el índice de precios de 2013 con el año 2000 como periodo base y 100 como valor base? Es 176.7, determinado por:

$$P = \frac{P_t}{P_0}(100) = \frac{\$795}{\$450}(100) = 176.7$$

La interpretación de este resultado es que el precio del paquete de fin de semana durante el otoño aumentó 76.7% de 2000 a 2013.

El periodo base no necesita ser un año individual. En la tabla 17.1 se observa que si se emplea 2000-2001 = 100, el precio base de la engrapadora sería de 21 dólares [determinado al calcular el precio medio de 2000 y 2001: ($20 + $22)/2 = $21]. Los precios $20, $22 y $23 se promedian si se selecciona 2000-2002 como base. El precio medio sería de 21.67 dólares. Los índices elaborados con los tres periodos bases distintos se reportan en la tabla 17.1 (observe que, cuando 2000-2002 = 100, los números índices de 2000, 2001 y 2002 promedian 100.0, como cabría esperar). Como es lógico, los números índices de 2010 con las tres bases distintas son diferentes.

TABLA 17.1 Precios de una engrapadora automática Benson, modelo 3, convertidos en índices con tres periodos bases distintos

Año	Precio de la engrapadora	Índice de precios (2000 = 100)	Índice de precios (2000-01 = 100)	Índice de precios (2000-02 = 100)
1995	$18	90.0	$\frac{18}{21} \times 100 = 85.7$	$\frac{18}{21.67} \times 100 = 83.1$
2000	20	100.0	$\frac{20}{21} \times 100 = 95.2$	$\frac{20}{21.67} \times 100 = 92.3$
2001	22	110.0	$\frac{22}{21} \times 100 = 104.8$	$\frac{22}{21.67} \times 100 = 101.5$
2002	23	115.0	$\frac{23}{21} \times 100 = 109.5$	$\frac{23}{21.67} \times 100 = 106.1$
2010	38	190.0	$\frac{38}{21} \times 100 = 181.0$	$\frac{38}{21.67} \times 100 = 175.4$

AUTOEVALUACIÓN 17-1

1. En la tabla superior de la derecha se presentan las principales naciones productoras de acero en 2012; (http//www.worldsteel.org) exprese la cantidad que produjeron China, Japón, India, y Rusia como índice, y utilice a Estados Unidos como base. ¿Qué porcentaje produce China más que Estados Unidos?
2. En la tabla inferior de la derecha se presentan los salarios por hora promedio de obreros durante años seleccionados.

(a) Con 1995 como periodo base y 100 como valor base, determine los índices de otros años e interprete el índice.

(b) Utilice el promedio de 1995 y 2000 como base, determine los índices de los demás años e interprete el índice.

País	Cantidad (millones de toneladas)
China	716.5
Japón	107.2
Estados Unidos	88.7
India	77.6
Rusia	70.4

Año	Salarios por hora promedio
1995	$11.65
2000	14.02
2005	16.13
2013	19.97

EJERCICIOS

1. PNC Bank, Inc., con sede en Pittsburgh, Pennsylvania, reportó, en millones de dólares, 17 446 en concepto de préstamos comerciales en 1995, 19 989 en 1997, 21 468 en 1999, 21 685 en 2000, 15 922 en 2002, 18 375 en 2004, 54 818 en 2009 y 83 040 en 2012; utilice 1995 como base y desarrolle un índice simple que muestre el cambio de la cantidad de préstamos comerciales durante los años 2002, 2004, 2009 y 2012.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2. En la siguiente tabla se reportan las ganancias de cada una de las acciones comunes de Home Depot, Inc., de 2003 a 2012. Desarrolle un índice, con 2003 como base, que muestre el cambio de las ganancias por acción durante los años 2004 a 2012.

Año	Ganancias por acción	Año	Ganancias por acción
2003	1.56	2008	2.37
2004	1.88	2009	1.34
2005	2.26	2010	1.57
2006	2.72	2011	2.01
2007	2.79	2012	2.47

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

3. En la página siguiente se enumeran las ventas netas de un minorista de ventas por correo de San Francisco durante los años de 2003 a 2012; utilice las ventas medias de los primeros tres años para determinar una base y luego determine el índice de 2011 y 2012. ¿En cuánto aumentaron las ventas netas desde el periodo base?

Año	Ventas (millones de dólares)	Año	Ventas (millones de dólares)
2003	$486.6	2008	$568.5
2004	506.8	2009	581.9
2005	522.2	2010	496.1
2006	574.6	2011	456.6
2007	580.7	2012	433.3

4. En enero de 2003, el precio del pollo fresco entero fue 1.004 dólares por libra; en enero de 2013, de 1.497 dólares por libra; utilice el precio de enero de 2003 como periodo base y 100 como valor base para desarrollar un índice simple. ¿Qué porcentaje aumentó el precio del pollo?

Índices no ponderados

OA17-2 Calcular e interpretar un índice agregado y no ponderado.

En muchas situaciones se desea combinar varios artículos y elaborar un índice para comparar el costo de este agregado de artículos en dos periodos distintos; por ejemplo, podría necesitarse un índice que englobe los artículos que se relacionan con el gasto de operación y mantenimiento de un automóvil; cuyos artículos abarquen los precios de los neumáticos, aceite y gasolina. O bien, podría necesitarse un índice para estudiantes universitarios; este índice puede abarcar el costo de libros, colegiatura, alojamiento, alimentos y entretenimiento. Hay varias formas de combinar los artículos para determinar un índice.

Promedio simple de los índices de precios

En la tabla 17.2 se reportan los precios de varios artículos de alimentos de 2003 a 2013. Usted desea elaborar con ellos el índice de 2013 usando 2003 como base; esto se expresa con el código abreviado 2003 = 100.

TABLA 17.2 Cálculo del índice de precios de varios alimentos 2003 = 100

Artículo	Precio en 2003	Precio en 2013	Índice simple
Pan blanco (libra)	$1.042	$1.422	136.5
Huevos (docena)	1.175	1.933	164.5
Leche blanca (galón)	2.686	3.526	131.3
Manzanas, Red Delicious (libra)	0.911	1.350	148.2
Jugo de naranja concentrado (12 onzas)	1.848	2.511	135.9
Café, 100% grano tostado (libra)	2.999	5.902	196.8
Total	$10.661	$16.644	

En primer lugar, calcule un **promedio simple de los índices de precios de** cada artículo empleando 2003 como base; el índice simple del pan es 136.5, que se determina con la fórmula [17.1].

$$P = \frac{P_t}{P_0}(100) = \frac{1.422}{1.042}(100) = 136.5$$

El índice simple de los demás artículos de la tabla 17.2 se calcula de manera similar; el mayor aumento de precio fue 96.8% para el café, seguido por los huevos, con 64.5%.

También se puede determinar el porcentaje de cambio en el grupo de alimentos promediando los índices simples; cuya fórmula es:

PROMEDIO SIMPLE DE LOS PRECIOS RELATIVOS $$P = \frac{\Sigma p_i}{n} \qquad [17.2]$$

donde P_i representa el índice simple de cada artículo, y n, el número de artículos; en este ejemplo, el índice es 152.2, determinado por:

$$P = \frac{\Sigma p_i}{n} = \frac{136.5 + \cdots + 196.8}{6} = \frac{913.2}{6} = 152.2$$

Esto indica que la media del grupo de índices aumentó 52.2% de 2003 a 2013.

Una característica positiva del promedio simple de índices de precios es que se obtiene el mismo valor del índice sin importar las unidades de medida. Si en el índice anterior las manzanas estuvieran en toneladas, en lugar de libras, el impacto de estas en el índice combinado no cambiaría; es decir, la mercancía "manzanas" representa uno de seis artículos incluido en el índice, por lo cual el efecto del artículo no se relaciona con las unidades. Una característica negativa de este índice es que no considera la importancia relativa de los artículos que se consideran; por ejemplo, la leche y los huevos reciben la misma ponderación, aunque una familia común gaste mucho más durante el año en leche que en huevos.

Índice agregado simple

Una segunda probabilidad es sumar los precios (en lugar de los índices) de los dos periodos y luego determinar el índice con base en los totales. La fórmula es:

ÍNDICE AGREGADO SIMPLE $$P = \frac{\Sigma p_t}{\Sigma p_0} \times 100 \qquad \textbf{[17.3]}$$

A este se le denomina índice agregado simple. El índice de los artículos de alimentos que antes se presenta se determina dividiendo la suma de los precios de 2013 entre la suma de los precios de 2003. La suma de los precios del periodo base es 10.661 dólares, y la del periodo dado, 16.44 dólares. El índice agregado simple es 156.1; esto significa que el grupo de precios agregado aumentó 56.1% en el periodo de 10 años.

$$P = \frac{\Sigma p_t}{\Sigma P_0}(100) = \frac{\$16.644}{\$10.661}(100) = 156.1$$

Como las unidades de medición pueden influir en el valor de un índice agregado simple, este no se emplea con frecuencia. En el ejemplo, el valor del índice diferiría de manera significativa si el precio de las manzanas se reportara en toneladas en lugar de libras. También observe el efecto del café en el índice total. Tanto para el año actual como para el año base, el café es un contribuyente importante al índice total, por lo que un cambio de su precio incidirá en el índice mucho más que cualquier otro artículo. En consecuencia, es necesario encontrar una forma de "ponderar" de manera aproximada los artículos de acuerdo con su importancia relativa.

OA17-3
Calcular e interpretar un índice agregado y ponderado.

Índices ponderados

Los dos métodos más conocidos para calcular el índice de precios ponderado son el de **Laspeyres** y el de **Paasche**; los cuales difieren solo en el periodo de la ponderación. Cuando se emplea el método de Laspeyres se aplican *ponderaciones en el periodo base*; es decir, los precios y las cantidades originales de los artículos comprados se utilizan para encontrar el cambio porcentual durante un periodo, ya sea en el precio o en la cantidad consumida, según el problema. En el método de Paasche se aplican *ponderaciones en el año en curso*.

Índice de precios de Laspeyres

A finales del siglo XVIII, Etienne Laspeyres desarrolló un método para determinar un índice de precios ponderado con las cantidades del periodo base como ponderaciones; según dicho método, un índice de precios ponderado se calcula mediante:

ÍNDICE DE PRECIOS DE LASPEYRES $$P = \frac{\Sigma p_t q_0}{\Sigma p_0 q_0} \times 100 \qquad \textbf{[17.4]}$$

donde

P es el índice de precios;
P_t es el precio actual;
p_0 es el precio en el periodo base;
q_0 es la cantidad en el periodo base.

EJEMPLO

A continuación, los precios de los seis artículos de alimentos de la tabla 17.2 se repiten en la tabla 17.3; también se incluye el número típico de unidades consumidas por una familia promedio en 2003 y 2013.

TABLA 17.3 Precio y cantidad de artículos de alimentos en 2003 y 2013

	2003		2013	
Artículo	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Pan blanco (libra)	$1.042	50	$1.422	55
Huevos (docena)	1.175	26	1.933	20
Leche blanca (galón)	2.686	102	3.526	130
Manzanas, Red Delicious (libra)	0.911	30	1.350	40
Jugo de naranja concentrado (12 onzas)	1.848	40	2.511	41
Café, 100% grano tostado (libra)	2.999	12	5.902	12

Determine un índice de precios ponderado con el método de Laspeyres e interprete el resultado.

SOLUCIÓN

Primero determine la cantidad total que se gastó en los seis artículos en el periodo base (2003). Para encontrar este valor multiplique el precio, en el periodo base, del pan (1.042 dólares) por la cantidad en el periodo base (50), el resultado es 52.10 dólares; esta cifra indica que se gastó un total de 52.10 dólares en el periodo base en pan. Continúe de la misma manera con todos los artículos y sume los resultados; el total del periodo base es de 493.86 dólares. El total del periodo actual se calcula de manera similar: en el caso del primer artículo (pan), multiplique la cantidad en 2003 por el precio del pan en 2013, es decir, $1.422(50), el resultado es 71.10 dólares. Haga el mismo cálculo con cada artículo y sume el resultado; el total es 692.77 dólares. Debido a la naturaleza repetitiva de estas operaciones, se recomienda utilizar una hoja de cálculo para crear la siguiente tabla.

Índice Laspeyres

	2003		Precio 03	2013		Precio 13
Artículo	**Precio**	**Cantidad**	**Cantidad 03**	**Precio**	**Cantidad**	**Cantidad 03**
Pan blanco (libra)	$1.042	50	$ 52.10	$1.422	55	$ 71.10
Huevos (docena)	1.175	26	30.55	1.933	20	50.26
Leche blanca (galón)	2.686	102	273.97	3.526	130	359.65
Manzanas, Red Delicious (libra)	0.911	30	27.33	1.350	40	40.50
Jugo de naranja concentrado (12 onzas)	1.848	40	73.92	2.511	41	100.44
Café, 100% grano tostado (libra)	2.999	12	35.99	5.902	12	70.82
			$493.86			$692.77
				Índice	140.28	

El índice de precios ponderado de 2013 es 140.28, determinado por:

$$P = \frac{\Sigma p_t q_0}{\Sigma p_0 q_0}(100) = \frac{\$692.77}{\$493.86}(100) = 140.28$$

Con base en este análisis se concluye que el precio de este grupo de artículos aumentó 40.28% en el periodo de diez años. La ventaja de este método sobre el índice agregado simple es que se considera la ponderación de cada artículo. En el índice agregado simple, el café representaba alrededor de 40% de la ponderación para determinar el índice; pero en el índice de Laspeyres, el artículo con la mayor ponderación es la leche, debido a que el precio del producto y las unidades que se vendieron son mayores.

Índice de precios de Paasche

La desventaja principal del índice de Laspeyres es que se supone que las cantidades en el periodo base aún son reales en el periodo dado; es decir, las cantidades empleadas de los seis artículos son casi las mismas en 2003 y 2013. En este caso, la cantidad de huevos comprados declinó 23%, mientras que la cantidad de leche aumentó casi 28% y el número de manzanas subió 33%.

El índice de Paasche es una alternativa; su procedimiento es similar, pero en lugar de emplear cantidades del periodo base como ponderaciones, se utilizan cantidades del periodo actual. Se emplea la suma de los productos de los precios en 2003 y las cantidades en 2013; esto tiene la ventaja de emplear las cantidades más recientes. Si hubiera un cambio en las cantidades consumidas desde el periodo base, se reflejaría en el índice Paasche.

ÍNDICE DE PRECIOS DE PAASCHE $$P = \frac{\Sigma p_t q_t}{\Sigma p_0 q_t} \times 100 \quad \text{[17.5]}$$

EJEMPLO

Utilice la información de la tabla 17.3 para determinar el índice de Paasche y analice cuál de los índices debe usar.

SOLUCIÓN

En la tabla siguiente se muestran los cálculos para determinar el índice de Paasche. Debido a la naturaleza repetitiva de los cálculos, Excel es útil para determinar el índice.

Índice Paasche

	2003		Precio 03	2013		Precio 13
Artículo	**Precio**	**Cantidad**	**Cantidad 13**	**Precio**	**Cantidad**	**Cantidad 13**
Pan blanco (libra)	$1.042	50	$ 57.31	$1.422	55	$ 78.21
Huevos (docena)	1.175	26	23.50	1.933	20	38.66
Leche blanca (galón)	2.686	102	349.18	3.526	130	458.38
Manzanas, Red Delicious (libra)	0.911	30	36.44	1.350	40	54.00
Jugo de naranja concentrado (12 onzas)	1.848	40	75.77	2.511	41	102.95
Café, 100% grano tostado (libra)	2.999	12	35.99	5.902	12	70.82
			$578.19			$803.03
				Índice	138.9	

El índice de Paasche es 138.9, determinado por

$$P = \frac{\Sigma p_t q_t}{\Sigma p_0 q_t}(100) = \frac{\$803.03}{\$578.19}(100) = 138.9$$

Este resultado indica 38.9% de aumento en el precio de esta “canasta básica” de artículos entre 2003 y 2013; es decir, costó 38.9% más comprar estos artículos en 2013 que en 2003.

¿Cómo decidir cuál índice emplear? ¿Cuándo es más adecuado el índice de Laspeyres y cuándo lo es el de Paasche?

Laspeyres

Ventajas — Requiere datos sobre cantidades solo del periodo base, lo que permite una comparación más significativa en el transcurso del tiempo. Los cambios en el índice se pueden atribuir a cambios de precio.

Desventajas — No refleja cambios que el tiempo genera en los patrones de compra; además, puede ponderar demasiado los artículos cuyos precios aumentan.

Paasche

Ventajas — Como utiliza cantidades del periodo actual, refleja los hábitos actuales de compra.

Desventajas — Requiere datos de cantidades del año actual; como se utilizan cantidades diferentes cada año, es imposible atribuir cambios en el índice a cambios solo en el precio. Tiende a ponderar demasiado los artículos cuyos precios declinaron. Necesita que los precios se vuelvan a calcular cada año.

Índice ideal de Fisher

Como se observó antes, el índice de Laspeyres tiende a ponderar demasiado los artículos cuyos precios aumentaron; por otro lado, el de Paasche pondera demasiado los artículos cuyos precios disminuyeron. En un intento para compensar estas desventajas, Irving Fisher, en *The Making of Index Numbers*, publicado en 1922, propone un índice ideal de Fisher**,** compuesto por las medias geométricas de los índices de Laspeyres y Paasche. La media geométrica, descrita en el capítulo 3, se determina con la raíz k-ésima del producto de k números positivos.

$$\text{Índice ideal de Fisher} = \sqrt{(\text{Índice de Laspeyres})(\text{Índice de Paasche})} \quad \textbf{[17.6]}$$

En teoría, el índice de Fisher parece ideal porque combina las mejores características de los índices de Laspeyres y Paasche, es decir, equilibra los efectos de ambos índices; sin embargo, casi no se utiliza en la práctica debido a que tiene el mismo conjunto básico de problemas que el índice de Paasche, y precisa determinar un conjunto nuevo de cantidades en cada periodo.

EJEMPLO

Determine el índice ideal de Fisher con los datos de la tabla 17.3.

SOLUCIÓN

El índice ideal de Fisher es 139.6.

$$\begin{aligned}\text{Índice ideal de Fisher} &= \sqrt{(\text{Índice de Laspeyres})(\text{Índice de Paasche})}\\ &= \sqrt{(140.3)(138.9)} = 139.6\end{aligned}$$

AUTOEVALUACIÓN 17-2

Elabore el índice de precios de ropa de 2013 con base en 2000; considere zapatos y vestidos. Los precios y las cantidades de los dos años se dan en la siguiente tabla; utilice 2000 como periodo base y 100 como valor base.

	2000		2013	
Artículo	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Vestido (pieza)	$75	500	$85	520
Zapatos (par)	40	1 200	45	1 300

(a) Determine el promedio simple de los índices de precios.
(b) Determine el índice de precios agregado de los dos años.
(c) Determine el índice de precios de Laspeyres.
(d) Determine el índice de precios de Paasche.
(e) Determine el índice de precios ideal de Fisher.

EJERCICIOS

En los ejercicios 5 a 8:

a. Determine los índices de precios simples.
b. Determine el índice de precios agregado simple de los dos años.
c. Determine el índice de precios de Laspeyres.

d. Determine el índice de precios de Paasche.
e. Determine el índice ideal de Fisher.

5. A continuación se presentan los precios de dentífrico (9 onzas), champú (7 onzas), pastillas para la tos (paquete de 100), y antitranspirante (2 onzas) de agosto de 2000 y agosto de 2013; además, se incluyen las cantidades compradas; utilice agosto de 2000 como base.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Agosto de 2000		Agosto de 2013	
Artículo	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Dentífrico	$2.49	6	$3.35	6
Champú	3.29	4	4.49	5
Pastillas para la tos	1.59	2	4.19	3
Antitranspirante	1.79	3	2.49	4

6. En la siguiente tabla se reportan los precios de frutas y las cantidades consumidas en 2000 y 2014; utilice 2000 como base.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	2000		2014	
Fruta	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Plátanos (libra)	$0.23	100	$0.69	120
Toronja (pieza)	0.29	50	1.00	55
Manzanas (libra)	0.35	85	1.89	85
Fresas (canasta)	1.02	8	3.79	10
Naranjas (costal)	0.89	6	2.99	8

7. En la siguiente tabla se reportan los precios y los números de varios artículos que produce una máquina pequeña y una planta troqueladora; utilice 2000 como base.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	2000		2014	
Artículo	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Arandela	$0.07	17 000	$0.10	20 000
Chaveta	0.04	125 000	0.03	130 000
Perno para estufa	0.15	40 000	0.15	42 000
Tuerca hexagonal	0.08	62 000	0.10	65 000

8. A continuación se mencionan las cantidades y los precios de los años 2000 y 2014 de Kinzua Valley Geriatrics; utilice 2000 como periodo base.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	2000		2014	
Artículo	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Jeringas (docena)	$ 6.10	1 500	$ 6.83	2 000
Termómetros	8.10	10	9.35	12
Analgésico Advil (frasco)	4.00	250	4.62	250
Formas para historias clínicas (caja)	6.00	1 000	6.85	900
Papel para impresora (caja)	12.00	30	13.65	40

Índice de valores

Un índice de **valores** (como el índice de ventas en tiendas departamentales) mide cambios de precios y las cantidades implicadas; para su elaboración, se deben considerar los precios y las cantidades del año base, y los precios y las cantidades del año actual. Su fórmula es:

ÍNDICE DE VALORES

$$V = \frac{\Sigma p_t q_t}{\Sigma p_0 q_0} \times 100 \qquad \textbf{[17.7]}$$

EJEMPLO

Los precios y las cantidades que vendió Waleska Clothing Emporium de varias prendas de vestir en mayo de 2000 y mayo de 2014 son:

	2000		2014	
Artículo	Precio p_0	Cantidad q_0	Precio p_t	Cantidad q_t
Corbatas (pieza)	$ 1.00	1 000	$ 2.00	900
Trajes (pieza)	30.00	100	40.00	120
Zapatos (par)	10.00	500	8.00	500

Si mayo de 2000 es el periodo base, ¿cuál es el índice de valores de mayo de 2014?

SOLUCIÓN

Las ventas totales en mayo de 2014 ascendieron a 10 600 dólares, y la cifra comparable para 2000 es de 9 000 dólares (consulte la tabla 17.4); por lo tanto, el índice de valores de mayo de 2014 usando 2000 = 100 es 117.8. El valor de las ventas de ropa en 2014 fue 117.8% de las ventas en 2000; en otras palabras, el valor de las ventas de ropa aumentó 17.8% de mayo de 2000 a mayo de 2014.

$$V = \frac{\Sigma p_t q_t}{\Sigma p_0 q_0}(100) = \frac{\$10\ 600}{\$9\ 000}(100) = 117.8$$

TABLA 17.4 Elaboración de un índice de valores para 2014 (2000 = 100)

	2000			2014		
Artículo	Precio p_0	Cantidad q_0	$ p_0q_0	Precio p_t	Cantidad q_t	$ p_tq_t
Corbatas (pieza)	$ 1.00	1 000	$1 000	$ 2.00	900	$ 1 800.00
Trajes (pieza)	30.00	100	3 000	40.00	120	4 800.00
Zapatos (par)	10.00	500	5 000	8.00	500	4 000.00
Total			$9 000	Índice = 117.8		$10 600.00

AUTOEVALUACIÓN 17-3

El número de artículos que produjo Houghton Products en 2001 y 2013, y los precios al mayoreo de los dos periodos son:

(a) Encuentre el índice de valores de la producción de 2013 con 2001 como periodo base.

(b) Interprete el valor del índice.

	Precio		Número producido	
Artículo producido	2001	2013	2001	2013
Pernos de tijeras (caja)	$ 3	$4	10 000	9 000
Compuesto para corte (libra)	1	5	600	200
Varillas de tensión (pieza)	10	8	3 000	5 000

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

9. En seguida se reportan los precios y la producción de granos en Estados Unidos en 2002 y 2012 (**http://ers.usda.gov**).

Grano	2002 Precio por fanega	2002 Producción (millones de fanegas)	2012 Precio por fanega	2012 Producción (millones de fanegas)
Avena	$1.81	116	$3.8	64
Trigo	3.56	2	7.9	2
Maíz	2.32	8 967	6.9	10 780
Cebada	2.72	227	6.4	220

Con 2002 como periodo base, encuentre el índice de valores de los granos que se produjeron en agosto de 2012.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

10. Johnson Wholesale Company fabrica productos diversos. Los precios y las cantidades que produjo en abril de 2000 y abril de 2013 son:

Producto	Precio en 2000	Precio en 2013	Cantidad producida en 2000	Cantidad producida en 2013
Motor pequeño (pieza)	$23.60	$28.80	1 760	4 259
Compuesto depurador (galón)	2.96	3.08	86 450	62 949
Clavos (libra)	0.40	0.48	9 460	22 370

Con abril de 2000 como periodo base, encuentre el índice de valores de los artículos producidos en abril de 2013.

OA17-4
Enumerar y describir índices que tienen un propósito especial.

Índices para propósitos especiales

Muchos índices importantes se elaboran y publican mediante organizaciones privadas. J.D. Power & Associates realiza encuestas entre compradores de automóviles para determinar la satisfacción de los clientes con sus vehículos después de un año de poseerlo; este índice especial se denomina Índice de Satisfacción del Consumidor. Con frecuencia, instituciones financieras, compañías de servicios y centros de investigación de universidades elaboran índices sobre el empleo, jornadas laborales y salarios, y ventas al menudeo de las regiones donde se ubican. Muchas asociaciones comerciales elaboran índices de precios y cantidades vitales de su área particular de interés. ¿Cómo se elaboran? Un ejemplo, simplificado, por supuesto, ayudará a explicar algunos detalles.

EJEMPLO

La Cámara de Comercio de Seattle desea elaborar una medida de la actividad de negocios general de la zona noroeste de Estados Unidos; con este fin, al director de Desarrollo Económico se le asignó crear el Índice General de Actividades de Negocios del Noroeste.

SOLUCIÓN

Después de muchas ideas e investigaciones, el director concluyó que se deben considerar cuatro factores: 1) las ventas en tiendas departamentales de la región (que se reportan en millones de dólares), 2) el índice de empleo regional (que tiene como base el año 2005 y se reporta en la información del estado de Washington), 3) los embarques en transportes de carga (reportados en millones) y 4) las exportaciones del muelle de Seattle (reportadas en miles de toneladas). En la tabla 17.5 se incluye información reciente para los años 2005, 2010 y 2013.

TABLA 17.5 Datos para calcular el Índice General de Actividades de Negocios del Noroeste

Año	Ventas de tiendas departamentales	Índice de empleo	Embarques en transporte de carga	Exportaciones
2005	20	100	50	500
2010	41	110	30	900
2013	44	125	18	700

Después de revisar y consultar los datos, el director asignó ponderaciones de 40% a las ventas de tiendas departamentales, 30% al empleo, 10% a los embarques en transportes de carga y 20% a las exportaciones. Para elaborar el Índice General de Actividades de Negocios del Noroeste de 2013, con 2005 = 100, cada valor de 2013 se expresa como porcentaje, con el valor del periodo base como denominador. Para ilustrar esta operación, las ventas de tiendas departamentales en 2013 se convierten en porcentajes mediante ($44/$20)(100) = 220; esto significa que las ventas de tiendas departamentales aumentaron 120% en el periodo. Luego, este porcentaje se multiplica por la ponderación apropiada. En el caso de las ventas de tiendas departamentales es (220)(0.40) = 88.0; los detalles de los cálculos de 2010 y 2013 se muestran a continuación.

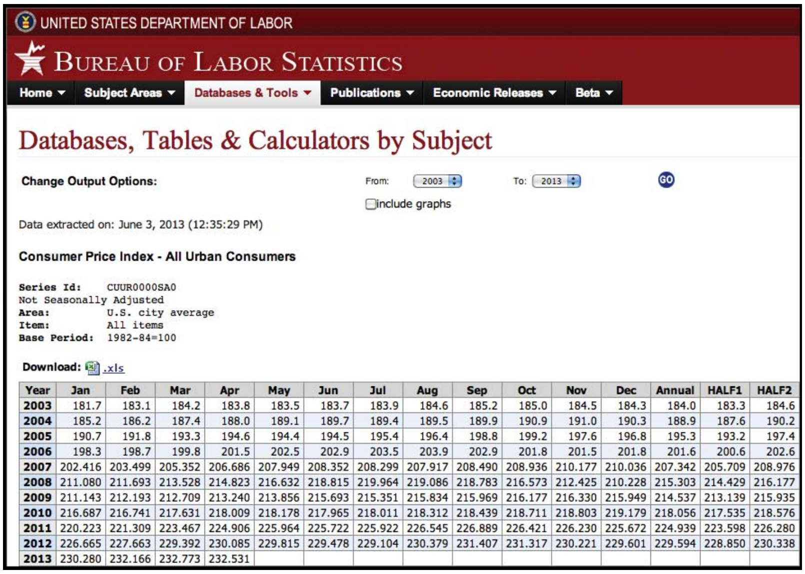
	2010	2013
Ventas de tiendas departamentales	[($41)/($20)][100][.40] = 82.0	[($44)/($20)][100][.40] = 88.0
Índice de empleo	[(110)/(100)][100][.30] = 33.0	[(125)/(100)][100][.30] = 37.5
Embarques en transporte de carga	[(30)/(50)][100][.10] = 6.0	[(18)/(50)][100][.10] = 3.6
Exportaciones	[(900)/(500)][100][.20] = 36.0	[(700)/(500)][100][.20] = 28.0
Total	157.0	157.1

El Índice General de Actividades de Negocios del Noroeste de 2010 es 157.0, mientras que el de 2013 es 157.1. La interpretación de estos índices es que la actividad de negocios aumentó 57.0% de 2005 a 2010, y 57.1% del periodo base de 2005 a 2013.

Como ya se dijo al inicio de esta sección, hay muchos índices para propósitos especiales. He aquí algunos ejemplos.

Índice de Precios al Consumidor

El Buró de Estadísticas Laborales de Estados Unidos reporta este índice cada mes; en el cual se describen los cambios de precios de un periodo a otro de una "canasta básica" de productos y servicios. En la siguiente sección se estudia su historia en detalle y se presentan algunas aplicaciones; esta información está disponible (en inglés) en http://www.bls.gov, en **Databases & Tools** seleccione **Inflation & Prices**, después seleccione **Top Picks** de **All Urban Consumers (Current Series),** haga clic en la casilla de **U.S all items 1982-84 = 100** y luego haga clic en **Retrieve data.** Quizá desee incluir periodos diferentes. A continuación se resume un informe reciente.


UNITED STATES DEPARTMENT OF LABOR
BUREAU OF LABOR STATISTICS
Home | Subject Areas | Databases & Tools | Publications | Economic Releases | Beta

Databases, Tables & Calculators by Subject

Change Output Options: From: 2003 To: 2013 GO
include graphs

Data extracted on: June 3, 2013 (12:35:29 PM)

Consumer Price Index - All Urban Consumers

```
Series Id:    CUUR0000SA0
Not Seasonally Adjusted
Area:         U.S. city average
Item:         All items
Base Period:  1982-84=100
```

Download: .xls

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual	HALF1	HALF2
2003	181.7	183.1	184.2	183.8	183.5	183.7	183.9	184.6	185.2	185.0	184.5	184.3	184.0	183.3	184.6
2004	185.2	186.2	187.4	188.0	189.1	189.7	189.4	189.5	189.9	190.9	191.0	190.3	188.9	187.6	190.2
2005	190.7	191.8	193.3	194.6	194.4	194.5	195.4	196.4	198.8	199.2	197.6	196.8	195.3	193.2	197.4
2006	198.3	198.7	199.8	201.5	202.5	202.9	203.5	203.9	202.9	201.8	201.5	201.8	201.6	200.6	202.6
2007	202.416	203.499	205.352	206.686	207.949	208.352	208.299	207.917	208.490	208.936	210.177	210.036	207.342	205.709	208.976
2008	211.080	211.693	213.528	214.823	216.632	218.815	219.964	219.086	218.783	216.573	212.425	210.228	215.303	214.429	216.177
2009	211.143	212.193	212.709	213.240	213.856	215.693	215.351	215.834	215.969	216.177	216.330	215.949	214.537	213.139	215.935
2010	216.687	216.741	217.631	218.009	218.178	217.965	218.011	218.312	218.439	218.711	218.803	219.179	218.056	217.535	218.576
2011	220.223	221.309	223.467	224.906	225.964	225.722	225.922	226.545	226.889	226.421	226.230	225.672	224.939	223.598	226.280
2012	226.665	227.663	229.392	230.085	229.815	229.478	229.104	230.379	231.407	231.317	230.221	229.601	229.594	228.850	230.338
2013	230.280	232.166	232.773	232.531											

Índice de Precios al Productor

Lo publica el Buró de Estadísticas Laborales; antes se denominaba Índice de Precios al Mayoreo, que data de 1890. Refleja los precios de más de 3 400 productos; los datos de precios se recopilan de los vendedores de los productos, y por lo general se refieren a la primera transacción de gran volumen de cada producto (es un índice tipo Laspeyres); para consultar esta información, visite http://www.bls.gov, luego en "Databases & Tools" seleccione "Inflation & Prices"**,** después seleccione **"**Top Picks" de "Commodity Data", haga clic en la casilla "Finished goods" y luego en "Retrieve data"**.** Quizá desee incluir periodos diferentes. En la página siguiente se muestra una captura de pantalla reciente.

UNITED STATES DEPARTMENT OF LABOR

BUREAU OF LABOR STATISTICS

Home | Subject Areas | Databases & Tools | Publications | Economic Releases | Beta

Databases, Tables & Calculators by Subject

Change Output Options: From: 2003 To: 2013

include graphs

Data extracted on: June 3, 2013 (12:33:45 PM)

Producer Price Index-Commodities

```
Series Id:  WPUSOP3000
Not Seasonally Adjusted
Group:      Stage of processing
Item:       Finished goods
Base Date:  198200
```

Download: .xls

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual
2003	140.8	142.3	144.2	142.1	142.0	143.0	143.0	143.7	144.0	145.5	144.5	144.5	143.3
2004	145.4	145.3	146.3	147.3	148.9	148.7	148.5	148.5	148.7	152.0	151.7	150.6	148.5
2005	151.4	152.1	153.6	154.4	154.3	154.2	155.5	156.3	158.9	160.9	158.3	158.7	155.7
2006	159.9	158.0	159.1	160.7	161.2	161.8	161.7	162.3	160.3	158.9	159.8	160.5	160.4
2007	160.1	161.8	164.1	165.9	167.5	167.2	168.5	166.1	167.4	168.6	171.4	170.4	166.6
2008	172.0	172.3	175.1	176.5	179.8	182.4	185.1	182.2	182.2	177.4	172.0	168.8	177.1
2009	170.4	169.9	169.1	170.3	171.1	174.3	172.4	174.2	173.2	173.8	175.7(R)	176.0	172.5
2010	178.0	177.0	179.1	179.5	179.8	179.0	179.5	179.9	180.0	181.2	181.6	182.6	179.8
2011	184.4	186.6	189.1	191.4	192.5	191.4	192.2	191.7	192.6	191.8	191.7	191.1	190.5
2012	192.0	192.9	194.4	194.9	193.7	192.8	193.2	195.4	196.7	196.3	194.5	193.7	194.2
2013	194.7(P)	196.2(P)	196.6(P)	196.0(P)									

R : Revised
P : Preliminary. All indexes are subject to revision four months after original publication.

Promedio Industrial Dow Jones

Es un índice de precios accionarios, pero tal vez sería mejor llamarlo “indicador” en lugar de índice. Se supone que es el precio medio de 30 acciones industriales específicas; sin embargo, al sumar los 30 precios accionarios y dividir entre 30 no se obtiene su valor; esto se debe a las divisiones accionarias, a las fusiones y a la adición y eliminación de acciones. Cuando ocurre algún cambio, se hacen ajustes en el denominador empleado con el promedio; en la actualidad, el Dow Jones Industrial Average (DJIA) es más un indicador psicológico que una representación del movimiento general de precios en la Bolsa de Valores de Nueva York. La falta de representatividad de las acciones en el DJIA es una de las razones para el desarrollo del Índice de la Bolsa de **Valores de Nueva York** (New York Stock Exchange Index); el cual se desarrolló como un precio promedio de *todas* las acciones que se cotizan en la bolsa de esa ciudad. Puede encontrar más información sobre el DJIA en el sitio web **http://www.dowjones.com**, y puede encontrar su valor actual visitando **http://www.marketwatch.com**. Ahí se muestran los cambios en el Dow, NASDAQ, S&P 500 y otros índices para el día vigente.

Haciendo clic en “Dow” puede encontrar detalles adicionales sobre los cambios. En la parte central inferior de la tabla se puede cambiar la escala horizontal para mostrar los cambios por un día, cinco días y así sucesivamente hasta cinco años; en este caso se cambió la escala para mostrar los cambios por un año. La información sobre el NASDAQ y el S&P 500 también está disponible haciendo clic en ellos.

Index Components

SYMBOL	LAST	CHANGE	% CHANGE
MRK	48.81	+2.11	+4.52%
INTC	25.19	+0.91	+3.74%
KO	40.55	+0.56	+1.40%
HPQ	24.78	+0.36	+1.45%
BA	100.36	+1.34	+1.35%
PFE	27.58	+0.35	+1.29%
MCD	97.50	+0.93	+0.96%
WMT	75.56	+0.72	+0.96%
CVX	123.81	+1.06	+0.86%
MSFT	35.20	+0.30	+0.86%

En la misma página, Marketwatch enlista las 30 acciones que constituyen el Dow, y su cambio por el día de hoy. A la derecha se presenta el cambio en el precio para las 10 acciones que tuvieron el mayor cambio positivo.

El símbolo MRK representa a Merk & Co., la compañía (en este caso, farmacéutica) que tuvo el mayor aumento en el día de la consulta. El incremento fue 2.11 dólares, o 4.52%. INTC, el símbolo de mercado para Intel Corporation, aumentó 0.91 dólares, o 3.74%.

AUTOEVALUACIÓN 17-4

Como pasante en la Fulton County Economic Development Office, le piden desarrollar un índice para propósitos especiales de su condado. Tres series económicas parecen prometedoras como bases de un índice; estos datos son el precio del algodón (por libra), el número de automóviles nuevos vendidos en el condado, y la tasa de movimientos de dinero (publicada por el banco local). Después de analizar el proyecto con su supervisor y el director, usted decide que la tasa de movimiento de dinero deberá tener una ponderación de 0.60; el número de automóviles nuevos vendidos, una ponderación de 0.30, y el precio del algodón, una de 0.10. El periodo base es 1999.

Año	Precio del algodón	Automóviles vendidos	Movimientos de dinero
2004	$0.20	1 000	80
2009	0.25	1 200	90
2014	0.50	900	75

(a) Elabore el índice de 2009 y 2014.
(b) Interprete el índice de 2009 y 2014.

EJERCICIOS

11. El índice de los principales indicadores económicos, compilado y publicado por el U.S. National Bureau of Economic Research, se compone de 12 series de tiempo, como las horas laborales promedio de producción en manufactura, los nuevos pedidos a los fabricantes y la oferta de dinero; este índice y otros similares se diseñan para fluctuar hacia arriba o hacia abajo antes de que la economía cambie de igual forma. Con esta herramienta, un economista tiene evidencia estadística para predecir tendencias.

Usted desea elaborar el indicador principal de Erie County, en el norte de Nueva York; cuya base estará compuesta por datos de 2000. Debido al tiempo y al trabajo implicado, decide emplear solo cuatro series de tiempo; como experimento, seleccione estas cuatro series: desempleo en el condado, el índice compuesto de precios accionarios del condado, el Índice de Precios del Condado y las ventas al menudeo. A la derecha están las cifras de 2000 y 2014.

	2000	2014
Tasa de desempleo	5.3	6.8
Acciones compuestas del condado	265.88	362.26
Índice de Precios del Condado (1982 = 100)	109.6	125.0
Ventas al menudeo (millones de dólares)	529 917.0	622 864.0

Las ponderaciones que asigna son: tasa de desempleo 20%, precios accionarios 40%, Índice de Precios del Condado 25% y ventas al menudeo 15%.

a. Con 2000 como periodo base, elabore un indicador económico principal para 2014.

b. Interprete su índice principal.

12. Usted es empleado en la oficina estatal de desarrollo económico, y necesita un índice económico principal para revisar la actividad económica pasada y predecir las tendencias económicas que afectarán al estado; por lo tanto, decide incluir varios factores clave: número de negocios nuevos iniciados durante el año, número de negocios fallidos, recibos de impuesto al ingreso en el estado, inscripciones en universidades y los recibos de los impuestos sobre las ventas en el estado; estos son los datos de 2000 y 2009.

	2000	2012
Negocios nuevos	1 088	1 162
Negocios fallidos	627	520
Recibos de impuestos al ingreso en el estado (en millones de dólares)	191.7	162.6
Inscripciones en las universidades	242 119	290 841
Impuesto sobre las ventas en el estado (en millones de dólares)	41.6	39.9

a. Establezca las ponderaciones que se van a aplicar a cada elemento del índice principal.

b. Calcule el indicador económico principal de 2012.

c. Interprete los índices.

OA17-5 Aplicar el Índice de Precios al Consumidor.

Índice de precios al consumidor

En páginas anteriores se menciona con frecuencia el IPC; este mide el cambio de precios de una canasta básica fija de bienes y servicios de un periodo a otro. En enero de 1978, el Bureau of Labor Statistics inició la publicación del IPC para dos grupos de la población. Un índice, denominado Índice de Precios al Consumidor para todos los Consumidores Urbanos, cubre casi 87% de la población total; el otro es para los asalariados urbanos y oficinistas, y cubre casi 32% de la población.

El IPC tiene varias funciones importantes; por ejemplo, permite que los consumidores determinen el grado en que se reduce su poder de compra debido a los incrementos de precios. En ese sentido, es una medida para revisar salarios, pensiones y otros pagos de ingresos a fin de ir a la par con los cambios de precios; también es un indicador económico de la tasa de inflación en Estados Unidos.

El índice se basa en los precios de 80 000 artículos, recabados cada mes por cerca de 250 agentes; estos se recopilan en miles de establecimientos minoristas, de servicio, unidades residenciales y consultorios médicos (**http://stats.bis.gov/cpi/cpifag.htm**). Pan, cerveza, gasolina, cortes de cabello, tasas de intereses hipotecarios, honorarios médicos, impuestos y tarifas de quirófanos son solo algunos de los artículos incluidos en lo que con frecuencia se conoce como “canasta básica” de los bienes y servicios que adquieren los consumidores típicos.

El IPC se originó en 1913 y se publica de forma regular desde 1921; el periodo estándar de referencia es 1982-1984. Los periodos base anteriores fueron: 1967, 1957-1959, 1947-1949, 1935-1939 y 1925-1929. ¿Por qué es necesario cambiar la base? Los patrones de gasto cambian de forma dramática, y esto se debe reflejar en el índice.

En realidad, el IPC no solo es un índice; hay Índices de Precios al Consumidor de Nueva York, Chicago, Seattle y Atlanta, así como de otras ciudades grandes. También hay índices de precios de alimentos, ropa, servicios médicos y otros artículos; algunos de ellos se muestran a la derecha, 1982-1984 = 100, de agosto de 2013.

Artículo	IPC-U
Todos los artículos	233.877
Alimentos y bebidas	237.348
Ropa	125.767
Transporte	219.217
Servicios médicos	426.866
Vivienda	228.564

Una lectura de esta lista muestra que el índice ponderado de todos los artículos aumentó 133.877% desde 1982-84; los servicios médicos aumentaron más (326.866%); y la ropa subió menos (25.767%).

ESTADÍSTICA EN ACCIÓN

¿Tiene la impresión de que los precios solo aumentan? El IPC, calculado y reportado por el U.S. Department of Labor, es una medida relativa de cambios de precios que proporciona información interesante sobre los precios en categorías de productos y servicios; por ejemplo, ¿sabía que el IPC para computadoras personales y dispositivos periféricos fue de 58.045 en abril de 2013? Usando como base diciembre de 2007, esto significa que los precios relativos de computadoras y periféricos disminuyeron casi 42% desde esa fecha.

Casos especiales del IPC

Además de medir los cambios de los precios de bienes y servicios, los dos índices de precios al consumidor tienen diversas aplicaciones. Con el IPC se determina el ingreso personal disponible, la deflación de las ventas u otras variables, el poder de compra del dólar y el aumento del costo de vida. Primero se analiza el uso del IPC para determinar el **ingreso real.**

Ingreso real Como ejemplo del significado y cálculo del *ingreso real*, suponga que el IPC actual es 200 con 1982-1984 = 100; y que la señora Watts ganó 20 000 dólares por año en el periodo base de 1982, 1983 y 1984. Ella tiene un ingreso actual de 40 000 dólares. Observe que aunque su *ingreso monetario* aumentó al doble desde el periodo base de 1982-1984, los precios que pagó por alimentos, gasolina, ropa y otros artículos también aumentaron el doble; por lo tanto, el estándar de vida de la señora Watts permaneció igual desde el periodo base hasta la actualidad. Los aumentos de precios compensaron exactamente el aumento del ingreso, por lo que su poder de compra actual (ingreso real) aún es de 20 000 dólares (consulte la tabla 17.6 para los cálculos). En general:

INGRESO REAL

$$\text{Ingreso real} = \frac{\text{Ingreso monetario}}{\text{IPC}} \times 100 \qquad \textbf{[17.8]}$$

TABLA 17.6 Cálculo del ingreso real en 1982-1984 y en el año en curso

Año	Ingreso monetario anual	Índice de Precios al Consumidor (1982-84 = 100)	Cálculo del ingreso real	Ingreso real
1982-84	\$20 000	100	$\frac{\$20\ 000}{100}(100)$	\$20 000
Año en curso	40 000	200	$\frac{\$40\ 000}{200}(100)$	20 000

Algunas veces, el concepto de ingreso real se denomina *ingreso deflacionado,* y el IPC se denomina índice de deflación; además, un término popular para designar el ingreso deflacionado es *ingreso expresado en dólares constantes*. Así, para determinar si el estándar de vida de la señora Watts cambió, su ingreso monetario se convirtió en dólares constantes en la tabla 17.6; y se determinó que su poder de compra, expresado en dólares de 1982-1984 (dólares constantes), permaneció en 20 000 dólares.

AUTOEVALUACIÓN 17-5

El salario neto de Jon Greene, y el IPC de 2000 y 2013 son:

Año	Pago neto	IPC (1982-84 = 100)
2000	\$25 000	170.8
2013	41 200	229.6

(a) ¿Cuál fue el ingreso real de Jon en 2000?
(b) ¿Cuál fue su ingreso real en 2009?
(c) Interprete sus resultados.

Ventas deflacionadas Un índice de precios también sirve para "deflacionar" las ventas o series monetarias similares. Las ventas deflacionadas se determinan mediante:

USO DE UN ÍNDICE COMO FACTOR DE DEFLACIÓN

$$\text{Ventas deflacionadas} = \frac{\text{Ventas reales}}{\text{Un índice apropiado}} \times 100 \qquad \textbf{[17.9]}$$

EJEMPLO

Las ventas de Hill Enterprises, pequeña compañía de moldeo por inyección al norte de Nueva York, aumentaron de 875 000 dólares en 1982 a 1 482 000 dólares en 1995, a 1 491 000 dólares en 2000, a 1 502 000 dólares en 2004, a 1 515 000 dólares en 2007, a 1 596 000 dólares en 2009 y a 1 697 000 dólares en 2013. El propietario, Harry Hill, observa que el precio de la materia prima para el proceso también aumentó durante el mismo periodo, por lo que desea deflacionar las ventas para tomar en cuenta el aumento de precios de las materias primas. ¿Cuáles son las ventas deflacionadas de 1995, 2000, 2004, 2007, 2009 y 2013 con base en dólares de 1982? Es decir, ¿cuáles son las ventas de 1995, 2000, 2004, 2007, 2009 y 2013 expresadas en dólares constantes de 1982?

SOLUCIÓN

El Índice de Precios al Productor (IPP) se publica cada mes en el *Monthly Labor Review*; también se encuentra disponible en el sitio web de la Bureau of Labor Statistics. Los precios del IPP reflejan lo que paga el fabricante por metales, caucho y otros artículos; por lo tanto, parece un índice apropiado para deflacionar las ventas del fabricante; las cuales se presentan en la segunda columna de la tabla 17.7, y el IPP de cada año se encuentra en la tercera columna. En la siguiente columna se muestran las ventas divididas entre el IPP, y en la columna de la derecha se dan los detalles de los cálculos. Los resultados se muestran en la siguiente salida de Excel.

TABLA 17.7 Cálculo de las ventas deflacionadas de Hill Enterprises

Año	Ventas	IPP	Dólares constantes	Calculados por:
1982	$ 875 000	100.0	$ 875 000.00	($875 000/100.0)*(100)
1995	1 482 000	127.9	1 158 717.75	($1 482 000/127.9)*(100)
2000	1 491 000	139.0	1 072 661.87	($1 491 000/139.0)*(100)
2004	1 502 000	148.5	1 011 447.81	($1 502 000/148.5)*(100)
2007	1 515 000	166.6	909 363.75	($1 515 000/166.6)*(100)
2009	1 596 000	172.5	925 217.39	($1 596 000/172.5)*(100)
2013	1 697 000	194.2	873 841.40	($1 697 000/194.2)*(100)

De hecho, las ventas aumentaron 93.9% desde 1982 a 2013 [($1 697 000/$875 000)*(100)]. Pero si compara las ventas deflacionadas con el IPP, los dólares constantes de las ventas declinaron desde 1982.

Poder de compra del dólar Con el IPC también se determina el *poder de compra del dólar*.

USO DE UN ÍNDICE PARA DETERMINAR EL PODER DE COMPRA

$$\text{Poder de compra del dólar} = \frac{\$1}{\text{IPC}} \times 100 \quad \textbf{[17.10]}$$

EJEMPLO

Suponga que el IPC de este mes es 200.0 (1982-1984 = 100). ¿Cuál es el poder de compra del dólar?

SOLUCIÓN

De acuerdo con la fórmula [17.10], es 50 centavos, determinado por

$$\text{Poder de compra del dólar} = \frac{\$1}{200.0}(100) = \$0.50$$

El IPC de 200.0 indica que los precios se incrementaron al doble desde 1982-1984 hasta este mes; por ello, el poder de compra de un dólar disminuyó a la mitad, es decir, un dólar de 1982-1984 vale solo 50 centavos este mes. En otras palabras, si usted perdió 1 000 dólares en el periodo 1982-1984 y los acaba de encontrar, con ellos solo podrá comprar la mitad de lo que pudo comprar en 1982, 1983 y 1984.

Ajustes del costo de vida El IPC también es la base para realizar ajustes del costo de vida (*Cost-of-Living Adjustments*, COLA) en muchos contratos entre empresas y sindicatos. Con frecuencia, al punto específico del contrato se le denomina "cláusula escaladora". Cerca de 31 millones de beneficiarios de la seguridad social, 2.5 millones de militares y empleados del servicio civil federal jubilados y pensionistas, y 600 000 trabajadores del servicio postal tienen sus ingresos o pensiones basadas en el IPC.

Este índice también se utiliza para ajustar los pagos de pensión alimentaria y manutención; honorarios de abogados; pagos de compensaciones a trabajadores; rentas de departamentos, casas y edificios de oficinas; pagos del seguro de desempleo, etcétera; por ejemplo, una persona jubilada recibe una pensión de 500 dólares al mes y el IPC aumenta 5 puntos (de 165 a 170); suponga que por cada punto de aumento del IPC los beneficios de la pensión aumentan 1%; por lo tanto, el aumento mensual de los beneficios será de 15.15 dólares, determinado por $[(170 - 165)/165]) \times 500$. Ahora la persona jubilada recibirá 515.15 dólares al mes.

AUTOEVALUACIÓN 17-6

Suponga que el IPC del mes anterior fue 195.4 (1982-1984) = 100). Determine el poder de compra del dólar e interprete su respuesta.

Cambio de base

Si dos o más series tienen el mismo periodo base, se pueden comparar de manera directa; por ejemplo, suponga que tiene interés en la tendencia de los precios de alimentos y bebidas, vivienda, servicios médicos, etcétera, desde el periodo base, 1982-1984. En la tabla 17.8 se observa que en todos los índices de precios al consumidor se utiliza la misma base.

TABLA 17.8 Tendencia de los precios al consumidor hasta 2012 (1982-1984 = 100)

Año	Todos los artículos	Alimentos y bebidas	Vivienda	Ropa y manutención	Servicios médicos
1982–84	100.0	100.0	100.0	100.0	100.0
1990	130.7	132.1	128.5	124.1	162.8
1995	152.4	148.9	148.5	132	220.5
2000	172.2	168.4	169.6	129.6	260.8
2005	195.3	191.2	195.7	119.5	323.2
2006	201.6	195.7	203.2	119.5	336.2
2010	218.056	219.984	216.256	119.503	388.436
2011	224.939	227.866	219.102	122.111	400.258
2012	229.594	233.67	222.715	126.265	414.924

En la tabla 17.8 se muestra que el precio combinado de todos los artículos de consumo aumentó 129.594% desde el periodo base (1982-84) hasta el año 2012 (a partir de enero de 2007, el INPC reporta con tres cifras decimales en vez de una). Del mismo modo, los precios promedio de los alimentos y bebidas aumentaron 133.67%; los precios de vivienda, 122.715%; ropa y mantenimiento, 26.265% y la atención médica, 314.924%.

Sin embargo, surge un problema cuando dos o más series que se comparan no tienen el mismo periodo base. En el siguiente ejemplo se comparan los dos índices de negocios que se reportan con más frecuencia, el DJIA y el Nasdaq.

EJEMPLO

Compare los cambios de precios entre el DJIA y el Nasdaq el primer día laboral del año 2004 a 2013.

Fecha	Precio de apertura	
	DJIA	NASDAQ
2-ene-04	$10 452.74	$2 011.08
3-ene-05	10 783.75	2 184.75
3-ene-06	10 718.30	2 216.53
3-ene-07	12 459.54	2 429.72
2-ene-08	13 261.82	2 653.91
2-ene-09	8 772.25	1 578.87
4-ene-10	10 430.69	2 294.41
3-ene-11	11 577.43	2 676.65
3-ene-12	12 221.19	2 657.39
2-ene-13	13 104.30	3 091.33

SOLUCIÓN

No es apropiado hacer una comparación directa de los precios de apertura del DJIA con los del NASDAQ; dado que usted quiere comparar los cambios en los precios de apertura de ambos mercados, un enfoque lógico es calcular los índices para cada mercado utilizando como base el precio de apertura de 2004. Para el DJIA, la base es 10 452.74 dólares, y para el NASDAQ, 2 011.08 dólares.

El cálculo del índice del DJIA en 2013 es:

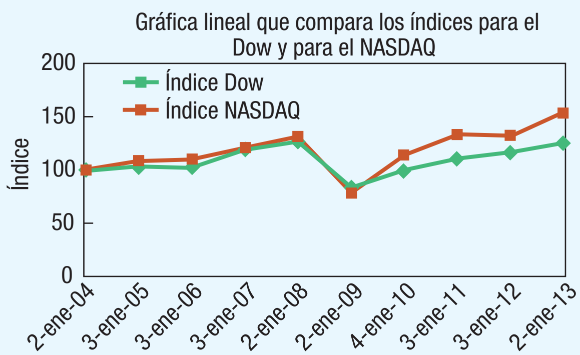
$$\text{Índice} = \frac{\$13\,104.30}{\$10\,452.74}(100) = 125.4$$

En la siguiente tabla se reporta el grupo completo de índices.

Fecha	DJIA		NASDAQ	
	Valor	Índice	Valor	Índice
2-ene-04	$10 452.74	100.0	$2 011.08	100.0
3-ene-05	10 783.75	103.2	2 184.75	108.6
3-ene-06	10 718.30	102.5	2 216.53	110.2
3-ene-07	12 459.54	119.2	2 429.72	120.8
2-ene-08	13 261.82	126.9	2 653.91	132.0
2-ene-09	8 772.25	83.9	1 578.87	78.5
4-ene-10	10 430.69	99.8	2 294.41	114.1
3-ene-11	11 577.43	110.8	2 676.65	133.1
3-ene-12	12 221.19	116.9	2 657.39	132.1
2-ene-13	13 104.30	125.4	3 091.33	153.7

Se concluye que los dos índices aumentaron durante este periodo (DJIA 25.4%; Nasdaq 53.7%).

En la siguiente gráfica se muestran los índices DJIA en verde y los índices NASDAQ en café; además, se indican los cambios en ambos índices con base en el 2 de enero de 2004. De ella se concluye que el NASDAQ alcanzó su pico más alto a comienzos de 2013, y el pico más alto del Dow fue en enero de 2008; en general, ambos índices parecen reflejarse mutuamente de manera bastante estrecha. Es preciso señalar que si se seleccionan periodos distintos como base, los resultados quizá no sean exactamente iguales. Siempre se debe tener cuidado al seleccionar el periodo base para una tabla o gráfica.



AUTOEVALUACIÓN 17-7

Usted desea comparar los cambios en la producción industrial y en los precios que pagaron los fabricantes por materias primas desde 1982; por desgracia, el índice de la producción industrial (mide los cambios en la producción) y el índice de precios del productor (mide los cambios de precios de las materias primas) tienen periodos base distintos. El índice de producción tiene como periodo base 2007, y el Índice de Precios al Productor, 1982. En la tabla de la derecha se reporta el valor de cada índice al comienzo del año; cambie la base a 2010, haga comparables ambas series e interprete los resultados.

Año	Índice de producción industrial (2007 = 100)	Índice de precios al productor (1982 = 100)
2007	98.16	160.1
2008	100.50	172.0
2009	87.55	170.4
2010	87.96	178.0
2011	92.61	184.4
2012	96.20	192.0
2013	98.06	194.7

EJERCICIOS

13. En abril de 2013, el salario medio de una supervisora de enfermeras con licenciatura fue 89 673 dólares, y el IPC en esa fecha fue 213.240 (1982-84 = 100); por otra parte, el salario medio anual de una enfermera en el periodo base de 1982-1984 fue 19 800 dólares. ¿Cuál fue el ingreso real de la enfermera en abril de 2013? ¿Cuánto aumentó el salario medio?

14. La Trade Union Association of Orlando, Florida, mantiene índices sobre los salarios por hora de diversos oficios; por desgracia, no todos los índices tienen el mismo periodo base. A continuación se presenta la información sobre plomeros y electricistas; cambie los periodos base a 2000 y compare los aumentos de los salarios por hora de 2000 a 2013.

Año	Plomeros (1995 = 100)	Electricistas (1998 = 100)
2000	133.8	126.0
2013	159.4	158.7

15. En 2000, el salario medio de los maestros del Distrito Escolar Tinora fue de 28 650 dólares; en 2009 aumentó a 33 972 dólares, y en 2013 aún más, a 37 382 dólares. La American Federation of Classroom Teachers recolecta información sobre las tendencias de los salarios de maestros en Estados Unidos; su índice, cuyo periodo base es 2000, fue 122.5 en 2009 y 136.9 en 2013. Compare los salarios de los maestros del Distrito Escolar Tinora con las tendencias nacionales.

16. Sam Steward es un diseñador de páginas web que trabaja de manera independiente. En la siguiente tabla se pueden observar sus salarios anuales durante varios años entre 2007 y 2013; en esta se incluye un índice industrial de diseñadores de páginas web que reporta la tasa de inflación de los salarios en la industria; 1995 es el periodo base de dicho índice.

Año	Salario (en miles de dólares)	Índice (1995 = 100)
2007	134.8	160.6
2009	145.2	173.6
2011	156.6	187.9
2013	168.8	203.3

Calcule el ingreso real de Sam en los años seleccionados durante el periodo de seis años. ¿Van a la par sus salarios con la inflación o ha perdido ingresos?

RESUMEN DEL CAPÍTULO

I. Un número índice mide un cambio relativo de un periodo a otro.
- **A.** Las características importantes de un índice son:
  - **1.** Es un porcentaje, pero en general se omite el signo correspondiente.
  - **2.** Se refiere a un periodo base.
- **B.** Las razones para calcular un índice son:
  - **1.** Facilita la comparación de series desiguales.
  - **2.** Si los números son muy grandes, con frecuencia es más fácil comprender el cambio del índice que las cifras reales.

ESTADÍSTICA EN ACCIÓN

En la década de 1920, en Alemania, los precios al mayoreo aumentaron de forma drástica; en 1920 crecieron casi 80%, en 1921 la tasa aumentó a 140%, y en 1922 fue un sorprendente 4 100%. Entre diciembre de 1922 y noviembre de 1923 los precios al mayoreo aumentaron otro 4 100%. En esa época, las prensas de impresión de papel moneda del Gobierno no podían mantener ese ritmo, ni siquiera con billetes con denominaciones tan grandes como 500 millones de marcos. Se cuenta que a los trabajadores se les pagaba todos los días, y luego dos veces al día, para que sus esposas pudieran hacer sus compras antes de que sus salarios se devaluaran demasiado.

II. Hay dos tipos de índices de precios: ponderados y no ponderados.

A. En un índice no ponderado, no se consideran las cantidades.

1. En un índice simple se compara el periodo base con el periodo dado.

$$P = \frac{P_t}{P_0} \times 100 \qquad \textbf{[17.1]}$$

donde p_t se refiere al precio del periodo actual, y p_0 es el precio del periodo base.

2. En el promedio simple de los índices de los precios, se suman los índices simples de cada artículo y el resultado se divide entre el número de artículos.

$$P = \frac{\Sigma p_i}{n} \qquad \textbf{[17.2]}$$

3. En un índice de precios agregado simple, se suma el precio de los artículos del grupo de los dos periodos y estos se comparan.

$$P = \frac{\Sigma p_t}{\Sigma p_0} \times 100 \qquad \textbf{[17.3]}$$

B. En un índice ponderado se consideran las cantidades.

1. Cuando se emplea el método de Laspeyres se utilizan las cantidades del periodo base tanto en el periodo base como en el dado.

$$P = \frac{\Sigma p_t q_0}{\Sigma p_0 q_0} \times 100 \qquad \textbf{[17.4]}$$

2. En el método de Paasche se utilizan las cantidades del periodo actual.

$$P = \frac{\Sigma p_t q_t}{\Sigma p_0 q_t} \times 100 \qquad \textbf{[17.5]}$$

3. El índice de precios ideal de Fisher es la media geométrica del índice de Laspeyres y del índice de Paasche.

$$\text{Índice ideal de Fisher} = \sqrt{(\text{Índice de Laspeyres})(\text{Índice de Paasche})} \qquad \textbf{[17.6]}$$

C. En el índice de valores se contemplan los precios y las cantidades del periodo base y del periodo actual.

$$V = \frac{\Sigma p_t q_t}{\Sigma p_0 q_0} \times 100 \qquad \textbf{[17.7]}$$

III. El Índice de Precios al Consumidor (IPC) es el que se reporta con más frecuencia.

A. Se utiliza con frecuencia para mostrar la tasa de inflación en Estados Unidos.

B. Lo elabora mensualmente el U.S. Bureau of Labor Statistics.

C. Su periodo base actual es 1982-1984.

D. El sistema de seguridad social lo utiliza, por lo que, cuando el IPC cambia, también lo hace el monto de las pensiones.

EJERCICIOS DEL CAPÍTULO

Para los ejercicios 17 a 22 utilice la siguiente información, que se obtuvo de los reportes anuales de Johnson & Johnson; cuya oficina matriz se encuentra en New Brunswick, Nueva Jersey. Sus acciones comunes se cotizan en la Bolsa de Valores de Nueva York, con el símbolo JNJ.

Año	Ventas nacionales (en millones de dólares)	Ventas internacionales (en millones de dólares)	Empleados (en miles)
2000	17 316	11 856	100.9
2001	19 825	12 492	101.8
2002	22 455	13 843	108.3
2003	25 274	16 588	110.6
2004	27 770	19 578	109.9
2005	28 377	22 137	115.6

(*continúa*)

(*continuación*)

Año	Ventas nacionales (en millones de dólares)	Ventas internacionales (en millones de dólares)	Empleados (en miles)
2006	29 775	23 549	122.2
2007	32 444	28 651	119.2
2008	32 309	31 438	118.7
2009	30 889	31 008	115.5
2010	29 437	32 124	114.0
2011	28 907	36 107	117.9
2012	29 832	37 371	118.7

17. Utilice 2000 como periodo base; calcule un índice simple de las ventas nacionales de cada año desde 2000 hasta 2012 e interprete la tendencia de las ventas nacionales.
18. Utilice el periodo 2000-2002 como periodo base y calcule un índice simple de las ventas nacionales de cada año de 2003 a 2012.
19. Utilice 2000 como periodo base; calcule un índice simple de las ventas internacionales de cada año de 2001 a 2012 e interprete la tendencia de las ventas internacionales.
20. Utilice el periodo 2000-2002 como periodo base y calcule un índice simple de las ventas internacionales de cada año de 2003 a 2012.
21. Utilice 2000 como periodo base; calcule un índice simple del número de empleados de cada año de 2001 a 2012 e interprete la tendencia del número de empleados.
22. Utilice el periodo 2000-2002 como periodo base y calcule un índice simple del número de empleados de cada año de 2003 a 2012.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Para los ejercicios 23 a 26 utilice la siguiente información, que se obtuvo de los reportes anuales de General Electric Corporation (GE).

Año	Ingreso (en millones de dólares)	Empleados (en miles)	Año	Ingreso (en millones de dólares)	Empleados (en miles)
2004	134	325	2009	183	323
2005	152	307	2010	149.6	304
2006	157	316	2011	147.3	287
2007	168	319	2012	147.3	301
2008	177	327			

23. Calcule un índice simple del ingreso de GE; utilice 2004 como periodo base. ¿Qué puede concluir acerca del cambio en el ingreso durante el periodo dado?
24. Calcule un índice simple del ingreso de GE con el periodo 2004-2006 como base. ¿Qué puede concluir acerca del cambio en el ingreso durante el periodo dado?
25. Calcule un índice simple del número de empleados de GE; utilice 2004 como periodo base. ¿Qué puede concluir acerca del cambio en el número de empleados de la empresa durante este periodo?
26. Calcule un índice simple del número de empleados de GE con el periodo 2004-2006 como base. ¿Qué puede concluir acerca del cambio en el número de empleados durante este periodo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Para los ejercicios 27 a 32 utilice la siguiente información sobre artículos de alimentos en 2000 y 2013.

	2000		2013	
Artículo	**Precio**	**Cantidad**	**Precio**	**Cantidad**
Margarina (libra)	$0.81	18	$2.00	27
Manteca (libra)	0.84	5	1.88	9
Leche (½ galón)	1.44	70	2.89	65
Papas (libra)	2.91	27	3.99	33

27. Calcule un índice de precios simple de cada uno de los cuatro artículos; considere el año 2000 como periodo base.
28. Calcule un índice de precios agregado simple; utilice 2000 como periodo base.
29. Calcule el índice de precios de Laspeyres de 2013 con el año 2000 como periodo base.
30. Calcule el índice de Paasche de 2013 con el año 2000 como periodo base.
31. Determine el índice ideal de Fisher con los valores de los índices de Laspeyres y Paasche que se calcularon en los dos problemas anteriores.
32. Determine el índice de valores de 2013 con el año 2000 como periodo base.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Para los ejercicios 33 a 38 utilice la siguiente información: Betts Electronics compra tres partes de repuesto para máquinas robóticas que utiliza en su proceso de manufactura; he aquí la información de los precios de las partes de repuesto y la cantidad que compró.

	Precio		Cantidad	
Parte	**2000**	**2013**	**2000**	**2013**
RC-33	$0.50	$0.60	320	340
SM-14	1.20	0.90	110	130
WC50	0.85	1.00	230	250

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

33. Calcule un índice de precios simple de cada uno de los tres artículos; utilice 2000 como periodo base.
34. Calcule un índice de precios agregado simple de 2013; utilice 2000 como periodo base.
35. Calcule el índice de precios de Laspeyres de 2013 con el año 2000 como periodo base.
36. Calcule el índice de Paasche de 2013 con el año 2000 como periodo base.
37. Determine el índice ideal de Fisher con los valores de los índices de Laspeyres y Paasche que se calcularon en los dos problemas anteriores.
38. Determine un índice de valores de 2013 con el año 2000 como periodo base.

Para los ejercicios 39 a 44 utilice la información de precios de alimentos seleccionados en 2000 y 2013 que se muestra en la siguiente tabla.

	Precio		Cantidad	
Artículo	**2000**	**2013**	**2000**	**2013**
Col (libra)	$0.06	$0.05	2 000	1 500
Zanahorias (racimo)	0.10	0.12	200	200
Chícharos (cuarto)	0.20	0.18	400	500
Endivia (racimo)	0.15	0.15	100	200

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

39. Calcule un índice de precios simple para cada uno de los artículos; utilice 2000 como periodo base.
40. Calcule un índice de precios agregado simple; utilice 2000 como periodo base.
41. Calcule el índice de precios de Laspeyres de 2013 con 2000 como periodo base.
42. Calcule el índice de Paasche de 2013 con 2000 como periodo base.
43. Determine el índice ideal de Fisher con los valores de los índices de Laspeyres y Paasche que se calcularon en los dos ejemplos anteriores.
44. Determine un índice de valores de 2013; utilice 2000 como periodo base.

Para los ejercicios 45 a 50 utilice la siguiente información de precios de artículos seleccionados en 1990 y 2013; también se proporcionan las cifras de la producción de ambos periodos.

	Precio		Cantidad	
Artículo	**1990**	**2013**	**1990**	**2013**
Aluminio (centavos por libra)	$ 0.287	$ 0.76	1 000	1 200
Gas natural (1 000 pies cúbicos)	0.17	2.50	5 000	4 000
Petróleo (barril)	3.18	26.00	60 000	60 000
Platino (onza troy)	133.00	490.00	500	600

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

45. Calcule un índice de precios simple de cada uno de los cuatro artículos; utilice 1990 como periodo base.
46. Calcule un índice de precios agregado simple; utilice 1990 como periodo base.
47. Calcule el índice de precios de Laspeyres de 2013; utilice 1990 como periodo base.
48. Calcule el índice de precios de Paasche de 2013; utilice 1990 como periodo base.
49. Determine el índice ideal de Fisher con los valores de los índices de Laspeyres y Paasche que se calcularon en los dos problemas anteriores.
50. Determine un índice de valores de 2013; utilice 1990 como periodo base.
51. Se diseñará un índice para propósitos especiales para vigilar la economía global del suroeste de Estados Unidos. Se seleccionaron cuatro series claves; después de una deliberación considerable

se decidió ponderar las ventas al menudeo 20%, los depósitos bancarios totales 10%, la producción industrial en el área 40%, y el empleo en el área no agrícola 30%. He aquí los datos de 1996 y 2013:

Año	Ventas al menudeo (en millones de dólares)	Depósitos bancarios (en miles de millones de dólares)	Producción industrial (1990 = 100)	Empleo
1996	1 159.0	87	110.6	1 214 000
2013	1 971.0	91	114.7	1 501 000

Elabore un índice para propósitos especiales de 2013; utilice 1996 como periodo base e interprete su resultado.

52. Se realizó un estudio histórico de la economía estadounidense de 1950 a 1980, para lo cual se recopilaron datos sobre precios, fuerza de trabajo, productividad y PIB. En la siguiente tabla se observa que el IPC tiene 1967 como periodo base, el empleo está en millones de personas, etcétera; por lo tanto, no es posible una comparación directa.

a. Realice los cálculos necesarios para comparar la tendencia en las cuatro series de 1950 a 1980.

b. Interprete sus resultados.

Año	Índice de Precios al Consumidor (1967 = 100)	Fuerza laboral total (millones)	Índice de productividad en la manufactura (1967 = 100)	Producto Interno Bruto (miles de millones de dólares)
1950	72.1	64	64.9	286.2
1967	100.0	81	100.0	789.6
1971	121.3	87	110.3	1 063.4
1975	161.2	95	114.9	1 516.3
1980	246.8	107	146.6	2 626.0

53. La gerencia de las tiendas Ingalls Super Discount, con varios locales en el área de Oklahoma City, desea elaborar un índice de la actividad económica del área metropolitana. La gerencia está de acuerdo en que, si este revela una economía en receso, el inventario se deberá mantener en un nivel bajo.

Tres series parecen prometedoras como factores de predicción de la actividad económica: las ventas al menudeo en el área, los depósitos bancarios y el empleo; toda esta información se puede obtener del Gobierno de Estados Unidos. Las ventas al menudeo tendrán una ponderación de 40%, los depósitos bancarios, 35%, y el empleo, 25%. Los datos ajustados por temporada del primer trimestre del año son:

Mes	Ventas al menudeo (millones)	Depósitos bancarios (miles de millones)	Empleo (miles)
Enero	8.0	20	300
Febrero	6.8	23	303
Marzo	6.4	21	297

Elabore un índice de la actividad económica de cada uno de los tres meses, con enero como periodo base.

54. En la siguiente tabla se da la información sobre el IPC y el ingreso neto mensual de Bill Martin, empleado de Ford Motor Corporation, de 1982-84 a 2012.

Año	Índice de Precios al Consumidor (1982-84 = 100)	Ingreso neto mensual de Martin
1982-84	100.0	$ 600
2012	229.594	2 000

a. ¿Cuál es el poder de compra del dólar en 2012 con base en el periodo 1982-1984?

b. Determine el ingreso mensual "real" de Martin en 2012.

55. Suponga que el Índice de Precios al Productor y las ventas de Hoskin's Wholesale Distributors de 1995 y 2012 aparecen a la derecha.

¿Cuáles son las ventas reales (o ventas deflacionadas) de Hoskin's Wholesale Distributors en ambos años?

Año	Índice de Precios al Productor	Ventas
1995	127.9	$2 400 000
2012	194.5	3 500 000

EJERCICIOS DE LA BASE DE DATOS

(Los datos para este ejercicio están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**)

56. Consulte los datos sobre Baseball 2012 que contienen información de la temporada 2012 de las Ligas Mayores de Béisbol y, además, incluye el salario medio por jugador desde 1989.

a. Utilice el salario de 2000 como el periodo base y 100 como el valor base para desarrollar un índice simple para los años desde 2000. ¿En qué porcentaje se ha elevado el salario típico?

b. Los valores del IPC para 2001, 2003, 2011 y 2012 son 177.1, 184.0, 224.94 y 229.594, respectivamente. ¿Cuál es el promedio del salario real (o deflacionado) para esos años? Describa la tendencia en salarios deflacionados en unas cuantas oraciones y compare estos resultados con su respuesta para el punto a.

Series de tiempo y proyección

18

TEAM SPORTS, INC., vende artículos deportivos a preparatorias y universidades por medio de un catálogo de distribución nacional, y la gerencia de la empresa estima que venderá 2 000 manoplas marca Wilson, modelo A2000, el próximo año. Las ventas desestacionalizadas proyectadas serán iguales en cada trimestre del año próximo; el factor estacional del segundo trimestre es 145. Determine las ventas ajustadas por temporada en el segundo trimestre del próximo año (vea el ejercicio 12 y el OA18-6).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA18-1 Definir y describir los componentes de una serie de tiempo.

OA18-2 Suavizar una serie de tiempo calculando un promedio móvil.

OA18-3 Suavizar una serie de tiempo calculando un promedio móvil ponderado.

OA18-4 Utilizar un análisis de regresión para ajustar una línea de tendencia lineal a una serie de tiempo.

OA18-5 Utilizar un análisis de regresión para ajustar una serie de tiempo no lineal.

OA18-6 Calcular y aplicar índices estacionales para realizar pronósticos ajustados a la estación.

OA18-7 Desestacionalizar una serie de tiempo utilizando índices estacionales.

OA18-8 Realizar una prueba de hipótesis de autocorrelación.

Introducción

En este capítulo se enfatizan el análisis y la proyección de las series de tiempo. Una **serie de tiempo** es un grupo de datos registrados durante un periodo semanal, mensual, trimestral o anual. Dos ejemplos de las series de tiempo son las ventas de Microsoft Corporation por trimestre desde 1985, y la producción anual de ácido sulfúrico desde 1970.

Un análisis de la historia, que es una serie de tiempo, es útil para que la administración tome decisiones en el presente y planee con base en una predicción, o proyección, de largo plazo; en general, se supone que los patrones pasados continuarán en el futuro. Las proyecciones de largo plazo se amplían a más de un año; son comunes las proyecciones de dos, cinco y diez años. Las proyecciones de largo plazo son esenciales a fin de dar tiempo suficiente para que los departamentos de compras, manufactura, ventas, finanzas y otros de una compañía elaboren planes para construir nuevas plantas, solicitar financiamiento, desarrollar productos nuevos y métodos de ensamble innovadores.

En Estados Unidos, la proyección del nivel de ventas, tanto a corto como a largo plazo, se rige casi en su totalidad por la propia naturaleza de las organizaciones de negocios. La competencia por el dinero de los consumidores, la presión para obtener utilidades para los accionistas, el deseo de obtener una mayor participación de mercado y las ambiciones de los ejecutivos son algunas fuerzas de motivación en los negocios; por lo tanto, se necesita una proyección (una declaración de los objetivos de la administración) para tener las materias primas, las instalaciones de producción y el personal para cumplir con la demanda.

En este capítulo se aborda el uso de los datos para proyectar eventos futuros: primero se analizan los componentes de una serie de tiempo; luego, algunas técnicas para analizar los valores y, por último, se proyectan eventos futuros.

OA18-1
Definir y describir los componentes de una serie de tiempo.

Componentes de una serie de tiempo

Una serie de tiempo consta de cuatro componentes: 1) tendencia, 2) variación cíclica, 3) variación estacional y 4) variación irregular.

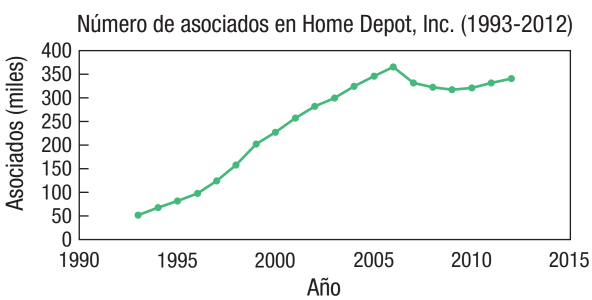
Tendencia secular

Las tendencias a largo plazo de las ventas, el empleo, los precios accionarios y de otras series de negocios y económicas siguen varios patrones; algunas se mueven hacia arriba de manera uniforme, otras declinan y otras más permanecen iguales con el paso del tiempo. El cambio a través del tiempo puede ser lineal y seguir una línea recta, o puede incrementarse a un ritmo exponencial; el cambio a largo plazo (o la ausencia de cambio) se denomina tendencia de la serie de tiempo o, de manera más precisa, la **tendencia secular**.

Años	Asociados	Años	Asociados
1993	50.6	2003	298.8
1994	67.3	2004	323.1
1995	80.8	2005	344.8
1996	98.1	2006	364.4
1997	124.4	2007	331.0
1998	156.7	2008	322.0
1999	201.4	2009	317.0
2000	227.3	2010	321.0
2001	256.3	2011	331.0
2002	280.9	2012	340.0

TENDENCIA SECULAR Dirección uniforme de una serie de tiempo de largo plazo.

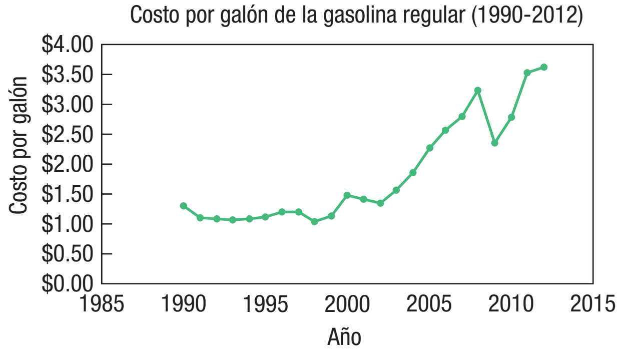
He aquí diversos ejemplos de una tendencia secular.



- Home Depot se fundó en 1978 y, en la actualidad, es el minorista más grande de Estados Unidos en artículos para mejorar el hogar. En la gráfica de la izquierda, se muestra el número de asociados de esta empresa; puede observar que este número aumentó con rapidez en los últimos 15 años, de 1993 a 2006. En 1993, había poco más de 50 000 asociados, mientras que en 2005 el número aumentó a más de 340 000; desde entonces, el número de asociados disminuyó drásticamente en 2007, después decreció ligeramente en 2008 y 2009, y permaneció bastante constante durante los últimos tres años.

- El precio promedio del galón de gasolina regular fue de casi 1.15 dólares durante la mayor parte de la década de 1990. Desde entonces, y salvo por un lapso entre 2009 y 2010, ha aumentado de manera uniforme: casi 0.20 dólar por año; esta información se muestra en la siguiente gráfica.

Año	Costo por galón	Año	Costo por galón
1990	$1.30	2002	$1.35
1991	1.10	2003	1.56
1992	1.09	2004	1.85
1993	1.07	2005	2.27
1994	1.08	2006	2.57
1995	1.11	2007	2.80
1996	1.20	2008	3.25
1997	1.20	2009	2.35
1998	1.03	2010	2.78
1999	1.14	2011	3.52
2000	1.48	2012	3.62
2001	1.42		



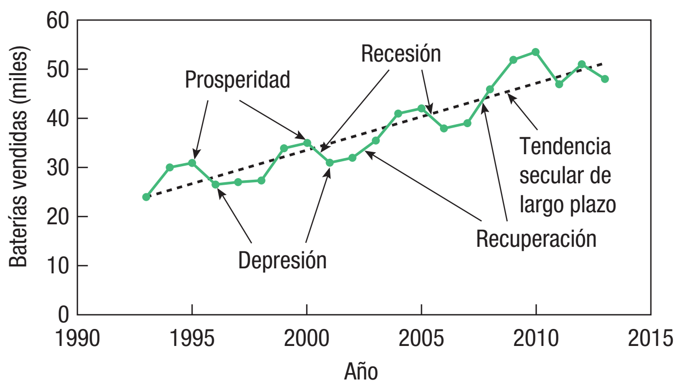
Variación cíclica

El segundo componente de una serie de tiempo es la **variación cíclica**. Un ciclo de negocios habitual está conformado por un periodo de prosperidad, seguido por periodos de recesión, depresión y luego recuperación. Hay fluctuaciones considerables que se desarrollan durante más de un año, arriba y debajo de la tendencia secular; por ejemplo, en una recesión, el empleo, la producción, el Promedio Industrial Dow Jones y muchas otras series de negocios y económicas se encuentran debajo de las líneas de las tendencias de largo plazo, por el contrario, en periodos de prosperidad se encuentran arriba de ellas.

VARIACIÓN CÍCLICA Aumento y reducción de una serie de tiempo durante periodos mayores de un año.

En la gráfica 18.1 se presentan las unidades anuales de baterías que vendió National Battery Retailers, Inc., desde 1993 hasta 2013; se resalta el ciclo natural del negocio. Los periodos, en orden cronológico, son: recuperación, prosperidad, recesión y, por último, el ciclo desciende con depresión.

Año	Ventas de baterías (miles)	Año	Ventas de baterías (miles)
1993	24.0	2004	41.0
1994	30.0	2005	42.0
1995	31.0	2006	38.0
1996	26.5	2007	39.0
1997	27.0	2008	46.0
1998	27.5	2009	52.0
1999	34.0	2010	53.5
2000	35.0	2011	47.0
2001	31.0	2012	51.0
2002	32.0	2013	48.0
2003	35.5		



GRÁFICA 18.1 Baterías vendidas por National Battery Retailers, Inc., de 1993 a 2013

Variación estacional

El tercer componente de una serie de tiempo es la **variación estacional**. Muchas series de ventas, de producción y de otros tipos fluctúan de acuerdo con las temporadas; la unidad de tiempo generalmente se reporta por trimestre o por mes pero también puede hacerse por semana.

ESTADÍSTICA EN ACCIÓN

Los profesionales en estadística, economistas y ejecutivos de negocios constantemente tratan de encontrar variables que les permitan proyectar la economía del país. La producción de petróleo crudo, el precio del oro en los mercados mundiales y el Promedio Industrial Dow Jones, así como muchos índices que publica el gobierno, son variables que han tenido cierto éxito. También se han probado variables como la longitud de los trajes y el ganador del Súper Tazón; pero la variable que, en general, parece más exitosa es el precio del metal de desecho. ¿Por qué? El metal de desecho es el inicio de la cadena de manufactura; cuando aumenta su demanda, se prevé que la manufactura también lo hará.

VARIACIÓN ESTACIONAL Patrones de cambio en una serie de tiempo en un año; estos tienden a repetirse cada año.

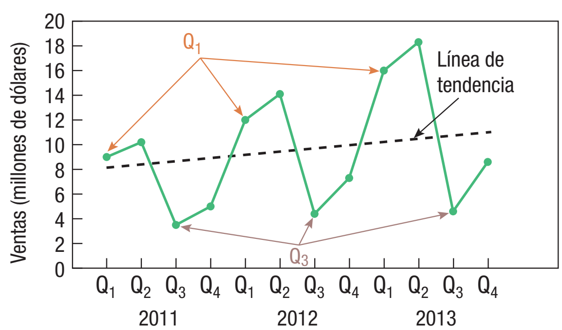
Casi todos los negocios suelen tener patrones estacionales recurrentes; por ejemplo, la ropa para caballeros y niños tiene ventas muy elevadas justo antes de Navidad, y relativamente bajas después de esa celebración y durante el verano. La renta de propiedades en la playa es otro ejemplo: hay muy pocas rentas en invierno, pero durante el verano, cuando el tiempo es cálido, aumentan significativamente. Las ventas de juguetes son otro ejemplo con un patrón estacional extremo: más de la mitad de los negocios del año se realizan, en general, en noviembre y diciembre. El negocio de jardinería es estacional en los estados del noreste y del centro-norte de Estados Unidos; por lo tanto, muchos negocios tratan de equilibrar los efectos estacionales y se dedican a otras actividades de compensación estacional. En el noreste de Estados Unidos es posible ver al encargado de un negocio de jardinería con un quitanieve al frente del camión, en un intento por obtener algún ingreso durante la temporada de invierno. Con frecuencia, en las cercanías de los centros de esquí de todo el país hay campos de golf; cuyos propietarios tratan de rentarlos a esquiadores en el invierno y a golfistas en el verano; este es un método eficaz para repartir los gastos fijos en todo el año, en lugar de distribuirlos solo en algunos meses.

En la gráfica 18.2 se registran las ventas trimestrales, en millones de dólares, de Hercher Sporting Goods, Inc. Dicha compañía de artículos deportivos del área de Chicago se especializa en la venta de equipo de béisbol y *softball* a preparatorias, universidades y ligas juveniles. También tiene varias tiendas de descuento en algunos de los centros comerciales más grandes, y su negocio tiene un patrón estacional distintivo. La mayoría de sus ventas son en el primero y segundo trimestres del año, cuando las escuelas y organizaciones compran equipo para la próxima temporada; durante principios del verano se mantiene ocupada con la venta de equipo de reemplazo. Esto hacen algunos negocios durante la temporada navideña (cuarto trimestre), mientras que las últimas semanas del verano (tercer trimestre) conforman su temporada baja.

Variación irregular

Muchos analistas prefieren subdividir la **variación irregular** en variaciones episódicas y residuales. Las fluctuaciones episódicas son impredecibles, pero es posible identificarlas; por otra parte, el efecto inicial de una huelga importante o de una guerra en la economía no se puede predecir. Después de eliminar las fluctuaciones episódicas, la variación restante se denomina variación residual. Las fluctuaciones residuales, con frecuencia denominadas fluctuaciones azarosas, son impredecibles y no se pueden identificar; por supuesto, no es posible proyectar a futuro ni la variación episódica ni la residual.

Año	Trimestre	Código	Ventas (millones de dólares)
2011	Q_1	*1*	9.0
	Q_2	*2*	10.2
	Q_3	*3*	3.5
	Q_4	*4*	5.0
2012	Q_1	*5*	12.0
	Q_2	*6*	14.1
	Q_3	*7*	4.4
	Q_4	*8*	7.3
2013	Q_1	*9*	16.0
	Q_2	*10*	18.3
	Q_3	*11*	4.6
	Q_4	*12*	8.6



GRÁFICA 18.2 Ventas de equipo de béisbol y softball en Hercher Sporting Goods, 2011-2013, por trimestre

OA18-2
Suavizar una serie de tiempo calculando un promedio móvil.

Promedio móvil

Un **promedio móvil** es útil para suavizar una serie de tiempo y apreciar su tendencia; además, es el método básico para medir la fluctuación estacional que se describe más adelante en el capítulo. En contraste con el método de mínimos cuadrados, que expresa la tendencia en términos de una ecuación matemática ($\hat{y} = a + bt$), el método del promedio móvil solo suaviza las fluctuaciones de los valores; este objetivo se logra al "desplazar" los promedios aritméticos en la serie de tiempo.

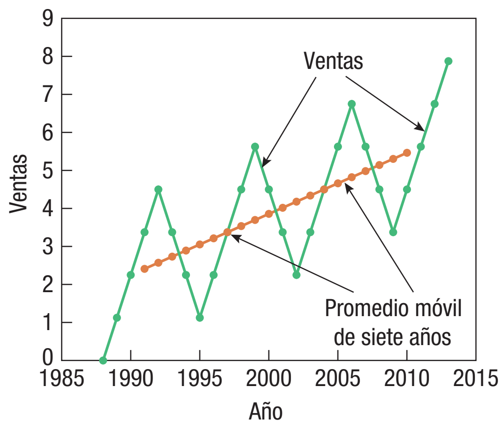
Para aplicar el promedio móvil a una serie de tiempo, los datos deben seguir una tendencia muy lineal y tener un patrón rítmico definido de las fluctuaciones (que se repita, por ejemplo, cada tres años); en los datos que se encuentran en el siguiente ejemplo hay tres componentes: tendencia, ciclo e irregularidad (abreviadas *T*, *C* e *I*). No hay variación estacional debido a que los datos se registran cada año. Lo que logra el promedio móvil es equilibrar *C* e *I*; lo que queda es la tendencia.

Si la duración de los ciclos es constante y sus amplitudes son iguales, las fluctuaciones cíclica e irregular se eliminan por completo con el promedio móvil, y el resultado es una recta; por ejemplo, en la siguiente serie de tiempo, el ciclo se repite cada siete años y la amplitud de cada ciclo es 4; es decir, hay exactamente cuatro unidades desde el valle (el periodo más bajo) hasta el pico, por lo tanto, el promedio móvil de siete años promedia a la perfección las fluctuaciones cíclicas e irregulares, y el residuo es una tendencia lineal.

El primer paso para calcular el promedio móvil de siete años es determinar los totales móviles de siete años. Las ventas totales de los primeros siete años (1984-1990, inclusive) son 22 millones de dólares, determinadas por 1 + 2 + 3 + 4 + 5 + 4 + 3 (consulte la tabla 18.1); ese total se divide entre 7 para determinar la media aritmética de las ventas anuales. El total de la suma de siete años (22) y la media de siete años (3.143) se colocan opuestos al año medio de ese grupo de siete, es decir, 1991, como se indica en la tabla 18.1. Luego se determinan las ventas totales de los siguientes siete años (1985-1995, inclusive; una forma conveniente para hacer este cálculo es restar las ventas de 1988 [$1 millón] al primer total de siete años [$22 millones] y sumar las ventas de 1995 [$2 millones], para obtener el nuevo total de 23 millones de dólares). La media de este total (3.286 millones de dólares), se coloca opuesta al año medio, 1992; los datos de las ventas y el promedio móvil de siete años se muestran en la gráfica 18.3.

TABLA 18.1 Cálculos para determinar el promedio móvil de siete años

Año	Ventas (en millones de dólares)	Total móvil de siete años	Promedio móvil de siete años
1988	1		
1989	2		
1990	3		
1991	4	22	3.143
1992	5	23	3.286
1993	4	24	3.429
1994	3	25	3.571
1995	2	26	3.714
1996	3	27	3.857
1997	4	28	4.000
1998	5	29	4.143
1999	6	30	4.286
2000	5	31	4.429
2001	4	32	4.571
2002	3	33	4.714
2003	4	34	4.857
2004	5	35	5.000
2005	6	36	5.143
2006	7	37	5.286
2007	6	38	5.429
2008	5	39	5.571
2009	4	40	5.714
2010	5	41	5.857
2011	6		
2012	7		
2013	8		



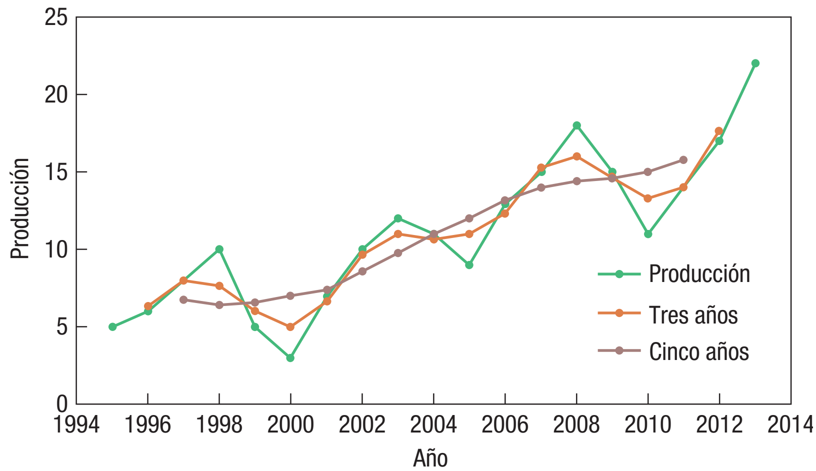
GRÁFICA 18.3 Ventas y promedio móvil de siete años

El número de valores que se incluirán en un promedio móvil depende del carácter de los que se recopilen. Si son trimestrales, puesto que hay cuatro trimestres en un año, sería adecuado tener

TABLA 18.2 Producción, promedio móvil de tres y cinco años

Año	Producción, *y*	Total móvil de tres años	Promedio móvil de tres años	Total móvil de cinco años	Promedio móvil de cinco años
1995	5				
1996	6	19	6.33		
1997	8	24	8.00	34	6.80
1998	10	23	7.67	32	6.40
1999	5	18	6.00	33	6.60
2000	3	15	5.00	35	7.00
2001	7	20	6.67	37	7.40
2002	10	29	9.67	43	8.60
2003	12	33	11.00	49	9.80
2004	11	32	10.67	55	11.00
2005	9	33	11.00	60	12.00
2006	13	37	12.33	66	13.20
2007	15	46	15.33	70	14.00
2008	18	48	16.00	72	14.40
2009	15	44	14.67	73	14.60
2010	11	40	13.33	75	15.00
2011	14	42	14.00	79	15.80
2012	17	53	17.67		
2013	22				

cuatro términos; si son diarios, como hay siete días en una semana, sería apropiado tener siete términos. También se puede emplear el método de ensayo y error para determinar un número que nivele mejor las fluctuaciones debidas al azar; por ejemplo, en la tabla 18.2 y en la gráfica 18.4 se muestran los promedios móviles a tres y cinco años por una serie de datos de producción.



GRÁFICA 18.4 Trazo de serie de tiempo de producción, promedio móvil de tres y cinco años

Las ventas, la producción y otras series económicas y de negocios en general no tienen 1) periodos de oscilación con igual longitud ni 2) oscilaciones con amplitudes iguales; por lo tanto, en la práctica, la aplicación de un promedio móvil no genera de manera precisa una recta, por ejemplo, la serie de producción que se encuentra en la tabla 18.2 se repite casi cada cinco años, pero la amplitud de los datos varía de una oscilación a otra. Como la tendencia parece ser ascendente y un tanto lineal, los dos promedios móviles (tres y cinco años) parecen adecuados para describir la tendencia en la producción desde 1995.

El promedio móvil de cuatro años, seis años y otros números de años pares presenta un problema menor respecto del centrado de los totales móviles y de los promedios móviles. Observe en la tabla 18.3, que se muestra a continuación, que no hay un periodo central, por lo que los totales móviles se colocan *entre* dos periodos. El total de los primeros cuatro años ($42) se coloca entre 2006 y 2007; el total de los siguientes cuatro años es 43 dólares. Se obtiene la media de los promedios de los primeros cuatro años y de los segundos cuatro años ($10.50 y $10.75, respectivamente), y la cifra resultante se centra en 2007; este procedimiento se repite hasta calcular todos los promedios posibles de cuatro años.

TABLA 18.3 Cálculo del promedio móvil de cuatro años

Año	Ventas	Total móvil de cuatro años	Promedio móvil de cuatro años	Promedio móvil centrado
2005	8			
2006	11			
		42 (8 + 11 + 9 + 14)	10.50 (42/4)	
2007	9			10.625
		43 (11 + 9 + 14 + 9)	10.75 (43/4)	
2008	14			10.625
		42	10.50	
2009	9			10.625
		43	10.75	
2010	10			10.000
		37	9.25	
2011	10			9.625
		40	10.00	
2012	8			
2013	12			

Promedio móvil ponderado

OA18-3
Suavizar una serie de tiempo calculando un promedio móvil ponderado.

En un promedio móvil se utiliza la misma ponderación para cada observación; por ejemplo, el total móvil de tres años se divide entre el valor 3 para producir el promedio móvil. En otras palabras, en este caso, cada valor tiene una ponderación de un tercio; de manera similar, en el caso de un promedio móvil de cinco años, cada uno tiene una ponderación de un quinto.

Una extensión natural de la media ponderada que se analizó en el capítulo 3 es calcular un promedio móvil ponderado; esto implica seleccionar una posible ponderación distinta para cada valor y luego calcular un promedio ponderado de los valores *n* más recientes como valor uniformizado. En la mayoría de las aplicaciones se emplea el valor uniformizado como una proyección al futuro; por lo tanto, a la observación más reciente se le da la ponderación mayor, la cual disminuye conforme estas son más antiguas; por supuesto, la suma de las ponderaciones debe ser igual a 1.

Por ejemplo, suponga que calcula un promedio móvil ponderado de dos años para los datos que se registran en la tabla 18.3, y se obtiene una ponderación del doble al valor más reciente; en otras palabras, se asigna una ponderación de 2/3 al último año y de 1/3 al valor inmediatamente anterior a ese. Luego, las ventas “proyectadas” para 2007 se determinan mediante (1/3)($8) + (2/3)($11) = $10. El siguiente promedio móvil se calcularía como (1/3)($11) + (2/3)($9) = $9.667; de la misma manera, el promedio móvil final, o ponderado, de 2014, sería (1/3)($8) + (2/3)($12) = $10.667. En resumen, la técnica de utilizar promedios móviles tiene el objetivo de identificar la tendencia de largo plazo en una serie de tiempo (pues suaviza las fluctuaciones de corto plazo), y se utiliza para revelar cualesquiera fluctuaciones cíclicas y estacionales.

EJEMPLO

Cedar Fair opera siete parques de diversiones y cinco parques acuáticos independientes; su asistencia combinada (en miles) durante los últimos 20 años se muestra en la tabla de la página siguiente.

Año	Asistencia (miles)	Año	Asistencia (miles)
1993	5 761	2003	12 181
1994	6 148	2004	12 557
1995	6 783	2005	12 700
1996	7 445	2006	19 300
1997	7 405	2007	22 100
1998	11 450	2008	22 720
1999	11 224	2009	21 136
2000	11 703	2010	22 785
2001	11 890	2011	23 377
2002	12 380	2012	23 300

Como un socio le pide estudiar la tendencia de la asistencia, usted debe calcular un promedio móvil de tres años y un promedio móvil ponderado de tres años con ponderaciones de 0.2, 0.3 y 0.5 para años sucesivos.

SOLUCIÓN

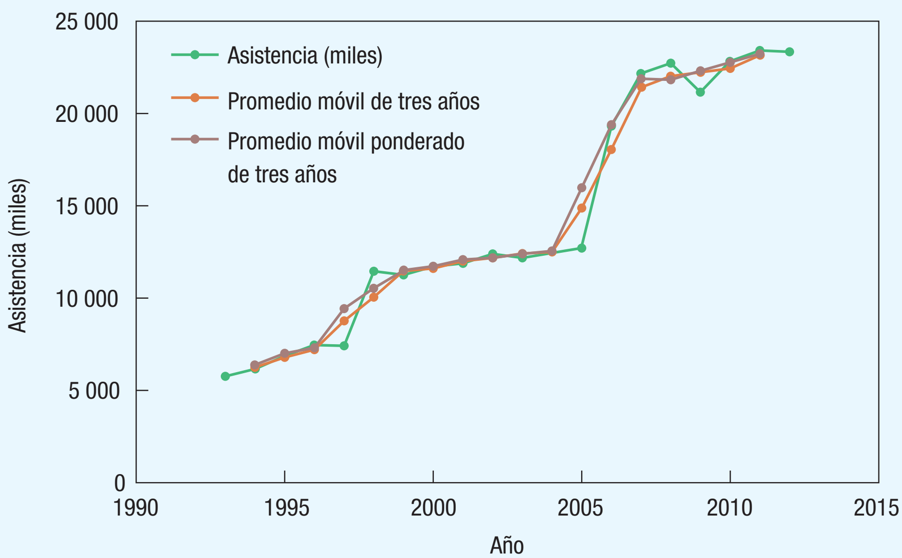
El promedio móvil de tres años es:

Año	Asistencia (miles)	Promedio móvil de tres años	Determinado por
1993	5 761		
1994	6 148	6 230.7	(5 761.0 + 6 148.0 + 6 783.0)/3
1995	6 783	6 792.0	(6 148.0 + 6 783.0 + 7 445.0)/3
1996	7 445	7 211.0	
1997	7 405	8 766.7	
1998	11 450	10 026.3	
1999	11 224	11 459.0	
2000	11 703	11 605.7	
2001	11 890	11 991.0	
2002	12 380	12 150.3	
2003	12 181	12 372.7	
2004	12 557	12 479.3	
2005	12 700	14 852.3	
2006	19 300	18 033.3	
2007	22 100	21 373.3	
2008	22 720	21 985.3	
2009	21 136	22 213.7	
2010	22 785	22 432.7	
2011	23 377	23 154.0	(22 785.0 + 23 377.0 + 23 300)/3
2012	23 300		

El promedio móvil *ponderado* de tres años es:

Año	Asistencia (miles)	Promedio móvil ponderado de tres años	Determinado por
1993	5 761		
1994	6 148	6 388.1	(0.2)(5 761) + (0.3)(6 148) + (0.5)(6 783)
1995	6 783	6 987.0	(0.2)(6 148) + (0.3)(6 783) + (0.5)(7 445)
1996	7 445	7 292.6	
1997	7 405	9 435.5	
1998	11 450	10 528.0	
1999	11 224	11 508.7	
2000	11 703	11 700.7	
2001	11 890	12 097.6	
2002	12 380	12 182.5	
2003	12 181	12 408.8	
2004	12 557	12 553.3	
2005	12 700	15 971.4	
2006	19 300	19 380.0	
2007	22 100	21 850.0	
2008	22 720	21 804.0	
2009	21 136	22 277.3	
2010	22 785	22 751.2	
2011	23 377	23 220.1	(0.2)(22 785) + (0.3)(23 377) +(0.5)(23 300)
2012	23 300		

Asistencia, promedio móvil de tres años y promedio móvil ponderado para Cedar Fair (1993 a 2012)



Si estudia la gráfica con cuidado, observará que la tendencia de la asistencia es ascendente de manera uniforme, con incrementos de 400 000 visitantes cada año; sin embargo, hay un “salto” de casi tres millones por año entre 1997 y 1998; es probable que esto refleje que Cedar Fair adquirió Knott’s Berry Farm (ubicado en el sur de California) a finales de 1997, lo que generó un incremento repentino de la asistencia. Ocurrió un auge similar en 2006 con la compra de King’s Island, cerca de Cincinnati, Ohio. También se puede observar la crisis financiera de 2009.

El promedio móvil ponderado sigue los datos de manera más cercana que el promedio móvil, lo que refleja la influencia adicional que recibe el periodo más reciente; en otras palabras, el método ponderado, conforme al cual se da la ponderación mayor al periodo más reciente, no será tan suave.

AUTOEVALUACIÓN 18-1

Determine el promedio móvil de tres años de las ventas de Waccamaw Machine Tool, Inc., y trace los datos originales y el promedio móvil.

Año	Número producido (miles)	Año	Número producido (miles)
2008	2	2011	5
2009	6	2012	3
2010	4	2013	10

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

1. Calcule un promedio móvil ponderado en cuatro trimestres del número de suscriptores de la Boxley Box Company durante los nueve trimestres que abarcan los datos (estos se reportan en miles. Aplique ponderaciones de 0.1, 0.2, 0.3 y 0.4, respectivamente, a los tres trimestres anteriores y al trimestre en curso; en pocas palabras, describa la tendencia en el número de suscriptores.

1er. trimestre 2011	28 766
2o. trimestre 2011	30 057
3er. trimestre 2011	31 336
4o. trimestre 2011	33 240
1er. trimestre 2012	34 610
2o. trimestre 2012	35 102
3er. trimestre 2012	35 308
4o. trimestre 2012	35 203
1er. trimestre 2013	34 386

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

2. En la siguiente tabla se registra el número de boletos para cine que vendió el Library Cinema-Complex, en miles, durante el periodo de 2001 a 2013. Calcule el promedio móvil ponderado de cinco años con ponderaciones de 0.1, 0.1, 0.2, 0.3 y 0.3, respectivamente. Describa la tendencia del rendimiento.

2001	8.61	2008	6.61
2002	8.14	2009	5.58
2003	7.67	2010	5.87
2004	6.59	2011	5.94
2005	7.37	2012	5.49
2006	6.88	2013	5.43
2007	6.71		

OA18-4 Utilizar un análisis de regresión para ajustar una línea de tendencia lineal a una serie de tiempo.

Tendencia lineal

La tendencia de largo plazo de muchas series de negocios, como ventas, exportaciones y producción, con frecuencia se aproxima a una recta; en este caso, la ecuación para describir este crecimiento es:

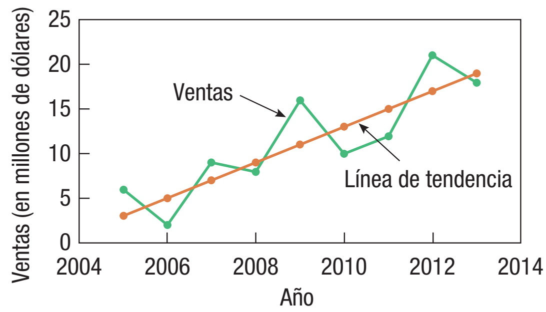
ECUACIÓN DE TENDENCIA LINEAL $$\hat{y} = a + bt \quad [18.1]$$

donde:

- $\hat{y}$, que se lee *y* testada, es el valor proyectado de la variable *y* de un valor seleccionado de *t*;
- *a* es la intersección con el eje *y*; es el valor estimado de *y* cuando $t = 0$; otra forma de expresarlo es: *a* es el valor estimado de *y* donde la línea cruza el eje *y* cuando *t* es cero;
- *b* es la pendiente de la recta, o el cambio promedio en $\hat{y}$ por cada aumento de una unidad en *t*;
- *t* es cualquier valor de tiempo seleccionado.

Para ilustrar el significado de $\hat{y}$, *a*, *b* y *t* en un problema de serie de tiempo, en la gráfica 18.5 se traza una recta para representar la tendencia habitual de las ventas. Suponga que esta compañía inició sus operaciones en 2005; este año inicial se codifica de manera arbitraria como año 1. Observe que las ventas aumentaron 2 millones de dólares en promedio cada año; es decir, con base en la recta trazada por los datos de ventas, estas aumentaron de 3 millones de dólares en 2005 a 5 millones en 2006, a 7 millones en 2007, a 9 millones en 2008, y así sucesivamente; por lo tanto, la pendiente (*b*) es 2. La pendiente es el cambio promedio en las ventas en cada aumento de unidades en un periodo. Además, observe que la recta interseca el eje *y* (cuando $t = 0$; es decir, el año 2004) en $ 1 millón; este punto es *a*. Otra manera de determinar *b* es ubicar el punto de partida de la recta en el año 1, (3, para 2005, en este ejemplo); luego se ubica el valor en la recta del último año (19, para 2013).

Periodo	Año	Ventas	Línea de tendencia
1	2005	6	3
2	2006	2	5
3	2007	9	7
4	2008	8	9
5	2009	16	11
6	2010	10	13
7	2011	12	15
8	2012	21	17
9	2013	18	19



GRÁFICA 18.5 Recta ajustada a los datos de ventas

He aquí la ecuación de la recta de la gráfica 18.5:

$$\hat{y} = 1 + 2t$$

donde:

- $\hat{y}$ representa las ventas en millones de dólares;
- 1 es la intercepción con el eje *Y*; también representa las ventas en millones de dólares del año 0 (2004);
- t es el periodo codificado para cada año.

En el capítulo 13 se trazó una recta por los puntos en un diagrama de dispersión para aproximar la recta de regresión; sin embargo, cabe observar que este método para determinar la ecuación de regresión tiene una desventaja importante: la posición de la recta depende del criterio del individuo que la trace. Es probable que tres personas tracen tres rectas distintas de las gráficas de dispersión; de igual forma, la recta que se traza según los datos de ventas en la gráfica 18.5 quizá no sea la de mejor ajuste. Debido al criterio subjetivo, este método solo se debe emplear cuando se necesite una aproximación rápida de la ecuación de la recta, o para verificar si la de mínimos cuadrados es razonable, tema que se analiza enseguida.

Método de los mínimos cuadrados

En el análisis de una regresión lineal simple, en el capítulo 13, se mostró el método de los mínimos cuadrados para determinar la mejor relación lineal entre dos variables. En los métodos de proyección, el tiempo es la variable independiente, y el valor de la serie de tiempo, la dependiente; además, con frecuencia se codifica la variable independiente (tiempo) para facilitar la interpretación de las ecuaciones. En otras palabras, se hace que *t* sea 1 en el primer año, 2 en el segundo, y así en lo sucesivo. Si una serie de tiempo incluye las ventas de General Electric para cinco años iniciando en 2002 hasta 2006, se codifica el año 2002 como 1, 2003 como 2, y 2006 como 5.

Cuando se utiliza el método de los mínimos cuadrados para encontrar la línea de tendencia para una serie de tiempo, los errores o residuos suelen estar correlacionados y no son independientes; por lo tanto, los resultados de la prueba de hipótesis para el análisis de regresión que se presentó en el capítulo 13 pueden no ser válidos en un análisis de serie de tiempo, sin embargo, la regresión lineal simple puede usarse aún para encontrar una línea de mejor ajuste para una línea de tiempo.

EJEMPLO

Las ventas de Jensen Foods, una pequeña cadena de abarrotes ubicada en el suroeste de Texas, de 2009 a 2013 se muestran a la derecha.

Determine la ecuación de regresión. ¿Cuál es el incremento anual de las ventas? ¿Cuál es la proyección de las ventas para 2015?

Año	Tiempo (t)	Ventas (en millones de dólares)
2009	1	7.0
2010	2	10.0
2011	3	9.0
2012	4	11.0
2013	5	13.0

SOLUCIÓN

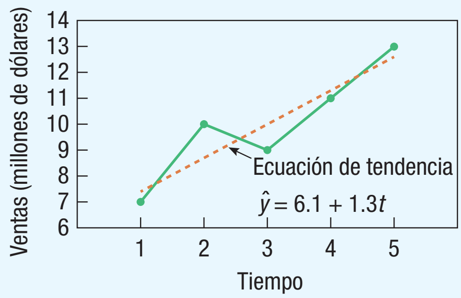
Para determinar la ecuación de la tendencia, puede utilizar la fórmula [13.4] para encontrar la pendiente (*b*), y la fórmula [13.5] para ubicar la intercepción (*a*). Se sustituye *t* (los valores codificados del

ESTADÍSTICA EN ACCIÓN

Con frecuencia los inversionistas emplean el análisis de regresión para estudiar la relación entre una acción en particular y la condición general del mercado. La variable dependiente es el cambio porcentual mensual del valor de la acción, y la variable independiente es el cambio porcentual mensual de un índice de mercado, como el Índice Compuesto 500, de Standard & Poor's. El valor de b en la ecuación de regresión es el *coeficiente beta*, o solo *beta*, de la acción en particular. Si b es mayor que 1, se deduce que la acción es sensible a los cambios que se producen en el mercado; si se encuentra entre 0 y 1, la implicación es que la acción no es sensible a los cambios del mercado.

año), por X en estas ecuaciones. También es posible utilizar un paquete de software. En la gráfica 18.6 se muestra el trazo de ventas y las ventas ajustadas o de tendencia para cada año.

A partir del análisis de regresión, la ecuación de la tendencia es $\hat{y} = 6.1 + 1.3t$. ¿Cómo se interpreta esta ecuación? Las ventas están en millones de dólares; por lo tanto, el valor 1.3 indica que estas aumentaron a una tasa de 1.3 millones de dólares por año. El valor 6.1 es el estimado de las ventas en el año 0; es decir, la estimación del año base (2008); por ejemplo, para determinar el punto en la recta de 2012, se sustituye el valor t de 4 en la ecuación. Entonces $\hat{y} = 6.1 + 1.3(4) = 11.3$.



GRÁFICA 18.6 Ventas y recta de tendencia para Jensen Foods (2009 a 2013)

Si los datos (ventas, producción u otros) aproximan a una tendencia lineal, se emplea la ecuación desarrollada mediante la técnica de mínimos cuadrados para estimar los siguientes. Es razonable que las ventas de Jensen Foods sigan una trayectoria lineal; por ello se utiliza la ecuación de la tendencia para proyectar las ventas futuras.

En la tabla 18.4, el año 2009 se codifica como 1, 2011 como 3, y 2013 como 5; es lógico codificar 2015 como 7 y 2016 como 8; por lo tanto, se sustituye el 7 en la ecuación de la tendencia y se despeja $\hat{y}$.

$$\hat{y} = 6.1 + 1.3t = 6.1 + 1.3(7) = 15.2$$

De esta manera, con base en las ventas pasadas, la estimación para 2015 es 15.2 millones de dólares.

TABLA 18.4 Cálculos para determinar los puntos de la recta de mínimos cuadrados con los valores codificados

Año	Ventas (millones de dólares)	Código (t)	y	Determinado por
2009	7.0	1	7.4	6.1 + 1.3(1)
2010	10.0	2	8.7	6.1 + 1.3(2)
2011	9.0	3	10.0	6.1 + 1.3(3)
2012	11.0	4	11.3	6.1 + 1.3(4)
2013	13.0	5	12.6	6.1 + 1.3(5)

En este ejemplo de serie de tiempo, había cinco años de datos de ventas; con base en esto, se estiman las ventas de 2016. Muchos investigadores sugieren que no se proyecten ventas, producción u otras series de negocios y económicas más que $n/2$ periodos de tiempo a futuro, donde n es el número de puntos de datos; por ejemplo, si hay 10 años de información, solo se estiman hasta 5 años a futuro ($n/2 = 10/2 = 5$). Otros sugieren que la proyección no puede ser mayor que dos años, en especial en tiempos de rápidos cambios económicos.

AUTOEVALUACIÓN 18-2

A continuación se muestra la producción anual de sillas mecedoras grandes de Wood Products, Inc., de 2006 a 2013.

Año	Producción (miles)	Año	Producción (miles)
2006	4	2010	11
2007	8	2011	9
2008	5	2012	11
2009	8	2013	14

(a) Trace los datos de la producción en un diagrama de dispersión.
(b) Determine la ecuación de mínimos cuadrados con un paquete de software.
(c) Utilizando la ecuación de los mínimos cuadrados, determine los puntos de la recta de 2006 y 2013. Conecte los puntos para llegar a la recta.
(d) Con base en la ecuación de la tendencia lineal, ¿cuál es la producción estimada de 2016?

EJERCICIOS

3. A continuación se reporta el número de habitaciones alquiladas en Plantation Resorts, Georgia, de los años de 2009 a 2013.

Año	Alquiladas	Año	Alquiladas	Año	Alquiladas
2003	6 714	2007	9 762	2011	6 162
2004	7 991	2008	10 180	2012	6 897
2005	9 075	2009	8 334	2013	8 285
2006	9 775	2010	8 272		

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Determine la ecuación de mínimos cuadrados. De acuerdo con esta información, ¿cuáles son los alquileres estimados para 2014?

4. En la siguiente tabla se muestran las ventas netas en millones de dólares de Home Depot, Inc., y sus subsidiarias de 1993 a 2012.

Año	Ventas netas	Año	Ventas netas	Año	Ventas netas
1993	$ 9 239	2000	$45 738	2007	$77 349
1994	12 477	2001	53 553	2008	71 288
1995	15 470	2002	58 247	2009	66 176
1996	19 535	2003	64 816	2010	67 997
1997	24 156	2004	73 094	2011	70 395
1998	30 219	2005	81 511	2012	74 754
1999	38 434	2006	90 837		

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Determine la ecuación de mínimos cuadrados. Con base en esta información, ¿cuáles son las ventas estimadas para 2014 y 2015?

5. En la siguiente tabla se registran las cantidades anuales de vidrio de desecho producido por Kimble Glass Works, Inc., de 2009 a 2013.

Año	Código	Desecho (toneladas)	Año	Código	Desecho (toneladas)
2009	1	2	2012	4	5
2010	2	4	2013	5	6
2011	3	3			

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Determine la ecuación de la tendencia de mínimos cuadrados y estime la cantidad de desecho que se generará en 2015.

6. En la siguiente tabla se muestran las ventas de Walder's Milk and Dairy Products, en millones de dólares, durante el periodo de 2007 a 2013.

Año	Código	Ventas (en millones de dólares)	Año	Código	Ventas (en millones de dólares)
2007	1	17.5	2011	5	24.5
2008	2	19.0	2012	6	26.7
2009	3	21.0	2013	7	27.3
2010	4	22.7			

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Determine la ecuación de tendencia de mínimos cuadrados y estime las ventas de 2015.

Tendencias no lineales

OA18-5 Utilizar un análisis de regresión para ajustar una serie de tiempo no lineal.

En el análisis anterior la atención se centró en una serie de tiempo cuyo crecimiento o declinación se aproximaban a una recta. Una ecuación de tendencia lineal se utiliza para representar la serie de tiempo cuando se considera que los datos aumentan (o disminuyen) en *cantidades iguales*, en promedio, de un periodo a otro.

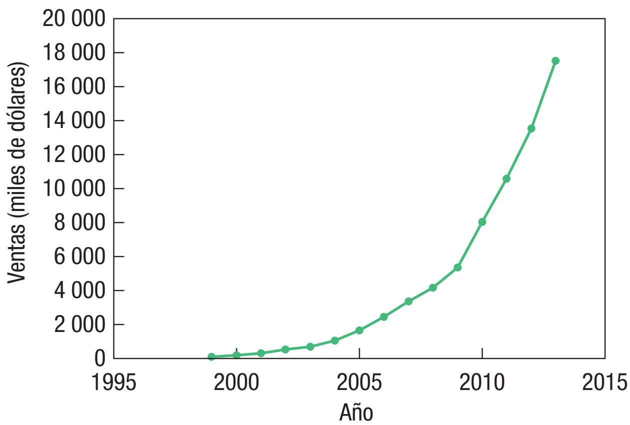
Los datos que aumentan (o disminuyen) en *cantidades cada vez mayores* durante un periodo aparecen *curvilíneos* cuando se trazan en una gráfica con escala aritmética; en otras palabras, si aumentan (o disminuyen) en *porcentajes* o *proporciones iguales* durante un periodo aparecen curvilíneos sobre un papel cuadriculado (consulte la gráfica 18.7).

La ecuación de la tendencia de una serie de tiempo que no se aproxime a una tendencia curvilínea, como la que se representa en la gráfica 18.7, se calcula con los logaritmos de los datos y el método de mínimos cuadrados. La ecuación general de la ecuación de la tendencia logarítmica es:

ECUACIÓN DE TENDENCIA LOGARÍTMICA $\log \hat{y} = \log a + \log b(t)$ **[18.2]**

La ecuación de la tendencia logarítmica se puede determinar, con los datos de Gulf Shores Importers que se muestran en la gráfica 18.7, utilizando Excel. El primer paso es capturar la información y después determinar el logaritmo base 10 de cada una de las importaciones del año; por último, se utiliza el procedimiento de regresión para encontrar la ecuación de mínimos cuadrados. En otras palabras, se toma el logaritmo de cada uno de los datos del año, y luego se utilizan los logaritmos como la variable dependiente y el año codificado como la variable independiente.

Año	Ventas (miles de dólares)	Año	Ventas (miles de dólares)
1999	124.2	2007	3 358.2
2000	175.6	2008	4 181.3
2001	306.9	2009	5 388.5
2002	524.2	2010	8 027.4
2003	714.0	2011	10 587.2
2004	1 052.0	2012	13 537.4
2005	1 638.3	2013	17 515.6
2006	2 463.2		



GRÁFICA 18.7 Ventas de Gulf Shores Importers, 1999-2013

Código	Año	Ventas (en miles de dólares)	Ventas log
1	1999	124.2	2.094122
2	2000	175.6	2.244525
3	2001	306.9	2.486997
4	2002	524.2	2.719497
5	2003	714.0	2.853698
6	2004	1 052.0	3.022016
7	2005	1 638.3	3.214393
8	2006	2 463.2	3.391500
9	2007	3 358.2	3.526107
10	2008	4 181.3	3.621311
11	2009	5 388.5	3.731468
12	2010	8 027.4	3.904575
13	2011	10 587.2	4.024781
14	2012	13 537.4	4.131535
15	2013	17 515.6	4.243425

RESUMEN DE SALIDA

Estadísticos de regresión	
Múltiple "R"	0.994
"R" cuadrada	0.988
"R" cuadrada ajustada	0.987
Error estándar	0.079
Observaciones	15

ANOVA

	gl	*SS*	*MS*	*F*
Regresión	1	6.585	6.585	1 065.2279
Residual	13	0.080	0.006	
Total	14	6.666		

	Coeficientes
Intercepción	2.053805
Código	0.153357

Ecuación de regresión
$\hat{y} = 2.053805 + 0.153357t$

La ecuación de regresión es $\hat{y} = 2.053805 + 0.153357t$, que es la forma logarítmica. Ahora se tiene una ecuación de la tendencia en términos del porcentaje de cambio; es decir, el valor 0.153357 es el porcentaje de cambio de $\hat{y}$ por cada aumento unitario de t. Este valor es similar a la media geométrica que se describió en el capítulo 3.

El logaritmo de b es 0.153357, y su antilogaritmo, o inverso, es 1.423498. Si a este valor se le resta 1, como se hizo en el capítulo 3, el valor 0.423498 indica la tasa anual de incremento de la media geométrica de 1999 a 2013. La conclusión es que las importaciones aumentaron a una tasa de 42.35% al año durante el periodo.

También se utiliza la ecuación de la tendencia logarítmica para hacer estimaciones de valores futuros. Suponga que desea estimar las importaciones de 2017. El primer paso es determinar el código de ese año, que es 19; para explicar esto, 2013 tiene un código de 15 y 2017 es cuatro años después; en consecuencia, 15 + 4 = 19. El logaritmo de las importaciones de 2017 es:

$$\hat{y} = 2.053805 + 0.153357t = 2.053805 + 0.153357(19) = 4.967588$$

Para encontrar las importaciones estimadas de 2017 necesita el antilogaritmo de 4.967588, que es 92 809; el cual, es la estimación del número de importaciones de 2017. Recuerde que los datos se dieron en miles de dólares, por lo que la estimación es 92 809 000 dólares.

AUTOEVALUACIÓN 18-3

Las ventas de Tomlin Manufacturing de 2009 a 2013 se muestran a la derecha.

(a) Determine la ecuación de la tendencia logarítmica de los datos de ventas.

(b) ¿Cuál fue el porcentaje de incremento anual de las ventas de 2009 a 2013?

(c) ¿Cuáles son las ventas proyectadas para 2014?

Año	Ventas (en millones de dólares)
2009	2.13
2010	18.10
2011	39.80
2012	81.40
2013	112.00

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

7. Sally's Software, Inc., es un proveedor de software en el área de Sarasota con un crecimiento rápido; las ventas de los últimos cinco años, de 2009 a 2013, se muestran a continuación.

Año	Ventas (en millones de dólares)
2009	1.1
2010	1.5
2011	2.0
2012	2.4
2013	3.1

a. Determine la ecuación de la tendencia logarítmica.
b. En promedio, ¿en qué porcentaje aumentaron las ventas durante el periodo?
c. Estime las ventas de 2016.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

8. Al parecer, las importaciones de carbón negro aumentaron casi 10% al año.

Año	Importaciones de carbón negro (miles de toneladas)	Año	Importaciones de carbón negro (miles de toneladas)
2006	92.0	2010	135.0
2007	101.0	2011	149.0
2008	112.0	2012	163.0
2009	124.0	2013	180.0

a. Determine la ecuación de la tendencia logarítmica.
b. En promedio, ¿en qué porcentaje aumentaron las importaciones durante el periodo?
c. Estime las importaciones durante 2016.

Variación estacional

OA18-6 Calcular y aplicar índices estacionales para realizar pronósticos ajustados a la estación.

Con anterioridad se mencionó que la *variación estacional* es otro componente de una serie de tiempo. Las series de negocios, como las ventas de automóviles, los embarques de botellas de bebidas de cola y la construcción residencial, tienen periodos de actividad superior e inferior al promedio cada año. En el área de producción, una razón para analizar las fluctuaciones estacionales es contar con un abastecimiento suficiente de materias primas que permita cumplir con la cambiante demanda estacional. La división de recipientes de vidrio de una compañía importante del sector, por ejemplo, fabrica botellas de cerveza no retornables, frascos para yodo, frascos para aspirina, botellas para cemento plastificado, etcétera. El departamento de programación de producción necesita saber cuántas botellas debe producir y cuándo debe producir

de cada tipo porque una corrida de demasiadas botellas de un tipo puede ocasionar un problema grave de almacenamiento; además, la producción no se puede basar por completo en los pedidos existentes, pues muchos pedidos se hacen por teléfono para su embarque inmediato. Como la demanda de muchas botellas varía de acuerdo con la temporada, una proyección con una anticipación de un año o dos, por mes, es esencial para lograr una programación adecuada.

Un análisis de las fluctuaciones estacionales durante un periodo de años también puede ayudar para evaluar las ventas actuales. Las ventas habituales de tiendas departamentales en Estados Unidos, salvo las ventas por correo, se muestran como índices en la tabla 18.5; cada uno representa las ventas promedio de un periodo de varios años. Las ventas reales de algunos meses estuvieron arriba del promedio (representado por un índice mayor que 100), y las ventas de los demás meses, debajo del promedio. El índice de 126.8 de diciembre indica que, por lo regular, las ventas de diciembre son 26.8% superiores al mes promedio; el índice de 86.0 de julio indica que las ventas departamentales de este mes casi siempre son 14% menores a las de un mes promedio.

TABLA 18.5 Índices estacionales habituales de ventas en tiendas departamentales en Estados Unidos, excluyendo las ventas por correo

Enero	87.0	Julio	86.0
Febrero	83.2	Agosto	99.7
Marzo	100.5	Septiembre	101.4
Abril	106.5	Octubre	105.8
Mayo	101.6	Noviembre	111.9
Junio	89.6	Diciembre	126.8

Suponga que un gerente de tienda emprendedor, en un esfuerzo por estimular las ventas durante diciembre, introdujo diversas promociones únicas, como coros de villancicos por toda la tienda, exhibiciones mecánicas y dependientes vestidos con trajes de Santa Claus. Cuando se calculó el índice de ventas de ese mes, fue 150.0; en comparación con las ventas de diciembre habituales de 126.8, se concluyó que el programa de promoción fue un gran éxito.

Determinación de un índice estacional

Un conjunto habitual de índices mensuales consta de 12 índices representativos de los datos de un periodo de 12 meses; es lógico que haya cuatro índices estacionales habituales para los datos que se reportaron en el trimestre. Cada índice es un porcentaje, cuyo promedio anual es igual a 100.0; es decir, cada índice mensual indica el nivel de ventas, producción u otra variable en relación con el promedio anual de 100.0. Un índice habitual de 96.0 en enero indica que las ventas (o cualquier otra variable) están, en general, 4% debajo del promedio del año; uno de 107.2 en octubre significa que la variable está, en general, 7.2% arriba del promedio.

Hay varios métodos para medir las fluctuaciones estacionales habituales en una serie de tiempo; el más común para calcular el patrón estacional habitual se denomina método de la **razón con el promedio móvil**. Este elimina los componentes de tendencia, cíclicos e irregulares de los datos originales (Y). En el siguiente análisis, T se refiere a la tendencia, S a la variación estacional, C a la variación cíclica e I a la variación irregular; los números resultantes son el *índice estacional habitual*.

A continuación, se estudian con detalle los pasos para obtener los índices estacionales habituales con el método de la razón con el promedio móvil; los índices de interés pueden ser mensuales o trimestrales. Para ilustrar este método, se eligieron las ventas trimestrales de Toys International; primero, se muestran los pasos necesarios para llegar al conjunto de índices estacionales habituales, luego se utiliza el software MegaStat de Excel para calcular los índices estacionales.

EJEMPLO

En la tabla 18.6 se registran (en millones de dólares) las ventas trimestrales de Toys International de 2008 a 2013. Determine un índice estacional trimestral con el método de la razón con el promedio móvil.

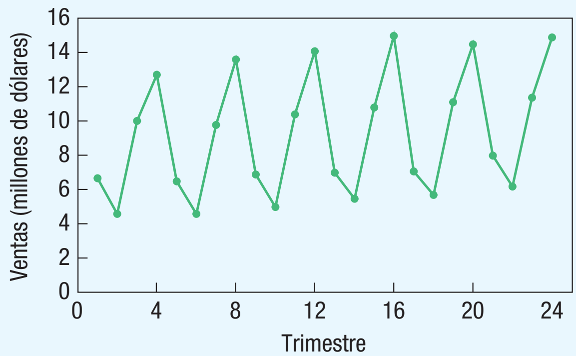
TABLA 18.6 Ventas trimestrales de Toys International (millones de dólares), 2008 a 2013

Año	Invierno	Primavera	Verano	Otoño
2008	6.7	4.6	10.0	12.7
2009	6.5	4.6	9.8	13.6
2010	6.9	5.0	10.4	14.1
2011	7.0	5.5	10.8	15.0
2012	7.1	5.7	11.1	14.5
2013	8.0	6.2	11.4	14.9

SOLUCIÓN

En la gráfica 18.8 se registran las ventas trimestrales de Toys International durante el periodo de seis años. Observe la naturaleza estacional de las ventas: en cada año, las ventas del cuarto trimestre son las mayores, y las del segundo trimestre, las menores; además, hay un aumento moderado de las ventas de un año al siguiente. Para detectar esta característica basta observar los valores de las ventas de todos los cuartos trimestres; en los cuales, durante el periodo de seis años, las ventas aumentaron; por lo tanto, se esperaría que hubiera un patrón estacional similar para 2014.

Año	Trimestre	Ventas (millones de dólares)	Año	Trimestre	Ventas (millones de dólares)
2008	1	6.7	2011	1	7.0
	2	4.6		2	5.5
	3	10.0		3	10.8
	4	12.7		4	15.0
2009	1	6.5	2012	1	7.1
	2	4.6		2	5.7
	3	9.8		3	11.1
	4	13.6		4	14.5
2010	1	6.9	2013	1	8.0
	2	5.0		2	6.2
	3	10.4		3	11.4
	4	14.1		4	14.9



GRÁFICA 18.8 Ventas trimestrales de Toys International (en millones de dólares), 2008-2013

Para determinar los índices estacionales trimestrales se deben dar seis pasos.

Paso 1: Para el siguiente análisis se debe consultar la tabla 18.7. El primer paso es determinar el total móvil de los cuatro trimestres de 2008. Inicie con el trimestre invernal de 2008, sume \$6.7, \$4.6, \$10.0 y \$12.7; el total es \$34.0 (millones). El total de los cuatro trimestres "se desplaza" al sumar las ventas de primavera, verano y otoño de 2008 a las ventas de invierno de 2009. El total es 33.8 millones de dólares, determinado por 4.6 + 10.0 + 12.7 + 6.5; este procedimiento se aplica a las ventas trimestrales de cada año. En la columna 2 de la tabla 18.7 se muestran los totales móviles; observe que el total móvil (34.0) se coloca entre las ventas de primavera y verano de 2008; el siguiente total móvil (33.8) se coloca entre las ventas del verano y otoño de 2008, etcétera; verifique los totales con frecuencia para evitar errores aritméticos.

Paso 2: Cada total móvil trimestral en la columna 2 se divide entre 4 para obtener el promedio móvil trimestral (consulte la columna 3). Todos los promedios móviles aún están colocados entre los trimestres; por ejemplo, el primer promedio móvil (8.500) se coloca entre la primavera y el verano de 2008.

Paso 3: Se centran los promedios móviles. El primer promedio móvil centrado se determina mediante (8.500 + 8.450)/2 = 8.475, y se centra en oposición al verano de 2008; el segundo se determina mediante (8.450 + 8.450)/2 = 8.450; los otros se determinan de manera similar. Observe en la columna 4 que cada promedio móvil centrado se coloca en un trimestre en particular.

Paso 4: Luego calcule el índice estacional específico por cada trimestre dividiendo las ventas en la columna 1 entre el promedio móvil centrado en la columna 4. El índice estacional específico reporta la razón del valor de la serie de tiempo original con el promedio móvil; para explicar esta cuestión un poco más, si representa la serie de tiempo con *TSCI* y el promedio móvil con *TC*, de manera algebraica, si calcula *TSCI*/*TC*, el resultado es el componente estacional específico *SI*. El índice estacional específico del trimestre del verano de 2008 es 1.180, determinado por 10.0/8.475.

Paso 5: Los índices estacionales específicos se organizan en la tabla 18.8; mediante esta se ubican los de los trimestres correspondientes. Los valores 1.180, 1.130, 1.141, 1.126 y 1.142 representan estimaciones del índice estacional habitual del trimestre de verano; un método razonable para encontrar un índice estacional habitual es promediarlos a fin de eliminar el componente irregular; por lo tanto, el índice habitual del trimestre de verano se determina mediante (1.180 + 1.130 + 1.141 + 1.126 + 1.143)/5 = 1.144. Se utilizó la media aritmética, aunque también sirve la mediana o una media modificada.

TABLA 18.7 Cálculos necesarios para índices estacionales específicos

Año	Trimestre	(1) Ventas (millones de dólares)	(2) Total de los cuatro trimestres	(3) Promedio móvil de los cuatro trimestres	(4) Promedio móvil centrado	(5) Estacional específico
2008	Invierno	6.7				
	Primavera	4.6				
			34.0	8.500		
	Verano	10.0			8.475	1.180
			33.8	8.450		
	Otoño	12.7			8.450	1.503
			33.8	8.450		
2009	Invierno	6.5			8.425	0.772
			33.6	8.400		
	Primavera	4.6			8.513	0.540
			34.5	8.625		
	Verano	9.8			8.675	1.130
			34.9	8.725		
	Otoño	13.6			8.775	1.550
			35.3	8.825		
2010	Invierno	6.9			8.900	0.775
			35.9	8.975		
	Primavera	5.0			9.038	0.553
			36.4	9.100		
	Verano	10.4			9.113	1.141
			36.5	9.125		
	Otoño	14.1			9.188	1.535
			37.0	9.250		
2011	Invierno	7.0			9.300	0.753
			37.4	9.350		
	Primavera	5.5			9.463	0.581
			38.3	9.575		
	Verano	10.8			9.588	1.126
			38.4	9.600		
	Otoño	15.0			9.625	1.558
			38.6	9.650		
2012	Invierno	7.1			9.688	0.733
			38.9	9.725		
	Primavera	5.7			9.663	0.590
			38.4	9.600		
	Verano	11.1			9.713	1.143
			39.3	9.825		
	Otoño	14.5			9.888	1.466
			39.8	9.950		
2013	Invierno	8.0			9.888	0.801
			40.1	10.025		
	Primavera	6.2			10.075	0.615
			40.5	10.125		
	Verano	11.4				
	Otoño	14.9				

Paso 6: En teoría, las cuatro medias trimestrales (0.767, 0.576, 1.144, y 1.522) deberán totalizar 4.00, pues el promedio se fija en 1.0. El total de las cuatro medias trimestrales quizá no sea exactamente igual a ese número debido al redondeo. En este problema, el total de las medias es 4.009; por lo tanto, se aplica un *factor de corrección* a cada una de las cuatro medias para que sumen 4.00.

TABLA 18.8 Cálculos necesarios para determinar índices trimestrales habituales

Año	Invierno	Primavera	Verano	Otoño	
2008			1.180	1.503	
2009	0.772	0.540	1.130	1.550	
2010	0.775	0.553	1.141	1.535	
2011	0.753	0.581	1.126	1.558	
2012	0.733	0.590	1.143	1.466	
2013	0.801	0.615			
Total	3.834	2.879	5.720	7.612	
Media	0.767	0.576	1.144	1.522	4.009
Ajustado	0.765	0.575	1.141	1.519	4.000
Índice	76.5	57.5	114.1	151.9	

FACTOR DE CORRECCIÓN PARA MEDIAS TRIMESTRALES

$$\text{Factor de corrección} = \frac{4.00}{\text{Total de cuatro medias}} \qquad \textbf{[18.3]}$$

En este ejemplo,

$$\text{Factor de corrección} = \frac{4.00}{4.009} = 0.997755$$

Por lo tanto, el índice trimestral ajustado de invierno es 0.767(0.997755) = 0.765. Cada una de las medias se ajusta hacia abajo de modo que el total de nuestras medias trimestrales sea 4.00. En general, los índices se reportan como porcentajes, por lo que cada valor en la última fila de la tabla 18.8 se multiplica por 100; así, el índice del trimestre de invierno es 76.5, y del verano, 151.9. ¿Cómo se interpretan estos valores? Las ventas del trimestre de otoño están 51.9% por arriba de un trimestre habitual, y del invierno, 23.5% por debajo (100 – 76.5); estos resultados no deben sorprender porque en el periodo anterior a Navidad (el trimestre de otoño) las ventas de juguetes se incrementan; después de Navidad (el trimestre de invierno), las ventas se reducen de forma drástica.

El software estadístico puede realizar estos cálculos; por ejemplo, la salida de MegaStat para Excel se muestra enseguida. El uso de software reducirá en gran medida el tiempo de cómputo y la probabilidad de cometer un error en los cálculos aritméticos, pero debe comprender los pasos del proceso; puede haber diferencias ligeras en las respuestas debido al número de dígitos manejados en los cálculos.

Promedio móvil centrado y desestacionalización

t	Año	Trimestre	Ventas	Promedio móvil centrado	Razón para el promedio móvil centrado	Índices estacionales	Ventas desestacionalizadas
1	2008	1	6.70			0.765	8.759
2	2008	2	4.60			0.575	8.004
3	2008	3	10.00	8.475	1.180	1.141	8.761
4	2008	4	12.70	8.450	1.503	1.519	8.361

(*continúa*)

(*continuación*)

Promedio móvil centrado y desestacionalización

t	Año	Trimestre	Ventas	Promedio móvil centrado	Razón para el promedio móvil centrado	Índices estacionales	Ventas desestacionalizadas
5	2009	1	6.50	8.425	0.772	0.765	8.498
6	2009	2	4.60	8.513	0.540	0.575	8.004
7	2009	3	9.80	8.675	1.130	1.141	8.586
8	2009	4	13.60	8.775	1.550	1.519	8.953
9	2010	1	6.90	8.900	0.775	0.765	9.021
10	2010	2	5.00	9.038	0.553	0.575	8.700
11	2010	3	10.40	9.113	1.141	1.141	9.112
12	2010	4	14.10	9.188	1.535	1.519	9.283
13	2011	1	7.00	9.300	0.753	0.765	9.151
14	2011	2	5.50	9.463	0.581	0.575	9.570
15	2011	3	10.80	9.588	1.126	1.141	9.462
16	2011	4	15.00	9.625	1.558	1.519	9.875
17	2012	1	7.10	9.688	0.733	0.765	9.282
18	2012	2	5.70	9.663	0.590	0.575	9.918
19	2012	3	11.10	9.713	1.143	1.141	9.725
20	2012	4	14.50	9.888	1.466	1.519	9.546
21	2013	1	8.00	9.988	0.801	0.765	10.459
22	2013	2	6.20	10.075	0.615	0.575	10.788
23	2013	3	11.40			1.141	9.988
24	2013	4	14.90			1.519	9.809

Cálculo de los índices estacionales

	1	2	3	4	
2008			1.180	1.503	
2009	0.772	0.540	1.130	1.550	
2010	0.775	0.553	1.141	1.535	
2011	0.753	0.581	1.126	1.558	
2012	0.733	0.590	1.143	1.466	
2013	0.801	0.615			
Media:	0.767	0.576	1.144	1.522	4.009
Ajustada:	0.765	0.575	1.141	1.519	4.000

A continuación se resume de forma breve el razonamiento de los cálculos anteriores. En los datos originales en la columna 1 de la tabla 18.7 se incluyen los componentes de tendencia (*T*), cíclica (*C*), estacional (*S*) e irregular (*I*). El objetivo principal es eliminar la variación estacional (*S*) de la valuación de las ventas originales.

De las columnas 2 y 3 de la tabla 18.7 se deriva el promedio móvil centrado dado en la columna 4; en esencia, “quedan fuera” las fluctuaciones estacional e irregular de los datos originales en la columna 1, por lo tanto, en la columna solo quedan las variaciones por tendencia y la cíclica (*TC*).

En seguida, se dividen los datos de ventas en la columna 1 (*TCSI*) entre el promedio móvil centrado del tercer trimestre en la columna 4 (*TC*) para llegar a las variaciones estacionales específicas en la columna 5 (*SI*). En términos de letras, $TCSI/TC = SI$; el cual se multiplica por 100.0 para expresar la variación estacional típica en forma de índice.

En el último paso se toma la medida de todos los índices comunes de invierno, de todos los índices de primavera, etcétera; este promedio elimina la mayoría de las fluctuaciones irregulares de las variaciones estacionales específicas, y los cuatro índices resultantes indican el patrón de ventas estacional típico.

AUTOEVALUACIÓN 18-4

En Teton Village, Wyoming, cerca del Grand Teton Park y Yellowstone Park, hay tiendas, restaurantes y moteles; el sitio tiene dos estaciones altas, una en invierno, para esquiar en las pendientes de 10 000 pies de declive, y la otra en verano, para los turistas que visitan los parques. He aquí el número de visitantes (en miles) por trimestre en cinco años, de 2009 a 2013.

	Trimestre			
Año	**Invierno**	**Primavera**	**Verano**	**Otoño**
2009	117.0	80.7	129.6	76.1
2010	118.6	82.5	121.4	77.0
2011	114.0	84.3	119.9	75.0
2012	120.7	79.6	130.7	69.6
2013	125.2	80.2	127.6	72.0

(a) Desarrolle el patrón estacional habitual de Teton Village con el método de la razón con promedio móvil.
(b) Explique el índice habitual de la temporada de invierno.

EJERCICIOS

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

9. Víctor Anderson, propietario de Anderson Belts, Inc., estudia el ausentismo entre sus empleados. Su fuerza laboral es pequeña, de solo cinco empleados, y durante los últimos tres años, de 2011 a 2013, registró el siguiente número de ausencias entre sus empleados, en días, por trimestre.

	Trimestre			
Año	**I**	**II**	**III**	**IV**
2011	4	10	7	3
2012	5	12	9	4
2013	6	16	12	4

Determine el índice estacional habitual de cada trimestre.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

10. Appliance Center vende diversos aparatos domésticos y equipo electrónico; a continuación se muestran las ventas trimestrales de los últimos cuatro años (en millones de dólares).

	Trimestre			
Año	**I**	**II**	**III**	**IV**
2010	5.3	4.1	6.8	6.7
2011	4.8	3.8	5.6	6.8
2012	4.3	3.8	5.7	6.0
2013	5.6	4.6	6.4	5.9

Determine un índice estacional habitual de cada trimestre.

Datos desestacionalizados

OA18-7 Desestacionalizar una serie de tempo utilizando índices estacionales.

Un conjunto de índices habituales es muy útil para ajustar las series de ventas de fluctuaciones estacionales. La serie de ventas resultantes se denomina **ventas desestacionalizadas**, o **estacionalmente ajustadas**. La razón para desestacionalizar la serie de ventas es eliminar las fluctuaciones estacionales de modo que sea posible estudiar la tendencia y el ciclo; para ilustrar el procedimiento, los totales de las ventas trimestrales de Toys International que se muestran en la tabla 18.6 también se registran en la columna 1 de la tabla 18.9.

Para eliminar el efecto de la variación estacional, la cantidad de ventas en cada trimestre (con los efectos de tendencia, cíclicos, irregulares y estacionales) se divide entre el índice estacional de ese trimestre, es decir, *TSCI/S*; por ejemplo, las ventas reales del primer trimestre de 2008 fueron de 6.7 millones de dólares. El índice estacional del trimestre de invierno es 76.5%, usando los resultados del ejemplo de la sección "Determinación de un índice estacional". El índice de 76.5 indica que las ventas del primer trimestre están habitualmente 23.5% por debajo del promedio de un trimestre típico; al dividir las ventas reales de 6.7 millones de dólares entre 76.5, y multiplicar el resul-

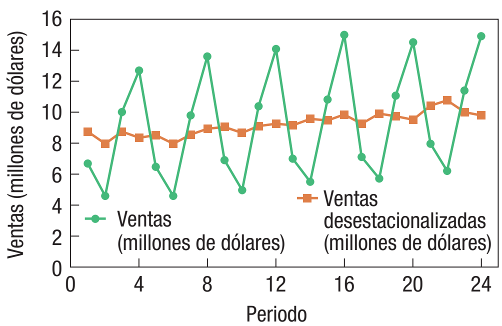
TABLA 18.9 Ventas reales y desestacionalizadas de Toys International

Año	Trimestre	(1) Ventas	(2) Índice estacional	(3) Ventas desestacionalizadas
2008	Invierno	6.7	0.765	8.759
	Primavera	4.6	0.575	8.004
	Verano	10.0	1.141	8.761
	Otoño	12.7	1.519	8.361
2009	Invierno	6.5	0.765	8.498
	Primavera	4.6	0.575	8.004
	Verano	9.8	1.141	8.586
	Otoño	13.6	1.519	8.953
2010	Invierno	6.9	0.765	9.021
	Primavera	5.0	0.575	8.700
	Verano	10.4	1.141	9.112
	Otoño	14.1	1.519	9.283

Año	Trimestre	(1) Ventas	(2) Índice estacional	(3) Ventas desestacionalizadas
2011	Invierno	7.0	0.765	9.151
	Primavera	5.5	0.575	9.570
	Verano	10.8	1.141	9.462
	Otoño	15.0	1.519	9.875
2012	Invierno	7.1	0.765	9.282
	Primavera	5.7	0.575	9.918
	Verano	11.1	1.141	9.725
	Otoño	14.5	1.519	9.546
2013	Invierno	8.0	0.765	10.459
	Primavera	6.2	0.575	10.788
	Verano	11.4	1.141	9.988
	Otoño	14.9	1.519	9.809

Año	Periodo	Ventas (millones de dólares)	Ventas desestacionalizadas (millones de dólares)
2008	1	6.7	8.759
	2	4.6	8.004
	3	10.0	8.761
	4	12.7	8.361
2009	6.5	8.498	
	6	4.6	8.004
	7	9.8	8.586
	8	13.6	8.953
2010	9	6.9	9.021
	10	5.0	8.700
	11	10.4	9.112
	12	14.1	9.283
2011	13	7.0	9.151
	14	5.5	9.570
	15	10.8	9.462
	16	15.0	9.875
2012	17	7.1	9.282
	18	5.7	9.918
	19	11.1	9.725
	20	14.5	9.546
2013	21	8.0	10.459
	22	6.2	10.788
	23	11.4	9.988
	24	14.9	9.809

tado por 100, se obtienen las *ventas desestacionalizadas*, es decir, se elimina el efecto estacional sobre las ventas, del primer trimestre de 2008. Este es 8 758 170 dólares, determinado mediante ($6 700 000/76.5)100. Continúe este proceso con los demás trimestres que se encuentran en la columna 3 de la tabla 18.9, con los resultados reportados en millones de dólares. Como ha eliminado (cancelado) el componente estacional de las ventas trimestrales, la cifra de las ventas desestacionalizadas solo contiene los componentes de tendencia (*T*), cíclica (*C*) e irregular (*I*). Al analizar las ventas desestacionalizadas que se registran en la columna 3 de la tabla 18.9, considere que en las ventas de juguetes se mostró un aumento moderado durante el periodo de seis años. En la gráfica 18.9 se registran tanto las ventas reales como las desestacionalizadas. Es claro que eliminar el factor estacional permite enfocarse en la tendencia general de largo plazo de las ventas; también puede determinar la ecuación de regresión de los datos de la tendencia y con ella proyectar ventas futuras.



GRÁFICA 18.9 Ventas reales y desestacionalizadas de Toys

Uso de datos desestacionalizados para proyección

El procedimiento para identificar la tendencia y los ajustes estacionales se combina para producir proyecciones estacionalmente ajustadas; para identificar la trayectoria, determine la ecuación de la tendencia de mínimos cuadrados en los datos históricos desestacionalizados; luego proyéctela en periodos futuros, y después ajuste sus tendencias para calcular los factores estacionales; esto se aclara mediante el siguiente ejemplo.

EJEMPLO

Toys International quiere proyectar sus ventas de cada trimestre de 2014. Determine la proyección considerando la información que se muestra en la tabla 18.9.

SOLUCIÓN

Los datos desestacionalizados, que se ilustran en la gráfica 18.9, parecen seguir una recta; de aquí que, con base en estos, sea razonable desarrollar una ecuación de tendencia lineal. La ecuación de la tendencia desestacionalizada es:

$$\hat{y} = a + bt$$

donde:

- $\hat{y}$ es el valor de la tendencia estimado de las ventas de Toys International durante el periodo t;
- a es la intersección de la recta de la tendencia en el tiempo 0;
- b es la pendiente de la recta;
- t es el periodo codificado.

El trimestre de invierno de 2008 es el primero, por lo cual se codifica como 1, el trimestre de primavera de 2008 se codifica como 2, etcétera; el último trimestre de 2013 se codifica como 24.

Utilice Excel para encontrar la ecuación de regresión; abajo a la derecha se muestra la salida, en la cual, se incluye un diagrama de dispersión de los periodos de tiempo codificados y las ventas desestacionalizadas, así como la recta de regresión.

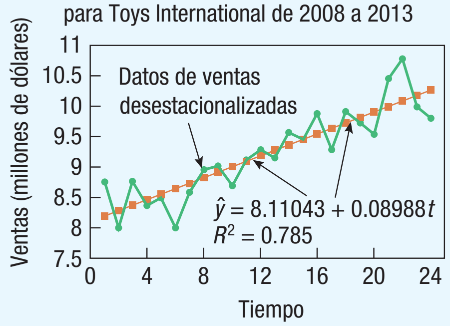
La ecuación de la recta de regresión es:

$$\hat{y} = 8.11043 + 0.08988t$$

La pendiente de la recta de tendencia es 0.08988; esto indica que durante los 24 trimestres las ventas desestacionalizadas aumentaron a una tasa de 0.08988 (millones de dólares) por trimestre, u 88 888 dólares por trimestre. El valor de 8.11043 es la intersección de la recta de tendencia con el eje Y (es decir, para $t = 0$).

El coeficiente de determinación es 0.785; el cual sirve como una indicación del ajuste de los datos. Como esta no es información de la muestra, técnicamente no debería utilizarse R^2 para juzgar una ecuación de regresión; sin embargo, servirá para evaluar de manera rápida el ajuste de los datos de ventas desestacionalizadas. En este caso, como R^2 es un tanto grande, se concluye que las ventas desestacionalizadas de Toys International se explican de manera clara mediante una ecuación de tendencia lineal.

Datos de ventas ajustadas y desestacionalizadas para Toys International de 2008 a 2013



Si se asume que los últimos 24 periodos son un buen indicador de las ventas futuras, utilice la ecuación de la tendencia para estimar las ventas futuras; por ejemplo, el valor de t en el trimestre de invierno de 2014 es 25; por lo tanto, las ventas estimadas de ese periodo son 10.35743, determinadas mediante:

$$\hat{y} = 8.11043 + 0.08988t = 8.11043 + 0.08988(25) = 10.35743$$

Las ventas desestacionalizadas estimadas del trimestre de invierno de 2014 alcanzan 10 357 430 dólares; esta es la proyección de ventas antes de considerar los efectos de la temporada.

Para determinar la proyección de cada trimestre de 2014 se utiliza el mismo procedimiento y una hoja de cálculo de Excel; a continuación se muestra una salida parcial en pantalla de dicho software.

Proyección por trimestre para Toys International, 2014

Trimestre	Tiempo	Ventas estimadas	Índice estacional	Proyección trimestral
Invierno	25	10.35743	0.765	7.92343
Primavera	26	10.44731	0.575	6.00720
Verano	27	10.53719	1.141	12.02293
Otoño	28	10.62707	1.519	16.14252

Ahora que ya se tienen las predicciones de los cuatro trimestres de 2014, las puede ajustar a las temporadas. El índice de un trimestre de invierno es 0.765; por ende, la proyección del trimestre de invierno de 2014 se puede ajustar por temporada mediante 10.35743(0.765) = 7.92343. Las estimaciones de cada trimestre de 2014 aparecen en la columna derecha de la salida en pantalla de Excel; observe cómo los ajustes estacionales aumentan de forma drástica las estimaciones de ventas de los dos últimos trimestres del año.

ESTADÍSTICA EN ACCIÓN

Las proyecciones no siempre son correctas; en realidad, una proyección puede ser solo una mejor suposición respecto de lo que sucederá. ¿Por qué no son correctas las proyecciones? Un experto enumera ocho errores comunes: 1) no examinar con cuidado las suposiciones, 2) experiencia limitada, 3) falta de imaginación, 4) olvido de las restricciones, 5) optimismo excesivo, 6) dependencia en la extrapolación mecánica, 7) cierre prematuro y 8) especificar demasiado.

AUTOEVALUACIÓN 18-5

Westberg Electric Company vende motores eléctricos a clientes en el área de Jamestown, Nueva Jersey. La ecuación de la tendencia mensual, con base en cinco años de datos mensuales, es

$$\hat{y} = 4.4 + 0.5t$$

El factor estacional de enero es 120, y el de febrero, 95. Determine la proyección estacional ajustada de enero y febrero del sexto año.

EJERCICIOS

11. El departamento de planeación de Padget and Kure Shoes, fabricante de una marca exclusiva de zapatos para mujeres, desarrolló la siguiente ecuación de la tendencia, en millones de pares, con base en cinco años de datos trimestrales.

$$\hat{y} = 3.30 + 1.75t$$

En la siguiente tabla se muestran los factores estacionales de cada trimestre.

	Trimestre			
	I	II	III	IV
Índice	110.0	120.0	80.0	90.0

Determine la proyección ajustada por temporada de cada trimestre de los seis años.

12. Team Sports, Inc., vende artículos deportivos a preparatorias y universidades por medio de un catálogo de distribución nacional, y la gerencia de la empresa estima que venderá 2 000 manoplas marca Wilson, modelo A2000, el próximo año. Las ventas desestacionalizadas proyectadas serán iguales en cada trimestre del año próximo; el factor estacional del segundo trimestre es 145. Determine las ventas ajustadas por temporada en el segundo trimestre del próximo año.

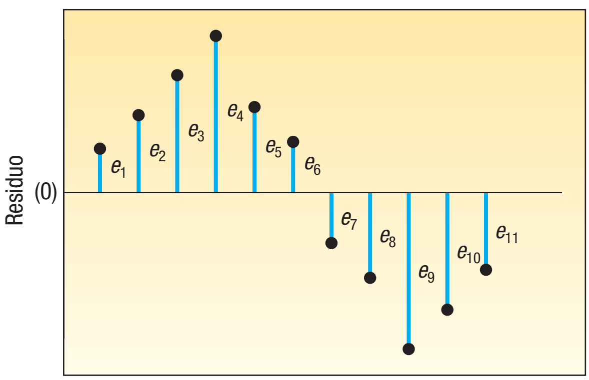
13. Consulte en el ejercicio 9 las ausencias en Anderson Belts, Inc., y utilice los índices estacionales que calculó para determinar las ausencias desestacionalizadas. Determine la ecuación de la tendencia lineal con base en los datos trimestrales de los tres años y proyecte las ausencias de 2014 ajustadas por temporada.

14. Consulte en el ejercicio 10 las ventas de Appliance Center, y utilice los índices estacionales que calculó para determinar las ventas desestacionalizadas. Determine la ecuación de la tendencia lineal de los cuatro años con base en los datos trimestrales y proyecte las ventas de 2014 ajustadas por temporada.

OA18-8 Realizar una prueba de hipótesis de autocorrelación.

El estadístico de Durbin-Watson

Los datos de series de tiempo recopiladas sucesivamente durante un periodo presentan una dificultad particular cuando se utiliza la regresión. Una de las suposiciones que por tradición se emplean en la regresión es que los residuos sucesivos son independientes; lo cual significa que estos no siguen un patrón ni están altamente correlacionados, y no hay corridas largas de residuos positivos o negativos. En la gráfica 18.10, los residuos se muestran a escala en el eje vertical, y los valores $\hat{y}$, a lo largo del eje horizontal; observe que hay "corridas" de residuos arriba y debajo de la recta 0. Si calcula la correlación entre residuos sucesivos, es probable que la correlación sea fuerte.



GRÁFICA 18.10 Residuos correlacionados

Esta condición se denomina **autocorrelación**, o correlación en serie.

AUTOCORRELACIÓN Los residuos sucesivos están correlacionados.

Los residuos sucesivos están correlacionados en datos de series de tiempo debido a que un evento de un periodo influye sobre el evento del siguiente; para explicar esto, el propietario de una mueblería decide obtener una venta especial este mes y gasta una cantidad considerable de dinero en publicidad. Esperaría una correlación entre las ventas y el gasto publicitario, pero no todos los resultados del aumento de publicidad se experimentarán este mes; es probable que una parte de su efecto se observe en el mes siguiente y, en consecuencia, espere una correlación entre los residuos.

La relación de regresión en una serie de tiempo se escribe:

$$Y_t = \alpha + \beta_1 X_t + \varepsilon_t$$

donde el subíndice t sustituye a i para sugerir que los datos se recopilaron en el tiempo.

Si los residuos están correlacionados, se originan problemas cuando se intenta realizar alguna prueba de hipótesis respecto de los coeficientes de regresión; asimismo, un intervalo de confianza o un de proyección, donde se use el error estándar de estimación múltiple, quizá no produzca los resultados correctos.

La autocorrelación, reportada como r, es la fuerza de la asociación entre residuos sucesivos; como la r tiene el mismo significado que el coeficiente de correlación, los valores cercanos a -1.00 o 1.00 indican una asociación fuerte, y los cercanos a 0, que no hay asociación. En lugar de realizar de manera directa una prueba de hipótesis en r, se emplea el **estadístico de Durbin-Watson**.

El estadístico de Durbin-Watson, identificado con la letra d, se calcula primero al determinar los residuos por cada observación; es decir, $e_t = (y_t - \hat{y}_t)$. Luego, d se calcula mediante la siguiente relación.

ESTADÍSTICO DE DURBIN-WATSON

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n}(e_t)^2} \qquad [18.4]$$

Para determinar el numerador de la fórmula [18.4], se "retarda" cada residuo un periodo y luego se eleva al cuadrado la diferencia entre residuos consecutivos; esta maniobra, a la que también se le puede llamar determinación de las diferencias, tome en cuenta la suma de las observaciones de 2, en lugar de 1, hasta n. En el denominador se elevan al cuadrado los residuos y se suman todas las observaciones n.

El valor del estadístico de Durbin-Watson, que varía de 0 a 4, es 2.00 cuando no hay autocorrelación entre los residuos; cuando el valor de d se acerca a 0, indica una autocorrelación positiva. Los valores mayores que 2 indican una autocorrelación negativa. En la práctica, la autocorrelación casi no se presenta; para que esto ocurra, los residuos sucesivos tenderían a ser grandes, pero con signos opuestos.

Para realizar una prueba de autocorrelación, las hipótesis nula y alternativa son:

$$H_0\text{: Sin correlación residual } (\rho = 0)$$
$$H_1\text{: Correlación residual positiva } (\rho > 0)$$

Recuerde, del capítulo anterior, que r se refiere a la correlación muestral, y que α es el coeficiente de correlación entre la población. Los valores críticos de d aparecen en el apéndice B.9; para determinar el valor crítico, necesita α (el nivel de significancia), n (el tamaño muestral) y k (el número de variables independientes). La regla de decisión de la prueba de Durbin-Watson difiere de lo acostumbrado y, como es común, hay un rango de valores donde la hipótesis nula se rechaza y otro donde no se rechaza; sin embargo, también hay un rango donde la prueba no es concluyente, es decir, en el rango no concluyente, la hipótesis nula no se rechaza ni se acepta; para expresarlo de manera más formal:

- Los valores menores que d_l obligan a rechazar la hipótesis nula.
- Los valores mayores que d_u indican que la hipótesis nula no se debe rechazar.
- Los valores de d entre d_l y d_u producen resultados no concluyentes.

El subíndice l se refiere al límite inferior de d, y el subíndice u, al límite superior.

¿Cómo interpretar las diversas decisiones de la prueba de correlación residual? Si no se rechaza la hipótesis nula, se concluye que no hay autocorrelación; si los residuos no están correlacionados, no hay autocorrelación y se cumple con la suposición de regresión; por lo tanto, no habrá problemas con el valor estimado del error estándar de estimación. Si la hipótesis nula se rechaza, se concluye que hay autocorrelación.

El remedio común de la autocorrelación es incluir otra variable de predicción que capture el orden de tiempo; por ejemplo, puede utilizar la raíz cuadrada de *y* en lugar de *y*; esta transformación generará un cambio en la distribución de los residuos. Si el resultado aparece en el rango no concluyente, será necesario recurrir a pruebas más elaboradas, o, de manera conservadora, considerar el rechazo de la hipótesis nula.

Mediante un ejemplo se ilustran los detalles de la prueba de Durbin-Watson y la manera en que se interpretan los resultados.

EJEMPLO

Banner Rocker Company, fabricante y comercializador de mecedoras, diseñó una mecedora especial para adultos mayores, que anuncia en la televisión. El mercado de la silla especial se encuentra en los estados de Carolina del Norte, Carolina del Sur, Florida y Arizona, donde viven muchos adultos mayores y jubilados. El presidente de la compañía estudia la asociación entre sus gastos en publicidad (*X*) y el número de mecedoras vendidas en los últimos 20 meses (*Y*), para lo cual recopiló los siguientes datos. A él le gustaría elaborar un modelo para proyectar las ventas, con base en la cantidad que la empresa gastó en publicidad, pero le preocupa que, como reunió la información durante meses consecutivos, pueda tener problemas con la autocorrelación.

Determine la ecuación de regresión. ¿Es la publicidad un buen factor de proyección de las ventas? Si el propietario gastara un millón de dólares más en publicidad, ¿cuántas sillas adicionales esperaría vender? Investigue la probabilidad de autocorrelación.

Mes	Ventas (en miles)	Publicidad (en millones de dólares)	Mes	Ventas (en miles)	Publicidad (en millones de dólares)
1	153	$5.5	11	169	$6.3
2	156	5.5	12	176	5.9
3	153	5.3	13	176	6.1
4	147	5.5	14	179	6.2
5	159	5.4	15	184	6.2
6	160	5.3	16	181	6.5
7	147	5.5	17	192	6.7
8	147	5.7	18	205	6.9
9	152	5.9	19	215	6.5
10	160	6.2	20	209	6.4

SOLUCIÓN

El primer paso es determinar la ecuación de regresión.

RESUMEN DE SALIDA

Estadísticos de regresión	
Múltiple *R*	0.828
"*R*" cuadrada	0.685
"*R*" cuadrada ajustada	0.668
Error estándar	12.347
Observaciones	20

ANOVA

	gl	*SS*	*MS*	*F*
Regresión	1	5967.731456	5967.73146	39.1430957
Residuo	18	2744.268544	152.459364	
Total	19	8712		

	Coeficientes	*Error estándar*	*Estadístico t*	*Valor p*
Intercepción	−43.80	34.44	−1.27	0.22
Código	35.95	5.75	6.256	0.00

El coeficiente de determinación es 68.5%; por lo tanto, hay una asociación positiva fuerte entre las variables. La conclusión es que, conforme aumenta la cantidad gastada en publicidad, se venderán más mecedoras; por supuesto, esto es lo que se esperaba.

¿Cuántas mecedoras más se venderán si los gastos en publicidad aumentan un millón de dólares? Debe tener cuidado con las unidades de los datos porque las ventas están en miles de mecedoras, y el gasto en publicidad, en millones de dólares. La ecuación de regresión es:

$$\hat{y} = -43.80 + 35.95x$$

Esta ecuación indica que un aumento de 1 en X dará como resultado un aumento de 35.95 en y. En consecuencia, un aumento de un millón de dólares en publicidad aumentará las ventas en 35 950 mecedoras; en otras palabras, costará 27.82 dólares en gastos publicitarios adicionales vender una mecedora, lo cual se determina por $1 000 000/35 950.

¿Qué sucede con el problema potencial de autocorrelación? Muchos paquetes de software calcularán el valor de la prueba de Durbin-Watson y darán salida a los resultados; para comprender la naturaleza de la prueba y ver los detalles de la fórmula [18.4], se utiliza una hoja de cálculo de Excel.

A	B	C	D	E	F	G	H
Mes	Ventas (en miles de dólares) y	Publicidad (en millones de dólares) x	Ventas pronosticadas $\hat{y}$	Residuales $e_t = y - \hat{y}$	Residuales retrasados e_{t-1}	$(e_t - e_{t-1})^2$	e_t^2
1	153	5.5	153.925	−0.925			0.8556
2	156	5.5	153.925	2.075	−0.925	9.000	4.3056
3	153	5.3	146.735	6.265	2.075	17.556	39.2502
4	147	5.5	153.925	−6.925	6.265	173.976	47.9556
5	159	5.4	150.330	8.670	−6.925	243.204	75.1689
6	160	5.3	146.735	13.265	8.670	21.114	175.9602
7	147	5.5	153.925	−6.925	13.265	407.636	47.9556
8	147	5.7	161.115	−14.115	−6.925	51.696	199.2332
9	152	5.9	168.305	−16.305	−14.115	4.796	265.8530
10	160	6.2	179.090	−19.090	−16.305	7.756	364.4281
11	169	6.3	182.685	−13.685	−19.090	29.214	187.2792
12	176	5.9	168.305	7.695	−13.685	457.104	59.2130
13	176	6.1	175.495	0.505	7.695	51.696	0.2550
14	179	6.2	179.090	−0.090	0.505	0.354	0.0081
15	184	6.2	179.090	4.910	−0.090	25.000	24.1081
16	181	6.5	189.875	−8.875	4.910	190.026	78.7656
17	192	6.7	197.095	−5.065	−8.875	14.516	25.6542
18	205	6.9	204.255	0.745	−5.065	33.756	0.5550
19	215	6.5	189.875	25.125	0.745	594.384	631.2656
20	209	6.4	186.280	22.720	25.125	5.784	516.1984
						2338.570	2744.26858

$\sum_{t=1}^{n}(e_t)^2$

$\sum_{t=2}^{n}(e_t - e_{t-1})^2$

Para investigar la probabilidad de autocorrelación es necesario determinar los residuos de cada observación y encontrar los valores ajustados (es decir, $\hat{y}$) en cada uno de los 20 meses; esta información aparece en la cuarta columna (D). Luego se encuentra el residuo, que es la diferencia entre el valor real y los valores ajustados; por lo tanto, en el primer mes:

$$\hat{y} = -43.80 + 35.950x = -43.80 + 35.950(5.5) = 153.925$$
$$e_1 = y_1 - \hat{y}_1 = 153 - 153.925 = -0.925$$

El residuo, reportado en la columna E, es un poco diferente debido al redondeo del software. Observe, en particular, la serie de cinco residuos negativos para los meses 9 a 13. En la columna F los residuos se retrasan un periodo; en la columna G se determina la diferencia entre el residuo actual y el anterior, y se la eleva al cuadrado. Con los valores del software:

$$(e_t - e_{t-1})^2 = (e_2 - e_{2-1})^2 = [2.075 - (-0.925)]^2 = (3.000)^2 = 9.000$$

El resto de los valores de la columna G se determina de igual forma; los de la columna H son los cuadrados de los de la columna E.

$$(e_1)^2 = (-0.925)^2 = 0.8556$$

Para encontrar el valor d se necesitan las sumas de las columnas G y H; las cuales, están resaltadas en color azul en la hoja de cálculo.

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n}(e_t)^2} = \frac{2\ 338.570}{2\ 744.269} = 0.8522$$

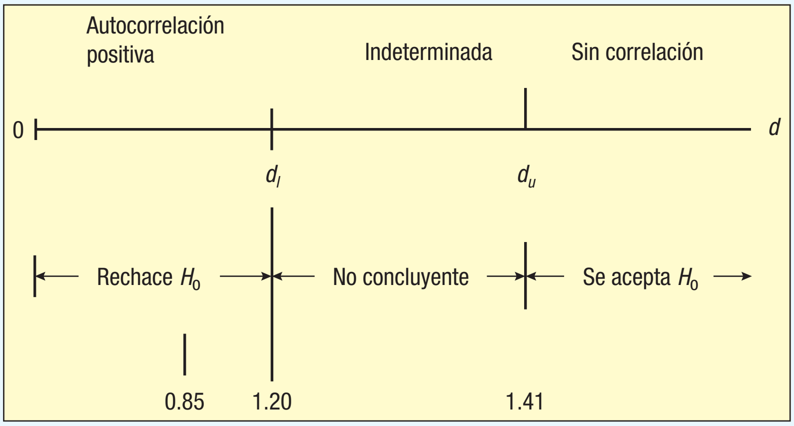
Ahora, para responder si la autocorrelación es significativa, las hipótesis nula y alternativa se formulan como sigue.

H_0: Sin correlación residual

H_1: Correlación residual positiva

El valor crítico de d aparece en el apéndice B.9, del cual se muestra una parte a la izquierda. Hay una variable independiente, por lo que $k = 1$, el nivel de significancia es 0.05 y el tamaño de la muestra, 20. En la tabla 0.05, ahora hay que desplazarse a la columna de $k = 1$ y hasta la fila de 20, donde los valores reportados son $d_l = 1.20$ y $d_u = 1.41$; se rechaza la hipótesis nula si $d < 1.20$ y no se rechaza si $d > 1.41$. No hay una conclusión si d se encuentra entre 1.20 y 1.41.

n \ k	1		2	
	d_l	d_u	d_l	d_u
15	1.08	1.36	0.95	1.54
16	1.10	1.37	0.98	1.54
17	1.13	1.38	1.02	1.54
18	1.16	1.39	1.05	1.53
19	1.18	1.40	1.08	1.53
20	1.20	1.41	1.10	1.54
21	1.22	1.42	1.13	1.54
22	1.24	1.43	1.15	1.54
23	1.26	1.44	1.17	1.54
24	1.27	1.45	1.19	1.55
25	1.29	1.45	1.21	1.55



Puesto que el valor calculado de d es 0.8522, que es menor que d_l, rechace la hipótesis nula y acepte la hipótesis alternativa; por lo tanto, se concluye que los residuos están autocorrelacionados. Se violó una de las suposiciones de regresión. ¿Qué hacer? La existencia de autocorrelación en general significa que el modelo de regresión no se especificó de manera correcta; es probable que necesite agregar una o más variables independientes que tengan algunos efectos en el orden del tiempo sobre la variable dependiente. La variable independiente más simple que aún se debe agregar es una que represente los periodos.

EJERCICIOS

15. Retome el ejercicio 9 del capítulo 14 y la ecuación de regresión para predecir el desempeño en el trabajo
   a. Trace la gráfica de los residuos en el orden de los datos.
   b. Pruebe por autocorrelación con el nivel de significancia 0.05.

16. Considere los datos del ejercicio 10 del capítulo 14 y la ecuación de regresión para predecir las comisiones ganadas
   a. Trace la gráfica de los residuos en el orden de los datos.
   b. Pruebe la autocorrelación con el nivel de significancia 0.01.

RESUMEN DEL CAPÍTULO

I. Una serie de tiempo es un conjunto de datos durante un periodo.
   A. La tendencia es la dirección de largo plazo de la serie de tiempo.
   B. El componente cíclico es la fluctuación por arriba y por debajo de la recta de tendencia de largo plazo durante un periodo mayor.
   C. La variación estacional es el patrón en una serie de tiempo en un año; estos patrones tienden a repetirse año tras año en la mayoría de los negocios.
   D. La variación irregular se divide en dos componentes.

1. Las variaciones episódicas son impredecibles, pero en general se pueden identificar; por ejemplo, una inundación.
2. Las variaciones residuales son de naturaleza aleatoria.

II. Un promedio móvil se utiliza para suavizar la tendencia en una serie de tiempo.

III. La ecuación de la tendencia lineal es $\hat{y} = a + bt$, donde a es la intersección con el eje Y, b es la pendiente de la recta y t es el tiempo codificado.

A. La ecuación de la tendencia se determina mediante el principio de los mínimos cuadrados.

B. Si, en vez de que la tendencia sea lineal, los incrementos tienden a ser un porcentaje constante, los valores y se convierten en logaritmos y con estos se determina la ecuación de mínimos cuadrados.

IV. El factor estacional se estima con el método de la razón con el promedio móvil.

A. El procedimiento de seis pasos produce el índice estacional de cada periodo.

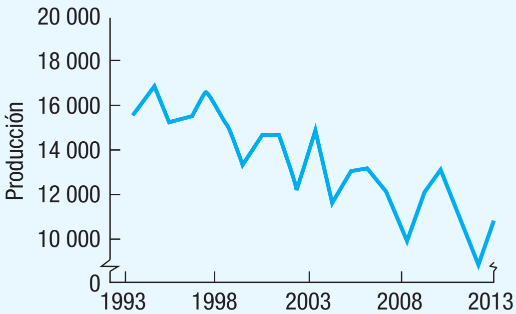
1. En general, los factores estacionales se calculan por mes o trimestre.
2. El factor estacional se utiliza para ajustar las proyecciones, tomando en cuenta los efectos de la temporada.

V. El estadístico de Durbin-Watson [18.4] se utiliza para probar si hay autocorrelación.

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n}(e_t)^2} \qquad \textbf{[18.4]}$$

EJERCICIOS DEL CAPÍTULO

17. Consulte el siguiente diagrama.



a. Estime la ecuación de la tendencia lineal de la serie de producción trazando la recta que mejor ajusta a los datos.

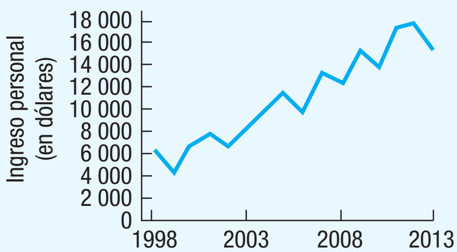
b. ¿Cuál es el decremento anual promedio de la producción?

c. Con base en la ecuación de la tendencia, ¿cuál es la proyección para 2019?

18. Consulte el siguiente diagrama.

a. Estime la ecuación de la tendencia lineal de la serie de ingreso personal.

b. ¿Cuál es el aumento anual promedio del ingreso personal?



19. El movimiento de los activos, excepto inversiones en efectivo y de corto plazo, de RNC Company de 2004 a 2014 es:

2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
1.11	1.28	1.17	1.10	1.06	1.14	1.24	1.33	1.38	1.50	1.65

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

a. Trace la gráfica de los datos.
b. Determine la ecuación de la tendencia de mínimos cuadrados.
c. Calcule los puntos de la recta de tendencia de 2007 y 2012, y trace la recta en la gráfica.
d. Estime el movimiento de los activos en 2019.
e. ¿Cuánto aumentó el movimiento de activos por año, en promedio, de 2004 a 2014?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

20. Las ventas, en miles de millones de dólares, de Keller Overhead Door, Inc., de 2009 a 2014 son:

Año	Ventas	Año	Ventas
2009	7.45	2012	7.94
2010	7.83	2013	7.76
2011	8.07	2014	7.90

a. Trace la gráfica de los datos.
b. Determine la ecuación de la tendencia de mínimos cuadrados.
c. Utilice la ecuación de la tendencia para calcular los puntos de 2011 y 2014. Trace los puntos en la gráfica y la recta de regresión.
d. Estime las ventas netas de 2017.
e. ¿Cuánto aumentaron (o disminuyeron) las ventas por año, en promedio, durante el periodo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

21. El número de empleados, en miles, de Keller Overhead Door, Inc., de 2009 a 2014 es:

Año	Empleados	Año	Empleados
2009	45.6	2012	39.3
2010	42.2	2013	34.0
2011	41.1	2014	30.0

a. Trace la gráfica de los datos.
b. Determine la ecuación de la tendencia de mínimos cuadrados.
c. Con la ecuación de la tendencia, calcule los puntos de 2011 y 2014. Trace los puntos en la gráfica y la recta de regresión.
d. Estime el número de empleados en 2017.
e. ¿En cuánto aumentó (o disminuyó) el número de empleados por año, en promedio, durante el periodo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

22. En la siguiente tabla se muestra el precio de venta de las acciones de PepsiCo, Inc., al cierre de año.

Año	Precio	Año	Precio	Año	Precio	Año	Precio
1990	12.9135	1996	29.0581	2002	42.22	2008	54.77
1991	16.8250	1997	36.0155	2003	46.62	2009	60.80
1992	20.6125	1998	40.6111	2004	52.20	2010	65.33
1993	20.3024	1999	35.0230	2005	59.85	2011	66.35
1994	18.3160	2000	49.5625	2006	62.00	2012	68.43
1995	27.7538	2001	48.68	2007	77.51		

a. Trace la gráfica de los datos.
b. Determine la ecuación de la tendencia de mínimos cuadrados.
c. Calcule los puntos de 1995 y 2000.
d. Calcule el precio de venta en 2016. ¿Parece una estimación razonable con base en los datos históricos?
e. ¿En cuánto aumentó o disminuyó (por año) el precio accionario, en promedio, durante el periodo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

23. Si se graficara la siguiente serie de ventas, aparecería curvilínea, lo cual indicaría que las ventas aumentan a una tasa (porcentaje) anual un tanto constante; en consecuencia, para ajustar las ventas se deberá utilizar una ecuación logarítmica.

Año	Ventas (millones de dólares)	Año	Ventas (millones de dólares)
2004	8.0	2010	39.4
2005	10.4	2011	50.5
2006	13.5	2012	65.0
2007	17.6	2013	84.1
2008	22.8	2014	109.0
2009	29.3		

a. Determine la ecuación logarítmica.
b. Determine las coordenadas de los puntos de la recta logarítmica de 2005 y 2011.
c. ¿Cuál es el aumento porcentual anual de las ventas, en promedio, durante el periodo de 2004 a 2014?
d. Con base en la ecuación, ¿cuáles son las ventas estimadas para 2015?

24. A la derecha se muestran las cantidades que gasta en publicidad (millones de dólares) una empresa grande de 2004 a 2014.
a. Determine la ecuación de la tendencia logarítmica.
b. Estime los gastos en publicidad en 2017.
c. ¿Cuál es el aumento porcentual anual del gasto en publicidad durante el periodo?

Año	Cantidad	Año	Cantidad
2004	88.1	2010	132.6
2005	94.7	2011	141.9
2006	102.1	2012	150.9
2007	109.8	2013	157.9
2008	118.1	2014	162.6
2009	125.6		

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

25. A continuación se muestran los precios de venta de las acciones de Oracle, Inc., al cierre de año desde 1990 hasta 2012.

Año	Precio	Año	Precio	Año	Precio	Año	Precio
1990	0.1944	1996	4.6388	2002	10.80	2008	17.73
1991	0.3580	1997	3.7188	2003	13.23	2009	24.53
1992	0.7006	1998	7.1875	2004	13.72	2010	31.30
1993	1.4197	1999	28.0156	2005	12.21	2011	26.33
1994	2.1790	2000	29.0625	2006	19.11	2012	34.08
1995	3.1389	2001	13.81	2007	20.23		

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

a. Trace la gráfica de los datos.
b. Determine la ecuación de la tendencia de mínimos cuadrados. Utilice el precio accionario actual y el logaritmo del precio. ¿Cuál parece producir una proyección más precisa?
c. Utilizando la ecuación con el logaritmo del precio, calcule los puntos de los años de 1993 y 1998.
d. Utilizando la ecuación con el logaritmo del precio, estime el precio de venta en 2015. ¿Parece una estimación razonable con base en los datos históricos?
e. Utilizando la ecuación con el logaritmo del precio, ¿cuánto aumentó o disminuyó el precio accionario (por año), en promedio, durante el periodo?

26. A continuación se muestra la producción de Reliable Manufacturing Company de 2012 y parte de 2013.

Mes	Producción en 2012 (miles)	Producción en 2013 (miles)	Mes	Producción en 2012 (miles)	Producción en 2013 (miles)
Enero	6	7	Julio	3	4
Febrero	7	9	Agosto	5	
Marzo	12	14	Septiembre	14	
Abril	8	9	Octubre	6	
Mayo	4	5	Noviembre	7	
Junio	3	4	Diciembre	6	

a. Con el método de razón con el promedio móvil, determine los índices específicos estacionales de julio, agosto y septiembre de 2012.
b. Suponga que los índices específicos estacionales que se registran en la siguiente tabla son correctos; inserte en esta los índices específicos estacionales que calculó en el punto a (julio, agosto y septiembre de 2012) y determine los 12 índices estacionales habituales.

Año	Ene.	Feb.	Mar.	Abr.	Mayo	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2012							?	?	?	92.1	106.5	92.9
2013	88.9	102.9	178.9	118.2	60.1	43.1	44.0	74.0	200.9	90.0	101.9	90.9
2014	87.6	103.7	170.2	125.9	59.4	48.6	44.2	77.2	196.5	89.6	113.2	80.6
2015	79.8	105.6	165.8	124.7	62.1	41.7	48.2	72.1	203.6	80.2	103.0	94.2
2016	89.0	112.1	182.9	115.1	57.6	56.9						

c. Interprete el índice estacional habitual.

27. Las ventas de Andre's Boutique en 2012 y parte de 2013 son:

Mes	Ventas en 2012 (miles)	Ventas en 2013 (miles)	Mes	Ventas en 2012 (miles)	Ventas en 2013 (miles)
Enero	78	65	Julio	81	65
Febrero	72	60	Agosto	85	61
Marzo	80	72	Septiembre	90	75
Abril	110	97	Octubre	98	
Mayo	92	86	Noviembre	115	
Junio	86	72	Diciembre	130	

a. Con el método de la razón con promedio móvil, determine los índices estacionales específicos de julio, agosto, septiembre y octubre de 2012.
b. Suponga que los índices estacionales específicos que se muestran en la siguiente tabla son correctos; inserte en esta los índices estacionales específicos que calculó en el inciso punto a (julio, agosto, septiembre y octubre de 2012) y determine los 12 índices estacionales habituales.

Año	Ene.	Feb.	Mar.	Abr.	Mayo	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2012							?	?	?	?	123.6	150.9
2013	83.9	77.6	86.1	118.7	99.7	92.0	87.0	91.4	97.3	105.4	124.9	140.1
2014	86.7	72.9	86.2	121.3	96.6	92.0	85.5	93.6	98.2	103.2	126.1	141.7
2015	85.6	65.8	89.2	125.6	99.6	94.4	88.9	90.2	100.2	102.7	121.6	139.6
2016	77.3	81.2	85.8	115.7	100.3	89.7						

c. Interprete el índice estacional típico.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

28. La producción trimestral de madera de pino, en millones de pies-tabla, de Northwest Lumber para 2010 a 2014 es:

	Trimestre			
Año	Invierno	Primavera	Verano	Otoño
2010	7.8	10.2	14.7	9.3
2011	6.9	11.6	17.5	9.3
2012	8.9	9.7	15.3	10.1
2013	10.7	12.4	16.8	10.7
2014	9.2	13.6	17.1	10.3

a. Determine el patrón estacional habitual de los datos de la producción con el método de razón con promedio móvil.
a. Interprete el patrón.
b. Desestacionalice los datos y determine la ecuación de la tendencia lineal.
c. Proyecte la producción estacionalmente ajustada de los cuatro trimestres de 2015.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

29. Work Gloves, Corp., estudia sus ventas trimestrales de Toughie, el tipo de guantes más durables que produce; los números de pares producidos (en miles) por trimestre para 2009 a 2014 son:

	Trimestre			
Año	I Ene.-Mar.	II Abr.-Jun.	III Jul.-Sep.	IV Oct.-Dic.
2009	142	312	488	208
2010	146	318	512	212
2011	160	330	602	187
2012	158	338	572	176
2013	162	380	563	200
2014	162	362	587	205

a. Con el método de la razón con promedio móvil, determine los cuatro índices trimestrales habituales.
b. Interprete el patrón estacional típico.

30. Las ventas de material para techos, por trimestre, de 2007 a 2013 de Carolina Home Construction, Inc., se muestran en la siguiente tabla (en miles de dólares).

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Trimestre			
Año	**I**	**II**	**III**	**IV**
2007	210	180	60	246
2008	214	216	82	230
2009	246	228	91	280
2010	258	250	113	298
2011	279	267	116	304
2012	302	290	114	310
2013	321	291	120	320

a. Determine los patrones estacionales habituales de las ventas con el método de la razón con promedio móvil.
b. Desestacionalice los datos y determine la ecuación de la tendencia.
c. Proyecte las ventas de 2014 utilizando la ecuación de la tendencia y ajuste estacionalmente cada trimestre.

31. Blueberry Farms Golf and Fish Club de Hilton Head, Carolina del Sur, quiere encontrar los índices estacionales mensuales del juego en paquete, juego sin paquete y juego total. El primero se refiere a los golfistas que visitan el área como parte de un paquete para jugar golf; en general, este incluye las tarifas del *green*, carrito, alojamiento, servicio al cuarto y alimentos; y el campo gana un porcentaje de este total. Cuando es *sin paquete* solo se incluye el juego de los residentes locales y visitantes en el área que deseen hacerlo. Los siguientes datos se registraron en julio de 2007 y en estos se reportan los juegos en paquete y sin paquete por mes, así como la cantidad total, en miles de dólares.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Año	Mes	Paquete	Local	Total	Año	Mes	Paquete	Local	Total
2010	Julio	$ 18.36	$43.44	$ 61.80	2012	Enero	30.60	9.48	40.08
	Agosto	28.62	56.76	85.38		Febrero	63.54	30.96	94.50
	Septiembre	101.34	34.44	135.78		Marzo	167.67	47.64	215.31
	Octubre	182.70	38.40	221.10		Abril	299.97	59.40	359.37
	Noviembre	54.72	44.88	99.60		Mayo	173.61	40.56	214.17
	Diciembre	36.36	12.24	48.60		Junio	64.98	63.96	128.94
2011	Enero	25.20	9.36	34.56		Julio	25.56	67.20	92.76
	Febrero	67.50	25.80	93.30		Agosto	31.14	52.20	83.34
	Marzo	179.37	34.44	213.81		Septiembre	81.09	37.44	118.53
	Abril	267.66	34.32	301.98		Octubre	213.66	62.52	276.18
	Mayo	179.73	40.80	220.53		Noviembre	96.30	35.04	131.34
	Junio	63.18	40.80	103.98		Diciembre	16.20	33.24	49.44
	Julio	16.20	77.88	94.08	2013	Enero	26.46	15.96	42.42
	Agosto	23.04	76.20	99.24		Febrero	72.27	35.28	107.55
	Septiembre	102.33	42.96	145.29		Marzo	131.67	46.44	178.11
	Octubre	224.37	51.36	275.73		Abril	293.40	67.56	360.96
	Noviembre	65.16	25.56	90.72		Mayo	158.94	59.40	218.34
	Diciembre	22.14	15.96	38.10		Junio	79.38	60.60	139.98

Con software estadístico:
a. Determine el índice estacional de cada mes de las ventas de los paquetes. ¿Qué observa en el transcurso del periodo?
b. Desarrolle un índice estacional de cada mes de las ventas sin paquete. ¿Qué observa en el transcurso del periodo?
c. Elabore un índice estacional de cada mes de las ventas totales. ¿Qué observa en el transcurso del periodo?
d. Compare los índices de las ventas de paquetes, ventas sin paquete y ventas totales. ¿Son iguales los meses más ocupados?

32. En la tabla de la página siguiente se muestran los números de jubilados que reciben beneficios del State Teachers Retirement System de Ohio de 1991 a 2012.
a. Trace la gráfica de los datos.
b. Determine la ecuación de tendencia de mínimos cuadrados. Utilice una ecuación lineal.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Año	Servicio	Año	Servicio	Año	Servicio	Año	Servicio
1991	58 436	1997	72 601	2003	89 257	2008	106 099
1992	59 994	1998	75 482	2004	92 574	2009	109 031
1993	61 515	1999	78 341	2005	95 843	2010	112 483
1994	63 182	2000	81 111	2006	99 248	2011	117 138
1995	67 989	2001	83 918	2007	102 771	2012	112 136
1996	70 448	2002	86 666				

c. Calcule los puntos de 2000 y 2010.
d. Estime el número de jubilados que recibirán beneficios en 2015. ¿Parece razonable la estimación con base en los datos históricos?
e. ¿Cuánto aumentó o disminuyó el número de jubilados (por año), en promedio, durante el periodo?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

33. Ray Anderson, propietario de Anderson Ski Lodge, firma que opera en el norte de Nueva York, tiene interés en proyectar el número de visitantes del próximo año. Él dispone de los siguientes datos, por trimestre, desde el primero de 2007 al cuarto de 2013. Elabore el índice estacional de cada trimestre. ¿Cuántos visitantes esperaría Ray en cada trimestre de 2014, si proyecta que en 2013 el número total de visitantes aumentará 10%? Determine la ecuación de tendencia, proyecte el número de visitantes de 2011 y ajuste estacionalmente la proyección. ¿Qué proyección elegiría?

Año	Trimestre	Visitantes	Año	Trimestre	Visitantes
2007	I	86	2011	I	188
	II	62		II	172
	III	28		III	128
	IV	94		IV	198
2008	I	106	2012	I	208
	II	82		II	202
	III	48		III	154
	IV	114		IV	220
2009	I	140	2013	I	246
	II	120		II	240
	III	82		III	190
	IV	154		IV	252
2010	I	162			
	II	140			
	III	100			
	IV	174			

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

34. Las inscripciones en la Facultad de Administración de Midwestern University, por trimestre, desde 2009 a la primavera de 2013 son:

	Trimestre			
Año	Invierno	Primavera	Verano	Otoño
2009	2 033	1 871	714	2 318
2010	2 174	2 069	840	2 413
2011	2 370	2 254	927	2 704
2012	2 625	2 478	1 136	3 001
2013	2 803	2 668	—	—

Con el método de la razón con promedio móvil:
a. Determine los cuatro índices trimestrales.
b. Interprete el patrón trimestral de las inscripciones. ¿Le sorprende la variación estacional?
c. Calcule la ecuación de tendencia y proyecte las inscripciones para 2011 por trimestre.

35. Al comienzo de este capítulo, se presentó una gráfica que muestra el precio por galón de gasolina de 1990 a 2012. En la página siguiente se dan los datos para cada año.

Determine la recta de tendencia utilizando ecuaciones lineales y no lineales. ¿Qué sugeriría y por qué? Basándose en su recomendación de la mejor ecuación, ¿qué costo por galón estimaría para 2013 y 2014?

Año	Costo/galón	Año	Costo/galón
1990	$1.30	2002	$1.35
1991	1.10	2003	1.56
1992	1.09	2004	1.85
1993	1.07	2005	2.27
1994	1.08	2006	2.57
1995	1.11	2007	2.80
1996	1.20	2008	3.25
1997	1.20	2009	2.35
1998	1.03	2010	2.78
1999	1.14	2011	3.52
2000	1.48	2012	3.62
2001	1.42		

Utilice la siguiente información, obtenida de los reportes anuales de Home Depot para resolver los ejercicios 36, 37 y 38; necesitará la ayuda de un paquete de software como Excel y tal vez un paquete adicional como MegaStat; para los años 1993 a 2012, en los datos se incluye una variable de tiempo codificado (1 a 20), el número de asociados en miles, las ventas netas en millones de dólares, la cantidad media de dólares por transacción, y el Índice de Precios al Consumidor (IPC) para cada año.

Año	Tiempo	Asociados (miles)	Ventas netas (millones de dólares)	Cantidad media por transacción	IPC
1993	1	50.6	9 239	39.13	144.500
1994	2	67.3	12 477	41.29	148.200
1995	3	80.8	15 470	41.78	152.400
1996	4	98.1	19 535	42.09	156.900
1997	5	124.4	24 156	43.63	160.500
1998	6	156.7	30 219	45.05	163.000
1999	7	201.4	38 454	47.87	166.000
2000	8	227.3	45 738	48.65	172.200
2001	9	256.3	53 553	48.64	177.100
2002	10	280.9	58 247	49.43	179.900
2003	11	298.8	64 816	51.15	184.000
2004	12	323.1	73 094	54.89	188.900
2005	13	344.8	81 511	57.98	195.300
2006	14	364.4	79 022	58.90	201.600
2007	15	331.0	77 349	57.48	207.342
2008	16	322.0	71 288	55.61	215.303
2009	17	317.0	66 176	51.76	214.537
2010	18	321.0	67 997	51.93	218.056
2011	19	331.0	70 395	53.28	224.939
2012	20	340.0	74 754	54.89	229.594

36. Desarrolle una ecuación de tendencia para las ventas netas (millones de dólares) para Home Depot. Considere tanto una tendencia lineal como una no lineal ¿Cuál elegiría y por qué? Dada la ecuación de tendencia que seleccionó, proyecte las ventas netas para 2014 y 2015.

37. Considere la variable cantidad media por transacción en los datos de Home Depot presentados arriba; esta indica, por ejemplo, que el cliente promedio gastó 39.13 dólares en bienes durante una visita a la tienda en 1993; para 2012, esta cantidad aumentó a 54.89 dólares. Durante el mismo periodo, el IPC reportado por el Buró de Estadísticas Laborales aumentó de 144.5 a 229.594. Convierta el IPC a la base de 1993 como se describió en el capítulo previo, y convierta la cantidad media por transacción a dólares de 1993. Desarrolle una ecuación de tendencia lineal para la constante “dólares de 1993” de la cantidad media por transacción. ¿Es razonable que la tendencia sea lineal? ¿Es posible concluir que el valor de la cantidad que el cliente gasta es menor?

38. Utilice las variables “cantidad media por transacción” y “número de asociados” para proyectar las ventas netas. ¿Ambas variables independientes predicen de manera razonable las ventas netas? (*Sugerencia:* ¿Cuál es el valor R^2? ¿Es grande? Estudie los valores p de cada variable independiente. ¿Este valor es menos de 0.05 en cada caso?) Como todas estas variables están asociadas con el tiempo, puede haber una autocorrelación ahí; por lo tanto, realice la prueba de hipótesis adecuada para determinar si existe una correlación. ¿Cuál es su conclusión? De nuevo, use el nivel de significancia 0.05.

EJERCICIOS DE LA BASE DE DATOS

(Los datos para este ejercicio están disponibles en el sitio web del libro: **www.mhhe.com/uni/lind_ae16e**).

39. Consulte los datos sobre Baseball 2012 que contienen información de la temporada 2012 de las Ligas Mayores de Béisbol y, además, incluye el salario medio por jugador desde 1989. Trace la gráfica de la información, elabore una ecuación de tendencia lineal y escriba un reporte breve de sus averiguaciones.

REPASO DE LOS CAPÍTULOS 17 y 18

En el capítulo 17 se presentaron los *números índices*; los cuales describen el cambio relativo de valor de un periodo, denominado periodo base, a otro denominado periodo dado; en realidad es un porcentaje, pero, en general, el signo de porcentaje se omite. Los índices se utilizan para comparar el cambio en series desiguales en el tiempo; por ejemplo, una compañía podría querer comparar el cambio en las ventas con el cambio en el número de vendedores empleados durante el mismo periodo. Una comparación directa no es significativa porque las unidades de un conjunto de datos son dólares, y las del otro, personas. Los números índice también facilitan la comparación de valores muy grandes, donde la cantidad de cambio en los actuales es muy grande y, por lo tanto, es difícil de interpretar.

Hay dos tipos de índices de precios. Un índice de precios no ponderado no considera las cantidades; para formarlo se divide el valor del periodo base entre el periodo actual (también denominado periodo dado) y se reporta el cambio porcentual; por lo tanto, si las ventas fueron de 12 000 000 de dólares en 2008 y de 18 600 000 dólares en 2014, el índice de precios sin ponderar simple de 2014 es:

$$p = \frac{p_t}{p_0}(100) = \frac{\$18\,600\,000}{\$12\,000\,000}(100) = 155.0$$

Se concluye que las ventas aumentaron 55% durante el periodo de seis años.

Un *índice de precios ponderado considera las cantidades*. El índice ponderado más común es el *índice de precios de Laspeyres*; en él se utilizan las cantidades del periodo base como ponderaciones para comparar cambios de precios. Se calcula al multiplicar las cantidades del periodo base por el precio del periodo base por cada producto considerado, y se suma el total; este resultado es el denominador de la fracción. El numerador de la fracción es el producto de las cantidades del periodo base por el precio actual; por ejemplo, una tienda de aparatos electrónicos vendió 50 computadoras a 1 000 dólares y 200 reproductores de DVD a 150 dólares cada uno en el año 2008; en 2014, la misma tienda vendió 60 computadoras a 1 200 dólares y 230 reproductores de DVD a 175 dólares. El índice de precios de Laspeyres es:

$$p = \frac{\Sigma p_t q_0}{\Sigma p_0 q_0}(100) = \frac{\$1\,200 \times 50 + \$175 \times 200}{\$1\,000 \times 50 + \$150 \times 200}(100)$$

$$= \frac{\$95\,000}{\$80\,000}(100) = 118.75$$

Observe que se utilizan las mismas cantidades del periodo base como ponderaciones tanto en el numerador como en el denominador. El índice indica 18.75% de aumento del valor de las ventas durante el periodo de seis años.

El índice de uso y reporte más frecuente es el *Índice de Precios al Consumidor (IPC)*; el cual, es del tipo de Laspeyres. Lo elabora cada mes el U.S. Department of Labor para reportar la tasa de inflación de los precios de bienes y servicios en Estados Unidos. El periodo base actual es 1982-1984.

En el capítulo 18 se estudiaron las series de tiempo y los pronósticos (proyección). Una *serie de tiempo* es un conjunto de datos durante un periodo; las ganancias por cada una de las acciones comunes de General Electric durante los últimos diez años es un ejemplo de una serie de tiempo. Una serie de tiempo consta de cuatro componentes: tendencia, efectos cíclicos, efectos estacionales y efectos irregulares.

La *tendencia* es la dirección de largo plazo de la serie de tiempo; puede aumentar o disminuir.

El *componente* cíclico es la fluctuación por arriba y por debajo de la recta de tendencia durante un periodo de varios años; por ejemplo, los ciclos económicos. La mayoría de los negocios cambian entre periodos de expansión relativa y reducción durante un ciclo de varios años.

La *variación estacional* es el patrón recurrente de la serie de tiempo en un año. El consumo de muchos productos y servicios se realiza por temporadas; por ejemplo, las casas de playa a lo largo de la Costa del Golfo casi no se rentan durante el invierno, y los albergues de ski en Wyoming no se utilizan en los meses de verano; de aquí que la renta de propiedades frente a la playa y los albergues de ski sean estacionales.

El *componente irregular* incluye cualesquiera eventos impredecibles; en otras palabras, incluye eventos que no se pueden prever. Hay dos tipos de componentes irregulares; por ejemplo, las variaciones episódicas son impredecibles, pero en general se pueden identificar, como la inundación de Nashville en el verano de 2010. La variación residual es de naturaleza aleatoria y no se puede predecir ni identificar.

La tendencia lineal de una serie de tiempo se obtiene por medio de la ecuación $\hat{y} = a + bt$, donde $\hat{y}$ es el valor estimado de la tendencia, *a* es la intersección con el eje *Y*, *b* es la pendiente de la recta de tendencia (la tasa de cambio) y *t* se refiere a los valores codificados de los periodos. El método de mínimos cuadrados que se describió en el capítulo 13 se emplea para determinar la recta de la tendencia. Con frecuencia, la autocorrelación (es decir, los valores sucesivos de la serie de tiempo están correlacionados) es un problema cuando se utiliza la ecuación de tendencia.

PROBLEMAS

1. En la siguiente tabla se registran las ventas netas (miles de millones de euros) de Adidas Group de 2008 a 2012.

Año	Ventas netas
2008	10 799
2009	10 381
2010	11 990
2011	13 332
2012	14 883

a. Determine el índice de 2012, con 2008 como periodo base.
b. Utilice el periodo 2008 a 2010 como periodo base y encuentre el índice de 2012.
c. Con 2008 como año base, utilice el método de mínimos cuadrados para encontrar la ecuación de tendencia. ¿Cuál es el ingreso consolidado estimado para 2014? ¿Cuál es la tasa de incremento por año?

2. En la siguiente tabla se registra la tasa de desempleo y la fuerza laboral disponible en tres condados en el noroeste de Pennsylvania en mayo de 2010 y abril de 2013.

	Mayo 2010		Abril 2013	
Condado	Fuerza laboral	Porcentaje de desempleo	Fuerza laboral	Porcentaje de desempleo
Erie	141 800	10.0	140 700	7.5
Warren	22 700	4.7	21 200	6.6
McKean	22 200	4.9	22 000	7.7

a. En mayo de 2010, el índice nacional de desempleo en Estados Unidos fue de 9.7%. Calcule, para ese mes, el índice simple del promedio de desempleo de la región utilizando el índice nacional de desempleo como base. Interprete el índice simple promedio.
b. En abril de 2013, el índice nacional de desempleo de Estados Unidos fue de 7.6%. Calcule, para ese mes, el índice simple del promedio de desempleo de la región utilizando el índice nacional de desempleo como base. Interprete el índice simple promedio.
c. Utilice los datos de esta región del noroeste de Pennsylvania para elaborar un índice ponderado de desempleo con el método de Laspeyres. Emplee la información de mayo de 2010 como periodo base. Interprete el índice.

3. Con base en cinco años de datos mensuales (de enero de 2008 a diciembre de 2012), la ecuación de tendencia de una compañía pequeña es $\hat{y} = 3.5 + 0.7t$. El índice estacional de enero es 120, y el de junio, 90. ¿Cuál es la proyección de las ventas ajustadas por temporada de enero de 2013 y junio de 2013?

TEST DE PRÁCTICAS

Parte 1: Objetivo

1. Para calcular un índice, el periodo base es el ________ (numerador, denominador, cualquiera de los dos, siempre 100). **1.** ____________________
2. Un número que mide el cambio relativo de un periodo a otro se denomina ________. **2.** ____________________
3. Un índice ponderado considera tanto el precio como ________. **3.** ____________________
4. Un índice de Laspeyres utiliza cantidades ________ tanto en el numerador como en el denominador (elija una: periodo base, periodo dado, las más antiguas, las más recientes). **4.** ____________________
5. El periodo base actual del Índice de Precios al Consumidor es ________. **5.** ____________________

6. La dirección a largo plazo de una serie de tiempo se denomina ________. 6. ________________
7. Uno de los métodos que se usan para suavizar la tendencia en una serie de tiempo es ________. 7. ________________
8. Cuando residuos sucesivos están correlacionados, la condición se denomina ________. 8. ________________
9. La variación irregular en una serie de tiempo, que es de naturaleza aleatoria, se denomina ________. 9. ________________
10. En un promedio móvil de tres años, las ponderaciones dadas a cada periodo son (las mismas, el año más lejano tiene más peso, el año más lejano tiene el menor peso). 10. ________________

Parte 2: Problemas

1. A continuación se reportan las ventas de Roberta's Ice Cream Stand de 2009 a 2013.

Año	Ventas
2009	$130 000
2010	145 000
2011	120 000
2012	170 000
2013	190 000

a. Calcule el índice simple de cada año, usando 2009 como periodo base.
b. Calcule el índice simple de cada año, usando 2009-2010 como periodo base.

2. A continuación se muestran el precio y la cantidad de diversos artículos de golf que compraron algunos miembros de la liga varonil de este deporte en el Indigo Creek Golf and Tennis Club durante 2009 y 2013.

	2009		2013	
	Precio	Cantidad	Precio	Cantidad
Driver	$250.00	5	$275.00	6
Putter	60.00	12	75.00	10
Fierros	700.00	3	750.00	4

a. Determine el índice simple agregado del precio, con 2006 como periodo base.
b. Determine el índice Laspeyres del precio.
c. Determine el índice de Paasche del precio.
d. Determine el índice de valor.

3. La ecuación lineal de tendencia mensual de la Hoopes ABC Beverage Store es:

$$\hat{y} = 5.50 + 1.25t$$

La ecuación se basa en cuatro años de datos mensuales, y se reporta en miles de dólares. El índice de enero es 105.0 y, el de febrero, 98.3. Determine la proyección estacionalmente ajustada de enero y febrero en el quinto año.

Control estadístico del proceso y administración de calidad

19

CADA DÍA, UN FABRICANTE de bicicletas selecciona al azar 10 cuadros y realiza pruebas para detectar defectos; el número de cuadros defectuosos determinado durante los últimos 14 días es 3, 2, 1, 3, 2, 2, 8, 2, 0, 3, 5, 2, 0 y 4. Elabore el diagrama de control de este proceso y comente si está "bajo control" (vea el ejercicio 11 y el OA19-6).

OBJETIVOS DE APRENDIZAJE

Al terminar este capítulo, usted será capaz de:

OA19-1 Explicar la función del control de calidad en operaciones de producción y servicio.

OA19-2 Definir las dos causas de la variación de los procesos y explicar cómo se usan para monitorear la calidad.

OA19-3 Explicar el uso de diagramas para investigar las fuentes de la variación de los procesos.

OA19-4 Calcular los límites de control de gráficas medias y de rango para una medición variable de calidad.

OA19-5 Evaluar los diagramas de control para determinar si un proceso está fuera de control.

OA19-6 Calcular los límites de los diagramas de control para un atributo con una medida de calidad variable.

OA19-7 Analizar el muestreo de aceptación.

Introducción

A lo largo de este libro se han presentado muchas aplicaciones de las pruebas de hipótesis. En el capítulo 10 se describieron métodos para probar una hipótesis respecto de un valor único de la población; en el capítulo 11 se abordaron los métodos para probar una hipótesis acerca de dos poblaciones. En este se presenta otra aplicación, distinta de la prueba de hipótesis, denominada **control estadístico del proceso** (*statistical process control*, **SPC**).

El control estadístico del proceso está conformado por un grupo de estrategias, técnicas y acciones de una organización para asegurar que fabrica un producto o proporciona un servicio de calidad. El SPC se inicia en la etapa de planeación del producto o servicio, cuando se especifican los atributos de ambos, y continúa en la etapa de producción. Cada atributo durante el proceso contribuye a incrementar la calidad general del producto; para un uso eficaz del control de calidad, se desarrollan atributos y especificaciones mensurables con las cuales se comparan los atributos reales del producto o servicio.

OA19-1
Explicar la función del control de calidad en operaciones de producción y servicio.

Breve historia del control de calidad

Antes del siglo XX, la industria estadounidense se caracterizaba por tiendas pequeñas que hacían productos relativamente simples, como velas o muebles; en estas tiendas, el trabajador era un artesano totalmente responsable de la calidad del trabajo, y podía asegurar la calidad mediante la selección personal de los materiales, su habilidad en la fabricación, colocación y ajuste selectivos.

A principios del siglo XX comenzaron a surgir las fábricas, donde se alineaban personas con capacitación limitada en largas líneas de ensamblado. Los productos se hicieron mucho más complejos, y el trabajador ya no tenía el control total de la calidad del producto. Un grupo de personal semiprofesional, en general llamado departamento de inspección, se responsabilizaba de la calidad del producto; en general, la responsabilidad por la calidad se lograba mediante una inspección de todas las características importantes. Si había alguna discrepancia, el supervisor del departamento de manufactura se encargaba del problema; en esencia, la calidad se lograba "con la inspección de la calidad del producto".

Durante la década de 1920, el doctor Walter A. Shewhart, de Bell Telephone Laboratories, desarrolló los conceptos del control estadístico de la calidad, e introdujo la idea de "controlar" la calidad de un producto a medida que se fabricaba, en lugar de inspeccionar la calidad del producto terminado; para hacerlo, Shewhart desarrolló técnicas de representación para controlar las operaciones de la manufactura en proceso. Además, introdujo el concepto de inspección estadística de la muestra para estimar la calidad de un producto a medida que se fabricaba; mediante este enfoque se reemplazó el método anterior de inspeccionar cada parte después de finalizar el proceso productivo.

El reconocimiento pleno del control estadístico de la calidad ocurrió durante la Segunda Guerra Mundial; debido a la necesidad de producir artículos bélicos en masa, como visores de bombardeo, radares precisos y demás equipo electrónico con el menor costo posible, se aceleró el uso del muestreo estadístico y de las tablas de control de calidad. Desde entonces, estas técnicas estadísticas se refinaron y perfeccionaron. El uso de computadoras también amplió la aplicación de dichas técnicas.

La Segunda Guerra Mundial destruyó virtualmente la capacidad de producción japonesa; sin embargo, en lugar de rediseñar los métodos de producción anteriores, los japoneses consiguieron la ayuda del ahora fallecido doctor W. Edwards Deming, del Departamento de Agricultura de Estados Unidos, para elaborar un plan global. En una serie de seminarios con planificadores japoneses, destacó la filosofía que en la actualidad se conoce como los 14 puntos de Deming; estos se presentan en seguida. El doctor Edwards recalcó que la calidad tiene su origen en la mejora del proceso, no en la inspección, y que son los clientes quienes determinan la calidad. El fabricante debe adquirir capacidad por medio de una investigación de mercado y al anticipar las necesidades de los clientes; además, la gerencia general tiene la responsabilidad de hacer mejoras de largo plazo. Otro de sus puntos, al que los japoneses respaldan en gran medida, es que cada miembro de la compañía debe contribuir a la mejora de largo plazo; para lograr este objetivo, es necesario implementar una educación y capacitación continuas.

Deming tenía algunas ideas que no concordaban con las filosofías contemporáneas de la administración en Estados Unidos. Dos áreas donde sus ideas diferían de la perspectiva administrativa en ese país fueron las cuotas de producción y las clasificaciones de excelencia; afirmó que ambas prácticas, comunes en ese país, no eran productivas y se debían eliminar. También señaló que en Estados Unidos los gerentes tienen mucho interés en recibir buenas noticias, pero estas no dan oportunidad de mejorar; por otro lado, las malas noticias abren la puerta para nuevos productos y permiten que la compañía mejore.

A continuación se resumen los 14 puntos del doctor Deming, quien afirmaba de manera categórica que debían adoptarse como un paquete para tener éxito. El tema central es la cooperación, el trabajo en equipo y la convicción de que los trabajadores quieren que su trabajo sea de calidad.

LOS 14 PUNTOS DE DEMING

1. Crear un propósito constante de mejora continua de productos y servicio para la sociedad.
2. Adoptar como filosofía que ya no es posible vivir con los niveles de retrasos, errores, materiales defectuosos y mano de obra deficiente comúnmente aceptados.
3. Eliminar la necesidad de la inspección masiva como manera de lograr calidad; para obtenerla se debe fabricar el producto en forma correcta desde el principio.
4. Terminar con la práctica de ganar negocios solo con base en el precio: es necesario incluir medidas de calidad significativas junto con este.
5. Mejorar de manera constante, y por siempre, cada proceso de planeación, producción y servicio.
6. Implementar métodos modernos de capacitación en el trabajo para todos los empleados, incluso para los administradores; esto generará un mejor aprovechamiento de cada empleado.
7. Adoptar e instituir un liderazgo dirigido a ayudar a que la gente haga un mejor trabajo.
8. Fomentar la comunicación bidireccional eficaz y otros medios para ahuyentar el miedo en la organización, de modo que todos trabajen de manera más eficiente y productiva para la compañía.
9. Romper las barreras entre los departamentos y las áreas de personal.
10. Eliminar el uso de lemas, carteles y exhortaciones que exijan cero defectos y nuevos niveles de productividad sin proporcionar los métodos para lograrlos.
11. Eliminar los estándares de trabajo que fijan cuotas para la fuerza de trabajo y metas numéricas para el personal administrativo; sustituir los apoyos y el liderazgo conveniente a fin de lograr una mejora permanente en la calidad y la productividad.
12. Eliminar las barreras que roban a los empleados y al personal administrativo su derecho a enorgullecerse del fruto de su trabajo.
13. Instituir un programa educativo riguroso y fomentar la superación personal de todos; lo que una organización necesita es buen personal que se supere con la educación. El ascenso a un puesto competitivo tendrá sus raíces en el conocimiento.
14. Definir con claridad el compromiso permanente de la administración para mejorar la calidad y la productividad y aplicar todos estos principios.

Los 14 puntos de Deming no ignoraron el control estadístico de la calidad, que con frecuencia se abrevia SQC, por sus siglas en inglés. El objetivo del control estadístico de la calidad es supervisar la producción en muchas etapas de la manufactura; se emplean las herramientas del control estadístico de la calidad, como las gráficas de barras *X* y *R*, para supervisar la calidad de muchos procesos y servicios. Las tablas de control permiten identificar cuándo un proceso o servicio está “fuera de control”, es decir, cuándo se produce un número excesivo de unidades defectuosas.

El interés en la calidad se aceleró de forma impresionante en Estados Unidos desde finales de la década de 1980. Encienda la televisión y vea los comerciales de Ford, Nissan y GM, donde destacan el control de calidad en sus líneas de ensamble; en la actualidad es uno de los temas “de moda” en todas las facetas de los negocios. Daniel Hunt, un connotado asesor estadounidense en control de calidad, reporta que en la actualidad, en Estados Unidos, de 20% a 25% del costo de producción se gasta en detectar y corregir errores; además, agrega que el costo adicional de reparar o reemplazar productos defectuosos sobre la marcha ocasiona que el costo total de la calidad deficiente sea de casi 30%. En Japón, indicó, este costo es de apenas 3%.

En años recientes, las compañías se motivaron para mejorar la calidad en un esfuerzo por obtener reconocimiento en este renglón. El Malcolm Baldrige National Quality Award, establecido en 1988, se otorga anualmente a compañías estadounidenses que demuestren excelencia en el logro y administración de la calidad. Las categorías del premio son manufactura, servicios, negocios pequeños, cuidado de la salud y educación. Los ganadores de años recientes fueron, entre otros, Xerox, IBM, la University of Wisconsin-Stout, Ritz-Carlton Hotel Corporation, Federal Express y Cadillac. Los ganadores en 2012 fueron:

- Lockheed Martin Missiles and Fire Control (MFC) es una empresa que diseña, desarrolla, fabrica y apoya sistemas avanzados de combate, misiles, cohetes y sensores para el ejército estadounidense y para el extranjero. La fuerza de trabajo de la compañía, de 10 688 empleados, produce y entrega más de 100 productos a través de 825 contratos en más de 60 países. La sede de MFC está en Dallas, Texas, y posee sus segundas mayores instalaciones en Orlando, Florida; sus otras instalaciones se localizan en nueve estados de Estados Unidos, así como en Japón y en el Reino Unido.
- MESA Products, Inc., es una pequeña compañía privada (139 empleados) que diseña, fabrica e instala sistemas catódicos de protección que controlan la corrosión de las superficies de metal en estructuras subterráneas y sumergidas, tales como tanques y tuberías. MESA vende productos y materiales en todo Estados Unidos, y proporciona servicios técnicos y de instalación en forma regional. Las oficinas centrales de la compañía y sus instalaciones de producción están en Tulsa, Oklahoma; pero posee sucursales en Fort Worth, Texas; Houston, Texas; Tallahassee, Florida; Huntington Beach, California; Wapakoneta, Ohio; Charleston, Carolina del Sur; y Freeland, Washington. Sus ganancias en 2011 fueron de 45 millones de dólares y, en ese entonces, proyectó que serían de 55 millones en 2012. Este es el segundo premio Baldrige de MESA; la compañía lo recibió en 2006 en la categoría de pequeñas empresas.
- North Mississippi Health Services (NMHS) es un sistema de salud integral, no lucrativo, propiedad de la comunidad, que atiende a 24 condados rurales al noreste del Mississippi y al noroeste de Alabama. La organización comprende seis hospitales, cuatro guarderías y 34 clínicas; proporciona servicios preventivos y de bienestar, de hospitalización de emergencia, terapia intensiva y post-terapia intensiva; además de una organización de proveedores preferentes. La fuerza de trabajo de NMHS (6 226 empleados y 491 médicos) atiende en instalaciones de salud localizadas en seis comunidades de Mississippi (Tupelo, Baldwyn, Eupora, Luka, Pontotoc y West Pont) y en una ciudad de Alabama (Hamilton). Sus ganancias netas en 2011 fueron de 730 millones de dólares. North Mississippi Medical Center (NNMC), el hospital principal y centro de referencia del sistema NMHS, fue premiado con un premio Baldrige en la categoría de cuidados a la salud en 2006.
- La ciudad de Irving, localizada entre Dallas y Fort Worth, es la trigésima ciudad más poblada en Texas y la nonagésima cuarta en Estados Unidos. La ciudad, hogar de aproximadamente 217 700 habitantes, comprende un área de 68 millas cuadradas, que incluye al Aeropuerto Internacional Dallas-Fort Worth; sus principales servicios incluyen aplicación y cumplimiento de la ley, protección contra incendios y emergencias médicas, agua y alcantarillado, recolección de basura, mantenimiento de calles y manejo de tráfico, parques, librerías, programas culturales y recreativos y mejoras capitales.

Hay más información sobre estos y otros ganadores en **http://www. nist.gov/baldrige**.

ESTADÍSTICA EN ACCIÓN

¿La excelencia en la administración de la calidad permite un mejor desempeño financiero? En una investigación reciente se comparó el desempeño financiero de las compañías que recibieron el Baldrige National Quality Award con compañías similares que no fueron premiadas; la investigación reveló que las compañías que lo recibieron tenían un promedio de 39% de ingreso operativo más alto y 26% más ventas, y su costo por dólar de venta fue 1.22% menor.

Six Sigma

Muchas organizaciones de servicio, manufactura y no lucrativas están comprometidas con la mejora de la calidad de sus productos y servicios. “Six Sigma” es el nombre que se le dio a un programa organizacional diseñado para mejorar la calidad y el desempeño de la totalidad de una corporación, cuyo enfoque se concentra en reducir la variación en cualquier proceso que se utilice para producir y entregar productos y servicios a los clientes. Los programas Six Sigma se aplican tanto a procesos de producción como a procesos contables y otros de apoyo organizacional; sus resultados finales son: reducir los costos de los errores y defectos, aumentar la satisfacción del cliente y las ventas de productos y servicios, e incrementar los rendimientos.

Six Sigma obtiene su nombre de la distribución normal; *sigma* significa “desviación estándar”, y “más o menos” tres desviaciones estándar dan un rango total de seis desviaciones estándar; por lo tanto, Six Sigma significa no tener más de 3.4 defectos por millón en cualquier proceso, producto o servicio. Muchas empresas se esfuerzan por tener aún menos defectos; para lograr esta meta, el programa Six Sigma capacita a cada miembro de la organización que participa en los procesos para que identifique las fuentes de variación que afectan significativamente la calidad. El proceso incluye localizar y definir el problema, mejorar el proceso para reducir su variación, e implementar procedimientos para mejorarlo.

Six Sigma utiliza muchas técnicas estadísticas para recabar y analizar los datos necesarios para reducir la variación de un proceso. En este libro se incluyen los siguientes: histogramas, análisis de variación, prueba de *ji* cuadrada de la independencia, regresión y correlación.

General Electric, Motorola y AlliedSignal (en la actualidad parte de Honeywell) son compañías grandes que lograron una mejora relevante de calidad y ahorros en costos al utilizar los métodos Six Sigma; incluso ciudades como Fort Wayne, Indiana, emplean las técnicas Six Sigma para mejorar sus operaciones. La ciudad ahorró 10 millones de dólares desde 2000 y mejoró el servicio a sus clientes; por ejemplo, redujo 50% la generación de basura y el tiempo de respuesta para reparar baches (de 21 a 3 horas). Puede aprender más acerca de las ideas, métodos y capacitación Six Sigma en **http://www.6sigma.us**.

Fuentes de variación

OA19-2
Definir las dos causas de la variación de los procesos y explicar cómo se usan para monitorear la calidad.

No hay dos productos *exactamente* iguales; siempre hay alguna variación. El peso de cada hamburguesa Quarter Pounder de McDonald´s no es exactamente 0.25 libras; algunas pesan más de eso, otras menos. El tiempo estándar para que el autobús de TARTA (Toledo Area Regional Transit Authority) haga su recorrido desde el centro de Toledo, Ohio, hasta Perrysburg es de 25 minutos; sin embargo, no todos los recorridos duran *exactamente* 25 minutos; algunos tardan más y, en otras ocasiones, el conductor de TARTA debe esperar en Perrysburg antes de regresar a Toledo. En algunos casos existe una razón de la demora, como un accidente en la vía rápida o una tormenta de nieve; en otros, el conductor quizá no alcance los semáforos en verde o el tráfico esté inusualmente congestionado y lento sin razón aparente. En un proceso hay dos fuentes generales de variación: aleatoria y asignable.

> **VARIACIÓN ALEATORIA** Variación atribuible al azar; esta no se elimina por completo a menos que haya un cambio importante en las técnicas, tecnologías, métodos, equipamiento o materiales propios del proceso.

Algunos ejemplos de fuentes de variación aleatoria son la fricción interna en una máquina, cambios ligeros en las condiciones del material o del proceso (como la temperatura del molde para hacer botellas de vidrio), condiciones atmosféricas (como temperatura, humedad y el contenido de polvo del aire) y vibraciones transmitidas a una máquina por un montacargas que pasa a su lado.

Si el agujero taladrado en una pieza de acero es demasiado grande debido a una broca sin filo, esta se debe afilar, o será preciso insertar una nueva. Un operador que calibra la máquina de manera incorrecta se puede reemplazar o volver a capacitar. Si el rollo de acero que se utilizará en el proceso no tiene la resistencia a la tensión adecuada, se debe rechazar. Estos son ejemplos de variación asignable.

> **VARIACIÓN ASIGNABLE** Variación que no es aleatoria; esta se elimina o reduce cuando se investiga el problema y se encuentra la causa.

Hay varias razones a las que se debe poner atención respecto de la variación; dos de ellas se mencionan a continuación:

1. Cambiará la forma, dispersión y ubicación central de la distribución de la característica del producto que se mide.
2. Por lo general, la variación asignable es corregible, en tanto que, por lo normal, la variación aleatoria no se puede corregir o estabilizar de manera económica.

Diagramas de diagnóstico

OA19-3
Explicar el uso de diagramas para investigar las fuentes de la variación de los procesos.

Existen diversas técnicas de diagnóstico para investigar problemas de calidad; dos de las más relevantes son los **diagramas de Pareto** y los **diagramas de esqueleto de pez**.

Diagramas de Pareto

El análisis de Pareto es una técnica para llevar la cuenta del número de defectos de un producto o servicio; su nombre le fue impuesto en honor de un científico italiano del siglo XIX, Vilfredo Pareto, quien observó que la mayor parte de la "actividad" de un proceso se debe a relativamente pocos "factores". Su concepto, con frecuencia denominado regla 80-20, es que 80% de la actividad se debe a 20% de los factores; así, al concentrarse en 20% de los factores, los gerentes pueden dedicarse a 80% del problema, por ejemplo, Emily´s Family Restaurant, ubicado en el cruce de las carreteras interestatales 75 y 70, investiga las "quejas de los clientes". Las cinco quejas escuchadas con más frecuencia son: servicio descortés, comida fría, larga espera por una mesa, pocas opciones en el menú y niños indisciplinados; suponga que el servicio descortés es lo más frecuente y la comida fría aparece en segundo lugar. Los datos revelan que ambos factores representan más de 85% de las quejas, y de aquí que sean los que se deben atender primero, pues producirán la mayor reducción de las quejas y el mayor incremento en la satisfacción del cliente.

Para elaborar un diagrama de Pareto, inicie con la cuenta del tipo de defectos; luego, clasifique los defectos en términos de la frecuencia en que ocurren de mayor a menor y, por último, elabore una tabla de barras verticales, cuya altura corresponda a la frecuencia de cada defecto. En el siguiente ejemplo se ilustran estas ideas.

EJEMPLO

La administradora de la ciudad de Grove City, Utah, está preocupada por el consumo del agua, en particular en los hogares unifamiliares. Le gustaría desarrollar un plan para reducirlo; para investigar este problema, selecciona una muestra de 100 hogares y determina el consumo normal de agua diario para diversos fines; a la derecha se muestran los resultados de la muestra. ¿Cuál es el área con mayor consumo? ¿Dónde debe concentrar sus esfuerzos para reducir el consumo de agua?

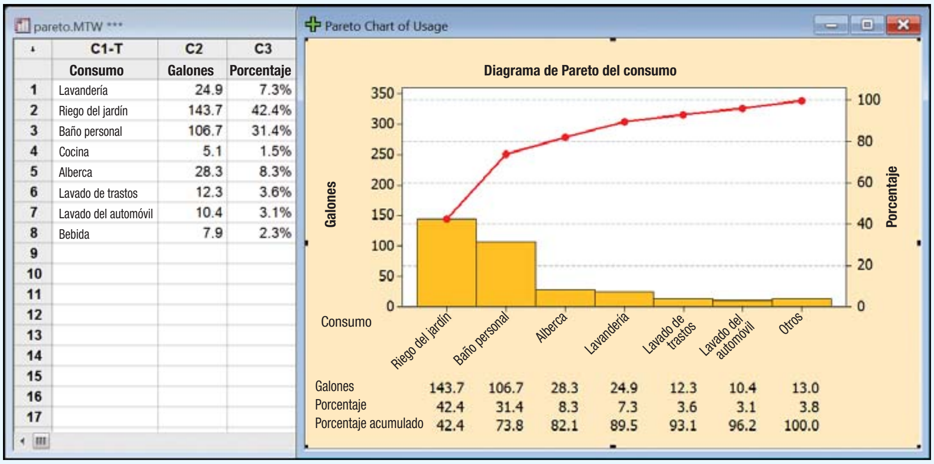
Consumo de agua	Galones por día
Lavandería	24.9
Riego del jardín	143.7
Baño personal	106.7
Cocina	5.1
Alberca	28.3
Lavado de trastos	12.3
Lavado del automóvil	10.4
Bebida	7.9

SOLUCIÓN

Un diagrama de Pareto es útil para identificar las áreas principales de consumo de agua y enfocarse en aquellas donde se pueda lograr la mayor reducción. El primer paso es convertir cada actividad en un porcentaje y luego ordenarlas de mayor a menor; en este caso, el consumo total de agua por día es de 339.3 galones, que se determinó al sumar el total de galones que consumen las ocho actividades. La actividad que consume más es el riego del jardín, que corresponde a 143.7 galones por día, o 42.4% de la cantidad de agua. La siguiente categoría mayor es el baño personal, que representa 31.4% del agua; ambas actividades representan 73.8% del consumo.

Consumo de agua	Galones por día	Porcentaje
Lavandería	24.9	7.3
Riego del jardín	143.7	42.4
Baño personal	106.7	31.4
Cocina	5.1	1.5
Alberca	28.3	8.3
Lavado de trastos	12.3	3.6
Lavado del automóvil	10.4	3.1
Bebida	7.9	2.3
Total	339.3	100.0

Para trazar el diagrama de Pareto, inicie con la representación a escala del número de galones que se consumen en el eje vertical izquierdo, y el porcentaje correspondiente en el eje vertical derecho; luego trace una barra vertical con la altura de la barra correspondiente a la actividad con el número mayor de eventos. En el ejemplo de Grove City, trace una barra vertical de la actividad de riego a una altura de 143.7 galones (llamado conteo). Continúe este procedimiento con las demás actividades, como se muestra en la salida en pantalla de Minitab (gráfica 19.1, página siguiente).



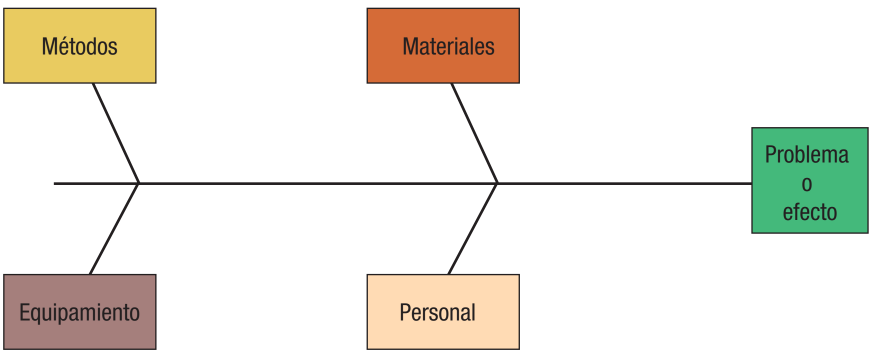
GRÁFICA 19.1 Diagrama de Pareto del consumo de agua en Grove City, Utah

Debajo del diagrama enumere las actividades, su frecuencia y el porcentaje de tiempo en que se realizan. En el último renglón liste el porcentaje acumulado; este permite determinar con rapidez qué conjunto de actividades representa el mayor consumo de agua. Los porcentajes acumulados se trazan arriba de las barras verticales. En el ejemplo de Grove City, las actividades de riego, baño personal y albercas representan 82.1% del consumo de agua; entonces, la administradora de la ciudad puede incrementar la ganancia si reduce el uso del agua en estas tres áreas.

Diagramas de esqueleto de pez

Otra tabla de diagnóstico es el **diagrama de causa y efecto** o **diagrama de esqueleto de pez**; su primer nombre destaca la relación entre un efecto particular y un conjunto de causas posibles que lo producen; además, es útil para organizar ideas e identificar relaciones, y es una herramienta que fomenta la generación de ideas. Identificar las relaciones permite determinar factores que son causa de variabilidad en algún proceso. El nombre *esqueleto de pez* proviene de la manera en que se organizan las diversas causas y efectos en el diagrama. El efecto, por lo general un problema particular, o tal vez un objetivo, se muestra a la derecha del diagrama; las causas principales se enumeran del lado izquierdo.

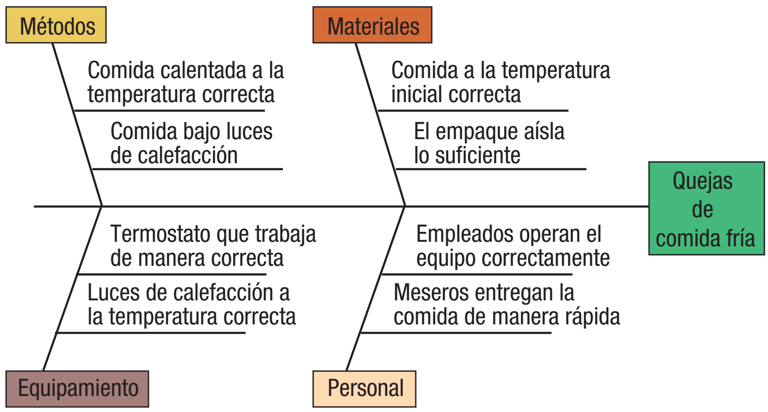
El enfoque habitual de un diagrama de esqueleto de pez es que permite considerar cuatro áreas del problema: métodos, materiales, equipamiento y personal. El problema, o el efecto, es la cabeza del pez (consulte la gráfica 19.2).



GRÁFICA 19.2 Diagrama de esqueleto de pez

En cada causa posible se encuentran motivos derivados que se deben identificar e investigar, los cuales son factores que quizás estén provocando el efecto particular. Se recopila la información concerniente al problema y con ella se completa el diagrama de esqueleto de pez; se investiga cada causa y se eliminan las que no son importantes, hasta identificar la causa real.

Para ilustrar un diagrama de esqueleto de pez, se investigan las causas de la comida fría que sirve Emily's Family Restaurant (recuerde que un análisis de Pareto mostró que la comida fría era una de las dos quejas principales). En la gráfica 19.3 se observa que cada causa derivada se enumera como suposición, y cada una debe investigarse para encontrar el problema real sobre la comida fría. En un diagrama de esqueleto de pez no hay ponderación de las causas derivadas.



GRÁFICA 19.3 Diagrama de esqueleto de pez para investigar quejas de comida fría en un restaurante

AUTOEVALUACIÓN 19-1

Rose Home, al sur de Chicago, es una institución de salud mental. Hace poco hubo quejas sobre sus condiciones, y el administrador quiere utilizar un diagrama de Pareto para investigar la situación. Cuando se queja un paciente o familiar, se le pide llenar un formato. He aquí el resumen de los formatos de quejas de los últimos 12 meses.

Queja	Número	Queja	Número
Nada que hacer	45	Condiciones insalubres	63
Atención deficiente del personal	71	Mala calidad de los alimentos	84
Error en los medicamentos	2	Personal irrespetuoso	35

Elabore un diagrama de Pareto. ¿Cuáles son las causas que el administrador debe resolver primero para lograr la mejora más significativa?

EJERCICIOS

1. Tom Sharkey es el propietario de Sharkey Chevy, Buick, GMC. A principios del año, él implementó un programa de opinión de los clientes a fin de determinar formas para mejorar el servicio. Una semana después de realizar el servicio, el asistente administrativo de Tom llama al cliente para averiguar si se efectuó de manera satisfactoria y cómo se puede mejorar. A continuación se incluye un resumen de las quejas de los primeros seis meses. Elabore un diagrama de Pareto. ¿Cuáles son las quejas que le sugeriría a Tom que resolviera primero para mejorar la calidad del servicio?

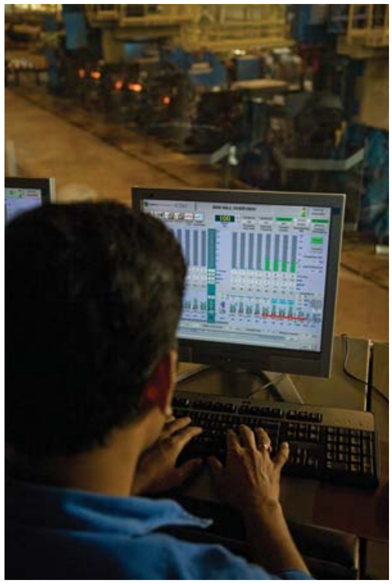
Queja	Frecuencia	Queja	Frecuencia
Problema sin corregir	38	Precio demasiado alto	23
Error en la factura	8	Mucho tiempo para prestar el servicio	10
Ambiente poco sociable	12		

2. En un taller de reparaciones se descubrió que de 110 motores que funcionan con diesel, 9 tenían bombas de agua con fugas, 15 presentaban cilindros defectuosos, 4 padecían problemas de encendido, 52 tenían fugas de aceite y 30 tenían bloques agrietados. Trace un diagrama de Pareto para identificar el problema clave de los motores.

Objetivo y tipos de diagramas de control de calidad

OA19-4
Calcular los límites de control de gráficas medias y de rango para una medición variable de calidad.

Los diagramas de control identifican el momento en que entran en el proceso las causas asignables de variación o los cambios; por ejemplo, Wheeling Company fabrica ventanas de aluminio recubiertas con vinilo para casas antiguas. El recubrimiento de vinilo debe tener un espesor comprendido entre ciertos límites; si es demasiado grueso, provocará que las ventanas se atoren, pero si es demasiado delgado, la ventana no sellará bien. El mecanismo que determina cuánto recubrimiento se pone en cada ventana se desgasta y comienza a engrosarlo demasiado; por lo tanto, ocurrió un cambio en el proceso. Los diagramas de control son útiles para detectar el cambio en las condiciones del proceso, pero es importante saber cuándo se produjeron cambios en el proceso, de modo que la causa se identifique y corrija antes de que se produzca un gran número de artículos inaceptables.



Los diagramas de control se parecen a la pizarra del marcador de un juego de béisbol. Al ver la pizarra, los fanáticos, entrenadores y jugadores saben qué equipo va ganando; sin embargo, esta no hace nada para ganar o perder el juego. Los diagramas de control tienen una función similar: indican a los trabajadores, líderes de grupos, ingenieros de control de calidad, supervisores de producción y gerentes si la producción de la parte o el servicio está “bajo control” o “fuera de control”. En este último caso, el diagrama de control no solucionará la situación; solo es una hoja de papel con cifras y puntos; en cambio, la persona responsable ajustará la máquina, fabricará la pieza o hará lo que sea necesario para poner la producción “bajo control”.

Hay dos tipos de diagramas de control. Un **diagrama de control de variables** representa mediciones, como la cantidad de refresco de cola en una botella de dos litros o el diámetro exterior de una tubería. Un diagrama de control de variables requiere un intervalo o escala de razón de medición. Un **diagrama de control de atributos** clasifica un producto o servicio como aceptable o inaceptable; se basa en la escala de medición nominal. A los infantes de marina estacionados en Camp Lejeune se les pide calificar los alimentos que se les sirven como aceptables o inaceptables; los préstamos bancarios se pagan o se dejan de pagar.

Diagramas de control de variables

Para elaborar diagramas de control de variables se depende de la teoría de muestreo que se analizó, junto con el teorema central del límite, en el capítulo 8; suponga que selecciona una muestra de cinco piezas cada hora del proceso de producción y calcula la media de cada muestra. Las medias de la muestra son $\overline{x}_1$, $\overline{x}_2$, $\overline{x}_3$, etcétera; la media de estas medias de las muestras se denota como $\overline{\overline{x}}$. Utilice k para indicar el número de medias de la muestra. La media general o media total se determina mediante:

MEDIA TOTAL
$$\overline{\overline{x}} = \frac{\Sigma \text{ de las medias de las muestras}}{\text{Número de medias muestrales}} = \frac{\Sigma \overline{x}_i}{k} \qquad \textbf{[19.1]}$$

El error estándar de la distribución de las medias de las muestras se designa mediante $s_{\overline{x}}$; el cual se determina por:

ERROR ESTÁNDAR DE LA MEDIA
$$s_{\overline{x}} = \frac{s}{\sqrt{n}} \qquad \textbf{[19.2]}$$

Estas relaciones permiten establecer límites respecto de las medias de las muestras para mostrar cuánta variación se espera en un tamaño determinado de la muestra; estos límites esperados se denominan **límite de control superior** (*LCS*) y **límite de control inferior** (*LCI*). Mediante un ejemplo se ilustrará el uso de los límites de control y la forma de determinarlos.

EJEMPLO

Statistical Software, Inc., ofrece un número telefónico de larga distancia sin costo al cual los clientes pueden llamar todos los días, de 7 a.m. a 11 p.m., para resolver problemas con sus productos; es imposible que un representante técnico conteste de inmediato, pero es importante que los clientes no esperen demasiado en línea para que les contesten porque los clientes se molestan cuando escuchan demasiadas veces el mensaje: "Su llamada es importante para nosotros; en breve le contestará un representante". Para comprender el proceso, Statistical Software decidió elaborar una tabla de control con el tiempo total desde el momento en que se recibe una llamada hasta que el representante la responde y soluciona el problema. El día de ayer se tomó una muestra de cinco llamadas cada hora durante las 16 horas de operación del servicio de atención al cliente y se registró el tiempo total para resolver el problema de un cliente, en minutos, para cada llamada. En la tabla siguiente se refleja esta información.

	Número de muestra				
Hora	**1**	**2**	**3**	**4**	**5**
a.m. 7	8	9	15	4	11
8	7	10	7	6	8
9	11	12	10	9	10
10	12	8	6	9	12
11	11	10	6	14	11
p.m. 12	7	7	10	4	11
1	10	7	4	10	10
2	8	11	11	7	7
3	8	11	8	14	12
4	12	9	12	17	11
5	7	7	9	17	13
6	9	9	4	4	11
7	10	12	12	12	12
8	8	11	9	6	8
9	10	13	9	4	9
10	9	11	8	5	11

Con base en esta información, elabore una tabla de control para determinar la duración media de la llamada. ¿Parece existir una tendencia en las horas de las llamadas? ¿Hay algún periodo donde parece que los clientes esperan más que en otros?

SOLUCIÓN

Una tabla para el control de la media tiene dos límites: *LCS* y *LCI*; estos se calculan mediante:

LÍMITES DE CONTROL DE LA MEDIA

$$LCS = \bar{\bar{x}} + 3\frac{s}{\sqrt{n}} \quad \text{y} \quad LCI = \bar{\bar{x}} - 3\frac{s}{\sqrt{n}} \qquad \textbf{[19.3]}$$

donde *s* es una estimación de la desviación estándar de la población, σ. Observe que en el cálculo de los límites de control superior e inferior aparece el número 3 representando 99.74% de los límites de confianza. Con frecuencia, a los límites se les denomina 3-sigma; sin embargo, se pueden utilizar otros límites de confianza (como 90% o 95%).

Esta aplicación se desarrolló antes del extenso acceso a las computadoras y era difícil calcular las desviaciones estándar; en vez de calcular la desviación estándar de cada muestra como una medida de variación, es más fácil utilizar el rango. En el caso de muestras de tamaño fijo hay una relación constante entre el rango y la desviación estándar, por lo que es apropiado que se utilicen las fórmulas siguientes para determinar 99.74% de los límites de control de la media; se puede demostrar que el término $3(s/\sqrt{n})$ de la fórmula [19.3] equivale a $A_2\bar{R}$ en la siguiente fórmula.

LÍMITES DE CONTROL DE LA MEDIA

$$LCS = \bar{\bar{x}} + A_2\bar{R} \quad LCI = \bar{\bar{x}} - A_2\bar{R} \qquad \textbf{[19.4]}$$

donde:

A_2 es una constante al calcular los límites de control superior e inferior; se basa en el rango promedio, $\overline{R}$. Los factores de varios tamaños de muestras aparecen en el apéndice B.10 (*nota:* en esta tabla, *n* se refiere al número de elementos de la muestra). A continuación se presenta una parte del apéndice B.8; para ubicar el factor A_2 de este problema, encuentre el tamaño de *n* en el margen izquierdo, que es 5. Luego continúe con un movimiento horizontal hasta la columna A_2; el factor es 0.577.

n	A_2	d_2	D_3	D_4
2	1.880	1.128	0	3.267
3	1.023	1.693	0	2.575
4	0.729	2.059	0	2.282
5	$\boxed{0.577}$	2.326	0	2.115
6	0.483	2.534	0	2.004

$\overline{\overline{x}}$ es la media de las medias de las muestras, que se calcula mediante $\Sigma\overline{x}/k$, donde k es el número de muestras seleccionadas. En este problema se toma una muestra de cinco observaciones cada hora durante 16 horas, por lo que $k = 16$.

$\overline{R}$ es la media de los rangos de la muestra, que es $\Sigma R/k$. Recuerde que el rango es la diferencia entre el valor mayor y el menor de cada muestra, y describe la variabilidad que ocurre en *esa* muestra (consulte la tabla 19.1).

El valor de la media total $\overline{\overline{x}}$ en la tabla es 9.413 minutos, determinado mediante 150.6/16. La media de los rangos ($\overline{R}$) es 6.375 minutos, que se determinó mediante 102/16; por lo tanto, el límite de control superior es:

$$LCS = \overline{\overline{x}} + A_2\overline{R} = 9.413 + 0.577(6.375) = 13.091$$

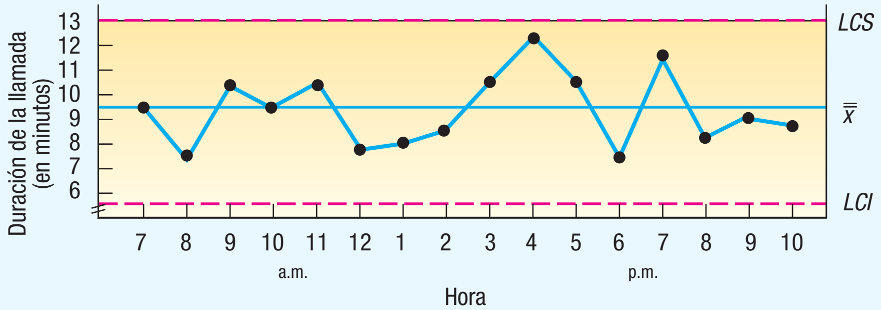
El límite de control inferior es:

$$LCI = \overline{\overline{x}} - A_2\overline{R} = 9.413 - 0.577(6.375) = 5.735$$

TABLA 19.1 Dieciséis muestras de duración de llamada en Statistical Software, Inc.

Hora	1	2	3	4	5	Media	Rango
a.m. 7	8	9	15	4	11	9.4	11
8	7	10	7	6	8	7.6	4
9	11	12	10	9	10	10.4	3
10	12	8	6	9	12	9.4	6
11	11	10	6	14	11	10.4	8
p.m. 12	7	7	10	4	11	7.8	7
1	10	7	4	10	10	8.2	6
2	8	11	11	7	7	8.8	4
3	8	11	8	14	12	10.6	6
4	12	9	12	17	11	12.2	8
5	7	7	9	17	13	10.6	10
6	9	9	4	4	11	7.4	7
7	10	12	12	12	12	11.6	2
8	8	11	9	6	8	8.4	5
9	10	13	9	4	9	9.0	9
10	9	11	8	5	11	8.8	6
Total						150.6	102

$\overline{\overline{x}}$, *LCS* y *LCI* y las medias de las muestras se presentan en la gráfica 19.4 (página siguiente). La media, $\overline{\overline{x}}$, es 9.413 minutos, el límite de control superior se ubica en 13.091 minutos, y el límite de control inferior, en 5.735 minutos. Hay una variación en la duración de las llamadas, pero todas las medias de la muestra están dentro de los límites de control; por lo tanto, con base en 16 muestras de cinco llamadas, la conclusión es que 99.74% de las veces, la duración media de dicha muestra estará entre 5.735 minutos y 13.091 minutos.



GRÁFICA 19.4 Diagrama de control de la duración media de las llamadas de clientes a Statistical Software, Inc.

ESTADÍSTICA EN ACCIÓN

Con ayuda de los diagramas de control, se consignó a una persona que sobornaba a jugadores de jai-alai para que perdieran. En las gráficas de la media y el rango se revelaron patrones de apuestas inusuales y que algunos jugadores no ganaron cuando hicieron ciertas apuestas. Un experto en calidad "bajo control" identificó las ocasiones en que cesó la variación asignable y los fiscales las relacionaron con la detención del sospechoso.

Puesto que la teoría estadística se basa en la normalidad de muestras grandes, los diagramas de control deben tener como base un proceso estable, es decir, una muestra muy grande tomada durante un periodo extenso. Una regla básica es diseñar el diagrama después de seleccionar al menos 25 muestras.

Diagrama de rangos

Además de la ubicación central en una muestra, también debe supervisar la cantidad de variación de muestra en muestra. Un **diagrama de rangos** presenta la variación de los rangos de las muestras; si los puntos que representan los rangos se encuentran entre los límites superior e inferior, concluya que la operación está bajo control. De acuerdo con la casualidad, casi 997 de 1 000 veces el rango de las muestras estará dentro de los límites; si el rango cae arriba de los límites, concluya que una causa asignable afectó la operación y es necesario ajustar el proceso. ¿Por qué no interesa el límite de control inferior del rango? Con frecuencia, en muestras pequeñas el límite inferior es cero. En realidad, en cualquier muestra de seis o menos, el límite de control inferior es 0; si tal es el rango, entonces por lógica todas las partes son iguales y no hay problema con la variabilidad de la operación.

Los límites de control superior e inferior del diagrama de rangos se determinan a partir de las siguientes ecuaciones.

DIAGRAMA DE CONTROL DE RANGOS $LCS = D_4\overline{R} \quad LCI = D_3\overline{R}$ **[19.5]**

Los valores de D_3 y D_4, que reflejan los límites habituales 3σ (sigma) de varios tamaños de la muestra, aparecen en el apéndice B.10 o después de la fórmula [19.4].

EJEMPLO

El tiempo que los clientes de Statistical Software, Inc., esperaron desde que entró su llamada hasta que un representante técnico respondió su pregunta o resolvió su problema se encuentra registrado en la tabla 19.1. Elabore un diagrama de control de rangos. ¿Parece que hay algún momento en el que es demasiada la variación en la operación?

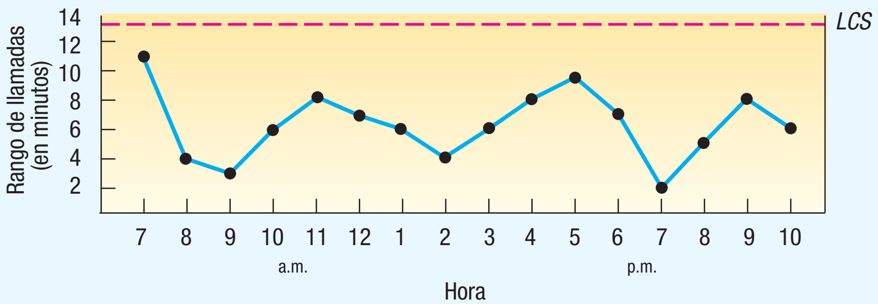
SOLUCIÓN

El primer paso es encontrar la media de los rangos de la muestra; el de las cinco llamadas de la muestra de las 7:00 a.m. es 11 minutos. La llamada de mayor duración seleccionada en esa hora fue de 15 minutos, y la más breve, de 4; la diferencia es 11 minutos. A las 8 a.m., el rango es de 4 minutos. El total de los 16 rangos es 102 minutos, por lo que el rango promedio es de 6.375 minutos, determinado por $\overline{R} = 102/16$. Con referencia al apéndice B.10 o a la tabla parcial que aparece después de la fórmula [19.4], D_3 y D_4 son 0 y 2.115, respectivamente. Los límites de control superior e inferior son 0 y 13.483.

$$LCS = D_4\overline{R} = 2.115(6.375) = 13.483$$

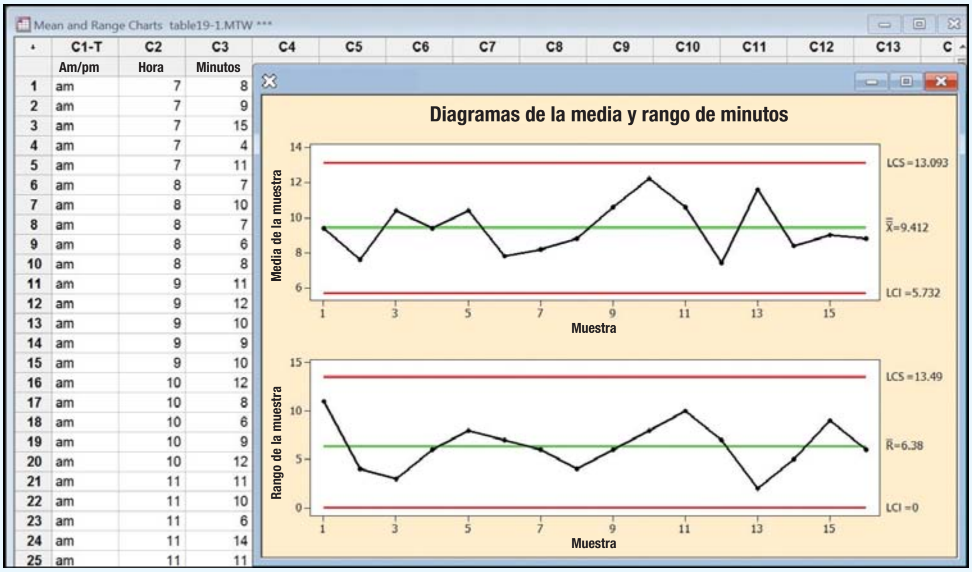
$$LCI = D_3\overline{R} = 0(6.375) = 0$$

El diagrama del trazo de los 16 rangos de las muestras se registra en la gráfica 19.5; este diagrama indica que todos los rangos están dentro de los límites de control. De aquí, se concluye que la variación en el tiempo para atender las llamadas de los clientes está dentro de los límites normales, es decir, "bajo control"; por supuesto, debe determinar los límites de control con base en un conjunto de datos y luego aplicarlos para evaluar los siguientes, no los que ya conoce.



GRÁFICA 19.5 Diagrama de control de rangos de la duración de las llamadas de los clientes a Statistical Software, Inc.

Minitab presenta un diagrama de control de la media y el rango. A continuación se incluye la salida en pantalla del ejemplo de Statistical Software. Los datos se muestran en la tabla 19.1 (las pequeñas diferencias entre los límites de control se deben al redondeo).

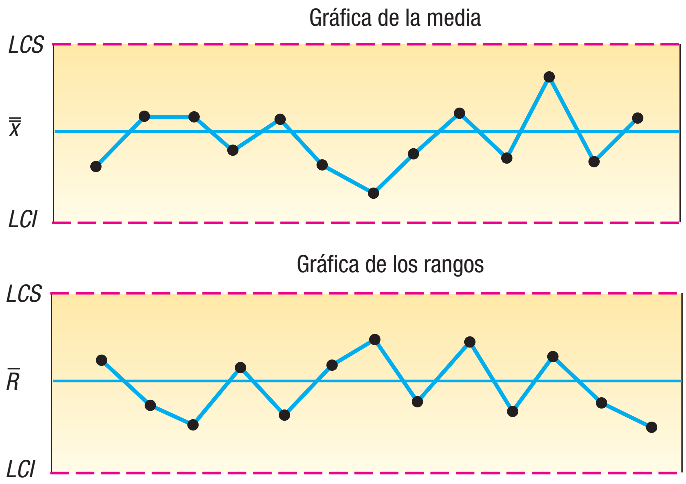


Situaciones de bajo control y fuera de control

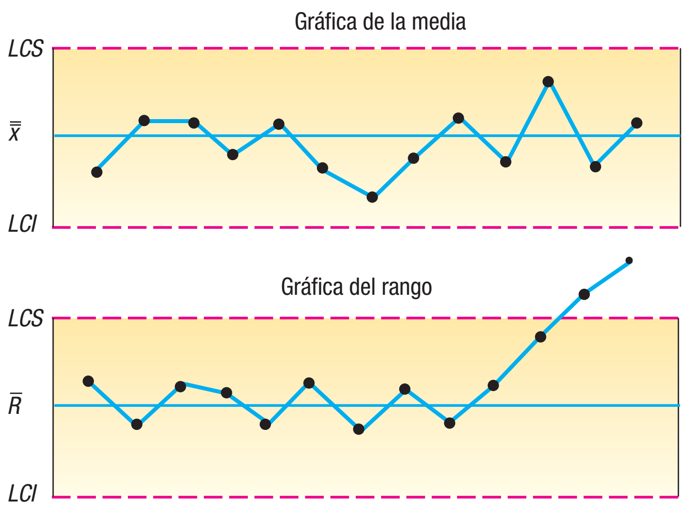
OA19-5 Evaluar los diagramas de control para determinar si un proceso está fuera de control.

He aquí tres ilustraciones de procesos bajo control y fuera de control:

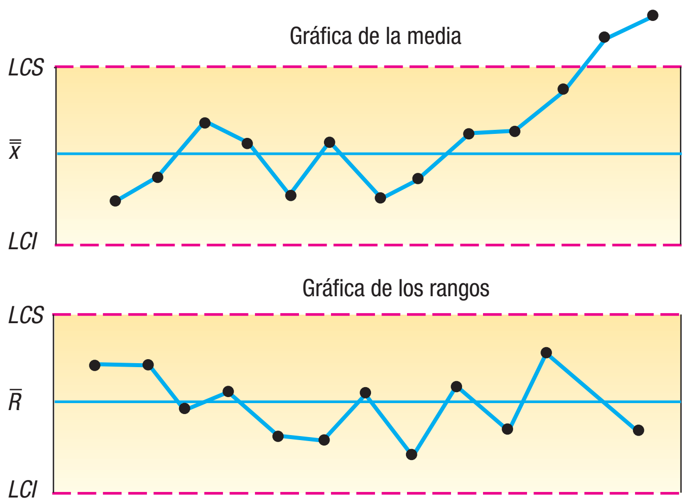
1. El diagrama de la media y el de rangos en conjunto que se muestran en la página siguiente, indican que el proceso está bajo control; observe que la media y los rangos de las muestras se agrupan cerca de las líneas centrales. Algunos están arriba y otros debajo de las líneas centrales, lo que indica que el proceso es muy estable; es decir, no hay una tendencia visible para que la media y los rangos se desplacen hacia las áreas fuera de control.



2. La media de las muestras está bajo control, pero los rangos de las últimas dos muestras no lo están; esto indica que hay una variación considerable en las muestras. Algunos rangos de las muestras son grandes, y otros, pequeños; es probable que se requiera ajustar el proceso.



3. La media está bajo control en las primeras muestras, pero hay una tendencia ascendente hacia el *LCS*. Las dos últimas medias de las muestras están fuera de control; probablemente sea necesario ajustar el proceso.



En la gráfica anterior de la media se aprecia un ejemplo de una gráfica de control que ofrece cierta información adicional. Observe la dirección de las últimas cinco observaciones de la media; todas están arriba de $\overline{\overline{x}}$, y, de hecho, las últimas dos están fuera de control; es poco probable que las medias de la muestra aumentaran durante seis observaciones consecutivas, lo cual es otra indicación de que el proceso está fuera de control.

AUTOEVALUACIÓN 19-2

La gerente de River City McDonald's selecciona al azar cuatro clientes durante tres horas del día: 9:00 a.m., 10:00 a.m. y 11:00 a.m. Luego, mide el tiempo, en minutos, entre la entrada de la orden que ellos solicitan y su entrega. A continuación se incluyen los resultados.

	Tiempos de la muestra			
Hora	**1**	**2**	**3**	**4**
9 a.m.	1	4	5	2
10 a.m.	2	3	2	1
11 a.m.	1	7	3	5

(a) Calcule el tiempo medio de espera, el rango medio y determine los límites de control de la media y el rango, y trace con ellos un diagrama.
(b) ¿Las mediciones están dentro de los límites de control? Interprete la gráfica.

EJERCICIOS

3. Describa la diferencia entre variación asignable y variación aleatoria.
4. Describa la diferencia entre una gráfica de control de atributos y una de control de variables.
5. De una línea de producción se toman muestras de tamaño $n = 4$.
  a. ¿Cuál es el valor del factor A_2 para determinar los límites de control superior e inferior de la media?
  b. ¿Cuáles son los valores de los factores D_3 y D_4 para determinar los límites de control superior e inferior de la media?
6. De un proceso de manufactura se seleccionan muestras de cinco. La media de los rangos de la muestra es 0.50; estime la desviación estándar de la población.
7. En Piatt Bakery se acaba de instalar un nuevo horno industrial; para conocer su temperatura, un inspector la lee en cuatro lugares distintos dentro del horno cada media hora, comenzando a las 8:00 a.m. La primera lectura fue de 340 grados Fahrenheit; la última fue a las 10:30 a.m., para un total de seis muestras (para facilitar los cálculos que se encuentran en la siguiente tabla, solo se dan los primeros dos dígitos).

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Lectura			
Hora	**1**	**2**	**3**	**4**
8:00 a.m.	40	50	55	39
8:30 a.m.	44	42	38	38
9:00 a.m.	41	45	47	43
9:30 a.m.	39	39	41	41
10:00 a.m.	37	42	46	41
10:30 a.m.	39	40	39	40

  a. Con base en esta experiencia inicial, determine los límites de control de la temperatura media y la media total. Trace la experiencia en una gráfica.
  b. Interprete la gráfica. ¿Parece haber una hora en que la temperatura está fuera de control?
8. Consulte el ejercicio 7.
  a. Con base en esta experiencia inicial, determine los límites de control del rango. Trace la experiencia en una gráfica.
  b. ¿Parece haber una hora en la que hay demasiada variación de temperatura?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Diagramas de control de atributos

OA19-6 Calcular los límites de los diagramas de control para un atributo con una medida de calidad variable.

Con frecuencia, los datos que se recopilan son el resultado de contar en vez de medir; es decir, se observa la presencia o ausencia de algún atributo; por ejemplo, la taparrosca de un frasco de champú se ajusta sin dejar salir líquido (una condición "aceptable") o bien no sella y deja salir líquido (una condición “inaceptable”), o un banco otorga un préstamo a un cliente, quien le paga o no le paga; en otros casos, interesa el número de defectos de una muestra. British Airways puede contar el número de sus vuelos demorados por día en Gatwick Airport, en Londres. En esta sección se estudian dos tipos de diagramas de atributos: la tabla *p* (porcentaje defectuoso) y la gráfica de barras *c* (número de defectos).

Diagramas p

Si el artículo registrado es la fracción de partes inaceptables existentes en un lote grande, el diagrama de control apropiado es el **diagrama p,** cuya base es la distribución binomial, que se analizó en el capítulo 6, y las proporciones, en el capítulo 15. La línea central está en p, la proporción media de defectos; p reemplaza a la $\bar{\bar{x}}$ del diagrama de control de variables. La proporción media de defectos se estima mediante:

PROPORCIÓN MEDIA DE DEFECTOS

$$p = \frac{\text{Número total de defectos}}{\text{Número total de artículos de la muestra}} \quad \textbf{[19.6]}$$

La variación en la proporción muestral se estima por medio del error estándar de una proporción muestral; se determina por medio de:

ERROR ESTÁNDAR DE LA PROPORCIÓN MUESTRAL

$$s_p = \sqrt{\frac{p(1-p)}{n}} \quad \textbf{[19.7]}$$

Por lo tanto, el *LCS* y el *LCI* se calculan como el porcentaje medio más o menos tres veces el error estándar de los porcentajes (proporciones). La fórmula de los límites de control es:

LÍMITES DE CONTROL DE PROPORCIONES

$$LCI, LCS = p \pm 3\sqrt{\frac{p(1-p)}{n}} \quad \textbf{[19.8]}$$

Mediante un ejemplo se ilustrarán los detalles de los cálculos y las conclusiones.

EJEMPLO

Jersey Glass Company, Inc., produce espejos pequeños de mano. La compañía opera un turno diurno y uno vespertino cada día laboral de la semana, y el departamento de control de calidad (QA) supervisa la calidad de los espejos dos veces durante cada turno. El departamento de calidad selecciona e inspecciona minuciosamente una muestra aleatoria de 50 espejos cada cuatro horas. Cada espejo se clasifica como aceptable o inaceptable; por último, se cuenta el número de espejos incluidos en la muestra que no cumplen con las especificaciones de calidad. A continuación se indican los resultados de estas verificaciones durante los últimos 10 días laborables.

Fecha	Número muestreado	Defectos	Fecha	Número muestreado	Defectos
10-Oct	50	1	12-Oct	50	9
	50	0		50	3
	50	9		50	10
	50	9		50	2
11-Oct	50	4	13-Oct	50	2
	50	4		50	4
	50	5		50	9
	50	3		50	4

(*continúa*)

(*continuación*)

Fecha	Número muestreado	Defectos	Fecha	Número muestreado	Defectos
14-Oct	50	6	19-Oct	50	4
	50	9		50	5
	50	2		50	2
	50	4		50	5
17-Oct	50	7	20-Oct	50	0
	50	9		50	0
	50	0		50	4
	50	8		50	7
18-Oct	50	6	21-Oct	50	5
	50	9		50	1
	50	8		50	9
	50	1		50	9

Elabore el diagrama del porcentaje defectuoso de este proceso. ¿Cuáles son los límites de control superior e inferior? Interprete los resultados. ¿Parece que el proceso está fuera de control durante el periodo?

SOLUCIÓN

El primer paso es estimar la proporción media de defectos mediante la fórmula [19.6].

$$p = \frac{\text{Número total de defectos}}{\text{Número total de artículos muestreados}} = \frac{196}{2\ 000} = 0.098$$

Por lo tanto, se estima que 0.098 de los espejos producidos durante el periodo no cumplen las especificaciones.

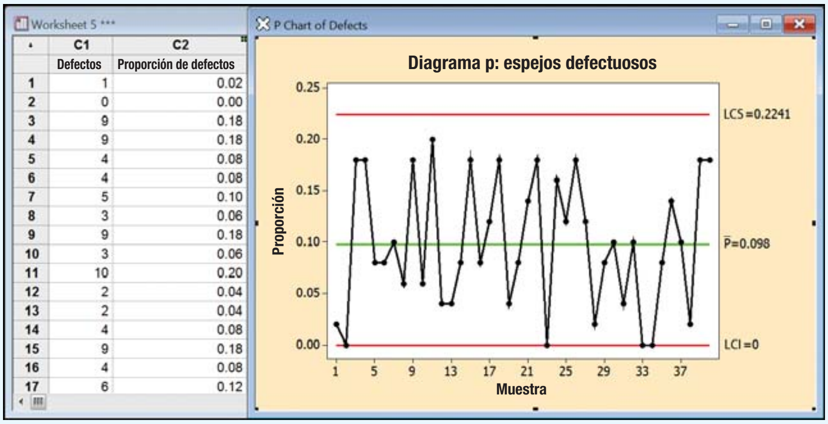
Fecha	Número muestreado	Defectos	Fracción defectuosa	Fecha	Número muestreado	Defectos	Fracción defectuosa
10-Oct	50	1	0.02	17-Oct	50	7	0.14
	50	0	0.00		50	9	0.18
	50	9	0.18		50	0	0.00
	50	9	0.18		50	8	0.16
11-Oct	50	4	0.08	18-Oct	50	6	0.12
	50	4	0.08		50	9	0.18
	50	5	0.10		50	6	0.12
	50	3	0.06		50	1	0.02
12-Oct	50	9	0.18	19-Oct	50	4	0.08
	50	3	0.06		50	5	0.10
	50	10	0.20		50	2	0.04
	50	2	0.04		50	5	0.10
13-Oct	50	2	0.04	20-Oct	50	0	0.00
	50	4	0.08		50	0	0.00
	50	9	0.18		50	4	0.08
	50	4	0.08		50	7	0.14
14-Oct	50	6	0.12	21-Oct	50	5	0.10
	50	9	0.18		50	1	0.02
	50	2	0.04		50	9	0.18
	50	4	0.08		50	9	0.18
				Total	2 000	196	

Los límites de control superior e inferior se calculan con la fórmula [19.8]

$$LCI,\ LCS = p \pm 3\sqrt{\frac{p(1-p)}{n}} = 0.098 \pm 3\sqrt{\frac{0.098(1-0.098)}{50}} = 0.098 \pm 0.1261$$

A partir de los cálculos anteriores, el límite de control superior se determina en 0.2241, mediante 0.098 + 0.1261. El límite de control inferior se establece en 0 porque el límite inferior calculado con la fórmula es 0.098 − 0.1261 = −0.0281; sin embargo, no es posible que haya una proporción negativa de defectos, por lo que el valor menor es 0. Entonces, los límites de control son 0 y 0.2241. Cualquier muestra fuera de estos límites indica que cambió el nivel de calidad del proceso.

Esta información se resume en la gráfica 19.6, que es una salida del software Minitab.



GRÁFICA 19.6 Diagrama p de espejos en Jersey Glass

Después de establecer los límites, el proceso se supervisa durante la siguiente semana, cinco días, dos turnos por día, con dos verificaciones de calidad por turno. He aquí los resultados.

Fecha	Número muestreado	Defectos	Fracción defectuosa	Fecha	Número muestreado	Defectos	Fracción defectuosa
24-Oct	50	1	0.02	27-Oct	50	2	0.04
	50	13	0.26		50	1	0.02
	50	10	0.20		50	7	0.14
	50	7	0.14		50	12	0.24
25-Oct	50	4	0.08	28-Oct	50	5	0.10
	50	5	0.10		50	5	0.10
	50	6	0.12		50	10	0.20
	50	10	0.20		50	9	0.18
26-Oct	50	6	0.12				
	50	1	0.02				
	50	8	0.16				
	50	4	0.08				

El proceso estuvo fuera de control en dos ocasiones, el 24 de octubre, cuando el número de defectos fue 13, y el 27 de octubre, cuando hubo 12 defectos. El departamento de control de calidad debe reportar esta información al equipo de producción para tomar las medidas pertinentes. A continuación se muestra la salida de Minitab.

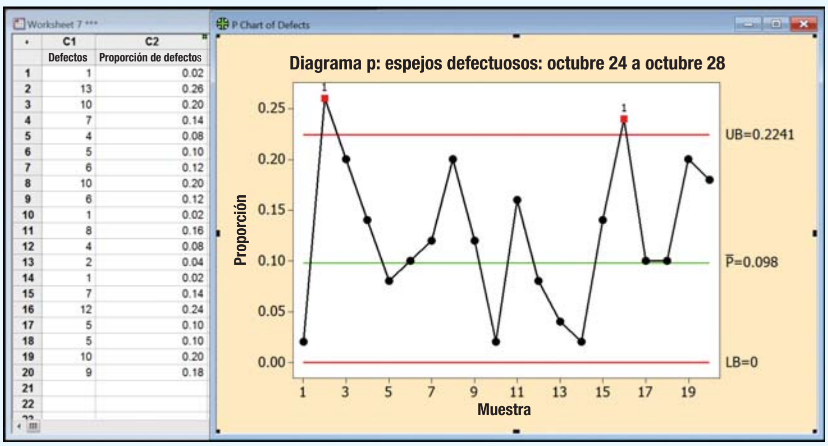


Diagrama de líneas *c*

En la gráfica de líneas *c* se traza el número de defectos o fallas por unidad; se basa en la distribución de Poisson, que se estudió en el capítulo 6. El número de maletas maltratadas en un vuelo por Southwest Airlines se puede supervisar mediante una gráfica de barras *c*; la "unidad" en consideración es el vuelo. En la mayoría de los vuelos no hay maletas maltratadas; en otros puede haber una, y en algunos más, dos, etcétera. El Internal Revenue Service puede contar y elaborar un diagrama de control del número de errores aritméticos en las declaraciones de impuestos; la mayoría de las declaraciones de impuestos no tendrán ningún error, algunas tendrán un solo error, otras tendrán dos, etcétera. Designe $\bar{c}$ como el número medio de defectos por unidad; así, $\bar{c}$ es el número medio de maletas maltratadas por Southwest Airlines por vuelo o el número medio de errores aritméticos por declaración de impuestos. Recuerde que en el capítulo 6 se vio que la desviación estándar de una distribución de Poisson es la raíz cuadrada de la media; por lo tanto, es posible determinar los límites de 3 sigma o 99.74% en un diagrama de barras *c* mediante:

LÍMITES DE CONTROL DEL NÚMERO DE DEFECTOS POR UNIDAD

$$LCI, LCS = \bar{c} \pm 3\sqrt{\bar{c}} \quad [19.9]$$

EJEMPLO

El editor del *Oak Harbor Daily Telegraph* está preocupado por el número de palabras mal escritas en el periódico. En un esfuerzo por controlar el problema y fomentar la buena ortografía, utilizó un diagrama de control; el número de palabras mal escritas que determinó en la edición final del periódico de los últimos 10 días es: 5, 6, 3, 0, 4, 5, 1, 2, 7 y 4. Determine los límites de control apropiados e interprete el diagrama. ¿Hubo algunos días durante el periodo en que el número de palabras mal escritas estuvo fuera de control?

SOLUCIÓN

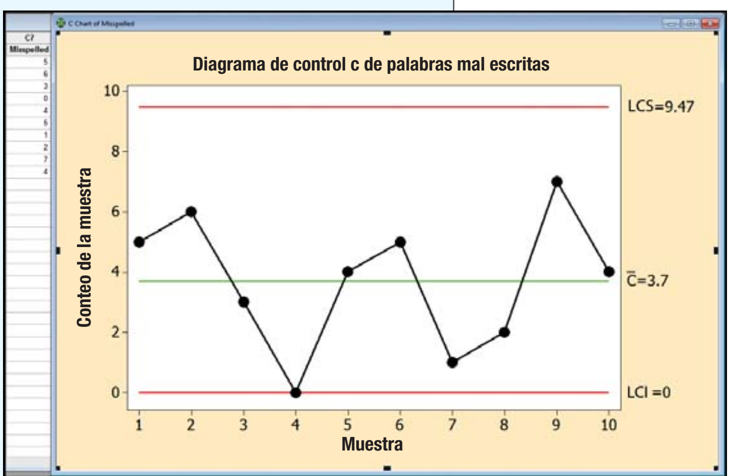
Durante el periodo de 10 días hubo un total de 37 palabras mal escritas. El número medio de palabras mal escritas por edición es 3.7, y sigue la distribución de probabilidad de Poisson. La desviación estándar es la raíz cuadrada de la media.

$$\bar{c} = \frac{\Sigma x}{n} = \frac{5 + 6 + \cdots + 4}{10} = \frac{37}{10} = 3.7 \qquad s = \sqrt{\bar{c}} = \sqrt{3.7} = 1.924$$

Para encontrar el límite de control superior utilice la fórmula [19.9]. El límite de control inferior es cero.

$$LCS = \bar{c} + 3\sqrt{\bar{c}} = 3.7 + 3\sqrt{3.7} = 3.7 + 5.77 = 9.47$$

El límite de control inferior calculado sería $3.7 - 3(1.924) = -2.07$; sin embargo, el número de palabras mal escritas no puede ser menor que 0, por lo que debe emplear 0 como límite inferior; por otra parte, el límite superior es 9.47. Cuando se compara cada uno de los puntos de datos con 9.47, resulta que todos son menores que el límite de control superior; el número de palabras mal escritas "está bajo control"; por supuesto, el periódico hará un esfuerzo para eliminar todas las palabras mal escritas, pero las técnicas de los diagramas de control ofrecen un medio para dar seguimiento a los resultados diarios y determinar si hay un cambio; por ejemplo, si se contrata una nueva correctora de pruebas, se puede comparar su trabajo con el de otros; los resultados se resumen en la gráfica 19.7, que es la salida del sistema Minitab.



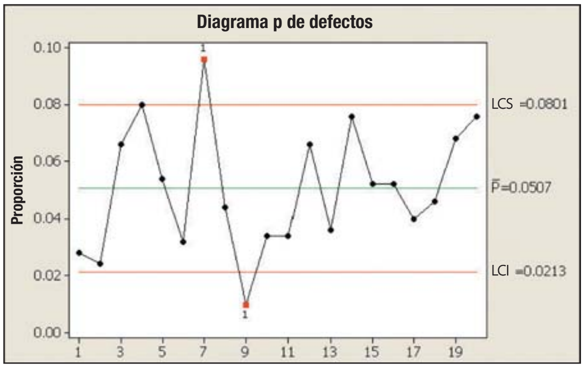
GRÁFICA 19.7 Diagrama de control *c* de las palabras mal escritas por edición del Oak Harbor Daily Telegraph

AUTOEVALUACIÓN 19-3

Auto-Lite Company fabrica baterías para automóviles. Al final de cada turno, el departamento de calidad selecciona una muestra de baterías para probarlas; el número de unidades defectuosas durante los últimos 12 turnos es 2, 1, 0, 2, 1, 1, 7, 1, 1, 2, 6 y 1. Elabore un diagrama de control del proceso y comente si está bajo control.

EJERCICIOS

9. A continuación se registra un diagrama *c* de un proceso de manufactura.



a. ¿Cuál es la media del porcentaje de defectos? ¿Cuáles son los límites de control superior e inferior?
b. ¿Hay algunas observaciones en la muestra que indiquen que el proceso está fuera de control? ¿Cuáles números de muestra son?
c. ¿Parece que hay alguna tendencia en el proceso? Es decir, ¿parece que el proceso mejora, empeora o permanece igual?

10. Inter State Moving and Storage Company establece un diagrama de control para supervisar la proporción de mudanzas residenciales que generan quejas por escrito por tardanzas, o artículos perdidos o dañados; se selecciona una muestra de 50 mudanzas de cada uno de los últimos 12 meses. El número de quejas en cada muestra es 8, 7, 4, 8, 2, 7, 11, 6, 7, 6, 8 y 12.
a. Diseñe un diagrama *p* y etiquete la media del porcentaje de defectos en el rango *LCS* y *LCI*.
b. Grafique la proporción de quejas por escrito en los últimos 12 años.
c. Interprete el diagrama. ¿Parece que el número de quejas está fuera de control en algún mes?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

11. Cada día, un fabricante de bicicletas selecciona al azar 10 cuadros y realiza pruebas para detectar defectos; el número de cuadros defectuosos determinado durante los últimos 14 días es 3, 2, 1, 3, 2, 2, 8, 2, 0, 3, 5, 2, 0 y 4. Elabore el diagrama de control de este proceso y comente si está "bajo control".

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

12. Durante el proceso de fabricar papel higiénico, Scott Paper selecciona aleatoriamente un rollo cinco veces a lo largo del día y somete a cada uno a una prueba de resistencia para ver qué tan a menudo se rasga. En un lapso de tres días, por medio de la prueba de 15 rollos se encontraron las siguientes cantidades de defectos en cada rollo: 2, 3, 1, 2, 2, 1, 3, 2, 2, 1, 2, 2, 1, 0 y 0. Elabore el diagrama de control del proceso y comente si está "bajo control".

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

13. Sam's Supermarkets prueba sus cajeros al examinar al azar los recibos impresos para detectar errores de exploración de precios. Los siguientes números corresponden a cada recibo del 27 de octubre: 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0. Elabore el diagrama de control del proceso y comente si está "bajo control".

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14. Dave Christi dirige una cadena de autolavado con sucursales en todo Chicago y le preocupa que algunos gerentes locales laven gratis los automóviles de sus amigos, por lo que decide recopilar datos sobre el número de recibos de venta "anulados"; por supuesto, algunos son legítimos, pero 3, 8, 3, 4, 6, 5, 0, 1, 2, 4, ¿indicarían un número razonable de anulaciones en sus instalaciones? Elabore un diagrama de control del proceso y comente si está "bajo control".

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

OA19-6
Analizar el muestreo de aceptación.

Muestreo de aceptación

La sección anterior se trató acerca de mantener la *calidad del producto a medida que se fabrica*. En muchas situaciones de negocios también interesa la *calidad del producto terminado que se recibe*. ¿Qué tienen en común los siguientes casos?

- Sims Software, Inc., es cliente de DVD International. La orden de compra normal es de 100 000 DVD, empacados en lotes de 1 000. Todd Sims, el presidente, no espera que todos los DVD sean perfectos; en realidad, ha aceptado lotes de 1 000 hasta con 10% de defectos, y quiere desarrollar un plan para inspeccionar los lotes que le llegan, para estar seguro de que se cumple con el estándar de calidad. El propósito del procedimiento de inspección es separar los lotes aceptables de los inaceptables.
- Zenith Electric compra tubos magnetrón de Bono Electronics para su nuevo horno de microondas. Los magnetrones se embarcan a Zenith en lotes de 10 000 unidades. Zenith permite que los lotes que recibe contengan hasta 5% de piezas defectuosas. Le gustaría elaborar un plan de muestreo para determinar los lotes que cumplen con el criterio y los que no lo hacen.
- General Motors compra parabrisas de muchos proveedores, e insiste en que los lotes sean de 1 000, y está dispuesto a aceptar 50 o menos defectos en cada lote, es decir, 5% de defectos. Le gustaría desarrollar un procedimiento de muestreo para verificar que los embarques que recibe cumplan con el criterio.

El hilo conductor en estos casos es la necesidad de verificar que un producto que llega a la planta cumpla con los requisitos estipulados. La situación es semejante a una puerta de mosquitero, que permite que entre el aire caliente del verano al recinto mientras mantiene afuera a los mosquitos. El muestreo de aceptación permite que entren los lotes con calidad aceptable al área de manufactura y se queden afuera los que no son aceptables.

Por supuesto, la situación en los negocios modernos es más compleja. El comprador quiere protección para no aceptar lotes inferiores al estándar de calidad. La mejor protección contra la calidad inferior es una inspección de 100%; desafortunadamente, con frecuencia el costo de una inspección de tal magnitud es prohibitivo. Otro problema con la verificación de cada artículo es que la prueba puede ser destructiva. Si se probaran todos los focos hasta que se fundieran antes de su embarque, no quedaría ninguno para vender; asimismo, la inspección de 100% quizá permita identificar todos los defectos, por lo tanto, en situaciones prácticas, pocas veces se lleva a cabo una inspección completa.

El procedimiento habitual es examinar la calidad de las partes de entrada mediante un plan de muestreo estadístico; así, se selecciona al azar una muestra de n unidades de los lotes de N unidades (la población); esto se denomina **muestreo de aceptación.** La inspección determinará el número de defectos que hay en la muestra; este número se compara con uno predeterminado, denominado **número crítico** o **número de aceptación**; por lo general, el número de aceptación se designa c. Si el número de defectos en la muestra de tamaño n es menor o igual a c, el lote se acepta; si el número de defectos excede c, el lote se rechaza y se regresa al proveedor, o tal vez se someta a una inspección completa.

El muestreo de aceptación es un proceso de toma de decisiones; hay dos decisiones posibles: aceptar o rechazar el lote. Además, hay dos situaciones en las cuales se toma la decisión: el lote es bueno o el lote es malo. Estos son estados de la naturaleza; si el lote es bueno y la inspección de la muestra revela que el lote es bueno, o si el lote es malo y la inspección de la muestra indica que es malo, se toma una decisión correcta; sin embargo, hay otras dos probabilidades. El lote puede contener más defectos que los aceptables, pero se acepta (a esto se denomina **riesgo del consumidor**). De manera similar, el lote puede estar dentro de los límites acordados, pero se rechaza durante la inspección de la muestra (a esto se le denomina **riesgo del productor**). En la siguiente tabla se resumen las decisiones de aceptación presentes en estas probabilidades. Observe cómo esta decisión es muy similar a las ideas de los errores de tipo I y tipo II que se vieron al inicio del capítulo 10.

	Estados de la naturaleza	
Decisión	**Lote bueno**	**Lote malo**
Aceptar el lote	Correcto	Riesgo del consumidor
Rechazar el lote	Riesgo del productor	Correcto

ESTADÍSTICA EN ACCIÓN

A finales de la década de 1980 se informó que una empresa canadiense ordenó algunas partes a una compañía japonesa con instrucciones de que no debería haber "más de tres partes defectuosas por millar". Cuando estas llegaron, había una nota que decía: "Sus tres partes defectuosas están envueltas por separado en el compartimiento superior izquierdo del embarque". Ha pasado mucho tiempo desde los días cuando "Hecho en Japón" significaba *barato, mas no confiable*.

Para evaluar un plan de muestreo y determinar que es justo tanto para el productor como para el consumidor, el procedimiento usual es desarrollar una **curva característica de operación,** o **curva CO,** como normalmente se denomina; esta reporta el porcentaje defectuoso en el eje horizontal, y la probabilidad de aceptar ese porcentaje defectuoso, en el vertical; por lo general, se traza una curva uniforme que conecta todos los niveles de calidad posibles. Se utiliza la distribución binomial para desarrollar las probabilidades de una curva CO.

EJEMPLO

Como se mencionó antes, Sims Software es cliente de DVD International. Los artículos se empacan en lotes de 1 000 cada uno. Todd Sims, presidente de la empresa, está de acuerdo en aceptar lotes con 10% o menos de DVD defectuosos. Todd indicó a su departamento de inspección que seleccione una muestra aleatoria de 20 DVD y los examine con detenimiento. Aceptará el lote si tiene dos o menos defectos en la muestra. Desarrolle la curva CO de este plan de aceptación. ¿Cuál es la probabilidad de aceptar un lote con 10% de DVD defectuosos?

SOLUCIÓN

Este tipo de muestreo se denomina **muestreo de atributos,** pues el artículo muestreado, en este caso un DVD, se clasifica como aceptable o inaceptable. No se obtiene una "lectura" o "medición" del DVD. Sea π la proporción actual defectuosa en la población.

El lote es bueno si $\pi \leq 0.10$.
El lote es malo si $\pi > 0.10$.

Sea x el número de defectos en la muestra. La regla de decisión es:

Aceptar el lote si $x \leq 2$.
Rechazar el lote si $x \geq 3$.

Aquí el lote aceptable es uno con 10% o menos de DVD defectuosos; si el lote es aceptable cuando tiene exactamente 10% de DVD defectuosos, sería aún más aceptable si contuviera menos de 10%; por lo tanto, la práctica usual es trabajar con el límite superior del porcentaje de defectos.

Mediante la distribución binomial se calculan los diversos valores en la CO. Recuerde que para emplear la distribución binomial hay cuatro requisitos:

1. Solo hay dos resultados posibles: el DVD es aceptable o inaceptable.
2. Hay un número fijo de ensayos; en este caso, es el tamaño de la muestra (20).
3. Existe una probabilidad constante de éxito; es decir, encontrar un DVD defectuoso. La probabilidad de éxito se supone de 0.10.
4. Los ensayos son independientes, y la probabilidad de obtener un DVD defectuoso en el tercero seleccionado no está relacionada con la probabilidad de encontrar un defecto en el cuarto.

En el apéndice B.1 se dan varias probabilidades binomiales; sin embargo, estas tablas solo llegan a 15, es decir, $n = 15$. En este problema $n = 20$; por lo tanto, utilice Excel para calcular las varias probabilidades binomiales. En la salida de Excel de la izquierda se muestran las probabilidades binomiales para $n = 20$ cuando π es igual a 0.05, 0.10, 0.15, 0.20, 0.25 y 0.30.

lot fraction defective.xlsx

Fracción defectuosa del lote entrante

Número de defectos	Probabilidad					
	0.05	0.10	0.15	0.20	0.25	0.30
0	0.358	0.122	0.039	0.012	0.003	0.001
1	0.377	0.270	0.137	0.058	0.021	0.007
2	0.189	0.285	0.229	0.137	0.067	0.028
3	0.060	0.190	0.243	0.205	0.134	0.072
4	0.013	0.090	0.182	0.218	0.190	0.130
5	0.002	0.032	0.103	0.175	0.202	0.179
6	0.000	0.009	0.045	0.109	0.169	0.192
7	0.000	0.002	0.016	0.055	0.112	0.164
8	0.000	0.000	0.005	0.022	0.061	0.114
9	0.000	0.000	0.001	0.007	0.027	0.065
10	0.000	0.000	0.000	0.002	0.010	0.031
11	0.000	0.000	0.000	0.000	0.003	0.012
12	0.000	0.000	0.000	0.000	0.001	0.004
13	0.000	0.000	0.000	0.000	0.000	0.001
14	0.000	0.000	0.000	0.000	0.000	0.000
•	•	•	•	•	•	•
20	0.000	0.000	0.000	0.000	0.000	0.000

Los términos del capítulo 6 se deben traducir al vocabulario de muestreo de aceptación: π representa la probabilidad de encontrar un defecto, c el número de defectos permitidos, y n el número de artículos muestreados. En este caso, permitirá hasta dos defectos, por lo que $c = 2$; esto significa que 0, 1 o 2 de los 20 artículos muestreados pueden ser defectuosos y aun así se aceptaría el embarque de entrada de DVD.

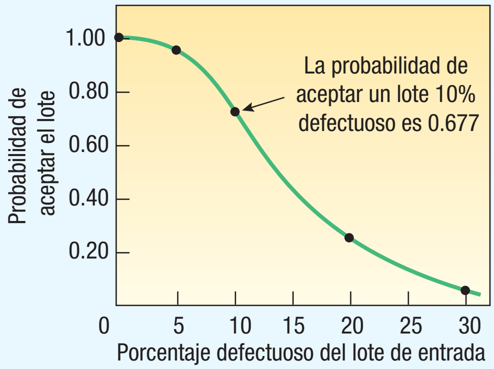
Para empezar, determine la probabilidad de aceptar un lote que sea 5% defectuoso; esto significa que $\pi = 0.05$, $c = 2$ y $n = 20$. De la salida de Excel, la probabilidad de seleccionar una muestra de 20 artículos de un embarque con 5% de defectos y encontrar exactamente 0 defectos es 0.358. La posibilidad de encontrar exactamente 1 defecto es 0.377, y la de encontrar 2 es 0.189. De aquí que la probabilidad de 2 o menos defectos sea 0.924, que se determina mediante 0.358 + 0.377 + 0.189; este resultado por lo general se escribe en notación abreviada, como sigue (recuerde que la barra "|" significa "dado que").

$$P(x \leq 2 \mid \pi = 0.5 \text{ y } n = 20) = 0.358 + 0.377 + 0.189 = 0.924$$

La probabilidad de aceptar un lote que en realidad tiene 10% de defectos es 0.677; es decir,

$$P(x \leq 2 \mid \pi = 0.10 \text{ y } n = 20) = 0.122 + 0.270 + 0.285 = 0.677$$

En la curva CO completa que se registra en la gráfica 19.8 se muestra la curva uniformizada para todos los valores de π entre 0 y casi 30%. No hay necesidad de mostrar los que son mayores que 30% porque su probabilidad es muy cercana a 0. La posibilidad de aceptar lotes con niveles de calidad seleccionados se indica en forma de tabla a la derecha de la gráfica 19.8. Con la curva CO, la gerencia de Sims Software podrá evaluar con rapidez las probabilidades de varios niveles de calidad.



Porcentaje defectuoso del lote	Probabilidad de aceptar el lote
0	1.000
5	0.924
10	0.677
20	0.207
30	0.036

GRÁFICA 19.8 Curva CO del plan de muestreo ($n = 20$, $c = 2$)

AUTOEVALUACIÓN 19-4

Calcule la probabilidad de aceptar un lote de DVD con 30% de artículos defectuosos, con el plan de muestreo de Sims Software.

EJERCICIOS

15. Determine la probabilidad de aceptar lotes con 10%, 20%, 30% y 40% de DVD defectuosos, una muestra de tamaño 12 y un número de aceptación de 2.

16. Determine la probabilidad de aceptar lotes con 10%, 20%, 30% y 40% de DVD defectuosos, una muestra de tamaño 14 y un número de aceptación de 3.

17. Warren Electric fabrica fusibles para muchos clientes; para asegurar la calidad del producto de salida, prueba 10 fusibles cada hora; si no más de un fusible es defectuoso, los empaca y prepara para su embarque. Desarrolle la curva CO de este plan de muestreo y calcule las probabilidades de aceptar lotes con 10%, 20%, 30% y 40% de unidades defectuosas. Trace la curva CO de este plan de muestreo con los cuatro niveles de calidad.

18. Grills Radio Products compra transistores de Mira Electronics. De acuerdo con su plan de muestreo, el propietario, Art Grills, aceptará un embarque de transistores si tres o menos son defectuosos en una muestra de 25. Elabore la curva CO de estos porcentajes de defectos: 10%, 20%, 30% y 40% (necesitará un paquete de software estadístico).

RESUMEN DEL CAPÍTULO

I. El objetivo del control estadístico de calidad es seguir de cerca la calidad del producto o servicio a medida que se elabora.

II. El diagrama de Pareto es una técnica para contar el número y tipo de defectos que se presentan en un producto o servicio.

A. A esta gráfica se le llama así en honor del científico italiano Vilfredo Pareto.

B. El concepto del diagrama es que 20% de los factores ocasiona 80% de la actividad.

III. Un diagrama de esqueleto de pez destaca la relación entre una posible causa de un problema que producirá el efecto particular.

A. También se denomina diagrama de causa y efecto.

B. El enfoque habitual es considerar cuatro áreas del problema: métodos, materiales, equipamiento y personal.

IV. El propósito de un diagrama de control es supervisar la calidad de un producto o servicio.

A. Hay dos tipos de diagramas de control.

1. Un diagrama de control de variables es el resultado de una medición.

2. Un diagrama de atributos indica si el producto o servicio es aceptable o no.

B. Existen dos fuentes de variación de la calidad de un producto o servicio.

1. La variación casual es de naturaleza aleatoria y no se puede controlar o eliminar.

2. La variación asignable no es por causas aleatorias y se puede eliminar.

C. En este capítulo se consideraron cuatro gráficas de control.

1. En una gráfica de la media se indica la media de una variable, y en una gráfica de rangos se presenta el rango de la variable.

a. Los límites de control superior e inferior se determinan en más o menos tres desviaciones estándar de la media.

b. Las fórmulas de los límites de control superior e inferior de la media son:

$$LCS = \bar{\bar{x}} + A_2\bar{R} \qquad LCI = \bar{\bar{x}} + A_2\bar{R} \qquad \textbf{[19.4]}$$

c. Las fórmulas de los límites de control superior e inferior del rango son:

$$LCS = D_4\bar{R} \qquad LCI = D_3\bar{R} \qquad \textbf{[19.5]}$$

2. Un diagrama p es un diagrama de atributos que presenta la proporción del producto o servicio que no cumple con el estándar.

a. El porcentaje defectuoso medio se determina mediante

$$p = \frac{\text{Número total de defectos}}{\text{Número total de artículos muestreados}} \qquad \textbf{[19.6]}$$

b. Los límites de control de la proporción defectuosa se estiman a partir de la ecuación

$$LCI, LCS = p \pm 3\sqrt{\frac{p(1-p)}{n}} \qquad \textbf{[19.8]}$$

3. Una gráfica de líneas c se refiere al número de defectos por unidad.

a. Se basa en la distribución de Poisson.

b. El número medio de defectos por unidad es $\bar{c}$.

c. Los límites de control se determinan a partir de la siguiente ecuación.

$$LCI, LCS = \bar{c} \pm 3\sqrt{\bar{c}} \qquad \textbf{[19.9]}$$

V. El muestreo de aceptación es un método para determinar si el lote de entrada de un producto cumple con los estándares especificados.

A. Se basa en técnicas de muestreo aleatorio.

B. Se selecciona una muestra de n unidades de una población de N unidades.

C. El número máximo de unidades defectuosas que se pueden encontrar en la muestra de n unidades, y aún considerar aceptable el lote, es c.

D. Una curva CO (característica de operación) se elabora con la distribución de probabilidad binomial para determinar la probabilidad de aceptar lotes con varios niveles de calidad.

CLAVE DE PRONUNCIACIÓN

Símbolo	**Significado**	**Pronunciación**
$\bar{\bar{x}}$	Media de las medias muestrales	*x doble barra*
$s_{\bar{x}}$	Error estándar de la media	*s subíndice x barra*
A_2	Constante de determinar los límites de control superior e inferior de la media	*A subíndice 2*
$\bar{R}$	Media de los rangos de las muestras	*R barra*
D_4	Constante para determinar el límite de control superior del rango	*D subíndice 4*
$\bar{c}$	Número medio de defectos por unidad	*c barra*

EJERCICIOS DEL CAPÍTULO

19. El supervisor de producción de Westburg Electric, Inc., observó un incremento del número de motores eléctricos rechazados en el momento de la inspección final. De los últimos 200 motores rechazados, 80 defectos se debieron a un cableado deficiente, 60 tenían un cortocircuito en la bobina, 50 bujías defectuosas y 10 padecían otras fallas. Desarrolle un diagrama de Pareto que muestre las principales áreas problemáticas.

20. Un fabricante de zapatos deportivos realizó un estudio acerca de sus nuevos zapatos para trotar. A continuación se mencionan tanto el tipo como la frecuencia de las discrepancias y fallas que se encontraron. Desarrolle el diagrama de Pareto que indique las principales áreas problemáticas.

Tipo de discrepancia	Frecuencia	Tipo de discrepancia	Frecuencia
Separación de la suela	34	Ruptura de agujetas	14
Separación del tacón	98	Defecto en ojal	10
Abertura en la suela	62	Otro	16

21. En Rumsey's Old Fashion las bebidas gaseosas se sirven con una máquina automática cuya operación se basa en el peso de la bebida. Cuando el proceso está bajo control, la máquina llena cada vaso de modo que la media total es de 10.0 onzas y el rango medio de 0.25 en el caso de muestras de cinco.

a. Determine los límites de control superior e inferior del proceso tanto de la media como del rango.

b. El gerente de la tienda I-280 probó cinco bebidas gaseosas servidas la hora pasada y encontró que la media fue de 10.16 onzas, y el rango, de 0.35 onzas. ¿Está bajo control el proceso? ¿Debe tomarse otra acción?

22. Recientemente se instaló una máquina nueva para cortar y desbastar piezas grandes; luego estas se transfieren a una pulidora de precisión. Una de las medidas críticas es el diámetro exterior. El inspector de calidad selecciona al azar cinco piezas cada media hora, mide el diámetro exterior y registra los resultados. Las mediciones (en milímetros) del periodo de las 8:00 a.m. a las 10:30 a.m. se registran a continuación.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

	Diámetro exterior (milímetros)				
Hora	1	2	3	4	5
8:00	87.1	87.3	87.9	87.0	87.0
8:30	86.9	88.5	87.6	87.5	87.4
9:00	87.5	88.4	86.9	87.6	88.2
9:30	86.0	88.0	87.2	87.6	87.1
10:00	87.1	87.1	87.1	87.1	87.1
10:30	88.0	86.2	87.4	87.3	87.8

a. Determine los límites de control de la media y del rango.

b. Trace los límites de control del diámetro exterior medio y del rango.

c. ¿Hay algunos puntos en la gráfica de la media o del rango fuera de control? Comente la gráfica.

23. Long Last Company, como parte de su proceso de inspección, prueba sus neumáticos para verificar el desgaste del área de contacto en condiciones de caminos simulados; se seleccionaron 20 muestras de tres neumáticos de turnos distintos durante el mes previo. El desgaste del área de contacto se muestra a continuación, en centésimos de pulgada.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

Muestra	Desgaste del área de contacto			Muestra	Desgaste del área de contacto		
1	44	41	19	11	11	33	34
2	39	31	21	12	51	34	39
3	38	16	25	13	30	16	30
4	20	33	26	14	22	21	35
5	34	33	36	15	11	28	38
6	28	23	39	16	49	25	36
7	40	15	34	17	20	31	33
8	36	36	34	18	26	18	36
9	32	29	30	19	26	47	26
10	29	38	34	20	34	29	32

a. Determine los límites de control de la media y del rango.
b. Trace los límites de control del desgaste del área de contacto medio y del rango.
c. ¿Hay algunos puntos en la gráfica de la media o del rango "fuera de control"? Comente la gráfica.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

24. Charter National Bank tiene un grupo de ejecutivos de préstamos en sus sucursales de todo el suroeste de Estados Unidos. Robert Kerns, vicepresidente de préstamos, quiere obtener información sobre la cantidad común de los préstamos y el rango de la cantidad de estos; su analista de personal seleccionó una muestra de 10 ejecutivos así como una muestra de cinco préstamos que cada uno de ellos otorgó el mes anterior. Los datos se registran en la siguiente tabla. Elabore una gráfica de control de la media y del rango. ¿Parece que alguno de los ejecutivos está "fuera de control"? Comente sus resultados.

	Cantidad del préstamo (miles de dólares)						Cantidad del préstamo (miles de dólares)				
Ejecutivo	**1**	**2**	**3**	**4**	**5**	**Ejecutivo**	**1**	**2**	**3**	**4**	**5**
Weinraub	59	74	53	48	65	Bowyer	66	80	54	68	52
Visser	42	51	70	47	67	Kuhlman	74	43	45	65	49
Moore	52	42	53	87	85	Ludwig	75	53	68	50	31
Brunner	36	70	62	44	79	Longnecker	42	65	70	41	52
Wolf	34	59	39	78	61	Simonetti	43	38	10	19	47

25. El fabricante de una barra de dulce, llamada "King James", informa en el paquete que el contenido calórico de una barra de dos onzas es de 420 unidades. Una muestra de cinco barras de cada uno de los últimos 10 días se somete a un análisis químico de contenido calórico. Los resultados se muestran en la siguiente tabla. ¿Parece que hay algunos días en los cuales el conteo de las calorías está fuera de control? Desarrolle una gráfica de control apropiada y analice sus resultados.

	Conteo calórico						Conteo calórico				
Muestra	**1**	**2**	**3**	**4**	**5**	**Muestra**	**1**	**2**	**3**	**4**	**5**
1	426	406	418	431	432	6	427	417	408	418	422
2	421	422	415	412	411	7	422	417	426	435	426
3	425	420	406	409	414	8	419	417	412	415	417
4	424	419	402	400	417	9	417	432	417	416	422
5	421	408	423	410	421	10	420	422	421	415	422

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

26. Early Mornig Delivery Service garantiza la entrega de paquetes pequeños a las 10:30 a.m. Por supuesto, algunos paquetes no se entregan a esa hora. En una muestra de 200 paquetes entregados cada uno de los últimos 15 días laborables, el siguiente número de paquetes se entregó después del límite de tiempo: 9, 14, 2, 13, 9, 5, 9, 3, 4, 3, 4, 3, 3, 8 y 4.
a. Estime la proporción media de los paquetes que se entregaron después de las 10:30 a.m.
b. Estime los límites de control de la proporción de paquetes que se entregaron después de las 10:30 a.m. ¿Hubo algunos días muestreados fuera de control?
c. En una muestra, si 10 paquetes de 200 se entregaron hoy después de las 10:30 a.m., ¿la muestra está dentro de los límites de control?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

27. Una máquina automática produce pernos de cinco milímetros a alta velocidad; se inició un programa de control de calidad para controlar el número de pernos defectuosos. El inspector de control de calidad selecciona 50 pernos al azar y determina cuántos son defectuosos. El número de pernos defectuosos en la primera de 10 muestras es 3, 5, 0, 4, 1, 2, 6, 5, 7 y 7.
a. Diseñe un diagrama *p*. Intercale el porcentaje medio defectuoso entre *LCS* Y *LCI*.
b. Trace en el diagrama el porcentaje defectuoso de las primeras 10 muestras.
c. Interprete el diagrama.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

28. Steele Breakfast Foods, Inc., produce una popular marca de cereal de salvado con pasas. El paquete indica que contiene 25.0 onzas de cereal; para asegurar la calidad, el departamento de control de calidad de Steele verifica cada hora el proceso de producción. Como parte de la verificación, se seleccionan cuatro cajas de cereal para pesar su contenido. En la página siguiente se muestran los resultados.

Muestra	Pesos				Muestra	Pesos			
1	26.1	24.4	25.6	25.2	14	23.1	23.3	24.4	24.7
2	25.2	25.9	25.1	24.8	15	24.6	25.1	24.0	25.3
3	25.6	24.5	25.7	25.1	16	24.4	24.4	22.8	23.4
4	25.5	26.8	25.1	25.0	17	25.1	24.1	23.9	26.2
5	25.2	25.2	26.3	25.7	18	24.5	24.5	26.0	26.2
6	26.6	24.1	25.5	24.0	19	25.3	27.5	24.3	25.5
7	27.6	26.0	24.9	25.3	20	24.6	25.3	25.5	24.3
8	24.5	23.1	23.9	24.7	21	24.9	24.4	25.4	24.8
9	24.1	25.0	23.5	24.9	22	25.7	24.6	26.8	26.9
10	25.8	25.7	24.3	27.3	23	24.8	24.3	25.0	27.2
11	22.5	23.0	23.7	24.0	24	25.4	25.9	26.6	24.8
12	24.5	24.8	23.2	24.2	25	26.2	23.5	23.7	25.0
13	24.4	24.5	25.9	25.5					

Elabore un diagrama de control apropiado. ¿Cuáles son los límites? ¿Está fuera de control el proceso en algún momento?

29. Un inversionista considera que hay una probabilidad de 50% de que una acción suba o baje en un día en particular; para investigar esta idea, durante 30 días consecutivos el inversionista selecciona una muestra de 50 acciones y cuenta el número de veces que aumenta. A continuación se registra el número de acciones de la muestra que aumentaron.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

14	12	13	17	10	18	10	13	13	14
13	10	12	11	9	13	14	11	12	11
15	13	10	16	10	11	12	15	13	10

Elabore un diagrama *p* y resuma sus resultados en un reporte breve; con base en los resultados, ¿es razonable concluir que las probabilidades de que la acción aumente son de 50%? ¿Qué porcentaje de las acciones necesitaría subir en un día para que el proceso esté "fuera de control"?

30. Lahey Motors se especializa en vender automóviles a compradores con un historial crediticio deficiente. He aquí los números de automóviles que se recuperaron debido a que los clientes de Lahey no cumplieron con sus pagos durante los últimos 36 meses.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

6	5	8	20	11	10	9	3	9	9
15	12	4	11	9	9	6	18	6	8
9	7	13	7	11	8	11	13	6	14
13	5	5	8	10	11				

Elabore un diagrama de líneas *c* del número de recuperaciones. ¿Hubo algunos meses en que el número estuvo fuera de control? Resuma sus resultados en un reporte breve.

31. Un ingeniero de proceso considera dos planes de muestreo. De acuerdo con el primero seleccionará una muestra de 10 y aceptará el lote si 3 o menos son defectuosas. En el segundo, el tamaño de la muestra es 20, y el número de aceptación, 5. Elabore la curva CO de cada uno. Compare la probabilidad de aceptación de lotes con 5%, 10%, 20% y 30% de unidades defectuosas; si usted fuera el proveedor, ¿qué plan recomendaría?

32. Christina Sanders es miembro del equipo femenil de basquetbol del Windy City College. La temporada pasada anotó 55% de sus intentos de tiros libres. En un esfuerzo por mejorar dicha estadística, asistió a un curso de verano dedicado a enseñar técnicas de tiros libres. Los siguientes 20 días hizo 100 tiros libres al día y, con minuciosidad, registró el número de tiros anotados cada día. He aquí los resultados.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

55	61	52	59	67	57	61	59	69	58
57	66	63	63	63	65	63	68	64	67

La tabla se interpreta así: el primer día anotó 55 tiros de 100, o 55%; el último día anotó 67 de 100, o 67%.

a. Elabore el diagrama de control de los tiros anotados. Durante los 20 días de práctica, ¿cuál fue el porcentaje de tiros que anotó? ¿Cuáles son los límites de control superior e inferior de la proporción de tiros anotados?

b. ¿Hay alguna tendencia en su proporción de tiros anotados? ¿Parece mejorar, empeorar o permanece igual?

c. Encuentre el porcentaje de intentos anotados durante los últimos cinco días de práctica. Utilice el procedimiento de prueba de hipótesis, fórmula [10.4], para determinar si hay una mejora a partir de 55%.

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

33. Eric's Cookie House vende galletas con chispas de chocolate en centros comerciales, y le interesa conocer el número de chispas de chocolate en cada galleta. Eric, propietario y presidente, quiere establecer un diagrama de control del número de chispas por galleta, para lo cual selecciona una muestra de 15 unidades de la producción de hoy y cuenta el número de chispas en cada una de ellas. A continuación se muestran los resultados: 6, 8, 20, 12, 20, 19, 11, 23, 12, 14, 15, 16, 12, 13 y 12.

a. Determine la línea central y los límites de control.

b. Desarrolle un diagrama de control y trace el número de chispas de chocolate por galleta.

c. Interprete el diagrama. ¿Parece que el número de chispas de chocolate está fuera de control en alguna de las galletas muestreadas?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

34. El número de ocasiones en que "los pasajeros casi pierden el vuelo" durante los últimos 20 meses en el Aeropuerto Internacional de Lima, Perú, es 3, 2, 3, 2, 2, 3, 5, 1, 2, 2, 4, 4, 2, 6, 3, 5, 2, 5, 1 y 3. Desarrolle un diagrama de control apropiado, determine el número medio de pasajeros que casi pierden el vuelo por mes y los límites del número de pasajeros que casi pierden el vuelo por mes. ¿Hay algún mes en que el número de pasajeros que casi pierden el vuelo esté fuera de control?

Para la **BASE DE DATOS** visite **www.mhhe.com/uni/lind_ae16e**

35. A continuación se indica el número de robos que se reportaron durante los últimos 10 días a la División de Robos de Metro City Police: 10, 8, 8, 7, 8, 5, 8, 5, 4 y 7. Elabore un diagrama de control apropiado, determine el número medio de robos reportado por día y los límites de control. ¿Hay días en que el número de robos reportado esté fuera de control?

36. Swiss Watches, Ltd., compra vástagos para relojes en lotes de 10 000; su plan de muestreo requiere 20 vástagos, y si tres o menos son defectuosos, se acepta el lote.

a. Con base en el plan de muestreo, ¿cuál es la probabilidad que se acepte un lote con 40% de defectos?

b. Diseñe la curva CO de lotes de entrada que tenga 0%, 10%, 20%, 30% y 40% de vástagos defectuosos.

37. Automatic Screen Door Manufacturing compra picaportes a diversos proveedores; el departamento de compras es el responsable de inspeccionar los picaportes de entrada. La compañía compra 10 000 picaportes por mes e inspecciona 20 al azar. Elabore una curva OC del plan de muestreo si tres picaportes son defectuosos y aún se acepta el lote de entrada.

38. Al inicio de cada temporada de futbol, Team Sports, tienda local de artículos deportivos, compra 5 000 balones; se selecciona una muestra de 25 balones y se inflan, prueban y luego se desinflan. Si más de dos son defectuosos, todo el lote se regresa al fabricante. Elabore la curva OC de este plan de muestreo.

a. ¿Cuáles son las probabilidades de aceptar lotes con 10%, 20%, 30% de unidades defectuosas?

b. Estime la probabilidad de aceptar un lote con 15% de unidades defectuosas.

c. John Brennen, propietario de Team Sports, quiere que la probabilidad de aceptar un lote con 5% de defectos sea de 90%. ¿Parece ser el caso con este plan de muestreo?

Apéndices

APÉNDICE A: CONJUNTO DE DATOS

APÉNDICE B: TABLAS

APÉNDICE C: COMANDOS DE SOFTWARE

APÉNDICE D: RESPUESTAS A LOS EJERCICIOS IMPARES, EJERCICIOS DE REPASO Y SOLUCIONES A LAS PRUEBAS DE PRÁCTICA

APÉNDICE E: RESPUESTAS A LAS AUTOEVALUACIONES

A.1 Conjunto de datos 1: Ventas inmobiliarias de Goodyear, Arizona

Variables

x_1 = Precio de venta (miles de dólares)
x_2 = Número de recámaras
x_3 = Tamaño de la casa en pies cuadrados
x_4 = Alberca (1 = sí; 0 = no)
x_5 = Distancia del centro de la ciudad en millas
x_6 = Colonia
x_7 = Cochera (1 = sí; 0 = no)
x_8 = Número de baños
105 casas vendidas

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
263.1	4	2 300	0	17	5	1	2.0
182.4	4	2 100	1	19	4	0	2.0
242.1	3	2 300	1	12	3	0	2.0
213.6	2	2 200	1	16	2	0	2.5
139.9	2	2 100	1	28	1	0	1.5
245.4	2	2 100	0	12	1	1	2.0
327.2	6	2 500	1	15	3	1	2.0
271.8	2	2 100	1	9	2	1	2.5
221.1	3	2 300	0	18	1	0	1.5
266.6	4	2 400	1	13	4	1	2.0
292.4	4	2 100	1	14	3	1	2.0
209.0	2	1 700	1	8	4	1	1.5
270.8	6	2 500	1	7	4	1	2.0
246.1	4	2 100	1	18	3	1	2.0
194.4	2	2 300	1	11	3	0	2.0
281.3	3	2 100	1	16	2	1	2.0
172.7	4	2 200	0	16	3	0	2.0
207.5	5	2 300	0	21	4	0	2.5
198.9	3	2 200	0	10	4	1	2.0
209.3	6	1 900	0	15	4	1	2.0
252.3	4	2 600	1	8	4	1	2.0
192.9	4	1 900	0	14	2	1	2.5
209.3	5	2 100	1	20	5	0	1.5
345.3	8	2 600	1	9	4	1	2.0
326.3	6	2 100	1	11	5	1	3.0
173.1	2	2 200	0	21	5	1	1.5
187.0	2	1 900	1	26	4	0	2.0
257.2	2	2 100	1	9	4	1	2.0
233.0	3	2 200	1	14	3	1	1.5
180.4	2	2 000	1	11	5	0	2.0
234.0	2	1 700	1	19	3	1	2.0
207.1	2	2 000	1	11	5	1	2.0
247.7	5	2 400	1	16	2	1	2.0
166.2	3	2 000	0	16	2	1	2.0
177.1	2	1 900	1	10	5	1	2.0
182.7	4	2 000	0	14	4	0	2.5
216.0	4	2 300	1	19	2	0	2.0
312.1	6	2 600	1	7	5	1	2.5
199.8	3	2 100	1	19	3	1	2.0
273.2	5	2 200	1	16	2	1	3.0

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
206.0	3	2 100	0	9	3	0	1.5
232.2	3	1 900	0	16	1	1	1.5
198.3	4	2 100	0	19	1	1	1.5
205.1	3	2 000	0	20	4	0	2.0
175.6	4	2 300	0	24	4	1	2.0
307.8	3	2 400	0	21	2	1	3.0
269.2	5	2 200	1	8	5	1	3.0
224.8	3	2 200	1	17	1	1	2.5
171.6	3	2 000	0	16	4	0	2.0
216.8	3	2 200	1	15	1	1	2.0
192.6	6	2 200	0	14	1	0	2.0
236.4	5	2 200	1	20	3	1	2.0
172.4	3	2 200	1	23	3	0	2.0
251.4	3	1 900	1	12	2	1	2.0
246.0	6	2 300	1	7	3	1	3.0
147.4	6	1 700	0	12	1	0	2.0
176.0	4	2 200	1	15	1	1	2.0
228.4	3	2 300	1	17	5	1	1.5
166.5	3	1 600	0	19	3	0	2.5
189.4	4	2 200	1	24	1	1	2.0
312.1	7	2 400	1	13	3	1	3.0
289.8	6	2 000	1	21	3	1	3.0
269.9	5	2 200	0	11	4	1	2.5
154.3	2	2 000	1	13	2	0	2.0
222.1	2	2 100	1	9	5	1	2.0
209.7	5	2 200	0	13	2	1	2.0
190.9	3	2 200	0	18	3	1	2.0
254.3	4	2 500	0	15	3	1	2.0
207.5	3	2 100	0	10	2	0	2.0
209.7	4	2 200	0	19	2	1	2.0
294.0	2	2 100	1	13	2	1	2.5
176.3	2	2 000	0	17	3	0	2.0
294.3	7	2 400	1	8	4	1	2.0
224.0	3	1 900	0	6	1	1	2.0
125.0	2	1 900	1	18	4	0	1.5
236.8	4	2 600	0	17	5	1	2.0
164.1	4	2 300	1	19	4	0	2.0
217.8	3	2 500	1	12	3	0	2.0
192.2	2	2 400	1	16	2	0	2.5
125.9	2	2 400	1	28	1	0	1.5

(*continúa*)

A.1 Conjunto de datos 1: Ventas inmobiliarias de Goodyear, Arizona (*continuación*)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
220.9	2	2 300	0	12	1	1	2.0
294.5	6	2 700	1	15	3	1	2.0
244.6	2	2 300	1	9	2	1	2.5
199.0	3	2 500	0	18	1	0	1.5
240.0	4	2 600	1	13	4	1	2.0
263.2	4	2 300	1	14	3	1	2.0
188.1	2	1 900	1	8	4	1	1.5
243.7	6	2 700	1	7	4	1	2.0
221.5	4	2 300	1	18	3	1	2.0
175.0	2	2 500	1	11	3	0	2.0
253.2	3	2 300	1	16	2	1	2.0
155.4	4	2 400	0	16	3	0	2.0
186.7	5	2 500	0	21	4	0	2.5
179.0	3	2 400	0	10	4	1	2.0
188.3	6	2 100	0	15	4	1	2.0
227.1	4	2 900	1	8	4	1	2.0
173.6	4	2 100	0	14	2	1	2.5
188.3	5	2 300	1	20	5	0	1.5
310.8	8	2 900	1	9	4	1	2.0
293.7	6	2 400	1	11	5	1	3.0
179.0	3	2 400	1	8	4	1	2.0
188.3	6	2 100	0	14	2	1	2.5
227.1	4	2 900	1	20	5	0	1.5
173.6	4	2 100	1	9	4	1	2.0
188.3	5	2 300	1	11	5	1	3.0

A.2 Conjunto de datos 2: Ligas Mayores de Béisbol, temporada 2012

Variables

x_1 = Equipo
x_2 = Liga (Americana = 1, Nacional = 0)
x_3 = Construcción (año en que se construyó el estadio)
x_4 = Tamaño (capacidad del estadio)
x_5 = Salario (total del equipo, en 2012, en millones de dólares)
x_6 = Victorias
x_7 = Asistencia (total anual del equipo)
x_8 = BA (promedio de bateo del equipo)
x_9 = ERA (promedio de carreras)
x_{10} = HR (cuadrangulares)
x_{11} = Errores
x_{12} = SB (bases robadas)
x_{13} = Año
x_{14} = Salario promedio por jugador (en dólares)

(*continúa*)

A.2 Conjunto de datos 2: Ligas Mayores de Béisbol, temporada 2012 (*continuación*)

Equipo x_1	Liga x_2	Construcción x_3	Tamaño x_4	Salario en 2012 x_5	Victorias x_6	Asistencia x_7	BA x_8	ERA x_9	HR x_{10}	Errores x_{11}	SB x_{12}	Salario promedio	
Arizona Diamondbacks	0	1998	48 633	74.3	81	2.18	0.26	3.93	165	90	93	1989	$ 512 930
Atlanta Braves	0	1996	49 586	83.3	94	2.42	0.25	3.42	149	86	101	1990	578 930
Baltimore Orioles	1	1992	45 971	81.4	93	2.1	0.25	3.9	214	106	58	1991	891 188
Boston Red Sox	1	1912	37 495	173.2	69	3.04	0.26	4.7	165	101	97	1992	1 084 408
Chicago Cubs	0	1914	41 009	88.2	61	2.88	0.24	4.51	137	105	94	1993	1 120 254
Chicago White Sox	1	1991	40 615	96.9	85	1.97	0.26	4.02	211	70	109	1994	1 188 679
Cincinnati Reds	0	2003	42 319	82.2	97	2.35	0.25	3.34	172	89	87	1995	1 071 029
Cleveland Indians	1	1994	43 429	78.4	68	1.6	0.25	4.78	136	96	110	1996	1 176 967
Colorado Rockies	0	1995	50 398	78.1	64	2.63	0.27	5.22	166	122	100	1997	1 383 578
Detroit Tigers	1	2000	41 255	132.3	88	3.03	0.27	3.75	163	99	59	1998	1 441 406
Houston Astros	0	2000	40 981	60.7	55	1.61	0.24	4.56	146	118	105	1999	1 720 050
Kansas City Royals	1	1973	37 903	60.9	72	1.74	0.27	4.3	131	113	132	2000	1 988 034
Los Angeles Angels	1	1966	45 957	154.5	89	3.06	0.27	4.02	187	98	134	2001	2 264 403
Los Angeles Dodgers	0	1962	56 000	95.1	86	3.32	0.25	3.34	116	98	104	2002	2 383 235
Miami Marlins	0	2012	36 742	118.1	69	2.22	0.24	4.09	137	103	149	2003	2 555 476
Milwaukee Brewers	0	2001	41 900	97.7	83	2.83	0.26	4.22	202	99	158	2004	2 486 609
Minnesota Twins	1	2010	39 504	94.1	66	2.78	0.26	4.77	131	107	135	2005	2 632 655
New York Mets	0	2009	41 922	93.4	95	2.24	0.25	4.09	139	101	79	2006	2 866 544
New York Yankees	1	2009	50 287	198	74	3.54	0.27	3.85	245	74	93	2007	2 944 556
Oakland Athletics	1	1966	35 067	55.4	94	1.68	0.24	3.48	195	111	122	2008	3 154 845
Philadelphia Phillies	0	2004	43 651	174.5	81	3.57	0.26	3.83	158	101	116	2009	3 240 206
Pittsburgh Pirates	0	2001	38 362	63.4	79	2.09	0.24	3.86	170	112	73	2010	3 297 828
San Diego Padres	0	2004	42 691	55.2	76	2.12	0.25	4.01	121	121	155	2011	3 305 393
San Francisco Giants	0	2000	41 915	117.6	94	3.38	0.27	3.68	103	115	118	2012	3 440 000
Seattle Mariners	1	1999	47 860	82	75	1.72	0.23	3.76	149	72	104		
St. Louis Cardinals	0	2006	43 975	110.3	88	3.26	0.27	3.71	159	107	91		
Tampa Bay Rays	1	1990	34 078	64.2	90	1.56	0.24	3.19	175	114	134		
Texas Rangers	1	1994	48 194	120.5	93	3.46	0.27	3.99	200	85	91		
Toronto Blue Jays	1	1989	49 260	75.5	73	2.1	0.25	4.64	198	101	123		
Washington Nationals	0	2008	41 487	81.3	98	2.37	0.26	3.33	194	94	105		

A.3 Conjunto de datos 3: Autobuses del Distrito Escolar Buena

Variables

x_1 = Número de autobuses
x_2 = Costo de mantenimiento (en dólares)
x_3 = Edad
x_4 = Millas
x_5 = Tipo de autobús (diésel o gasolina)
x_6 = Fabricante (Bluebird, Keiser, Thompson)
x_7 = Pasajeros

Número de autobuses x_1	Costo de mantenimiento x_2	Años de uso x_3	Millas x_4	Tipo de autobuses x_5	Fabricante x_6	Pasajeros x_7
135	329	7	853	Diésel	Bluebird	55
120	503	10	883	Diésel	Keiser	42
200	505	10	822	Diésel	Bluebird	55
40	466	10	865	Gasolina	Bluebird	55
427	359	7	751	Gasolina	Keiser	55
759	546	8	870	Diésel	Keiser	55
10	427	5	780	Gasolina	Keiser	14
880	474	9	857	Gasolina	Keiser	55
481	382	3	818	Gasolina	Keiser	6
387	422	8	869	Gasolina	Bluebird	55
326	433	9	848	Diésel	Bluebird	55
861	474	10	845	Gasolina	Bluebird	55
122	558	10	885	Gasolina	Bluebird	55
156	561	12	838	Diésel	Thompson	55
887	357	8	760	Diésel	Bluebird	6
686	329	3	741	Diésel	Bluebird	55
490	497	10	859	Gasolina	Bluebird	55
370	459	8	826	Gasolina	Keiser	55
464	355	3	806	Gasolina	Bluebird	55
875	489	9	858	Diésel	Bluebird	55
883	436	2	785	Gasolina	Bluebird	55
57	455	7	828	Diésel	Bluebird	55
482	514	11	980	Gasolina	Bluebird	55
704	503	8	857	Diésel	Bluebird	55
989	380	9	803	Diésel	Keiser	55
731	432	6	819	Diésel	Bluebird	42
75	478	6	821	Diésel	Bluebird	55
162	406	3	798	Gasolina	Keiser	55
732	471	9	815	Diésel	Keiser	42
751	444	2	757	Diésel	Keiser	14
600	493	10	1 008	Diésel	Bluebird	55
948	452	9	831	Diésel	Keiser	42
358	461	6	849	Diésel	Bluebird	55
833	496	8	839	Diésel	Thompson	55
692	469	8	812	Diésel	Bluebird	55
61	442	9	809	Diésel	Keiser	55
9	414	4	864	Gasolina	Keiser	55
314	459	11	859	Diésel	Thompson	6
396	457	2	815	Diésel	Thompson	55
365	462	6	799	Diésel	Keiser	55

(*continúa*)

A.3 Conjunto de datos 3: Autobuses del Distrito Escolar Buena (*continuación*)

Número de autobuses x_1	Costo de mantenimiento x_2	Años de uso x_3	Millas x_4	Tipo de autobuses x_5	Fabricante x_6	Pasajeros x_7
398	570	9	844	Diésel	Thompson	14
43	439	9	832	Gasolina	Bluebird	55
500	369	5	842	Gasolina	Bluebird	55
279	390	2	792	Diésel	Bluebird	55
693	469	9	775	Gasolina	Keiser	55
884	381	9	882	Diésel	Bluebird	55
977	501	7	874	Diésel	Bluebird	55
38	432	6	837	Gasolina	Keiser	14
725	392	5	774	Diésel	Bluebird	55
982	441	1	823	Diésel	Bluebird	55
724	448	8	790	Diésel	Keiser	42
603	468	4	800	Diésel	Keiser	14
168	467	7	827	Gasolina	Thompson	55
45	478	6	830	Diésel	Keiser	55
754	515	14	895	Diésel	Keiser	14
39	411	6	804	Gasolina	Bluebird	55
671	504	8	866	Gasolina	Thompson	55
418	504	9	842	Diésel	Bluebird	55
984	392	8	851	Diésel	Bluebird	55
953	423	10	835	Diésel	Bluebird	55
507	410	7	866	Diésel	Bluebird	55
540	529	4	846	Gasolina	Bluebird	55
695	477	2	802	Diésel	Bluebird	55
193	540	11	847	Diésel	Thompson	55
321	450	6	856	Diésel	Bluebird	6
918	390	5	799	Diésel	Bluebird	55
101	424	4	827	Diésel	Bluebird	55
714	433	7	817	Diésel	Bluebird	42
678	428	7	842	Diésel	Keiser	55
768	494	7	815	Diésel	Bluebird	42
29	396	6	784	Gasolina	Bluebird	55
554	458	4	817	Diésel	Bluebird	14
767	493	6	816	Diésel	Keiser	55
699	475	9	816	Gasolina	Bluebird	55
954	476	10	827	Diésel	Bluebird	42
705	403	4	806	Diésel	Keiser	42
660	337	6	819	Gasolina	Bluebird	55
520	492	10	836	Diésel	Bluebird	55
814	426	4	757	Diésel	Bluebird	55
353	449	4	817	Gasolina	Keiser	55

A.4 Conjunto de datos 4: Applewood Auto Group

x_1 = **Edad:** edad del comprador al tiempo de la adquisición
x_2 = **Ganancia:** la cantidad obtenida por el distribuidor sobre la venta de cada vehículo
x_3 = **Locación:** distribuidora donde fue comprado el vehículo
x_4 = **Tipo de vehículo:** SUV, sedán, compacto, híbrido o camión
x_5 = **Previo:** número de vehículos comprados por el cliente previamente en cualquiera de las cuatro distribuidoras de Applewood

Edad x_1	Ganancia x_2	Locación x_3	Tipo de vehículo x_4	Previo x_5
21	$1 387	Tionesta	Sedán	0
23	1 754	Sheffield	SUV	1
24	1 817	Sheffield	Híbrido	1
25	1 040	Sheffield	Compacto	0
26	1 273	Kane	Sedán	1
27	1 529	Sheffield	Sedán	1
27	3 082	Kane	Camión	0
28	1 951	Kane	SUV	1
28	2 692	Tionesta	Compacto	0
29	1 206	Sheffield	Sedán	0
29	1 342	Kane	Sedán	2
30	443	Kane	Sedán	3
30	754	Olean	Sedán	2
30	1 621	Sheffield	Camión	1
31	870	Tionesta	Sedán	1
31	1 174	Kane	Camión	0
31	1 412	Sheffield	Sedán	1
31	1 809	Tionesta	Sedán	1
31	2 415	Kane	Sedán	0
32	1 546	Sheffield	Camión	3
32	2 148	Tionesta	SUV	2
32	2 207	Sheffield	Compacto	0
32	2 252	Tionesta	SUV	0
33	1 428	Kane	SUV	2
33	1 889	Olean	SUV	1
34	1 166	Olean	Sedán	1
34	1 320	Tionesta	Sedán	1
34	2 265	Olean	Sedán	0
35	1 323	Olean	Sedán	2
35	1 761	Kane	Sedán	1
35	1 919	Tionesta	SUV	1
36	2 357	Kane	SUV	2
36	2 866	Kane	Sedán	1
37	732	Olean	SUV	1
37	1 464	Olean	Sedán	3
37	1 626	Tionesta	Compacto	4
37	1 761	Olean	SUV	1
37	1 915	Tionesta	SUV	2
37	2 119	Kane	Híbrido	1
38	1 766	Sheffield	SUV	0
38	2 201	Sheffield	Camión	2
39	996	Kane	Compacto	2
39	2 813	Tionesta	SUV	0
40	323	Kane	Sedán	0
40	352	Sheffield	Compacto	0
40	482	Olean	Sedán	1
40	1 144	Tionesta	Camión	0
40	1 485	Sheffield	Compacto	0
40	1 509	Kane	SUV	2
40	1 638	Sheffield	Sedán	0
40	1 961	Sheffield	Sedán	1
40	2 127	Olean	Camión	0
40	$2 430	Tionesta	Sedán	1
41	1 704	Sheffield	Sedán	1
41	1 876	Kane	Sedán	2
41	2 010	Tionesta	Sedán	1
41	2 165	Tionesta	SUV	0
41	2 231	Tionesta	SUV	2
41	2 389	Kane	Camión	1
42	335	Olean	SUV	1
42	963	Kane	Sedán	0
42	1 298	Tionesta	Sedán	1
42	1 410	Kane	SUV	2
42	1 553	Tionesta	Compacto	0
42	1 648	Olean	SUV	0
42	2 071	Kane	SUV	0
42	2 116	Kane	Compacto	2
43	1 500	Tionesta	Sedán	0
43	1 549	Kane	SUV	2
43	2 348	Tionesta	Sedán	0
43	2 498	Tionesta	SUV	1
44	294	Kane	SUV	1
44	1 115	Kane	Camión	0
44	1 124	Tionesta	Compacto	2
44	1 532	Tionesta	SUV	3
44	1 688	Kane	Sedán	4
44	1 822	Kane	SUV	0
44	1 897	Sheffield	Compacto	0
44	2 445	Kane	SUV	0
44	2 886	Olean	SUV	1
45	820	Kane	Compacto	1
45	1 266	Olean	Sedán	0
45	1 741	Olean	Compacto	2
45	1 772	Olean	Compacto	1
45	1 932	Tionesta	Sedán	1
45	2 350	Sheffield	Compacto	0
45	2 422	Kane	Sedán	1
45	2 446	Olean	Compacto	1
46	369	Olean	Sedán	1
46	978	Kane	Sedán	1
46	1 238	Sheffield	Compacto	1
46	1 818	Kane	SUV	0
46	1 824	Olean	Camión	0
46	1 907	Olean	Sedán	0
46	1 938	Kane	Sedán	0
46	1 940	Kane	Camión	3
46	2 197	Sheffield	Sedán	1
46	2 646	Tionesta	Sedán	2
47	1 461	Kane	Sedán	0
47	1 731	Tionesta	Compacto	0
47	2 230	Tionesta	Sedán	1
47	2 341	Sheffield	SUV	1
47	3 292	Olean	Sedán	2
48	1 108	Sheffield	Sedán	1

(*continúa*)

A.4 Conjunto de datos 4: Applewood Auto Group (*continuación*)

Edad x_1	Ganancia x_2	Locación x_3	Tipo de vehículo x_4	Previo x_5
48	$1 295	Sheffield	SUV	1
48	1 344	Sheffield	SUV	0
48	1 906	Kane	Sedán	1
48	1 952	Tionesta	Compacto	1
48	2 070	Kane	SUV	1
48	2 454	Kane	Sedán	1
49	1 606	Olean	Compacto	0
49	1 680	Kane	SUV	3
49	1 827	Tionesta	Camión	3
49	1 915	Tionesta	SUV	1
49	2 084	Tionesta	Sedán	0
49	2 639	Sheffield	SUV	0
50	842	Kane	SUV	0
50	1 963	Sheffield	Sedán	1
50	2 059	Sheffield	Sedán	1
50	2 338	Tionesta	SUV	0
50	3 043	Kane	Sedán	0
51	1 059	Kane	SUV	1
51	1 674	Sheffield	Sedán	1
51	1 807	Tionesta	Sedán	1
51	2 056	Sheffield	Híbrido	0
51	2 236	Tionesta	SUV	2
51	2 928	Kane	SUV	0
52	1 269	Tionesta	Sedán	1
52	1 717	Sheffield	SUV	3
52	1 797	Kane	Sedán	1
52	1 955	Olean	Híbrido	2
52	2 199	Tionesta	SUV	0
52	2 482	Olean	Compacto	0
52	2 701	Sheffield	SUV	0
52	3 210	Olean	Camión	4
53	377	Olean	SUV	1
53	1 220	Olean	Sedán	0
53	1 401	Tionesta	SUV	2
53	2 175	Olean	Sedán	1
54	1 118	Sheffield	Compacto	1
54	2 584	Olean	Compacto	2
54	2 666	Tionesta	Camión	0

Edad x_1	Ganancia x_2	Locación x_3	Tipo de vehículo x_4	Previo x_5
54	$2 991	Tionesta	SUV	0
55	934	Sheffield	Camión	1
55	2 063	Kane	SUV	1
55	2 083	Sheffield	Sedán	1
55	2 856	Olean	Híbrido	1
55	2 989	Tionesta	Compacto	1
56	910	Sheffield	SUV	0
56	1 536	Kane	SUV	0
56	1 957	Sheffield	SUV	1
56	2 240	Olean	Sedán	0
56	2 695	Kane	Sedán	2
57	1 325	Olean	Sedán	1
57	2 250	Sheffield	Sedán	2
57	2 279	Sheffield	Híbrido	1
57	2 626	Sheffield	Sedán	2
58	1 501	Sheffield	Híbrido	1
58	1 752	Kane	Sedán	3
58	2 058	Kane	SUV	1
58	2 370	Tionesta	Compacto	0
58	2 637	Sheffield	SUV	1
59	1 426	Sheffield	Sedán	0
59	2 944	Olean	SUV	2
60	2 147	Olean	Compacto	2
61	1 973	Kane	SUV	3
61	2 502	Olean	Sedán	0
62	783	Sheffield	Híbrido	1
62	1 538	Olean	Camión	1
63	2 339	Olean	Compacto	1
64	2 700	Kane	Camión	0
65	2 222	Kane	Camión	1
65	2 597	Sheffield	Camión	0
65	2 742	Tionesta	SUV	2
68	1 837	Sheffield	Sedán	1
69	2 842	Kane	SUV	0
70	2 434	Olean	Sedán	4
72	1 640	Olean	Sedán	1
72	1 821	Tionesta	SUV	1
73	2 487	Olean	Compacto	4

A.5 Conjunto de datos bancarios: Caso del Century National Bank

x_1 = Saldo en cuenta
x_2 = Número de operaciones en cajero automático en el mes
x_3 = Número de otros servicios bancarios utilizados
x_4 = Tiene tarjeta de débito (1 = sí; 0 = no)
x_5 = Recibe intereses sobre la cuenta (1 = sí; 0 = no)
x_6 = Ciudad donde se abrió la cuenta
60 cuentas

x_1	x_2	x_3	x_4	x_5	x_6	x_1	x_2	x_3	x_4	x_5	x_6
1 756	13	4	0	1	2	1 958	6	2	1	0	2
748	9	2	1	0	1	634	2	7	1	0	4
1 501	10	1	0	0	1	580	4	1	0	0	1
1 831	10	4	0	1	3	1 320	4	5	1	0	1
1 622	14	6	0	1	4	1 675	6	7	1	0	2
1 886	17	3	0	1	1	789	8	4	0	0	4
740	6	3	0	0	3	1 735	12	7	0	1	3
1 593	10	8	1	0	1	1 784	11	5	0	0	1
1 169	6	4	0	0	4	1 326	16	8	0	0	3
2 125	18	6	0	0	2	2 051	14	4	1	0	4
1 554	12	6	1	0	3	1 044	7	5	1	0	1
1 474	12	7	1	0	1	1 885	10	6	1	1	2
1 913	6	5	0	0	1	1 790	11	4	0	1	3
1 218	10	3	1	0	1	765	4	3	0	0	4
1 006	12	4	0	0	1	1 645	6	9	0	1	4
2 215	20	3	1	0	4	32	2	0	0	0	3
137	7	2	0	0	3	1 266	11	7	0	0	4
167	5	4	0	0	4	890	7	1	0	1	1
343	7	2	0	0	1	2 204	14	5	0	0	2
2 557	20	7	1	0	4	2 409	16	8	0	0	2
2 276	15	4	1	0	3	1 338	14	4	1	0	2
1 494	11	2	0	1	1	2 076	12	5	1	0	2
2 144	17	3	0	0	3	1 708	13	3	1	0	1
1 995	10	7	0	0	2	2 138	18	5	0	1	4
1 053	8	4	1	0	3	2 375	12	4	0	0	2
1 526	8	4	0	1	2	1 455	9	5	1	1	3
1 120	8	6	1	0	3	1 487	8	4	1	0	4
1 838	7	5	1	1	3	1 125	6	4	1	0	2
1 746	11	2	0	0	2	1 989	12	3	0	1	2
1 616	10	4	1	1	2	2 156	14	5	1	0	2

B.1 Distribución de probabilidad binomial

$n = 1$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.950	0.900	0.800	0.700	0.600	0.500	0.400	0.300	0.200	0.100	0.050
1	0.050	0.100	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	0.950

$n = 2$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.903	0.810	0.640	0.490	0.360	0.250	0.160	0.090	0.040	0.010	0.003
1	0.095	0.180	0.320	0.420	0.480	0.500	0.480	0.420	0.320	0.180	0.095
2	0.003	0.010	0.040	0.090	0.160	0.250	0.360	0.490	0.640	0.810	0.903

$n = 3$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.857	0.729	0.512	0.343	0.216	0.125	0.064	0.027	0.008	0.001	0.000
1	0.135	0.243	0.384	0.441	0.432	0.375	0.288	0.189	0.096	0.027	0.007
2	0.007	0.027	0.096	0.189	0.288	0.375	0.432	0.441	0.384	0.243	0.135
3	0.000	0.001	0.008	0.027	0.064	0.125	0.216	0.343	0.512	0.729	0.857

$n = 4$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.815	0.656	0.410	0.240	0.130	0.063	0.026	0.008	0.002	0.000	0.000
1	0.171	0.292	0.410	0.412	0.346	0.250	0.154	0.076	0.026	0.004	0.000
2	0.014	0.049	0.154	0.265	0.346	0.375	0.346	0.265	0.154	0.049	0.014
3	0.000	0.004	0.026	0.076	0.154	0.250	0.346	0.412	0.410	0.292	0.171
4	0.000	0.000	0.002	0.008	0.026	0.063	0.130	0.240	0.410	0.656	0.815

$n = 5$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.774	0.590	0.328	0.168	0.078	0.031	0.010	0.002	0.000	0.000	0.000
1	0.204	0.328	0.410	0.360	0.259	0.156	0.077	0.028	0.006	0.000	0.000
2	0.021	0.073	0.205	0.309	0.346	0.313	0.230	0.132	0.051	0.008	0.001
3	0.001	0.008	0.051	0.132	0.230	0.313	0.346	0.309	0.205	0.073	0.021
4	0.000	0.000	0.006	0.028	0.077	0.156	0.259	0.360	0.410	0.328	0.204
5	0.000	0.000	0.000	0.002	0.010	0.031	0.078	0.168	0.328	0.590	0.774

$n = 6$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.735	0.531	0.262	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000
1	0.232	0.354	0.393	0.303	0.187	0.094	0.037	0.010	0.002	0.000	0.000
2	0.031	0.098	0.246	0.324	0.311	0.234	0.138	0.060	0.015	0.001	0.000
3	0.002	0.015	0.082	0.185	0.276	0.313	0.276	0.185	0.082	0.015	0.002
4	0.000	0.001	0.015	0.060	0.138	0.234	0.311	0.324	0.246	0.098	0.031
5	0.000	0.000	0.002	0.010	0.037	0.094	0.187	0.303	0.393	0.354	0.232
6	0.000	0.000	0.000	0.001	0.004	0.016	0.047	0.118	0.262	0.531	0.735

(*continúa*)

B.1 Distribución de probabilidad binomial (*continuación*)

***n* = 7**
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.698	0.478	0.210	0.082	0.028	0.008	0.002	0.000	0.000	0.000	0.000
1	0.257	0.372	0.367	0.247	0.131	0.055	0.017	0.004	0.000	0.000	0.000
2	0.041	0.124	0.275	0.318	0.261	0.164	0.077	0.025	0.004	0.000	0.000
3	0.004	0.023	0.115	0.227	0.290	0.273	0.194	0.097	0.029	0.003	0.000
4	0.000	0.003	0.029	0.097	0.194	0.273	0.290	0.227	0.115	0.023	0.004
5	0.000	0.000	0.004	0.025	0.077	0.164	0.261	0.318	0.275	0.124	0.041
6	0.000	0.000	0.000	0.004	0.017	0.055	0.131	0.247	0.367	0.372	0.257
7	0.000	0.000	0.000	0.000	0.002	0.008	0.028	0.082	0.210	0.478	0.698

***n* = 8**
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.663	0.430	0.168	0.058	0.017	0.004	0.001	0.000	0.000	0.000	0.000
1	0.279	0.383	0.336	0.198	0.090	0.031	0.008	0.001	0.000	0.000	0.000
2	0.051	0.149	0.294	0.296	0.209	0.109	0.041	0.010	0.001	0.000	0.000
3	0.005	0.033	0.147	0.254	0.279	0.219	0.124	0.047	0.009	0.000	0.000
4	0.000	0.005	0.046	0.136	0.232	0.273	0.232	0.136	0.046	0.005	0.000
5	0.000	0.000	0.009	0.047	0.124	0.219	0.279	0.254	0.147	0.033	0.005
6	0.000	0.000	0.001	0.010	0.041	0.109	0.209	0.296	0.294	0.149	0.051
7	0.000	0.000	0.000	0.001	0.008	0.031	0.090	0.198	0.336	0.383	0.279
8	0.000	0.000	0.000	0.000	0.001	0.004	0.017	0.058	0.168	0.430	0.663

***n* = 9**
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.630	0.387	0.134	0.040	0.010	0.002	0.000	0.000	0.000	0.000	0.000
1	0.299	0.387	0.302	0.156	0.060	0.018	0.004	0.000	0.000	0.000	0.000
2	0.063	0.172	0.302	0.267	0.161	0.070	0.021	0.004	0.000	0.000	0.000
3	0.008	0.045	0.176	0.267	0.251	0.164	0.074	0.021	0.003	0.000	0.000
4	0.001	0.007	0.066	0.172	0.251	0.246	0.167	0.074	0.017	0.001	0.000
5	0.000	0.001	0.017	0.074	0.167	0.246	0.251	0.172	0.066	0.007	0.001
6	0.000	0.000	0.003	0.021	0.074	0.164	0.251	0.267	0.176	0.045	0.008
7	0.000	0.000	0.000	0.004	0.021	0.070	0.161	0.267	0.302	0.172	0.063
8	0.000	0.000	0.000	0.000	0.004	0.018	0.060	0.156	0.302	0.387	0.299
9	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.134	0.387	0.630

***n* = 10**
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.599	0.349	0.107	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000
1	0.315	0.387	0.268	0.121	0.040	0.010	0.002	0.000	0.000	0.000	0.000
2	0.075	0.194	0.302	0.233	0.121	0.044	0.011	0.001	0.000	0.000	0.000
3	0.010	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001	0.000	0.000
4	0.001	0.011	0.088	0.200	0.251	0.205	0.111	0.037	0.006	0.000	0.000
5	0.000	0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001	0.000
6	0.000	0.000	0.006	0.037	0.111	0.205	0.251	0.200	0.088	0.011	0.001
7	0.000	0.000	0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057	0.010
8	0.000	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.302	0.194	0.075
9	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.268	0.387	0.315
10	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.107	0.349	0.599

(*continúa*)

B.1 Distribución de probabilidad binomial (*continuación*)

$n = 11$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.569	0.314	0.086	0.020	0.004	0.000	0.000	0.000	0.000	0.000	0.000
1	0.329	0.384	0.236	0.093	0.027	0.005	0.001	0.000	0.000	0.000	0.000
2	0.087	0.213	0.295	0.200	0.089	0.027	0.005	0.001	0.000	0.000	0.000
3	0.014	0.071	0.221	0.257	0.177	0.081	0.023	0.004	0.000	0.000	0.000
4	0.001	0.016	0.111	0.220	0.236	0.161	0.070	0.017	0.002	0.000	0.000
5	0.000	0.002	0.039	0.132	0.221	0.226	0.147	0.057	0.010	0.000	0.000
6	0.000	0.000	0.010	0.057	0.147	0.226	0.221	0.132	0.039	0.002	0.000
7	0.000	0.000	0.002	0.017	0.070	0.161	0.236	0.220	0.111	0.016	0.001
8	0.000	0.000	0.000	0.004	0.023	0.081	0.177	0.257	0.221	0.071	0.014
9	0.000	0.000	0.000	0.001	0.005	0.027	0.089	0.200	0.295	0.213	0.087
10	0.000	0.000	0.000	0.000	0.001	0.005	0.027	0.093	0.236	0.384	0.329
11	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.020	0.086	0.314	0.569

$n = 12$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.540	0.282	0.069	0.014	0.002	0.000	0.000	0.000	0.000	0.000	0.000
1	0.341	0.377	0.206	0.071	0.017	0.003	0.000	0.000	0.000	0.000	0.000
2	0.099	0.230	0.283	0.168	0.064	0.016	0.002	0.000	0.000	0.000	0.000
3	0.017	0.085	0.236	0.240	0.142	0.054	0.012	0.001	0.000	0.000	0.000
4	0.002	0.021	0.133	0.231	0.213	0.121	0.042	0.008	0.001	0.000	0.000
5	0.000	0.004	0.053	0.158	0.227	0.193	0.101	0.029	0.003	0.000	0.000
6	0.000	0.000	0.016	0.079	0.177	0.226	0.177	0.079	0.016	0.000	0.000
7	0.000	0.000	0.003	0.029	0.101	0.193	0.227	0.158	0.053	0.004	0.000
8	0.000	0.000	0.001	0.008	0.042	0.121	0.213	0.231	0.133	0.021	0.002
9	0.000	0.000	0.000	0.001	0.012	0.054	0.142	0.240	0.236	0.085	0.017
10	0.000	0.000	0.000	0.000	0.002	0.016	0.064	0.168	0.283	0.230	0.099
11	0.000	0.000	0.000	0.000	0.000	0.003	0.017	0.071	0.206	0.377	0.341
12	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.014	0.069	0.282	0.540

$n = 13$
Probabilidad

x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.513	0.254	0.055	0.010	0.001	0.000	0.000	0.000	0.000	0.000	0.000
1	0.351	0.367	0.179	0.054	0.011	0.002	0.000	0.000	0.000	0.000	0.000
2	0.111	0.245	0.268	0.139	0.045	0.010	0.001	0.000	0.000	0.000	0.000
3	0.021	0.100	0.246	0.218	0.111	0.035	0.006	0.001	0.000	0.000	0.000
4	0.003	0.028	0.154	0.234	0.184	0.087	0.024	0.003	0.000	0.000	0.000
5	0.000	0.006	0.069	0.180	0.221	0.157	0.066	0.014	0.001	0.000	0.000
6	0.000	0.001	0.023	0.103	0.197	0.209	0.131	0.044	0.006	0.000	0.000
7	0.000	0.000	0.006	0.044	0.131	0.209	0.197	0.103	0.023	0.001	0.000
8	0.000	0.000	0.001	0.014	0.066	0.157	0.221	0.180	0.069	0.006	0.000
9	0.000	0.000	0.000	0.003	0.024	0.087	0.184	0.234	0.154	0.028	0.003
10	0.000	0.000	0.000	0.001	0.006	0.035	0.111	0.218	0.246	0.100	0.021
11	0.000	0.000	0.000	0.000	0.001	0.010	0.045	0.139	0.268	0.245	0.111
12	0.000	0.000	0.000	0.000	0.000	0.002	0.011	0.054	0.179	0.367	0.351
13	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.055	0.254	0.513

(*continúa*)

B.1 Distribución de probabilidad binomial (*continuación*)

$n = 14$

	Probabilidad										
x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.488	0.229	0.044	0.007	0.001	0.000	0.000	0.000	0.000	0.000	0.000
1	0.359	0.356	0.154	0.041	0.007	0.001	0.000	0.000	0.000	0.000	0.000
2	0.123	0.257	0.250	0.113	0.032	0.006	0.001	0.000	0.000	0.000	0.000
3	0.026	0.114	0.250	0.194	0.085	0.022	0.003	0.000	0.000	0.000	0.000
4	0.004	0.035	0.172	0.229	0.155	0.061	0.014	0.001	0.000	0.000	0.000
5	0.000	0.008	0.086	0.196	0.207	0.122	0.041	0.007	0.000	0.000	0.000
6	0.000	0.001	0.032	0.126	0.207	0.183	0.092	0.023	0.002	0.000	0.000
7	0.000	0.000	0.009	0.062	0.157	0.209	0.157	0.062	0.009	0.000	0.000
8	0.000	0.000	0.002	0.023	0.092	0.183	0.207	0.126	0.032	0.001	0.000
9	0.000	0.000	0.000	0.007	0.041	0.122	0.207	0.196	0.086	0.008	0.000
10	0.000	0.000	0.000	0.001	0.014	0.061	0.155	0.229	0.172	0.035	0.004
11	0.000	0.000	0.000	0.000	0.003	0.022	0.085	0.194	0.250	0.114	0.026
12	0.000	0.000	0.000	0.000	0.001	0.006	0.032	0.113	0.250	0.257	0.123
13	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.041	0.154	0.356	0.359
14	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.044	0.229	0.488

$n = 15$

	Probabilidad										
x	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95
0	0.463	0.206	0.035	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.366	0.343	0.132	0.031	0.005	0.000	0.000	0.000	0.000	0.000	0.000
2	0.135	0.267	0.231	0.092	0.022	0.003	0.000	0.000	0.000	0.000	0.000
3	0.031	0.129	0.250	0.170	0.063	0.014	0.002	0.000	0.000	0.000	0.000
4	0.005	0.043	0.188	0.219	0.127	0.042	0.007	0.001	0.000	0.000	0.000
5	0.001	0.010	0.103	0.206	0.186	0.092	0.024	0.003	0.000	0.000	0.000
6	0.000	0.002	0.043	0.147	0.207	0.153	0.061	0.012	0.001	0.000	0.000
7	0.000	0.000	0.014	0.081	0.177	0.196	0.118	0.035	0.003	0.000	0.000
8	0.000	0.000	0.003	0.035	0.118	0.196	0.177	0.081	0.014	0.000	0.000
9	0.000	0.000	0.001	0.012	0.061	0.153	0.207	0.147	0.043	0.002	0.000
10	0.000	0.000	0.000	0.003	0.024	0.092	0.186	0.206	0.103	0.010	0.001
11	0.000	0.000	0.000	0.001	0.007	0.042	0.127	0.219	0.188	0.043	0.005
12	0.000	0.000	0.000	0.000	0.002	0.014	0.063	0.170	0.250	0.129	0.031
13	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.092	0.231	0.267	0.135
14	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.031	0.132	0.343	0.366
15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.035	0.206	0.463

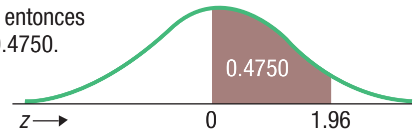
B.2 Distribución de Poisson

	μ								
x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	μ								
x	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149	0.0064	0.0027	0.0011
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446	0.0223	0.0107	0.0050
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892	0.0521	0.0286	0.0150
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339	0.0912	0.0573	0.0337
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606	0.1277	0.0916	0.0607
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606	0.1490	0.1221	0.0911
7	0.0001	0.0034	0.0216	0.0595	0.1044	0.1377	0.1490	0.1396	0.1171
8	0.0000	0.0009	0.0081	0.0298	0.0653	0.1033	0.1304	0.1396	0.1318
9	0.0000	0.0002	0.0027	0.0132	0.0363	0.0688	0.1014	0.1241	0.1318
10	0.0000	0.0000	0.0008	0.0053	0.0181	0.0413	0.0710	0.0993	0.1186
11	0.0000	0.0000	0.0002	0.0019	0.0082	0.0225	0.0452	0.0722	0.0970
12	0.0000	0.0000	0.0001	0.0006	0.0034	0.0113	0.0263	0.0481	0.0728
13	0.0000	0.0000	0.0000	0.0002	0.0013	0.0052	0.0142	0.0296	0.0504
14	0.0000	0.0000	0.0000	0.0001	0.0005	0.0022	0.0071	0.0169	0.0324
15	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0033	0.0090	0.0194
16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0014	0.0045	0.0109
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0021	0.0058
18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0029
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003
22	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

B.3 Áreas bajo la curva normal

Ejemplo:
Si $z = 1.96$, entonces
$P(0 \text{ a } z) = 0.4750$.

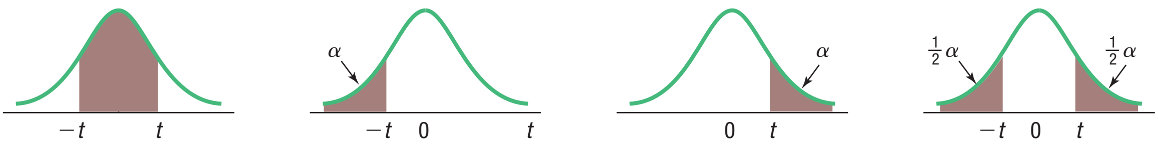


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	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

B.4 Tabla de números aleatorios

02711	08182	75997	79866	58095	83319	80295	79741	74599	84379
94873	90935	31684	63952	09865	14491	99518	93394	34691	14985
54921	78680	06635	98689	17306	25170	65928	87709	30533	89736
77640	97636	37397	93379	56454	59818	45827	74164	71666	46977
61545	00835	93251	87203	36759	49197	85967	01704	19634	21898
17147	19519	22497	16857	42426	84822	92598	49186	88247	39967
13748	04742	92460	85801	53444	65626	58710	55406	17173	69776
87455	14813	50373	28037	91182	32786	65261	11173	34376	36408
08999	57409	91185	10200	61411	23392	47797	56377	71635	08601
78804	81333	53809	32471	46034	36306	22498	19239	85428	55721
82173	26921	28472	98958	07960	66124	89731	95069	18625	92405
97594	25168	89178	68190	05043	17407	48201	83917	11413	72920
73881	67176	93504	42636	38233	16154	96451	57925	29667	30859
46071	22912	90326	42453	88108	72064	58601	32357	90610	32921
44492	19686	12495	93135	95185	77799	52441	88272	22024	80631
31864	72170	37722	55794	14636	05148	54505	50113	21119	25228
51574	90692	43339	65689	76539	27909	05467	21727	51141	72949
35350	76132	92925	92124	92634	35681	43690	89136	35599	84138
46943	36502	01172	46045	46991	33804	80006	35542	61056	75666
22665	87226	33304	57975	03985	21566	65796	72915	81466	89205
39437	97957	11838	10433	21564	51570	73558	27495	34533	57808
77082	47784	40098	97962	89845	28392	78187	06112	08169	11261
24544	25649	43370	28007	06779	72402	62632	53956	24709	06978
27503	15558	37738	24849	70722	71859	83736	06016	94397	12529
24590	24545	06435	52758	45685	90151	46516	49644	92686	84870
48155	86226	40359	28723	15364	69125	12609	57171	86857	31702
20226	53752	90648	24362	83314	00014	19207	69413	97016	86290
70178	73444	38790	53626	93780	18629	68766	24371	74639	30782
10169	41465	51935	05711	09799	79077	88159	33437	68519	03040
81084	03701	28598	70013	63794	53169	97054	60303	23259	96196
69202	20777	21727	81511	51887	16175	53746	46516	70339	62727
80561	95787	89426	93325	86412	57479	54194	52153	19197	81877
08199	26703	95128	48599	09333	12584	24374	31232	61782	44032
98883	28220	39358	53720	80161	83371	15181	11131	12219	55920
84568	69286	76054	21615	80883	36797	82845	39139	90900	18172
04269	35173	95745	53893	86022	77722	52498	84193	22448	22571
10538	13124	36099	13140	37706	44562	57179	44693	67877	01549
77843	24955	25900	63843	95029	93859	93634	20205	66294	41218
12034	94636	49455	76362	83532	31062	69903	91186	65768	55949
10524	72829	47641	93315	80875	28090	97728	52560	34937	79548
68935	76632	46984	61772	92786	22651	07086	89754	44143	97687
89450	65665	29190	43709	11172	34481	95977	47535	25658	73898
90696	20451	24211	97310	60446	73530	62865	96574	13829	72226
49006	32047	93086	00112	20470	17136	28255	86328	07293	38809
74591	87025	52368	59416	34417	70557	86746	55809	53628	12000
06315	17012	77103	00968	07235	10728	42189	33292	51487	64443
62386	09184	62092	46617	99419	64230	95034	85481	07857	42510
86848	82122	04028	36959	87827	12813	08627	80699	13345	51695
65643	69480	46598	04501	40403	91408	32343	48130	49303	90689
11084	46534	78957	77353	39578	77868	22970	84349	09184	70603

B.5 Distribución *t* de Student



Intervalo de confianza — Prueba de cola izquierda — Prueba de cola derecha — Prueba de dos colas

	Intervalos de confianza, *c*					
	80%	**90%**	**95%**	**98%**	**99%**	**99.9%**
	Nivel de significancia de una prueba de una cola, α					
gl	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
	Nivel de significancia de una prueba de dos colas, α					
	0.20	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
31	1.309	1.696	2.040	2.453	2.744	3.633
32	1.309	1.694	2.037	2.449	2.738	3.622
33	1.308	1.692	2.035	2.445	2.733	3.611
34	1.307	1.691	2.032	2.441	2.728	3.601
35	1.306	1.690	2.030	2.438	2.724	3.591
36	1.306	1.688	2.028	2.434	2.719	3.582
37	1.305	1.687	2.026	2.431	2.715	3.574
38	1.304	1.686	2.024	2.429	2.712	3.566
39	1.304	1.685	2.023	2.426	2.708	3.558
40	1.303	1.684	2.021	2.423	2.704	3.551
41	1.303	1.683	2.020	2.421	2.701	3.544
42	1.302	1.682	2.018	2.418	2.698	3.538
43	1.302	1.681	2.017	2.416	2.695	3.532
44	1.301	1.680	2.015	2.414	2.692	3.526
45	1.301	1.679	2.014	2.412	2.690	3.520
46	1.300	1.679	2.013	2.410	2.687	3.515
47	1.300	1.678	2.012	2.408	2.685	3.510
48	1.299	1.677	2.011	2.407	2.682	3.505
49	1.299	1.677	2.010	2.405	2.680	3.500
50	1.299	1.676	2.009	2.403	2.678	3.496
51	1.298	1.675	2.008	2.402	2.676	3.492
52	1.298	1.675	2.007	2.400	2.674	3.488
53	1.298	1.674	2.006	2.399	2.672	3.484
54	1.297	1.674	2.005	2.397	2.670	3.480
55	1.297	1.673	2.004	2.396	2.668	3.476
56	1.297	1.673	2.003	2.395	2.667	3.473
57	1.297	1.672	2.002	2.394	2.665	3.470
58	1.296	1.672	2.002	2.392	2.663	3.466
59	1.296	1.671	2.001	2.391	2.662	3.463
60	1.296	1.671	2.000	2.390	2.660	3.460
61	1.296	1.670	2.000	2.389	2.659	3.457
62	1.295	1.670	1.999	2.388	2.657	3.454
63	1.295	1.669	1.998	2.387	2.656	3.452
64	1.295	1.669	1.998	2.386	2.655	3.449
65	1.295	1.669	1.997	2.385	2.654	3.447
66	1.295	1.668	1.997	2.384	2.652	3.444
67	1.294	1.668	1.996	2.383	2.651	3.442
68	1.294	1.668	1.995	2.382	2.650	3.439
69	1.294	1.667	1.995	2.382	2.649	3.437
70	1.294	1.667	1.994	2.381	2.648	3.435

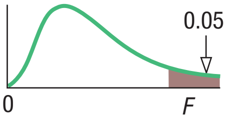
(*continúa*)

B.5 Distribución *t* de Student (*continuación*)

	Intervalos de confianza, *c*					
	80%	90%	95%	98%	99%	99.9%
	Nivel de significancia de una prueba de una cola, α					
gl	0.10	0.05	0.025	0.01	0.005	0.0005
	Nivel de significancia de una prueba de dos colas, α					
	0.20	0.10	0.05	0.02	0.01	0.001
71	1.294	1.667	1.994	2.380	2.647	3.433
72	1.293	1.666	1.993	2.379	2.646	3.431
73	1.293	1.666	1.993	2.379	2.645	3.429
74	1.293	1.666	1.993	2.378	2.644	3.427
75	1.293	1.665	1.992	2.377	2.643	3.425
76	1.293	1.665	1.992	2.376	2.642	3.423
77	1.293	1.665	1.991	2.376	2.641	3.421
78	1.292	1.665	1.991	2.375	2.640	3.420
79	1.292	1.664	1.990	2.374	2.640	3.418
80	1.292	1.664	1.990	2.374	2.639	3.416
81	1.292	1.664	1.990	2.373	2.638	3.415
82	1.292	1.664	1.989	2.373	2.637	3.413
83	1.292	1.663	1.989	2.372	2.636	3.412
84	1.292	1.663	1.989	2.372	2.636	3.410
85	1.292	1.663	1.988	2.371	2.635	3.409
86	1.291	1.663	1.988	2.370	2.634	3.407
87	1.291	1.663	1.988	2.370	2.634	3.406
88	1.291	1.662	1.987	2.369	2.633	3.405

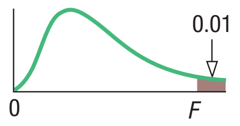
	Intervalos de confianza, *c*					
	80%	90%	95%	98%	99%	99.9%
	Nivel de significancia de una prueba de una cola, α					
gl	0.10	0.05	0.025	0.01	0.005	0.0005
	Nivel de significancia de una prueba de dos colas, α					
	0.20	0.10	0.05	0.02	0.01	0.001
89	1.291	1.662	1.987	2.369	2.632	3.403
90	1.291	1.662	1.987	2.368	2.632	3.402
91	1.291	1.662	1.986	2.368	2.631	3.401
92	1.291	1.662	1.986	2.368	2.630	3.399
93	1.291	1.661	1.986	2.367	2.630	3.398
94	1.291	1.661	1.986	2.367	2.629	3.397
95	1.291	1.661	1.985	2.366	2.629	3.396
96	1.290	1.661	1.985	2.366	2.628	3.395
97	1.290	1.661	1.985	2.365	2.627	3.394
98	1.290	1.661	1.984	2.365	2.627	3.393
99	1.290	1.660	1.984	2.365	2.626	3.392
100	1.290	1.660	1.984	2.364	2.626	3.390
120	1.289	1.658	1.980	2.358	2.617	3.373
140	1.288	1.656	1.977	2.353	2.611	3.361
160	1.287	1.654	1.975	2.350	2.607	3.352
180	1.286	1.653	1.973	2.347	2.603	3.345
200	1.286	1.653	1.972	2.345	2.601	3.340
∞	1.282	1.645	1.960	2.326	2.576	3.291

B.6A Valores críticos de la distribución F ($\alpha = 0.05$)



Grados de libertad en el denominador	Grados de libertad en el numerador															
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

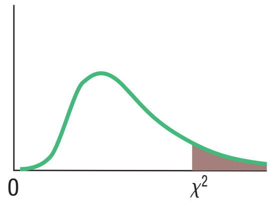
B.6B Valores críticos de la distribución F ($\alpha = 0.01$)



Grados de libertad en el denominador	Grados de libertad en el numerador															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59

B.7 Valores críticos de *ji* cuadrada

Esta tabla contiene los valores de χ^2 correspondientes a un área específica de la cola derecha y un número específico de grados de libertad.



Ejemplo: con 17 *gl* y un área de 0.02 en la cola superior, $\chi^2 = 30.995$

Grados de libertad, *gl*	Área de la cola derecha			
	0.10	**0.05**	**0.02**	**0.01**
1	2.706	3.841	5.412	6.635
2	4.605	5.991	7.824	9.210
3	6.251	7.815	9.837	11.345
4	7.779	9.488	11.668	13.277
5	9.236	11.070	13.388	15.086
6	10.645	12.592	15.033	16.812
7	12.017	14.067	16.622	18.475
8	13.362	15.507	18.168	20.090
9	14.684	16.919	19.679	21.666
10	15.987	18.307	21.161	23.209
11	17.275	19.675	22.618	24.725
12	18.549	21.026	24.054	26.217
13	19.812	22.362	25.472	27.688
14	21.064	23.685	26.873	29.141
15	22.307	24.996	28.259	30.578
16	23.542	26.296	29.633	32.000
17	24.769	27.587	30.995	33.409
18	25.989	28.869	32.346	34.805
19	27.204	30.144	33.687	36.191
20	28.412	31.410	35.020	37.566
21	29.615	32.671	36.343	38.932
22	30.813	33.924	37.659	40.289
23	32.007	35.172	38.968	41.638
24	33.196	36.415	40.270	42.980
25	34.382	37.652	41.566	44.314
26	35.563	38.885	42.856	45.642
27	36.741	40.113	44.140	46.963
28	37.916	41.337	45.419	48.278
29	39.087	42.557	46.693	49.588
30	40.256	43.773	47.962	50.892

B.8 Valores T de Wilcoxon

	2α						
	0.15	0.10	0.05	0.04	0.03	0.02	0.01
	α						
n	0.075	0.050	0.025	0.020	0.015	0.010	0.005
4	0						
5	1	0					
6	2	2	0	0			
7	4	3	2	1	0	0	
8	7	5	3	3	2	1	0
9	9	8	5	5	4	3	1
10	12	10	8	7	6	5	3
11	16	13	10	9	8	7	5
12	19	17	13	12	11	9	7
13	24	21	17	16	14	12	9
14	28	25	21	19	18	15	12
15	33	30	25	23	21	19	15
16	39	35	29	28	26	23	19
17	45	41	34	33	30	27	23
18	51	47	40	38	35	32	27
19	58	53	46	43	41	37	32
20	65	60	52	50	47	43	37
21	73	67	58	56	53	49	42
22	81	75	65	63	59	55	48
23	89	83	73	70	66	62	54
24	98	91	81	78	74	69	61
25	108	100	89	86	82	76	68
26	118	110	98	94	90	84	75
27	128	119	107	103	99	92	83
28	138	130	116	112	108	101	91
29	150	140	126	122	117	110	100
30	161	151	137	132	127	120	109
31	173	163	147	143	137	130	118
32	186	175	159	154	148	140	128
33	199	187	170	165	159	151	138
34	212	200	182	177	171	162	148
35	226	213	195	189	182	173	159
40	302	286	264	257	249	238	220
50	487	466	434	425	413	397	373
60	718	690	648	636	620	600	567
70	995	960	907	891	872	846	805
80	1 318	1 276	1 211	1 192	1 168	1 136	1 086
90	1 688	1 638	1 560	1 537	1 509	1 471	1 410
100	2 105	2 045	1 955	1 928	1 894	1 850	1 779

B.9A Valores críticos del estadístico *d* de Durbin-Watson ($\alpha = 0.05$)

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	$d_{L,0.05}$	$d_{U,0.05}$	$d_{L,0.05}$	$d_{U,0.05}$	$d_{L,0.05}$	$d_{U,0.05}$	$d_{L,0.05}$	$d_{U,0.05}$	$d_{L,0.05}$	$d_{U,0.05}$
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Fuente: J. Durbin y G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, II", *Biometrika* 30 (1951), pp. 159-178. Reproducido con el permiso de Biometrika Trustees.

B.9B Valores críticos del estadístico *d* de Durbin-Watson ($\alpha = 0.025$)

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	$d_{L,0.025}$	$d_{U,0.025}$	$d_{L,0.025}$	$d_{U,0.025}$	$d_{L,0.025}$	$d_{U,0.025}$	$d_{L,0.025}$	$d_{U,0.025}$	$d_{L,0.025}$	$d_{U,0.025}$
15	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
16	0.98	1.24	0.86	1.40	0.75	1.59	0.64	1.80	0.53	2.03
17	1.01	1.25	0.90	1.40	0.79	1.58	0.68	1.77	0.57	1.98
18	1.03	1.26	0.93	1.40	0.82	1.56	0.72	1.74	0.62	1.93
19	1.06	1.28	0.96	1.41	0.86	1.55	0.76	1.72	0.66	1.90
20	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
21	1.10	1.30	1.01	1.41	0.92	1.54	0.83	1.69	0.73	1.84
22	1.12	1.31	1.04	1.42	0.95	1.54	0.86	1.68	0.77	1.82
23	1.14	1.32	1.06	1.42	0.97	1.54	0.89	1.67	0.80	1.80
24	1.16	1.33	1.08	1.43	1.00	1.54	0.91	1.66	0.83	1.79
25	1.18	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
26	1.19	1.35	1.12	1.44	1.04	1.54	0.96	1.65	0.88	1.76
27	1.21	1.36	1.13	1.44	1.06	1.54	0.99	1.64	0.91	1.75
28	1.22	1.37	1.15	1.45	1.08	1.54	1.01	1.64	0.93	1.74
29	1.24	1.38	1.17	1.45	1.10	1.54	1.03	1.63	0.96	1.73
30	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
31	1.26	1.39	1.20	1.47	1.13	1.55	1.07	1.63	1.00	1.72
32	1.27	1.40	1.21	1.47	1.15	1.55	1.08	1.63	1.02	1.71
33	1.28	1.41	1.22	1.48	1.16	1.55	1.10	1.63	1.04	1.71
34	1.29	1.41	1.24	1.48	1.17	1.55	1.12	1.63	1.06	1.70
35	1.30	1.42	1.25	1.48	1.19	1.55	1.13	1.63	1.07	1.70
36	1.31	1.43	1.26	1.49	1.20	1.56	1.15	1.63	1.09	1.70
37	1.32	1.43	1.27	1.49	1.21	1.56	1.16	1.62	1.10	1.70
38	1.33	1.44	1.28	1.50	1.23	1.56	1.17	1.62	1.12	1.70
39	1.34	1.44	1.29	1.50	1.24	1.56	1.19	1.63	1.13	1.69
40	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
45	1.39	1.48	1.34	1.53	1.30	1.58	1.25	1.63	1.21	1.69
50	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
55	1.45	1.52	1.41	1.56	1.37	1.60	1.33	1.64	1.30	1.69
60	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
65	1.49	1.55	1.46	1.59	1.43	1.62	1.40	1.66	1.36	1.69
70	1.51	1.57	1.48	1.60	1.45	1.63	1.42	1.66	1.39	1.70
75	1.53	1.58	1.50	1.61	1.47	1.64	1.45	1.67	1.42	1.70
80	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
85	1.56	1.60	1.53	1.63	1.51	1.65	1.49	1.68	1.46	1.71
90	1.57	1.61	1.55	1.64	1.53	1.66	1.50	1.69	1.48	1.71
95	1.58	1.62	1.56	1.65	1.54	1.67	1.52	1.69	1.50	1.71
100	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72

Fuente: J. Durbin y G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, II", *Biometrika* 30 (1951), pp. 159-178. Reproducido con el permiso de Biometrika Trustees.

B.9C Valores críticos del estadístico *d* de Durbin-Watson ($\alpha = 0.01$)

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	$d_{L,0.01}$	$d_{U,0.01}$	$d_{L,0.01}$	$d_{U,0.01}$	$d_{L,0.01}$	$d_{U,0.01}$	$d_{L,0.01}$	$d_{U,0.01}$	$d_{L,0.01}$	$d_{U,0.01}$
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1,48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Fuente: J. Durbin y G. S. Watson, “Testing for Serial Correlation in Least Squares Regression, II”, *Biometrika* 30 (1951), pp. 159-178. Reproducción con el permiso de Biometrika Trustees.

B.10 Factores de las tablas de control

Número de elementos en la muestra, *n*	Tablas de promedios	Tablas de rangos		
	Factores de los límites de control	Factores de la línea central	Factores de los límites de control	
	A_2	d_2	D_3	D_4
2	1.880	1.128	0	3.267
3	1.023	1.693	0	2.575
4	.729	2.059	0	2.282
5	.577	2.326	0	2.115
6	.483	2.534	0	2.004
7	.419	2.704	.076	1.924
8	.373	2.847	.136	1.864
9	.337	2.970	.184	1.816
10	.308	3.078	.223	1.777
11	.285	3.173	.256	1.744
12	.266	3.258	.284	1.716
13	.249	3.336	.308	1.692
14	.235	3.407	.329	1.671
15	.223	3.472	.348	1.652

Fuente: Adaptado de American Society for Testing and Materials, *Manual on Quality Control of Materials*, 1951, tabla B2, p. 115. Para una tabla y una explicación más detalladas, vea Acheson J. Duncan, *Quality Control and Industrial Statistics*, 3a. ed., Homewood, Ill: Richard D. Irwin, 1974, tabla M, p. 927.

APÉNDICE C: COMANDOS DE SOFTWARE

CAPÍTULO 2

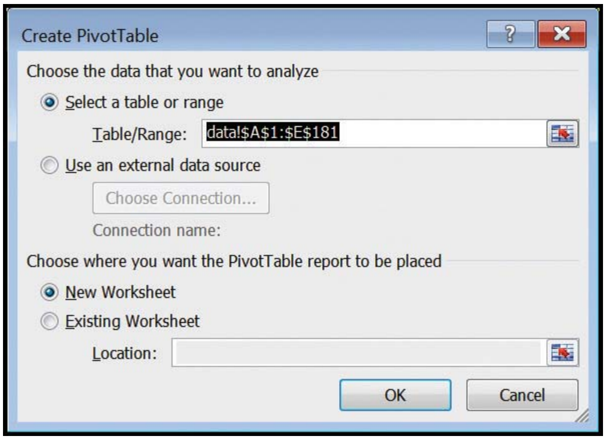
2.1 A continuación se muestran los comandos de Excel para utilizar el PivotTable Wizard para crear la tabla de frecuencias, la gráfica de barras y la gráfica de pastel de la página 20:

a. Abra el archivo con los datos del Applewood Auto Group.

b. Haga clic en alguna parte de la celda en el conjunto de datos, como en la celda *C5*.

c. Haga clic en el menú *Insert* en la barra de herramientas. Después haga clic en *PivotTable* en el extremo izquierdo de la barra.

d. Se abrirá la siguiente pantalla; haga clic en "Select a table or range" para seleccionar el rango de los datos como se muestra en la fila *Table/Range.* Después haga clic en *"*Existing Worksheet*"* y seleccione una celda, como *N1*, y haga clic en *OK.*



e. En el lado derecho de la hoja de cálculo se mostrará *PivotTable Field List*, que contiene una lista de las variables de la base de datos. Para resumir la variable "Vehicle-Type" (tipo de vehículo), haga clic en la variable "Vehicle-Type" y esta se verá en el cuadro inferior izquierdo, que se llama *Row Label*. Observará que la tabla de frecuencias comienza en la celda *N1* con las filas etiquetadas con los valores de la variable "Vehicle-Type"; ahora, regrese al cuadro superior y seleccione y arrastre la variable "Vehicle-Type" al cuadro "Σ Values". Se añadirá una columna de frecuencias a la tabla. Observe que puede formatear la tabla para centrar los valores y renombrar los encabezados de columna según lo necesite.

f. Para crear una gráfica de barras, seleccione cualquier celda de la PivotTable; enseguida, seleccione el menú *Insert* de la barra de herramientas y seleccione, dentro del grupo *Charts*, una gráfica de barras del menú *Column*; se abrirá una gráfica de barras. Haga clic en el encabezado de la barra y nómbrela como lo necesite.

g. Para crear una gráfica de pastel, las frecuencias se deben convertir a frecuencias relativas. Haga clic en el cuerpo de la PivotTable para que aparezca la *PivotTable Field List* a la derecha; en el cuadro "Σ Values", haga clic en el menú de "Count of Vehicle Type" y seleccione la opción *Value Field Settings.* Verá varias selecciones que pueden usarse para resumir las variables en una PivotTable; haga clic en la pestaña "Show Values As" y, en el menú que se mostrará, seleccione "% of Grand Total". Las frecuencias se convertirán a frecuencias relativas.

Para crear la gráfica de pastel, seleccione el menú *Insert* de la barra de herramientas y seleccione, dentro del grupo *Charts*, una gráfica de pastel del menú *Column*; se mostrará una gráfica de pastel. Haga clic en el encabezado de la barra y nómbrela como lo necesite; después haga clic en "Add Data Labels*"*.

2.2 A continuación se indican los comandos de Excel para utilizar el PivotTable Wizard para crear la tabla de frecuencias, la gráfica de barras y la gráfica de pastel de la página 20:

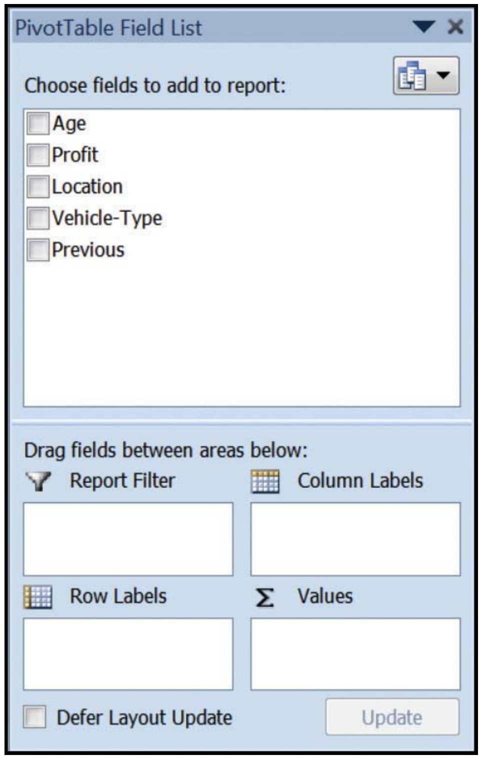
a. Abra el archivo con los datos del Applewood Auto Group.

b. Haga clic en alguna celda en el conjunto de datos, como en la celda *C5*.

c. Haga clic en el menú *Insert* en la barra de herramientas. Después haga clic en *PivotTable* en el extremo izquierdo de la barra.

d. Se mostrará la pantalla que aparece abajo. Haga clic en "*Select a table or range"* para seleccionar el rango de los datos como se muestra en la fila *Table/Range.* Después haga clic en *"New Worksheet"* y la PivotTable se creará en una nueva hoja de cálculo.

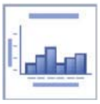
e. En el lado derecho de la hoja de cálculo se desplegará *PivotTable Field List*, que contiene una lista de las variables de la base de datos; para resumir la variable "Profit" (ganancia), haga clic en la variable "Profit" y arrástrela al cuadro "Row Labels". Regrese al cuadro superior y seleccione y arrastre la variable "Profits" al cuadro "Σ Values"; en este mismo cuadro, active el menú para "Sum of Profit". Verá algunas selecciones que puede usar para resumir las variables en una PivotTable; cuando haga clic en la pestaña "Show Values As" para crear frecuencias para la variable "Profit", se abrirá una PivotTable en la nueva hoja de cálculo.



f. En la PivotTable, la columna izquierda muestra cada valor para la variable "Profit". Para crear clases para ella, seleccione cualquier celda en la columna y haga clic con el botón derecho del mouse; del menú que aparezca, seleccione "Group" para crear clases. Primero, desmarque ambos cuadros; después, en el cuadro de diálogo, ingrese el límite inferior de la primera clase como el valor "Starting at"; ahora ingrese el intervalo de clase como el valor "By", y haga clic en *OK*; se mostrará una distribución de frecuencias.

g. Para crear una distribución de frecuencias relativas, señale y haga clic en una de las celdas de la PivotTable y se desplegará la "PivotTable Field List" a la derecha; haga clic y arrastre la variable "Profits" al cuadro "Σ Values"; se mostrará un segundo "Value Fields Setting"; en "Σ Values", haga clic en el segundo "Counts of Profit" y seleccione "Value Fields Setting". Verá varias selecciones distintas que puede usar para resumir las variables en la PivotTable; haga clic en la pestaña "*Show Values As*" y, en el menú que aparezca, seleccione "*% of Grand Total*". Se añadirán a la tabla las frecuencias relativas. Puede formatear la tabla renombrando los encabezados de columnas como "Frequency" y "Relative Frequency".

h. Para crear un histograma, seleccione una celda de la PivotTable, elija el menú *Insert* de la barra de herramientas, y luego, dentro del grupo de *Charts,* seleccione una gráfica de *columna* del menú secundario de *Column*; se desplegará un histograma en "Count of Profit" y en "Count of Profit2"; en el globo "Count of Profit2" en la parte superior de la gráfica, seleccione "Remove Field" con el botón derecho; ahora, tanto la gráfica como la PivotTable solo reportarán las frecuencias. Para eliminar el espacio entre barras, seleccione toda el área de la gráfica, y se abrirá "PivotChart Tools" en la parte superior; seleccione "Design". En las opciones de "Chart Layouts", seleccione la opción que no muestra espacios entre las barras (vea figura adjunta). Para añadir etiquetas a los datos, seleccione el histograma, haga clic con el botón derecho y seleccione "Add Data Labels". Renombre la gráfica y sus ejes según lo necesite.



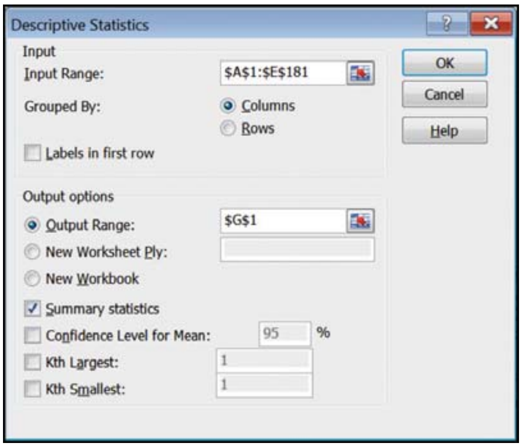
CAPÍTULO 3

3.1 Enseguida se indican los comandos de Excel de estadística descriptiva de la página 56:

a. Recupere el archivo de datos de Applewood del sitio **www.mhhe.com/uni/lind_ae16e**.

b. De la barra de menú, seleccione **Data** y, enseguida, **Data Analysis.** Seleccione **Descriptive Statistics** y haga clic en **OK.**

c. En **Input Range**, escriba *C1:C181,* indique que los datos se agrupan por columna y que las etiquetas se encuentran en la primera fila. Haga clic en **Output Range**, indique que la salida se debe incluir en *G1* (o en cualquier lugar que desee), haga clic en **Summary statistics** y luego en **OK**.



d. Cuando obtenga los resultados, verifique dos veces la cuenta en la salida para comprobar que contiene la cantidad correcta de elementos.

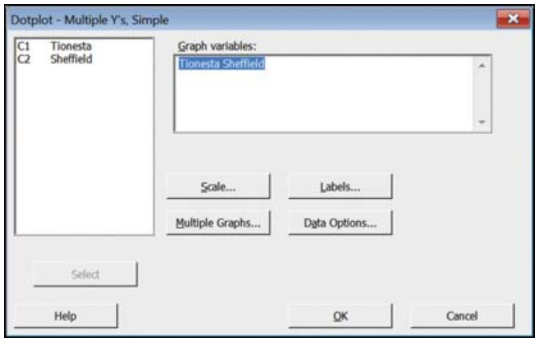
CAPÍTULO 4

4.1 He aquí los comandos de Minitab para elaborar el diagrama de puntos de la página 84:

a. Introduzca el número de vehículos que recibieron servicio en Tionesta Ford Lincoln Mercury en la columna *C1* y en Sheffield Motors en *C2.* Ponga el nombre adecuado a las variables.

b. Seleccione **Graph** y **Dotplot**; en el primer cuadro de diálogo, seleccione **Multiple Y's Simple** en la esquina inferior izquierda y haga clic en **OK**. En el siguiente cuadro de diálogo, seleccione **Tionesta** y **Sheffield** como variables para **Graph**, haga clic en **Labels** y escriba un título adecuado; haga clic en **OK**.

c. Para calcular las estadísticas descriptivas que aparecen en la pantalla, seleccione **Stat**, **Basic statistics** y, enseguida, **Display Descriptive statistics.** En el cuadro de diálogo, seleccione **Tionesta** y **Sheffield** como variables, haga clic en **Statistics**, seleccione las estadísticas que desee obtener y, finalmente, haga doble clic en **OK**.



4.2 Ahora se indican los comandos de Minitab para elaborar el diagrama de tallo y hojas de la página 86:

a. Importe los datos para **Table 4-1** de **www.mhhe.com/uni/lind_ae16e**.

b. Seleccione **Graph** y haga clic en **Stem-and-Leaf**.

c. Seleccione la variable **Spots**, introduzca *10* como **Increment** y haga clic enseguida en **OK**.

4.3 A continuación se muestran los comandos de Minitab para elaborar el resumen descriptivo de la página 91:

a. Importe los datos de las comisiones de Smith Barney del ejemplo en la página 89.

b. De la barra de herramientas, seleccione **Stat**, **Basic Statistics** y **Display Descriptive Statistics**; en el cuadro de diálogo seleccione **Commissions** como **Variable** y enseguida haga clic en **OK**.

4.4 Enseguida se indican los comandos de Excel para los cuartiles de la página 91:

a. Ingrese los datos de las comisiones de Smith Barney del ejemplo en la página 89 en la columna A.

Si está utilizando Excel 2010 y desea calcular los cuartiles utilizando la fórmula [4.1], siga los siguientes pasos:

b. En la celda *C3* escriba la **Formula (4.1)**, en la *C4* escriba **Quartile 1** y en la *C6* escriba **Quartile 3**.

c. En la celda *D4* escriba **"=QUARTILE.EXC(A2:A16,1)"** y presione **Enter**; en la celda *D6* escriba **"=QUARTILE.EXC (A2:A16,3)"** y presione **Enter**.

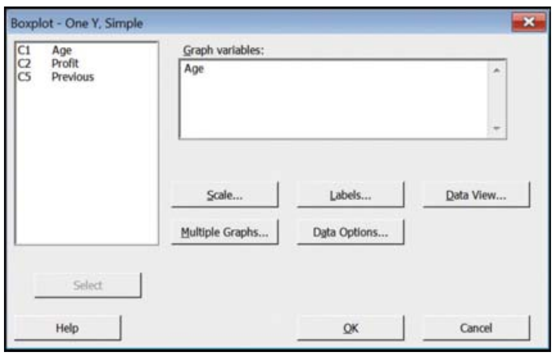
Si está usando Excel 2007 o 2010 y desea calcular los cuartiles mediante el *Método de Excel*:

b. En la celda *C8*, escriba **Excel Method**, en la celda *C9* escriba **Quartile 1** y en la celda *C11* teclee **Quartile 3**.

c. En la celda *DB*, escriba "**=QUARTILE(A2:A16,1)**" y oprima **Enter**. En la celda *D11* teclee "**=QUARTILE(A2:A16,3)**" y oprima **Enter**.

4.5 Ahora se señalan los comandos de Minitab para elaborar el diagrama de caja de la página 93:

a. Importe los datos de Applewood Auto Group de **www.mhhe.com/uni/lind_ae16e**.



b. Seleccione **Graph** y enseguida **Boxplot**. En el cuadro de diálogo seleccione **Simple** en la esquina superior izquierda y haga clic en **OK**; seleccione **Age** como **Graph Variable**, haga clic en **Labels**, incluya un encabezado adecuado y haga clic en **OK**.

4.6 He aquí los comandos de Minitab para construir el resumen descriptivo de la página 98:

a. Ingrese los datos en la primera columna; en la celda debajo de *C1*, ingrese la variable *Earnings.*

b. Seleccione **Stat**, **Basic Statistics** y haga clic en **Graphical Summary**; seleccione **Earnings** como variable y haga clic en **OK**.

4.7 A continuación se señalan los comandos de Excel para dibujar el diagrama de dispersión de la página 100:

a. Recupere los datos de Applewood Auto Group.

b. Resalte con el mouse la columna de edad y la de ganancia; incluya la primera fila.

c. Seleccione la pestaña **Insert** y seleccione **Scatter** en las opciones de **Chart**; seleccione la opción superior izquierda para que se muestre el diagrama de dispersión.

d. Con **Chart Tools** desplegada arriba, seleccione la pestaña **Layout**. Seleccione **Chart Title** y escriba un nombre para el diagrama; enseguida, bajo la misma pestaña **Layout**, seleccione **AxisTitles**. En **Primary Vertical Axis Title**, escriba *Profit* como el nombre del eje; en **Primary Horizontal Axis Title**, escriba *Age* como el nombre del eje, seleccione **Legend** y elija **None**.

CAPÍTULO 5

5.1 Enseguida se muestran los comandos de Excel para determinar el número de permutaciones de las páginas 145 y 146:

a. Haga clic en la pestaña en la barra de herramientas y seleccione **Insert Function *fx***.

b. En el cuadro **Insert Function**, seleccione **Statistical** como categoría; vaya al recuadro de abajo y busque **PERMUT** en la lista **Select a function** y haga clic en **OK**.

c. En el cuadro **PERMUT**, introduzca *8* en **Number** y en el cuadro de **Number_chosen**, inserte *3*. La respuesta correcta, *336*, aparece dos veces en el cuadro.

5.2 Ahora se señalan los comandos de Excel para determinar el número de combinaciones de las páginas 145 y 146:

a. Haga clic en **Formulas** en la barra de herramientas y seleccione **Insert Function *fx***.

b. En el cuadro **Insert function,** seleccione **Math & Trig** como categoría y vaya a **COMBIN** en la lista **Select a function**; haga clic en **OK**.

c. En el cuadro **COMBIN**, escriba *7* en **Number** y *3* en **Number_chosen**. La respuesta correcta, *35*, aparece dos veces en el cuadro.

CAPÍTULO 6

6.1 A continuación se indican los comandos de Excel para determinar la distribución de probabilidad binomial de la página 166:

a. En una hoja de cálculo de Excel en blanco escriba la palabra Éxito en la celda *A1*, y la palabra *Probabilidad* en la celda *B1*; de las celdas *A2* a *A17* escriba los números enteros *0* a *15* y active la celda *B2* haciendo clic en ella.

b. De la barra de herramientas seleccione **Formulas** y en el extremo izquierdo, seleccione **Insert Function *fx***.

c. En el primer cuadro de diálogo seleccione **Statistical** en la categoría de funciones, y **BINOM.DIST** en la categoría del nombre de la función; haga clic en **OK**.

d. En el segundo cuadro de diálogo introduzca los cuatro elementos que se requieren para calcular una probabilidad binomial.

1. Introduzca *0* en **Number_s**.
2. Introduzca *40* en **Trials**.
3. Introduzca *0.09* como probabilidad de un éxito.
4. Introduzca la palabra *false* o el número *0* en **Cumulative** y haga clic en **OK**.
5. Excel calculará la probabilidad de 0 éxitos en 40 ensayos, con una probabilidad de 0.09 de éxito; el resultado, 0.02299618, se almacena en la celda B2.

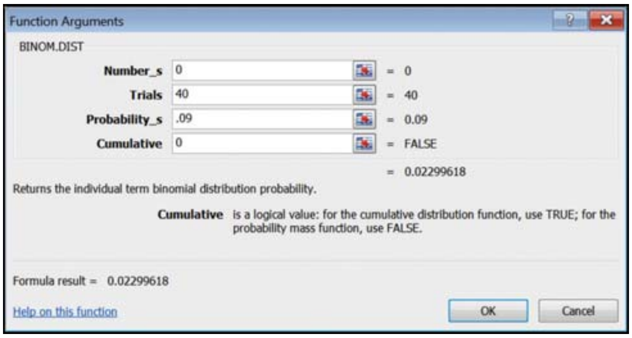
e. Para completar la distribución de probabilidad de éxito de 1 al 15, haga doble clic en la celda **B2**; se verá la función binomial. Reemplace el **0** a la derecha del paréntesis de apertura con la referencia de celda **A2**.

f. Arrastre el mouse a la esquina inferior izquierda de la celda B2 hasta que aparezca el símbolo + en líneas negras sólidas; enseguida haga clic sin soltar el botón y resalte la columna B, hasta la celda B17; se mostrará la probabilidad de un éxito para los diversos valores de la variable aleatoria.

6.2 A continuación se señalan los comandos de Excel para determinar la distribución hipergeométrica de la página 172:

a. En una hoja de cálculo en blanco de Excel, escriba las palabras *Miembros de un sindicato* en la celda A1 y la palabra *Probabilidad* en la celda B1; en las celdas A2 a A7 escriba los enteros *0* a *5*. Haga clic en **B2** como celda activa.

b. De la barra de herramientas elija **Formulas** y en el extremo izquierdo, **Insert Function *fx***.



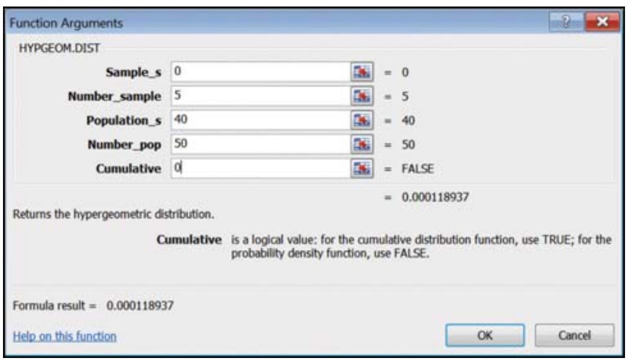
c. En el primer cuadro de diálogo, seleccione **Statistical** y **HYPERGEOM.DIST**, y haga clic en **OK**.

d. En el segundo cuadro de diálogo introduzca los cuatro elementos necesarios para calcular una probabilidad hipergeométrica.

1. Introduzca *0* en **Sample_s**.
2. Introduzca *5* en **Number_sample**.
3. Introduzca *40* en **Population_s**.
4. Introduzca *50* en **Number_pop**.
5. Introduzca 0 en **Cumulative** y haga clic en **OK**.
6. Excel calculará la probabilidad de 0 éxitos en 5 ensayos (0.000118937) y almacenará el resultado en la celda **B2**.

e. Para determinar la distribución de probabilidad completa, haga doble clic en la celda **B2**; se mostrará la función hipergeométrica. Reemplace el **0** a la derecha del paréntesis abierto con la referencia de la celda **A2**.

f. Arrastre el mouse a la esquina inferior derecha de la celda **B2** hasta que aparezca el símbolo + en líneas negras sólidas; enseguida haga clic, mantenga apretado el botón y resalte la columna **B**, a la celda **B7**; se indicará la probabilidad de un éxito para los diversos resultados.



6.3 Ahora se señalan los comandos de Excel para determinar la distribución de probabilidades de Poisson en la página 175:

a. En una hoja de cálculo de Excel en blanco, escriba la palabra ***Éxito*** en la celda **A1** y ***Probabilidad*** en **B1**. En las celdas **A2** a **A9**, escriba los números íntegros *0* a *7*; haga clic en **B2** para activarla.

b. En el menú superior, haga clic en la pestaña **Formulas** y seleccione **Insert Function *fx***.

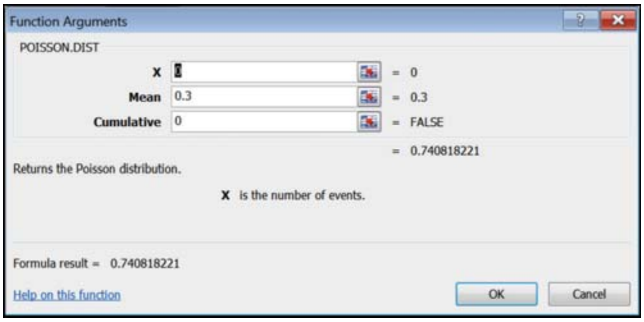
c. En el primer cuadro de diálogo, seleccione **Statistical** en la categoría de funciones y **POISSON.DIST** en la categoría del nombre de la función; haga clic en **OK**.

d. En el segundo cuadro de diálogo, ingrese los tres elementos necesarios para calcular una probabilidad de Poisson.

1. **Escriba** *0* **en X**.
2. Escriba *0.3* en **Media**.
3. Escriba la palabra *false* (falso) o el número *0* en **Cumulative** y haga clic en **OK**.
4. Excel calculará la probabilidad de 0 éxitos para una distribución de probabilidad de Poisson con una media de 0.3; el resultado, 0.74081822, se almacenará en la celda **B2**.

e. Para completar la distribución de probabilidad para éxitos de 1 a 7, haga doble clic en la celda **B2**; se mostrará la función Poisson. Reemplace el **0** a la derecha del paréntesis abierto con la celda de referencia **A2**.

f. Arrastre el mouse a la esquina inferior derecha de la celda **B2** hasta que aparezca un símbolo + en líneas negras sólidas, haga clic y mantenga apretado el botón, y resalte la columna **B** a la celda **B9**; se indicará la probabilidad de un éxito para los distintos valores de la variable aleatoria.



CAPÍTULO 7

7.1 He aquí los comandos de Excel que se requieren para generar la pantalla de la página 195:

a. Haga clic en la pestaña **Formulas** en la barra de herramientas, y seleccione **Insert Function *fx*** en el extremo izquierdo; del recuadro de categorías, seleccione **Statistical**, y debajo, **NORM.DIST**, y haga clic en **OK**.

b. En el cuadro de diálogo escriba *1100* en el cuadro correspondiente a X; *1000* para la **Mean**; *100* para la **Standard_dev**; *True* en el cuadro **Cumulative** y haga clic en **OK**.

c. El resultado se mostrará en el cuadro de diálogo; si hace clic en **OK**, la respuesta se presentará en su hoja de cálculo.

7.2 Ahora se indican los comandos de Excel que se requieren para generar la pantalla de la página 200:

a. Haga clic en la pestaña **Formulas** en la barra de herramientas, y seleccione **Insert Function *fx*** en el extremo izquierdo; del recuadro de categorías, seleccione **Statistical**, y debajo, **NORM.INV**, y haga clic en **OK**.

b. En el cuadro de diálogo, escriba *0.04* en **Probability**; *67900* en **Mean**, y *2050* en **Standard_dev**.

c. Los resultados se verán en el cuadro de diálogo. Observe que la respuesta es diferente a la de la página 200, como consecuencia del error de redondeo; si hace clic en **OK**, la respuesta también aparece en su hoja de cálculo.

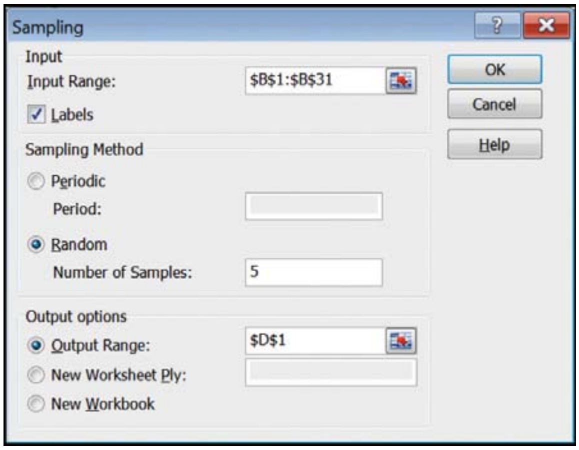
d. Intente introducir una **Probability** de *0.04*, una **Mean** de *0* y una **Standard_dev** de *1*. Se calculará el valor *z*.

CAPÍTULO 8

8.1 A continuación se muestran los comandos de Excel que se requieren en la página 223 para seleccionar una muestra aleatoria simple:

a. Seleccione **Data** en la barra de herramientas; en el extremo derecho seleccione **Data analysis** y enseguida **Sampling**, y haga clic en **OK**.

b. En **Input Range**, introduzca *B1:B31*. Como la columna tiene nombre, haga clic en el recuadro de **Labels**. Seleccione **Random** e introduzca el tamaño de la muestra como **Number of samples**, en este caso, *5*; haga clic en **Output Range** e indique el lugar de la hoja de cálculo en el que desea la información de la muestra. Observe que los resultados de su muestra diferirán de los del texto; asimismo, recuerde que Excel toma muestras con reemplazo, así que es posible que el valor de una población aparezca más de una vez en la muestra.



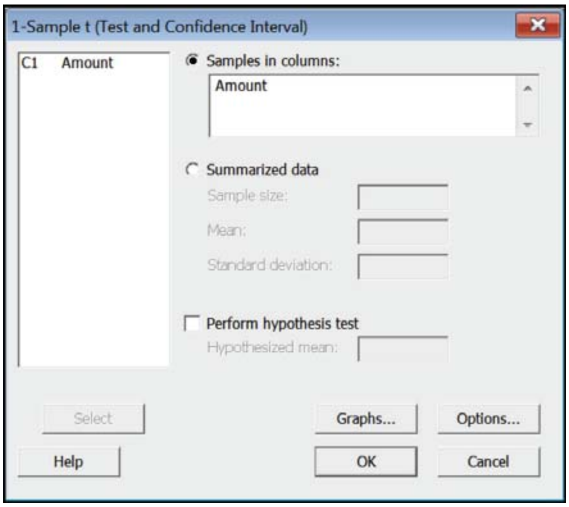
CAPÍTULO 9

9.1 Enseguida se muestran los comandos de Minitab para el intervalo de confianza para la cantidad gastada en el Inlet Square Mall en la página 262:

a. Ingrese las 20 cantidades gastadas en la columna **C1** y nombre la variable *Amount* (cantidad).

b. De la barra de herramientas, seleccione **Start**, **Basic Statistics** y haga clic en **1-Sample *t***.

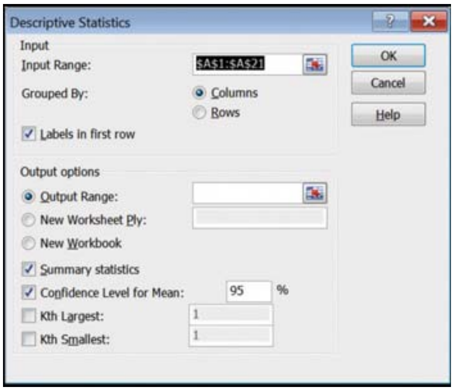
c. Seleccione **Samples in columns**; seleccione **Amount** y haga clic en **OK**.



9.2 A continuación se muestran los comandos de Excel para el intervalo de confianza para la cantidad gastada en el Inlet Square Mall en la página 262:

a. Seleccione la pestaña **Data** en el menú superior; después, seleccione **Data Analysis** en el extremo derecho y luego **Descriptive Statistics**; haga clic en **OK**.

b. En **Input Range**, escriba *A1:A21,* haga clic en **Labels in first row,** escriba *C1* en **Output Range**, haga clic en **Summary statistics** y **Confidence Level for Mean**, y haga clic en **OK**.

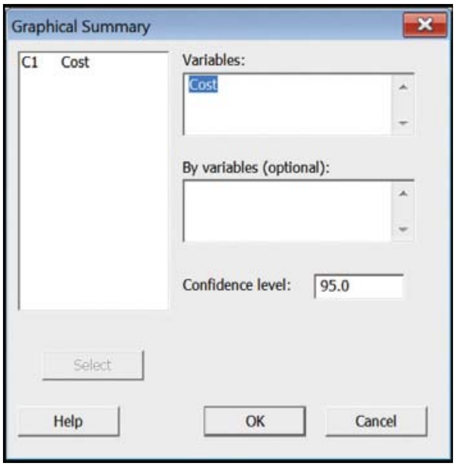


CAPÍTULO 10

10.1 A continuación se indican los comandos de Minitab para el histograma y la estadística descriptiva de la página 296:

a. Escriba las 26 observaciones de la muestra en la columna **C1** y nombre *Cost* a la variable.

b. En la barra de menú, seleccione **Stat**, **Basic Statistics** y **Graphical Summary**; en el cuadro de diálogo, seleccione **Cost** como variable y haga clic en **OK**.

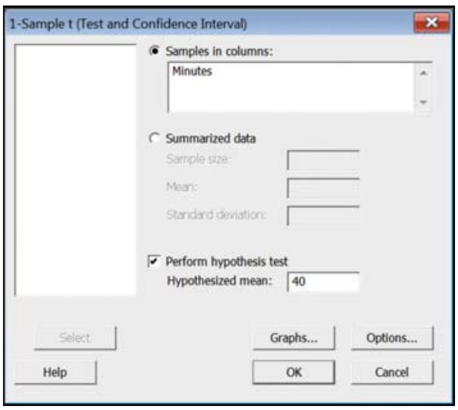


10.2 A continuación se muestran los comandos de Minitab para la prueba *t* de una muestra de la página 299:

a. Escriba los datos de la muestra en la columna **C1** y denomine *Length* a la variable.

b. En la barra de menú, seleccione **Stat**, **Basic Statistics**, **1-Simple *t*** y presione **Enter**.

c. Seleccione **Minutes** como variable, elija **Perform hypothesized mean**, introduzca el número *40*, haga clic en **Options**. En **Alternative**, seleccione **greater than**. Finalmente, haga doble clic en **OK**.



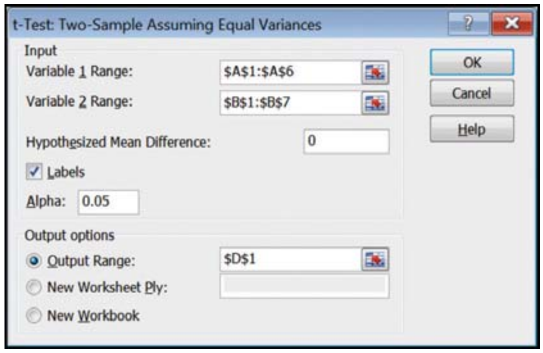
CAPÍTULO 11

11.1 Enseguida se indican los comandos de Excel para la prueba *t* de dos muestras de la página 319:

a. Ingrese los datos en las columnas A y B (o en cualquier otra columna) en la hoja de cálculo; utilice la primera fila de cada columna para escribir el nombre de la variable.

b. Seleccione la pestaña de **Data** en el menú superior; después, seleccione **Data Analysis** en el extremo derecho. Seleccione **t-Test: Two Sample Assuming Equal Variances**, y haga clic en **OK**.

c. En el cuadro de diálogo, indique que el rango de la **Variable 1** es de *A1* a *A6* y la **Variable 2** de *B1* a *B7*, ***Hypothesized Mean difference*** es *0*; haga clic en **Labels**, **Alpha** es *0.05*, y **Output Range** es *D1.* Haga clic en **OK**.

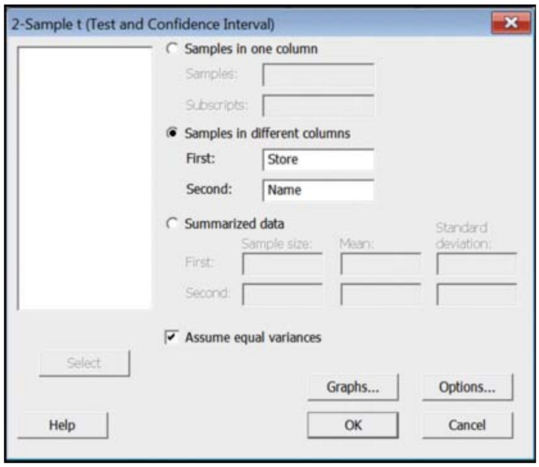


11.2 Ahora se señalan los comandos de Minitab para la prueba *t* de dos muestras de la página 323:

a. Coloque la cantidad absorbida por la marca de la tienda (Store) en *C1* y la cantidad absorbida por la otra marca (Name) de toallas de papel en *C2.*

b. De la barra de herramientas, seleccione **Stat**, **Basic Statistics** y después **2-Sample** y haga clic en **OK**.

c. En el siguiente cuadro de diálogo, seleccione **Samples in different columns**, seleccione *C1 Store* para la **Primera** columna y *C2 Name* para la **Segunda**, haga clic en el cuadro de diálogo siguiente para **Assume equal variances**, y haga clic en **OK**.

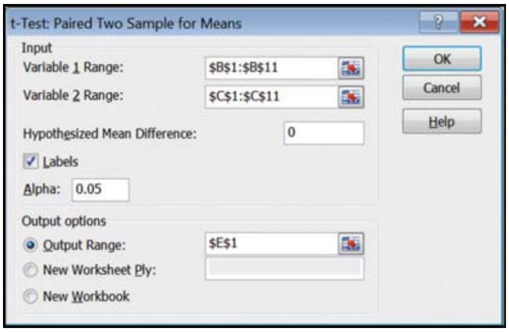


11.3 He aquí los comandos de Excel para la prueba *t* pareada de la página 327:

a. Ingrese los datos en las columnas B y C (o en cualquier otra) en la hoja de cálculo, con los nombres de las variables en la primera fila.

b. Seleccione la pestaña **Data** del menú superior. Después, seleccione **Data Analysis** en el extremo derecho; seleccione **t-Test: Paired Two Sample for Means**, y haga clic en **OK**.

c. En el cuadro de diálogo, indique que el rango de la **Variable 1** es de *B1* a *B11* y de la **Variable 2**, de *C1* a *C11;* **Hypothesized Mean Difference** es *0*; haga clic en **Labels**, **Alpha** es *0.05*, y **Output Range** es *E1.* Haga clic en **OK**.



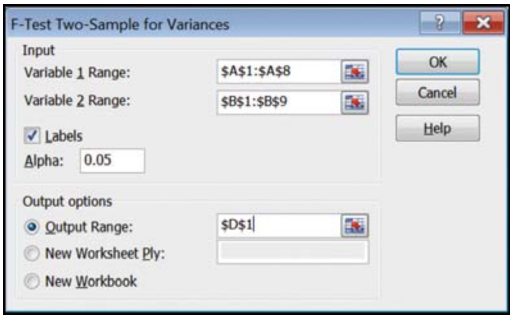
CAPÍTULO 12

12.1 A continuación se indican los comandos de Excel de la prueba de varianzas de la página 342:

a. Escriba los datos de la carretera U.S. 25 en la columna A y los de la I-75 en la columna B; ponga nombre a ambas columnas.

b. Seleccione la pestaña de **Data** en la barra de herramientas; en el extremo derecho, seleccione **Data Analysis**, seleccione **F-Test**: **Two-Sample for Variances** y haga clic en **OK**.

c. El rango de la primera variable es *A1:A8,* y *B1:B9* el de la segunda. Haga clic en **Labels**, escriba *0.05* para **Alpha**, seleccione *D1* para **Output Range** y haga clic en **OK**.

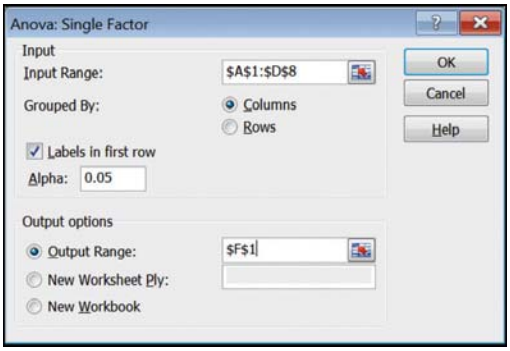


12.2 Enseguida se señalan los comandos de Excel de la prueba ANOVA de una vía de la página 348:

a. Escriba los datos en cuatro columnas identificadas: *Northern*, *TWA*, *Pocono* y *Branson*.

b. Seleccione la pestaña **Data** en la barra de herramientas; en el extremo derecho, seleccione **Data Analysis**; seleccione **ANOVA: Single Factor** y haga clic en **OK**.

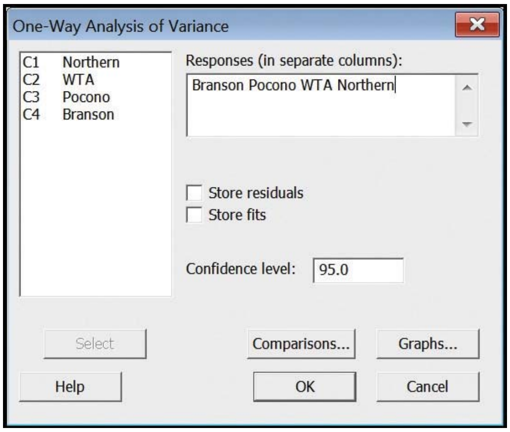
c. En el cuadro de diálogo siguiente establezca el rango de entrada *A1:D8*, haga clic en **Grouped by Columns**, seleccione **Labels in first row**, el cuadro de texto **Alpha** es *0.05,* y finalmente seleccione **Output Range** como *F1* y haga clic en **OK**.



12.3 Ahora se muestran los comandos en Minitab de las comparaciones por pares de la página 352:

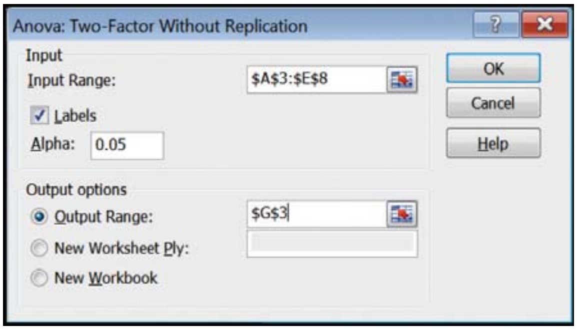
a. Escriba los datos en cuatro columnas e identifíquelas como *Northern*, *TWA*, *Pocono* y *Branson*.

b. Seleccione **Stat**, **ANOVA** y **One-way (Unstacked)**, seleccione e ingrese los nombres de las variables en el cuadro **Responses** haciendo clic en los nombres de las variables en el siguiente orden: *Branson*, *Pocono*, *TWA* y *Northern*; seleccione **Comparisons** y después **Fisher's, individual error rate**. Haga clic en **OK**.



12.4 He aquí los comandos de Excel de la prueba ANOVA de dos vías de la página 356:

a. En la primera fila de la primera columna escriba la palabra *Conductor*, después ingrese los cinco conductores en la primera columna. En la primera fila de las cuatro columnas siguientes escriba el nombre de las rutas; escriba los datos bajo cada nombre de la ruta.



b. Seleccione la pestaña **Data** en la barra de herramientas; en el extremo derecho, seleccione **Data Analysis**, seleccione **ANOVA: Two-Factor without Replication**, y después haga click en **OK**.

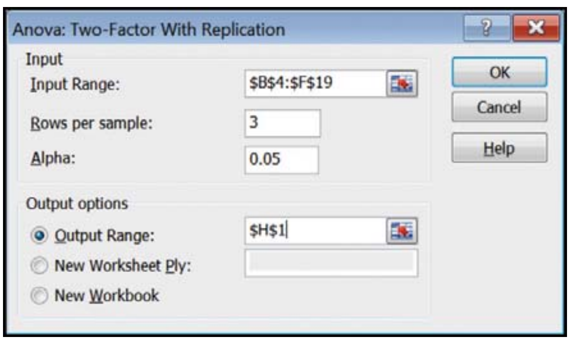
c. En el cuadro de diálogo, **Input Range** es *A3:E8*, haga clic en **Labels**, seleccione **G3** para **Output Range** y haga clic en **OK**.

12.5 A continuación se señalan los comandos de Excel de la prueba ANOVA de dos vías con interacción de la página 363:

a. Escriba los datos en Excel como se muestra en la página 361.

b. Seleccione la pestaña **Data** en la barra de herramientas; en el extremo derecho, seleccione **Data Analysis**. Seleccione **ANOVA: Two-Factor with Replication**, y después haga clic en **OK**.

c. En el cuadro de diálogo, en **Input Range**, resalte todo el rango de datos incluyendo las etiquetas de filas y columnas, escriba **Rows per simple** como *3,* ingrese *0.05* para **Alpha, Output Range** y la celda H1.



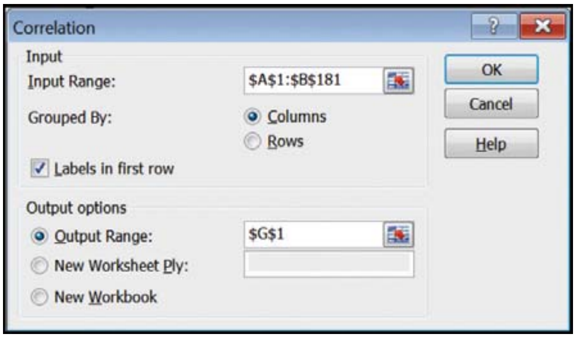
CAPÍTULO 13

13.1 Enseguida se muestran los comandos de Excel para calcular el coeficiente de correlación de la página 387:

a. Acceda a la base de datos de Applewood Auto Group (www.mhhe.com/uni/lind_ae16e).

b. Seleccione la pestaña **Data** en la parte superior; después, en el extremo derecho, seleccione **Data Analysis**. Seleccione **Correlation** y haga clic en **OK**.

c. Para **Input Range**, resalte las columnas **Age** y **Profit**, incluyendo las etiquetas en la fila 1; los datos están agrupados por **Columnas**. Marque el cuadro **Labels in first row**. Seleccione una celda en la hoja de cálculo como el comienzo del rango para obtener la correlación. Haga clic en **OK**.

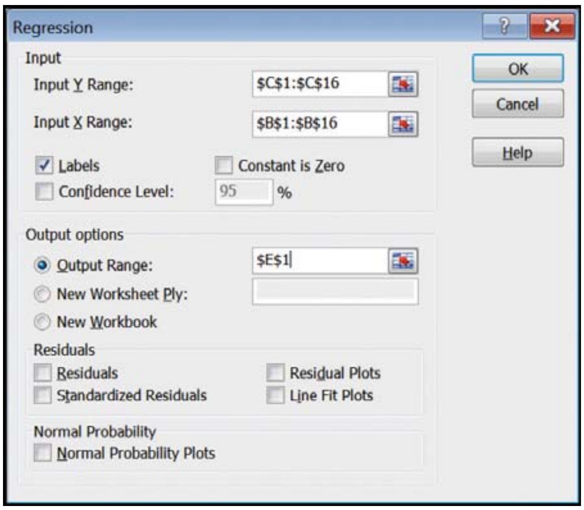


13.2 Ahora se indican los comandos para la salida en pantalla de Excel de la página 399:

a. Escriba los nombres de las variables en la fila 1 de las columnas A, B y C; escriba los datos en las filas 2 a 16 en las mismas columnas.

b. Seleccione **Data** en la barra de herramientas; en el extremo derecho, seleccione **Data Analysis**, **Regression** y haga clic en **OK**.

c. Para la hoja de cálculo tiene *Calls* (llamadas) en la columna B y *Sales* (ventas) en la columna C. **Input Y-Range** es *C1:C16,* e **Input X-Range**, *B1:B16*; haga *clic* en **Labels**, seleccione *E1* como **Output Range**, y haga *clic* en **OK**.

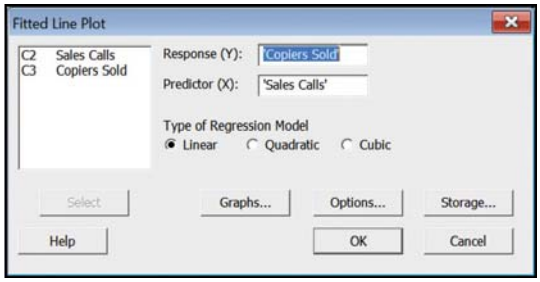


13.3 A continuación se muestran los comandos en Minitab para los intervalos de confianza y de predicción de la página 408:

a. Seleccione **Stat**, **Regression** y **Fitted line plot**.

b. En el siguiente cuadro de diálogo **Response (Y)** es *Copiers Sold* (copiadoras vendidas), y **Predictor (X)** es *Sales Calls* (llamadas de ventas). Seleccione **Linear** para el tipo de modelo de regresión y luego haga clic en **Options**.

c. En el cuadro de diálogo **Options** haga clic en **Display confidence and prediction interval**, utilice **95.0 para el nivel de confianza** y en el cuadro **Title** escriba el encabezado apropiado; luego haga clic en **OK** dos veces.



CAPÍTULO 14

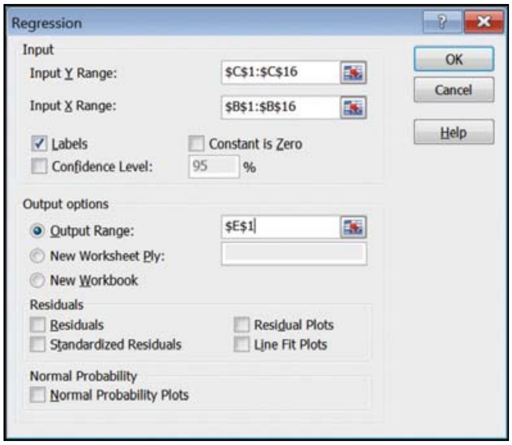
Nota: No se presentan todos los pasos para todo el software estadístico que se emplea en este capítulo; a continuación se muestran los pasos básicos.

14.1 Enseguida se indican los comandos de Excel para producir la salida de regresión múltiple en la página 427:

a. Importe los datos del sitio www.mhhe.com/uni/lind_ae16e; el nombre del archivo es **Tbl14-1**.

b. Seleccione la pestaña **Data** en el menú superior; en el extremo derecho, seleccione **Data Analysis**. Seleccione **Regression** y haga clic en **OK**.

c. **Input Y Range** es *A1:A21*, **Input X Range** es *B1:D21*; marque el cuadro **Labels**. **Output range** es *F1*. Haga clic en **OK**.

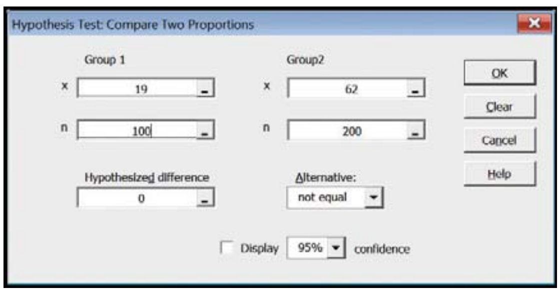


CAPÍTULO 15

15.1 Ahora se señalan los comandos de MegaStat para la muestra de proporciones de dos muestras en la página 481:

a. Seleccione **MegaStat** en la pantalla **Add-Ins**; del menú, seleccione **Hypothesis Tests** y después **Compare Two Independent Proportions**.

b. Ingrese los datos. Para **Group 1**, ingrese **x** como *19* y **n** como *100.* Para **Group 2,** ingrese **x** como *62* y **n** como *200.* Seleccione **OK**.

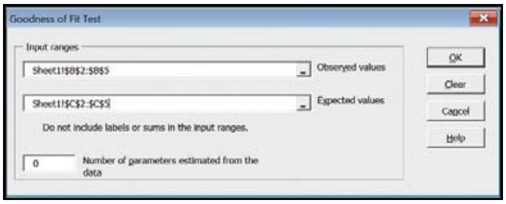


15.2 He aquí los comandos en MegaStat para elaborar la prueba de bondad de ajuste de *ji* cuadrada de la página 485:

a. Escriba la información de la tabla 15.2 en una hoja de cálculo, como se muestra.

b. Seleccione **MegaStat**, **Chi-Square/Crosstabs** y **Goodness of Fit Test**, y oprima **Enter**.

c. En el cuadro de diálogo seleccione *B2:B5* como **Observed values**, *C2:C5* como **Expected values** y escriba *0* como **Number of parameters estimated from the data**. Haga clic en **OK**.

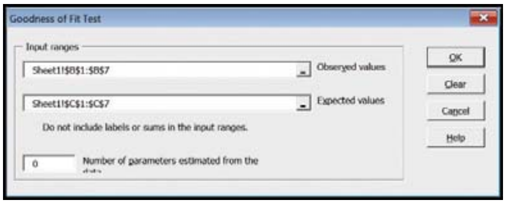


15.3 Los comandos de MegaStat para elaborar las pruebas de bondad de ajuste de *ji* cuadrada de las páginas 489 y 490 son los mismos, excepto por el número de artículos en las columnas de frecuencia observada y esperada. Solo se muestra un cuadro de diálogo.

a. Escriba la información sobre los niveles de administración de la página 489.

b. Seleccione **MegaStat**, **Chi-Square/Crosstabs** y **Goodness of Fit Test**, y oprima **Enter**.

c. En el cuadro de diálogo seleccione *B2:B7* como **Observed values**, *C1:C7* como **Expected values** y escriba *0* como **Number of parameters estimated from the data.** Haga clic en **OK.**

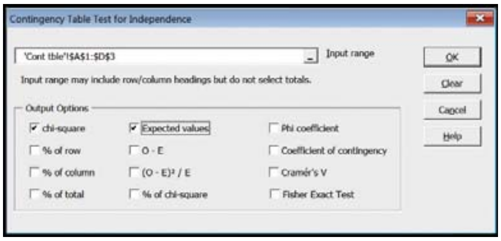


15.4 A continuación se indican los comandos de MegaStat para el análisis de tabla de contingencias en la página 496:

a. Ingrese la tabla 15.8 de la página 495 en las celdas **A1** a **D3**. Incluya las etiquetas de filas y columnas. NO INCLUYA la columna o fila Total.

b. Seleccione **MegaStat** de la pestaña de **Add-Ins**. Del menú, seleccione **Chi/square/Crosstab**, después **Contingency Table**.

c. Para **Input Range**, seleccione las celdas **A1** a **D3**. Marque los cuadros **Chi-square** y **Expected values**. Seleccione **OK**.



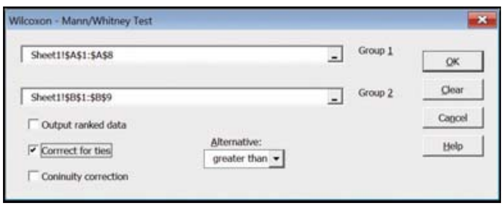
CAPÍTULO 16

16.1 Enseguida se señalan los comandos en MegaStat para Excel para la prueba de Wilcoxon de la suma de los rangos de la página 520:

a. Escriba el número de personas que no se presentaron en Atlanta en la columna A y en Chicago en la columna B.

b. Seleccione **MegaStat**, **Nonparametric Tests** y **Wilcoxon-Mann/Whitney Test**, y oprima **Enter**.

c. Para **Group 1**, utilice los datos sobre los vuelos de Atlanta (*A1:A9*), y para **Group 2,** los datos sobre los vuelos de Chicago (*B1:B8*); haga clic en **Correct for ties** y **one-tailed**, y *greater than* como **Alternative**; haga clic en **OK**.



16.2 Ahora se muestran los comandos de MegaStat para la prueba de Kruskal-Wallis de la página 523:

a. Ingrese los datos de la tabla 16.6 en una hoja de cálculo de Excel, incluyendo las etiquetas en las columnas A, B y C, comenzando en la fila 1.

b. Seleccione **MegaStat** en la pestaña **Add-Ins**; del menú, seleccione **Nonparametric Tests**, y después **Kruskal-Wallis Test**.

c. Para **Input Range**, seleccione las celdas *A1* a *C9.* Marque el cuadro **Correct for Ties**. Seleccione **OK**.

16.3 He aquí los comandos de Excel para la ANOVA de una vía de la página 524:

a. Ingrese los datos de la tabla 16.6 en una hoja de cálculo de Excel, incluyendo las etiquetas en las columnas A, B y C, comenzando en la fila 1.

b. Seleccione **Data** de la barra de herramientas; después, en el extremo derecho, seleccione **Data Análisis** y **ANOVA**: **Single Factor**, y luego haga clic en **OK**.

c. En el cuadro de diálogo, **Input Range** es *A1:C9*, haga clic en **Labels in First Row** y escriba *E1* como **Output Range**; haga clic en **OK**.

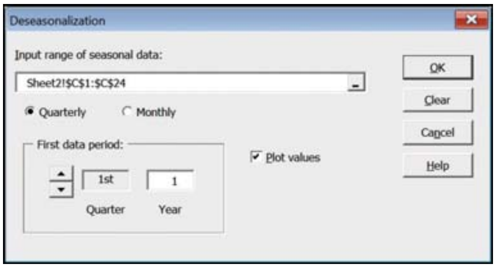
CAPÍTULO 18

18.1 A continuación se señalan los comandos de MegaStat para elaborar los índices estacionales de las páginas 585 y 586:

a. Escriba el periodo codificado y el valor de la serie de tiempo en dos columnas; quizá también desee incluir información sobre los años y trimestres.

b. Seleccione **MegaStat**, **Time/Forecasting** y **Deseasonalization**, y oprima **Enter**.

c. Escriba el rango de los datos, indique que los datos son del primer trimestre y haga clic en **OK**.



CAPÍTULO 19

19.1 Enseguida se muestran los comandos en Minitab para el diagrama de Pareto de la página 611:

a. Escriba las razones del consumo de agua en la columna **C1** y los galones consumidos en **C2**. Otorgue nombres apropiados a las columnas.

b. Haga clic en **Stat**, **Quality Tools**, **Pareto Chart** y luego oprima **Enter**.

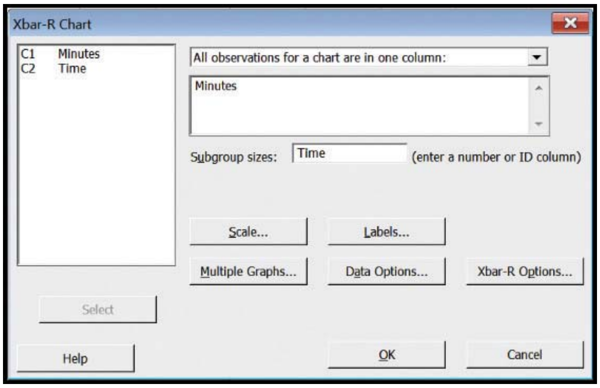
c. Indique **Defects of attribute data** como la variable *Usage* (uso) y **Frequencies** como la variable *Gallons* (galones); haga clic en **Options** y escriba un título para la gráfica; haga clic en **OK**.

19.2 Ahora se indican los comandos de Minitab para las gráficas de media y rango de la página 618:

a. Escriba la información de la tabla 19.1 o bájela del sitio www.mhhe.com/uni/lind_ae16e; el nombre del archivo es **Table 19-1**.

b. Haga clic en **Stat**, **Control Charts**, **Variables Charts for Subgroups**, **Xbar-R** y oprima **Enter**.

c. Seleccione **All observations for a chart are in one column**; en el cuadro inferior, seleccione la variable **Minutes**. Para **Subgroup sizes**, ingrese la variable *Time* (tiempo).

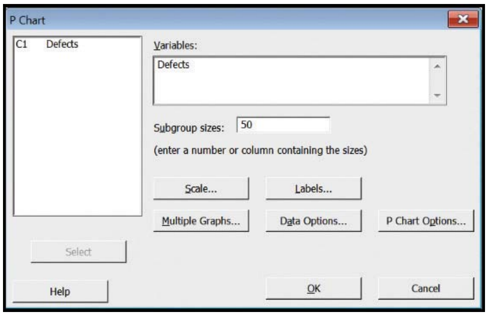


19.3 A continuación se muestran los comandos de Minitab de la gráfica p de la página 622:

a. Escriba los datos sobre el número de defectos de la página 621.

b. Haga clic en **Stat**, **Control Charts**, **Attribute Charts**, **P** y oprima **Enter**.

c. En **Variables**, seleccione *Defects*, luego escriba *50* para **Subgroup sizes**. Haga clic en **Labels**, escriba el título y haga clic en **OK** dos veces.

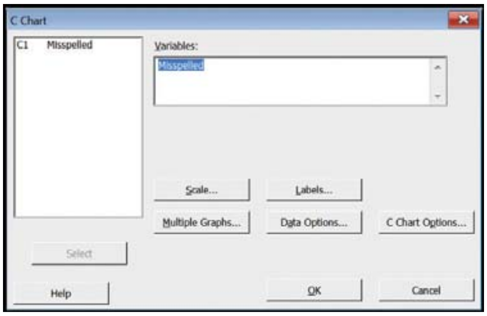


19.4 Enseguida se indican los comandos en Minitab para la gráfica de barras *c* de la página 623:

a. Escriba los datos del número de palabras mal escritas de la página 623.

b. Haga clic en **Stat**, **Control Charts**, **Attribute Charts**, **C** y oprima **Enter**.

c. Seleccione **Variable** e indique el número de palabras mal escritas, luego haga clic en **Labels** y escriba el título en el espacio proporcionado; después, haga clic en **OK** dos veces.



APÉNDICE D: RESPUESTAS A LOS EJERCICIOS IMPARES DE CADA CAPÍTULO, EJERCICIOS DE REVISIÓN Y SOLUCIONES A LOS TEST DE PRÁCTICA

Respuestas a los ejercicios impares de cada capítulo

CAPÍTULO 1

1. **a.** De intervalo **c.** Nominal **e.** Ordinal
b. De razón **d.** Nominal **f.** De razón

3. Las respuestas variarán

5. Los datos cualitativos no son numéricos, mientras que los cuantitativos sí lo son; los ejemplos varían según el estudiante

7. Una variable discreta puede asumir solo ciertos valores; una variable continua puede asumir una infinidad de valores dentro de cierto intervalo dado. El número de infracciones de tránsito que se levantaron diariamente durante el mes de febrero en Garden City Beach, Carolina del Sur, constituye una variable discreta. El peso de los camiones comerciales que pasan por la estación de pesaje ubicada en el kilómetro 195 en la autopista interestatal 95 en Carolina del Norte constituye una variable continua

9. **a.** Ordinal
b. De razón
c. El sistema más nuevo proporciona información sobre la distancia entre salidas

11. Si usted utilizara esta tienda como un establecimiento típico de Barnes & Noble, entonces serían datos simples; sin embargo, si usted la considerara como la única tienda de interés, los datos serían poblacionales

13.

	Variable discreta	Variable continua
Cualitativa	b. Género d. Preferencia por el refresco g. Posición del estudiante en clase h. Evaluación de un profesor de finanzas	
Cuantitativa	c. Volumen de ventas de reproductores MP3 f. Resultados del SAT i. Número de computadoras domésticas	a. Salario e. Temperatura

	Discreta	Continua
Nominal	b. Género	
Ordinal	d. Preferencia por el refresco g. Posición del estudiante en clase h. Evaluación de un profesor de finanzas	
De intervalo	f. Resultados del SAT	e. Temperatura
De razón	c. Volumen de ventas de reproductores MP3 i. Número de computadoras domésticas	a. Salario

15. Según la información de la muestra, 120/300 (40%) aceptarían una transferencia en el trabajo

17. **a.** Las ventas totales aumentaron a 736 725, calculado por 1 886 021 − 1 149 296, es decir, 64.1% de aumento
b. La participación de mercado en 2013 y 2009, respectivamente, es:

General Motors	22.2%	22.0%	American Honda	10.7%	12.4%
Ford Motor	19.2%	16.2%	Nissan NA	9.6%	9.4%
Toyota	17.2%	19.7%	Hyundai	5.1%	4.8%
Chrysler	13.6%	12.7%	Mazda	2.5%	2.8%

Ford ganó 3.0%, Toyota perdió 2.6% y Honda perdió 1.7% de sus participaciones en el mercado
c. Los cambios porcentuales son:

General Motors	165.8%
Ford Motor	197.7%
Toyota	142.9%
Chrysler	175.6%
American Honda	141.4%
Nissan NA	167.0%
Hyundai	174.2%
Mazda	145.4%

Ford tuvo el cambio más alto en su porcentaje (197.7%), mientras que Toyota, Honda y Mazda tuvieron los aumentos más pequeños en sus porcentajes (alrededor de 142%)

19. Las ganancias aumentaron aproximadamente 3 000 millones de dólares por año durante el periodo; sin embargo, hubo un gran aumento en 2008 y en 2009 una gran disminución. Quizás los 15 000 millones de dólares de las ganancias de 2008 fueron un tanto "avanzadas" durante los siguientes uno o dos años

21. **a.** "Liga" es una variable cualitativa; las otras son cuantitativas
b. "Liga" es una variable de nivel nominal; las otras son variables de nivel de razón

CAPÍTULO 2

1. 25% de participación de mercado

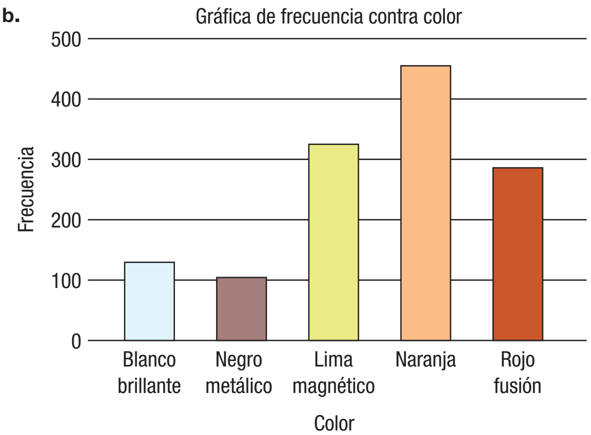
3.

Estación	Frecuencia	Frecuencia relativa
Invierno	100	0.10
Primavera	300	0.30
Verano	400	0.40
Otoño	200	0.20
	1 000	1.00

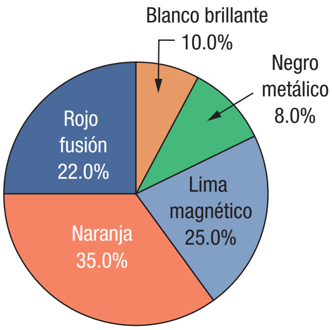
5. **a.** Tabla de frecuencias

Color	Frecuencia	Frecuencia relativa
Blanco brillante	130	0.10
Negro metálico	104	0.08
Lima magnético	325	0.25
Naranja	455	0.35
Rojo fusión	286	0.22
Total	1300	1.00

b.



c. Gráfica de pastel de frecuencias de color



d. 350 000 naranja, 250 000 lima, 220 000 rojo, 100 000 blanco y 80 000 negro, calculados multiplicando la frecuencia relativa por la producción total de 1 000 000 de unidades de producción

7. $2^5 = 32$, $2^6 = 64$, por lo tanto, 6 clases

9. $2^7 = 128$, $2^8 = 256$, sugiere 8 clases

$i \geq \frac{\$567 - \$235}{8} = 41$ Intervalos de clase de 40 o 50 serían aceptables

11. a. $2^4 = 16$ Sugiere 5 clases

b. $i \geq \frac{31 - 25}{5} = 1.2$ Utilice un intervalo de 1.5

c. 24

d.

Unidades	*f*	Frecuencia relativa
24.0 hasta 25.5	2	0.125
25.5 hasta 27.0	4	0.250
27.0 hasta 28.5	8	0.500
28.5 hasta 30.0	0	0.000
30.0 hasta 31.5	2	0.125
Total	16	1.000

e. La concentración más grande se encuentra en la clase de 27.0 hasta 28.5 (8)

13. a.

Número de visitas	*f*
0 hasta 3	9
3 hasta 6	21
6 hasta 9	13
9 hasta 12	4
12 hasta 15	3
15 hasta 18	1
Total	51

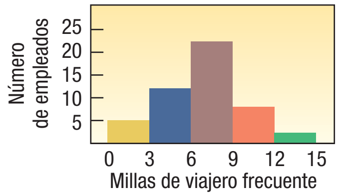
b. El grupo más grande de compradores (21) compra en el BiLo Supermarket tres, cuatro o cinco veces en un lapso de un mes. Algunos clientes visitan la tienda solo una vez durante el mes, pero otros compran tanto como 15 veces

c.

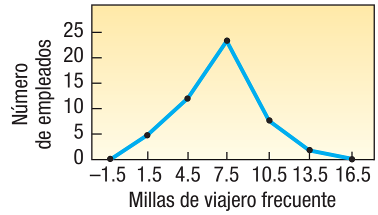
Número de visitas	Porcentaje del total
0 hasta 3	17.65
3 hasta 6	41.18
6 hasta 9	25.49
9 hasta 12	7.84
12 hasta 15	5.88
15 hasta 18	1.96
Total	100.00

15. a. Histograma
b. 100
c. 5
d. 28
e. 0.28
f. 12.5
g. 13

17. a. 50
b. 1 500 millas
c.



d. $X = 1.5$, $Y = 5$
e.



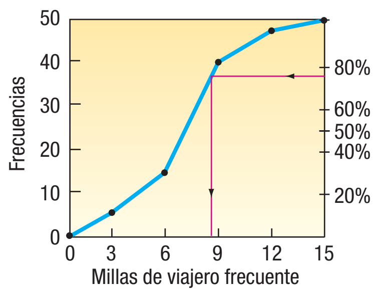
f. En el caso de los 50 empleados, alrededor de la mitad viajó entre 6 000 y 9 000 millas; cinco empleados viajaron menos de 3 000 millas y dos viajaron más de 12 000 millas

19. a. 40
b. 5
c. 11 o 12
d. Aproximadamente \$18/hr
e. Aproximadamente \$9/hr
f. Aproximadamente 75%

21. a. 5
b.

Millas	*CF*
Menos de 3	5
Menos de 6	17
Menos de 9	40
Menos de 12	48
Menos de 15	50

c.

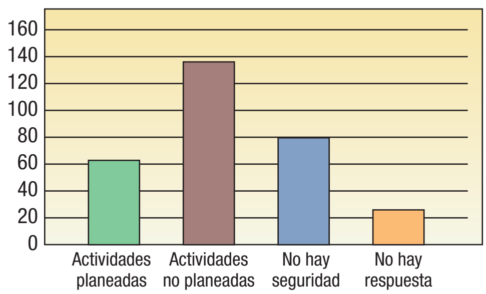


d. Aproximadamente 8 700 millas

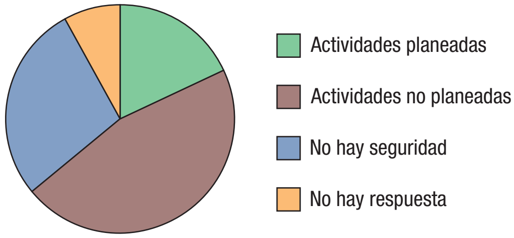
23. a. Una variable cualitativa utiliza tanto la escala de medición nominal como la ordinal; por lo general es resultado de conteos. Las variables cuantitativas son discretas o continuas, y existe un orden natural en el caso de los resultados de una variable cuantitativa. Las variables cuantitativas pueden utilizar la escala de medición de intervalo o de razón

b. Ambos tipos de variables se pueden utilizar para muestras y poblaciones

25. a. Tabla de frecuencias
b.



c.



d. Una gráfica de pastel sería mejor, ya que muestra con claridad que casi la mitad de los clientes prefieren las actividades no planeadas

27. $2^6 = 64$ y $2^7 = 128$, sugieren 7 clases

29. **a.** 5, porque $2^4 = 16 < 25$ y $2^5 = 32 > 25$

b. $i \geq \dfrac{48 - 16}{5} = 6.4$ Utilice un intervalo de 7

c. 15

d.

Clase	Frecuencia	
15 hasta 22	𝍷𝍷𝍷	3
22 hasta 29	𝍸𝍷𝍷𝍷	8
29 hasta 36	𝍸𝍷𝍷	7
36 hasta 43	𝍸	5
43 hasta 50	𝍷𝍷	2
		25

e. Es casi simétrica; la mayoría de los valores se encuentran entre 22 y 36

31. **a.** $2^5 = 32$, $2^6 = 64$, 6 clases recomendadas

b. $i = \dfrac{10 - 1}{6} = 1.5$, use un intervalo de 2

c. 0

d.

Clase	Frecuencia
0 hasta 2	1
2 hasta 4	5
4 hasta 6	12
6 hasta 8	17
8 hasta 10	8
10 hasta 12	2

e. La distribución es casi simétrica, o en forma de campana, con un gran pico en medio de ambas clases de 4 hasta 8

33.

Clase	Frecuencia
0 hasta 200	19
200 hasta 400	1
400 hasta 600	4
600 hasta 800	1
800 hasta 1 000	2

Esta distribución tiene un sesgo positivo, con una larga “cola” hacia la derecha, o valores positivos. Observe que las siete tonadas más populares representan 4 342 reproducciones de un total de 5 968, o cerca de 73% de todas las reproducciones

35. **a.** 56

b. 10 (calculado por 60 − 50)

c. 55

d. 17

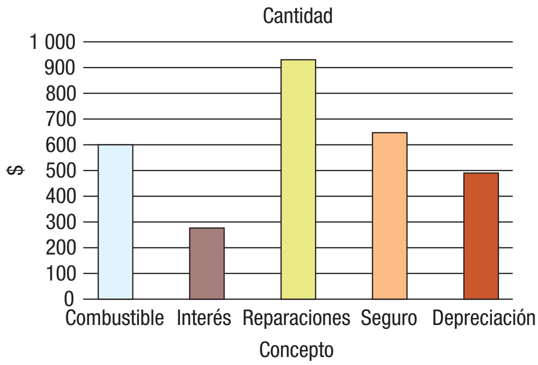
37. **a.** \$35, porque el mínimo es (\$265 − \$82)/6 = \$30.5

b.

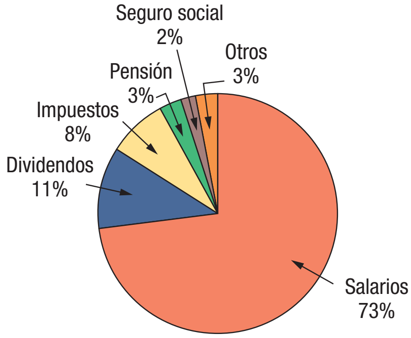
\$ 70 hasta \$105	4
105 hasta 140	17
140 hasta 175	14
175 hasta 210	2
210 hasta 245	6
245 hasta 280	1

c. Las compras variaron de cantidades bajas de alrededor de \$70 hasta alrededor de \$280. La concentración se encuentra en las clases \$105 hasta \$140 y \$140 hasta \$175

39. Las gráficas de barra son preferibles cuando el objetivo es comparar la cantidad actual en cada categoría



41.

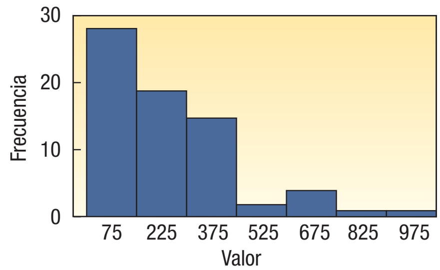


Ingreso	Porcentaje	Acumulado
Salarios	73	73
Dividendos	11	84
Impuestos	8	92
Pensiones	3	95
Seguro social	2	97
Otros	3	100

Por mucho, la mayor parte del ingreso en Carolina del Sur es el que se gana en el trabajo. Casi tres cuartas partes del ingreso bruto ajustado provienen de sueldos y salarios. Los dividendos y los impuestos contribuyen con otro 10% cada uno

43. **a.** Como $2^6 = 64 < 70 < 128 = 2^7$, se recomiendan 7 clases. El intervalo deberá ser $(1\,002.2 - 3.3)/7 = 142.7$, por lo menos. Utilice 150 como valor conveniente

b.

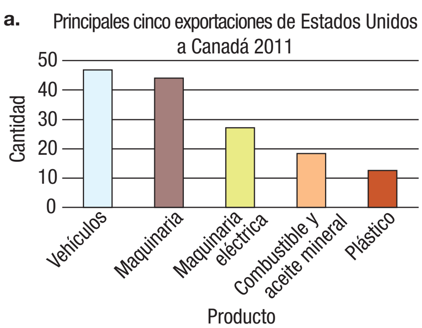


45. **a.** Gráfica de pastel

b. 700, calculado por 0.7(1 000)

c. Sí, 0.70 + 0.20 = 0.90

47. **a.**

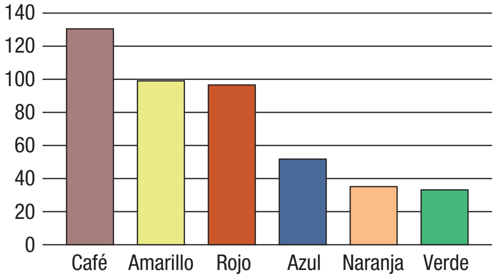


b. 23, calculado por (18.4 + 46.9)/281

c. 44, calculado por (18.4 + 46.9)/(46.9 + 442.2 + 27.1 + 18.4 + 12.6)

49.

Color	Frecuencia
Café	130
Amarillo	98
Rojo	96
Azul	52
Naranja	35
Verde	33
	444



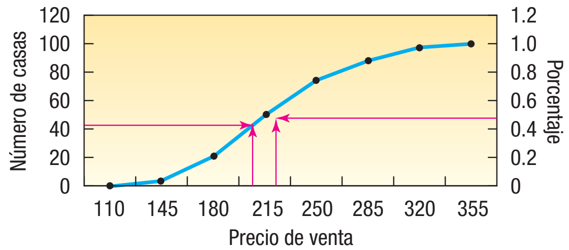
51. $i \geq \dfrac{345.3 - 125.0}{7} = 31.47$ Utilice un intervalo de 35

Precio de venta	*f*	*FC*
110 hasta 145	3	3
145 hasta 180	19	22
180 hasta 215	31	53
215 hasta 250	25	78
250 hasta 285	14	92
285 hasta 320	10	102
320 hasta 355	3	105

a. La mayoría de las casas (53%) se encuentran en el rango de 180 hasta 250

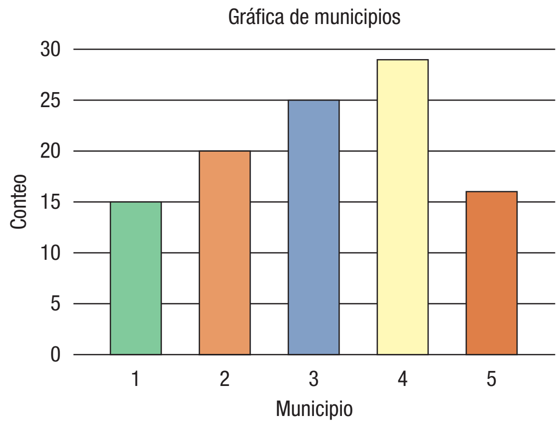
b. El valor más alto se encuentra cerca de 355; el más bajo, cerca de 110

c.



Alrededor de 42 casas se vendieron en menos de 200
Aproximadamente 55% de las casas se vendieron en menos de 220, así que 45% se vendió en más
Menos de 1% de las casas se vendió en menos de 125

d.



Los municipios 3 y 4 tienen más ventas que el promedio, y los 1 y 5 estuvieron un poco abajo del promedio

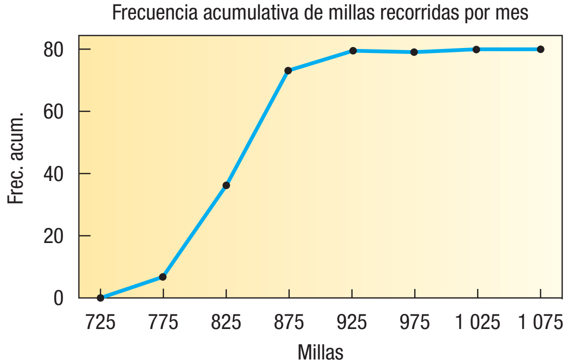
53. Como $2^6 = 64 < 80 < 128 = 2^7$, utilice 7 clases. El intervalo debe ser por lo menos (1 008 − 741)/7 = 38.14 millas (utilice 40); la distribución de frecuencia resultante es:

Clase	*f*
730 hasta 770	5
770 hasta 810	17
810 hasta 850	37
850 hasta 890	18
890 hasta 930	1
930 hasta 970	0
970 hasta 1 010	2

a. La cantidad típica de millas recorridas es 830. El rango es de 730 hasta 1 010 millas

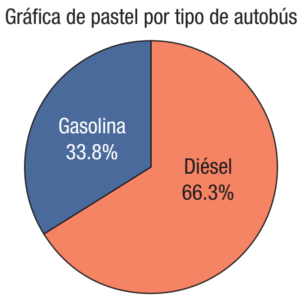
b. La distribución tiene forma de campana, alrededor de 830; sin embargo, hay dos datos atípicos de hasta alrededor de 1 000 millas

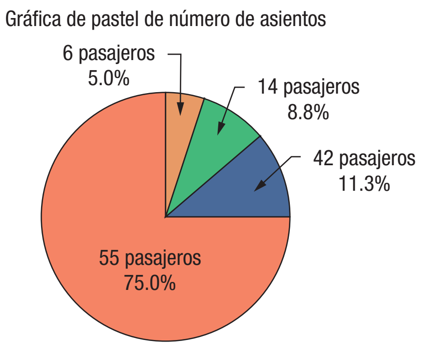
c.



40% de los autobuses recorrieron menos de 820 millas
59 autobuses recorrieron menos de 850 millas

d.





La primera gráfica muestra que alrededor de dos tercios de los autobuses son diésel; el segundo diagrama indica que cerca de tres cuartos de los autobuses tienen 55 asientos

CAPÍTULO 3

1. $\mu = 5.4$, calculado por 27/5

3. **a.** $\bar{x} = 7.0$, calculado por 28/4

b. $(5 - 7) + (9 - 7) + (4 - 7) + (10 - 7) = 0$

5. $\bar{x} = 14.58$, calculado por $43.74/3$

7. **a.** 15.4, calculado por $154/10$
b. Parámetro de la población, ya que incluye a todos los vendedores de Midtown Ford

9. **a.** \$54.55, calculado por \$1 091/20
b. Una estadística muestral, suponiendo que la compañía de electricidad atienda a más de 20 clientes

11. $\bar{x} = \dfrac{\Sigma x}{n}$ así que
$\Sigma x = \bar{x} \cdot n = (\$5\,430)(30) = \$162\,900$

13. **a.** Sin moda
b. El valor dado sería la moda
c. 3 y 4 bimodal

15. **a.** Media $= 3.583$
b. Mediana $= 5$
c. Moda $= 5$

17. **a.** Mediana $= 2.9$
b. Moda $= 2.9$

19. $\bar{x} = \dfrac{647}{11} = 58.82$
Mediana $= 58$, Moda $= 58$
Cualquiera de las tres medidas sería satisfactoria

21. **a.** $\bar{x} = \dfrac{90.4}{12} = 7.53$
b. Mediana $= 7.45$; hay varias modas: 6.5, 7.3 y 8.7
c. $\bar{x} = \dfrac{33.8}{4} = 8.45$
Mediana $= 8.7$
Alrededor de 1 punto porcentual más alto en invierno

23. \$22.91, determinado por $\dfrac{300(\$20) + 400(\$25) + 400(\$23)}{300 + 400 + 400}$

25. \$17.75, determinado por $(\$400 + \$750 + \$2\,400)/200$

27. 12.8 de incremento porcentual, determinado por
$\sqrt[5]{(1.08)(1.12)(1.14)(1.26)(1.05)} = 1.1228$

29. 12.28% de incremento porcentual, determinado por
$\sqrt[5]{(1.094)(1.138)(1.117)(1.119)(1.147)} = 1.1228$

31. 2.43%, calculado por $\sqrt[12]{\dfrac{229.6}{172.2}} - 1$

33. 29.0%, calculado por $\sqrt[27]{\dfrac{327\,577\,529}{340\,213}} - 1$

35. **a.** 7, determinado por $10 - 3$
b. 6, determinado por $30/5$
c. 6.8, determinado por $34/5$
d. La diferencia entre el número más alto vendido (10) y el número más bajo vendido (3) es 7. La desviación típica al cuadrado de 6 es 6.8

37. **a.** 30, determinado por $54 - 24$
b. 38, determinado por $380/10$
c. 74.4, determinado por $744/10$
d. La diferencia entre 54 y 24 es 30. El promedio de desviaciones al cuadrado de 38 es 74.4

39.

Estado	Media	Mediana	Rango
California	33.10	34.0	32
Iowa	24.50	25.0	19

Las puntuaciones de la media y la mediana fueron más altas, pero había aún más variación en California

41. **a.** 5
b. 4.4, determinado por
$$\frac{(8-5)^2 + (3-5)^2 + (7-5)^2 + (3-5)^2 + (4-5)^2}{5}$$

43. **a.** \$2.77
b. 1.26, determinado por
$$\frac{(2.68 - 2.77)^2 + (1.03 - 2.77)^2 + (2.26 - 2.77)^2 + (4.30 - 2.77)^2 + (3.58 - 2.77)^2}{5}$$

45. **a.** Rango: 7.3, determinado por $11.6 - 4.3$. Media aritmética: 6.94, determinada por $34.7/5$. Varianza: 6.5944, determinada por $32.972/5$. Desviación estándar: 2.568, determinada por $\sqrt{6.5944}$
b. Dennis tiene un rendimiento medio más alto ($11.76 > 6.94$); no obstante, tiene una mayor dispersión en sus rendimientos sobre el capital ($16.89 > 6.59$)

47. **a.** $\bar{x} = 4$
$$s^2 = \frac{(7-4)^2 + \cdots + (3-4)^2}{5-1} = \frac{22}{5-1} = 5.5$$
b. $s = 2.3452$

49. **a.** $\bar{x} = 38$
$$s^2 = \frac{(28-38)^2 + \cdots + (42-38)^2}{10-1} = \frac{744}{10-1} = 82.667$$
b. $s = 9.0921$

51. **a.** $\bar{x} = \dfrac{951}{10} = 95.1$
$$s^2 = \frac{(101-95.1)^2 + \cdots + (88-95.1)^2}{10-1} = \frac{1\,112.9}{9} = 123.66$$
b. $s = \sqrt{123.66} = 11.12$

53. Alrededor de 69%, determinado por $1 - 1/(1.8)^2$

55. **a.** Aproximadamente 95%
b. 47.5%, 2.5%

57. Como en una distribución de frecuencias no se conocen los valores exactos, se utiliza el punto medio para cada miembro de dicha clase

59.

Clase	f	M	fM	$(M - \bar{x})$	$f(M - \bar{x})^2$
20 hasta 30	7	25	175	−22.29	3 477.909
30 hasta 40	12	35	420	−12.29	1 812.529
40 hasta 50	21	45	945	−2.29	110.126
50 hasta 60	18	55	990	7.71	1 069.994
60 hasta 70	12	65	780	17.71	3 763.729
	70		3 310		10 234.287

$$\bar{x} = \frac{3\,310}{70} = 47.29$$
$$s = \sqrt{\frac{10\,234\,287}{70-1}} = 12.18$$

61.

Número de clientes	f	M	fM	$(M - \bar{x})$	$f(M - \bar{x})^2$
20 hasta 30	1	25	25	−19.8	392.04
30 hasta 40	15	35	525	−9.8	1 440.60
40 hasta 50	22	45	990	0.2	0.88
50 hasta 60	8	55	440	10.2	832.32
60 hasta 70	4	65	260	20.2	1 632.16
	50		2 240		4 298.00

$$\bar{x} = \frac{2\,240}{50} = 44.8$$
$$s = \sqrt{\frac{4\,298}{50-1}} = 9.37$$

63. **a.** Media $= 5$, determinada mediante $(6 + 4 + 3 + 7 + 5)/5$
La mediana es 5, calculada al volver a ordenar los valores y seleccionar el valor medio
b. Población, ya que se incluyen todos los patrones
c. $\Sigma(x - \mu) = (6-5) + (4-5) + (3-5) + (7-5) + (5-5) = 0$

65. $\bar{x} = \dfrac{545}{16} = 34.06$
Mediana $= 37.50$

67. La media es 37.675, calculada por $1\,427/40$
La mediana es 36, calculada ordenando los datos y promediando las observaciones 20 y 21

69. $\bar{x}_w = \dfrac{\$5.00(270) + \$6.50(300) + \$8.00(100)}{270 + 300 + 100} = \$6.12$

71. $\bar{x}_w = \dfrac{15\,300(4.5) + 10\,400(3.0) + 150\,600(10.2)}{176\,300} = 9.28$

73. $\text{GM} = \sqrt[21]{\dfrac{6\,286\,800}{5\,164\,900}} - 1 = 1.0094 - 1.0 = 0.0094$

75. **a.** 55, calculado por 72 − 17

b. 17.6245, calculado por la misma raíz cuadrada de 2795.6/9

77. **a.** Esta es una población, porque incluye a todas las universidades públicas de Ohio

b. La media es 21 534

c. La mediana es 17 529

d. El rango es 54 235

e. La desviación estándar es 13 101

79. **a.** Se llevaron a cabo 13 vuelos; se consideran todos los elementos

b. $\mu = \dfrac{2\,259}{13} = 173.77$

Mediana = 195

c. Rango = 310 − 7 = 294

$$s = \sqrt{\frac{133\,846}{13}} = 101.47$$

81. **a.** La media es $717.20, calculada por $17 930/25. La mediana es $717.00 y hay dos modas: $710 y $722

b. El rango es $90, calculado por $771 − $681, y la desviación estándar es $24.87, calculada por la raíz cuadrada de 14 850/24

c. De $667.46 hasta $766.94, calculado por $717.20 ± 2($24.87)

83. **a.** $\bar{x} = \dfrac{273}{30} = 9.1$, Mediana = 9

b. Rango = 18 − 4 = 14

$$s = \sqrt{\frac{368.7}{30-1}} = 3.57$$

c. $2^5 = 32$, de modo que se sugieren 5 clases

$i = \dfrac{18-4}{5} = 2.8$ Use $i = 3$

Clase	M	f	fM	$M-\bar{x}$	$(M-\bar{x})^2$	$f(M-\bar{x})^2$
3.5 hasta 6.5	5	10	50	−4	16	160
6.5 hasta 9.5	8	6	48	−1	1	6
9.5 hasta 12.5	11	9	99	2	4	36
12.5 hasta 15.5	14	4	56	5	25	100
15.5 hasta 18.5	17	1	17	8	64	64
			270			366

d. $\bar{x} = \dfrac{270}{30} = 9.0$

$$s = \sqrt{\frac{366}{30-1}} = 3.552$$

La media y la desviación estándar de los datos agrupados se estiman a partir de la media de las desviaciones estándar de los valores reales

85. $\bar{x} = 13 = \dfrac{910}{70}$

$s = 5.228 = \sqrt{1\,807.5/69}$

87. **a.** El salario medio del equipo es $98 020 000, y la mediana es $85 750 000. Como la distribución está sesgada, el valor mediano de $85 750 000 es más típico

b. El rango es $142 800 000, calculado por 198 000 000 − 55 200 000. La desviación estándar es $36 830 000. Cuando menos 75% de los salarios del equipo está entre $24 360 000 y $171 680 000, calculado por $98 020 000 más o menos 2($36 830 000)

c. 8.63% por año, calculado por $\sqrt[23]{\dfrac{3\,440\,000}{512\,930}} - 1$

La nómina media del equipo es $104 800 000, y la desviación estándar es $44 400 000 en la Liga Americana. La nómina media del equipo es $92 090 000 y la desviación estándar es $28 850 000 en la Liga Nacional. No hay mucha diferencia en las nóminas medias, pero hay mayor dispersión en la Liga Americana

CAPÍTULO 4

1. En un histograma las observaciones se encuentran agrupadas, así que pierden su identidad individual. Con un diagrama de puntos se conserva la identidad de cada observación

3. **a.** Diagrama de puntos **b.** 15 **c.** 1, 7 **d.** 2 y 3

5. **a.** De 620 a 629 **b.** 5 **c.** 621, 623, 623, 627, 629

7. **a.** 25 **b.** Uno **c.** 38 106 **d.** 60, 61, 63, 63, 65, 65, 69 **e.** Sin valor **f.** 9 **g.** 9 **h.** 76 **i.** 16

9.

Tallo	Hojas
0	5
1	28
2	
3	0024789
4	12366
5	2

Se estudiaron en total 16 llamadas; su número osciló entre 5 y 52; 7 de los 16 suscriptores hicieron entre 30 y 39 llamadas

11. Mediana = 53, calculada mediante $(11 + 1)(\frac{1}{2})$ ∴ sexto valor a partir del más bajo

$Q_1 = 49$, calculado mediante $(11 + 1)(\frac{1}{4})$ ∴ tercer valor a partir del más bajo

$Q_3 = 55$, calculado mediante $(11 + 1)(\frac{3}{4})$ ∴ noveno valor a partir del más bajo

13. **a.** $Q_1 = 33.25$, $Q_3 = 50.25$

b. $D_2 = 27.8$, $D_8 = 52.6$

c. $P_{67} = 47$

15. **a.** 350

b. $Q_1 = 175$, $Q_3 = 930$

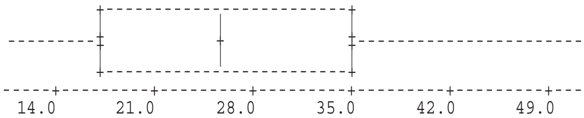
c. 930 − 175 = 775

d. Menos de 0, o más de 2 060

e. No hay datos atípicos

f. La distribución tiene un sesgo positivo

17.



La distribución tiene un sesgo ligeramente positivo; observe que la línea punteada sobre 35 es más larga que la que se encuentra debajo de 18

19. **a.** La media es 30.8, calculada mediante 154/5. La mediana es 31.0, y la desviación estándar es 3.96, calculada mediante

$$s = \sqrt{\frac{62.8}{4}} = 3.96$$

b. −0.15, calculado por $\dfrac{3(30.8 - 31.0)}{3.96}$

c.

Salario	$\left(\dfrac{x-\bar{x}}{s}\right)$	$\left(\dfrac{x-\bar{x}}{s}\right)^3$
36	1.313131	2.264250504
26	−1.212121	−1.780894343
33	0.555556	0.171467764
28	−0.707071	−0.353499282
31	0.050505	0.000128826
		0.301453469

0.125, calculado mediante [5/(4 × 3)] × 0.301

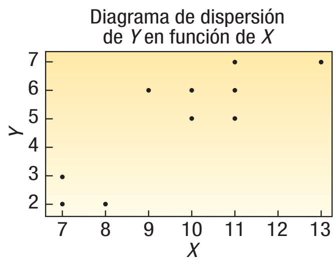
21. **a.** La media es de 21.93, calculada por medio de 328.9/15. La mediana es de 15.8, y la desviación estándar, de 21.18, calculada por medio de

$$s = \sqrt{\frac{6\,283}{14}} = 21.18$$

b. 0.868, calculado por [3(21.93 − 15.8)]/21.18

c. 2.444, calculado por [15/(14 × 13)] × 29.658

23.



Existe una relación positiva entre las variables

25. **a.** Ambas variables están en escala nominal
b. Tabla de contingencias
c. Es dos veces más probable que los hombres ordenen un postre. Según la tabla, 32% de los hombres pidieron postre y solo 15% de las mujeres lo hicieron

27. **a.** Diagrama de puntos
b. 15
c. 5

29. Tallo y hojas $N = 23$

3	3	222
5	3	77
11	4	000002
(6)	4	666666
6	5	222222

31. **a.** $L_{50} = (20 + 1)\frac{50}{100} = 10.50$

$$\text{Mediana} = \frac{83.7 + 85.6}{2} = 84.65$$

$L_{25} = (21)(0.25) = 5.25$
$Q_1 = 66.6 + 0.25(72.9 - 66.6) = 68.175$
$L_{75} = 21(0.75) = 15.75$
$Q_3 = 87.1 + 0.75(90.2 - 87.1) = 89.425$

b. $L_{26} = 21(0.26) = 5.46$
$P_{26} = 66.6 + 0.46(72.9 - 66.6) = 69.498$
$L_{83} = 21(0.83) = 17.43$
$P_{83} = 93.3 + 0.43(98.6 - 93.3)$
$= 95.579$

c.
```
                 ------------------------
    -------------I                 +     I--------------
                 ------------------------
  --------+---------+---------+---------+---------+--------C20
        64.0      72.0      80.0      88.0      96.0
```

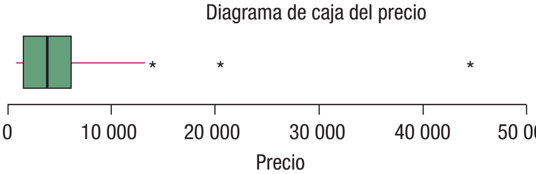
33. **a.** $Q_1 = 26.25$, $Q_3 = 35.75$, Mediana $= 31.50$
```
                  -----------------------
    -------------I          +           I------------------
                  -----------------------
  ------+---------+---------+---------+---------+---------+
      24.5      28.0      31.5      35.0      38.5      42.0
```

b. $Q_1 = 33.25$, $Q_3 = 38.75$, Mediana $= 37.50$
```
             ---------------------
      ------I                +   I----------------------
             ---------------------
  ------+---------+---------+---------+---------+---------+
      32.5      35.0      37.5      40.0      42.5      45.0
```

c. El tiempo mediano para el transporte público es de casi 6 minutos menos. Hay mayor variación en el transporte público; la diferencia entre Q_1 y Q_3 es de 9.5 minutos en el caso del transporte público y de 5.5 minutos en el del transporte privado

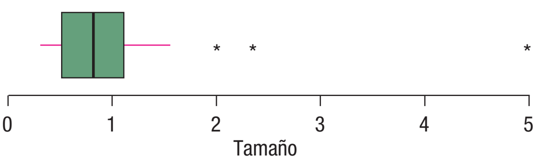
35. La distribución tiene un sesgo positivo; el primer cuartil es de aproximadamente \$20 y el tercero de casi \$90. Hay un extremo localizado en \$255; la mediana es de \$50 más o menos

37. **a.** Diagrama de caja del precio



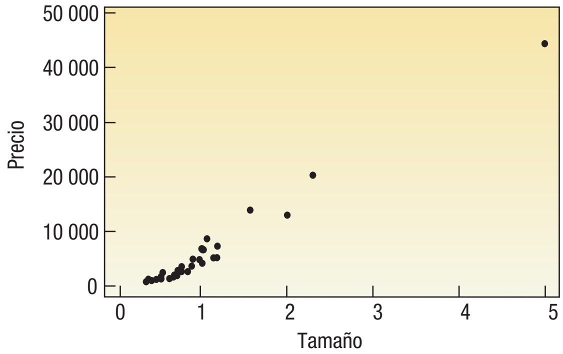
La mediana es de 3 373; el primer cuartil es de 1 478. El tercer cuartil es de 6 141; así que los precios sobre 13 135.5, calculados mediante 6 141 + 1.5(6 141 − 1 478), son atípicos. Hay tres (13 925; 20 413 y 44 312)

b. Gráfica de caja del tamaño



La mediana es de 0.84; el primer cuartil es de 0.515. El tercer cuartil es 1.12; así que los tamaños por encima de 2.0275, que se calcula mediante 1.12 + 1.5 (1.12 − 0.515), son extremos. Hay tres (2.03, 2.35 y 5.03)

c. Diagrama de dispersión del precio en función del tamaño



Existe una relación directa entre ellas. La primera observación es más grande en ambas escalas

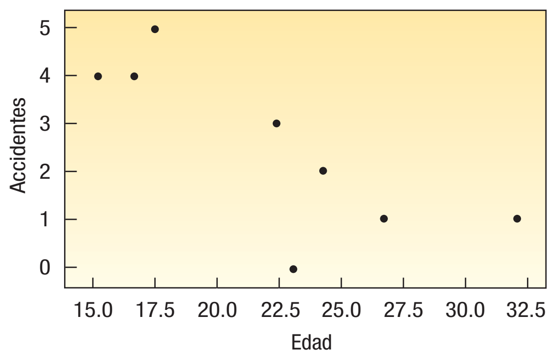
d.

Forma/ corte	Promedio	Bueno	Ideal	De alta calidad	Ultra ideal	Todos
Esmeralda	0	0	1	0	0	1
Marquesa	0	2	0	1	0	3
Oval	0	0	0	1	0	1
Princesa	1	0	2	2	0	5
Redondo	1	3	3	13	3	23
Total	2	5	6	17	3	33

La mayoría de los diamantes son redondos (23). El corte de alta calidad es el más común (17). La combinación entre “redondo” y “de alta calidad” se presenta con mayor frecuencia (13)

39. $sk = 0.065$ o $sk = \dfrac{3(7.7143 - 8.0)}{3.9036} = -0.22$

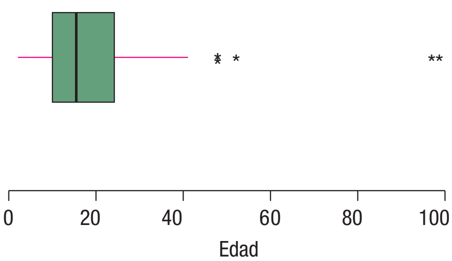
41. Diagrama de dispersión de accidentes en función de la edad



Conforme la edad aumenta, el número de accidentes se reduce

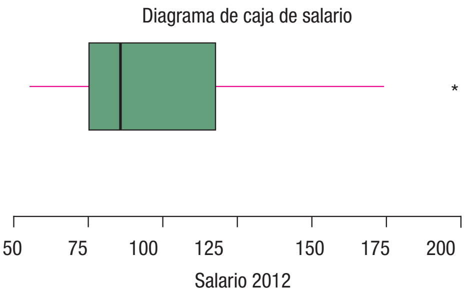
43. **a.** 139 340 000
b. 5.4% desempleados, determinados por (7 523/139 340)100
c. Hombres = 5.64%
d. Mujeres = 5.12%

45. **a.** Diagrama de caja de la edad



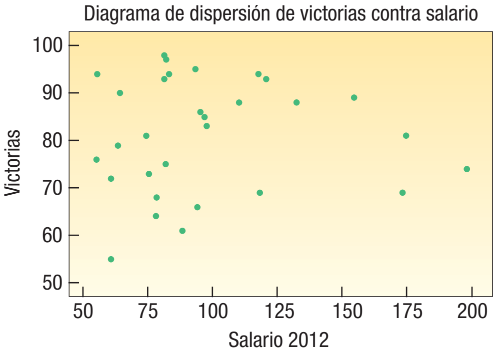
Hay cinco datos atípicos; hay un grupo de tres de alrededor de 40 años (Angels, Athletics y Dodgers) y un grupo de dos cercanos a cien años de edad (Cachorros y Medias Rojas)

b.



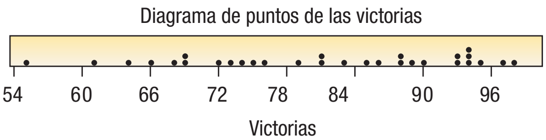
El primer cuartil es de $75 200 000, y el tercero, de $117 720 000. La distribución tiene un sesgo positivo, con los Yankees de Nueva York como un dato atípico definitivo

c.



Salarios más altos se traducen en más victorias

d.



La distribución es casi uniforme entre 55 y 98

CAPÍTULO 5

1.

	Persona	
Resultado	**1**	**2**
1	A	A
2	A	F
3	F	A
4	F	F

3. **a.** 0.176, calculado con $\frac{6}{34}$ **b.** Empírico

5. **a.** Empírico
b. Clásico
c. Clásico
d. Empírico, basado en los datos sismológicos

7. **a.** La encuesta entre 40 personas sobre los problemas del medio ambiente
b. Por ejemplo, 26 o más respondieron que sí
c. $10/40 = 0.25$
d. Empírico
e. Los eventos no son iguales, pero son mutuamente excluyentes

9. **a.** Las respuestas variarán; he aquí algunas probabilidades: 123, 124, 125, 999
b. $(1/10)^3$
c. Clásico

11. $P(A \text{ o } B) = P(A) + P(B) = 0.30 + 0.20 = 0.50$
$P(\text{ninguna}) = 1 - 0.50 = 0.50$

13. **a.** $102/200 = 0.51$
b. 0.49, calculado por $61/200 + 37/200 = 0.305 + 0.185$. Regla especial de la adición

15. $P(\text{sobre } C) = 0.25 + 0.50 = 0.75$

17. $P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) = 0.20 + 0.30 - 0.15 = 0.35$

19. Cuando dos eventos son mutuamente excluyentes, si uno ocurre, el otro no puede ocurrir; por lo tanto, la probabilidad de que se presenten de manera conjunta es cero

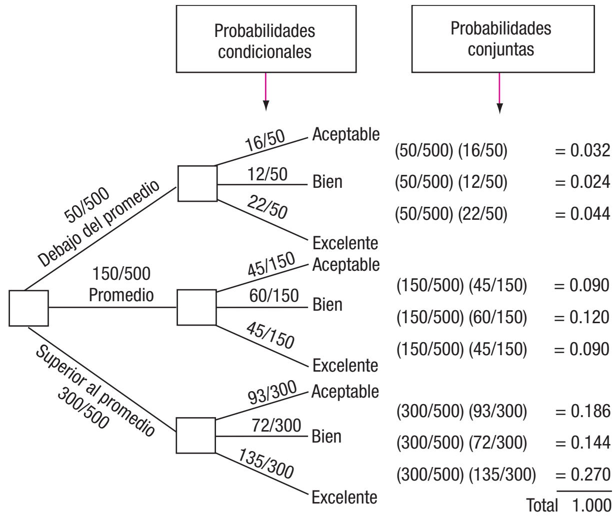
21. Sea que *A* defina el evento de que el pescado es verde y *B* el evento de que el pescado es macho
a. $P(A) = 80/140 = 0.5714$
b. $P(B) = 60/140 = 0.4286$
c. $P(A \text{ y } B) = 36/140 = 0.2571$
d. $P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) = 80/140 + 60/140 - 36/140 = 104/140 = 0.7429$

23. $P(A \text{ y } B) = P(A) \times P(B|A) = 0.40 \times 0.30 = 0.12$

25. 0.90, determinado por $(0.80 + 0.60) - 0.5$
0.10, determinado por $(1 - 0.90)$

27. **a.** $P(A_1) = 3/10 = 0.30$
b. $P(B_1|A_2) = 1/3 = 0.33$
c. $P(B_2 \text{ y } A_3) = 1/10 = 0.10$

29. **a.** Una tabla de contingencias
b. 0.27, calculado por $300/500 \times 135/300$
c. A continuación se muestra el diagrama de árbol:



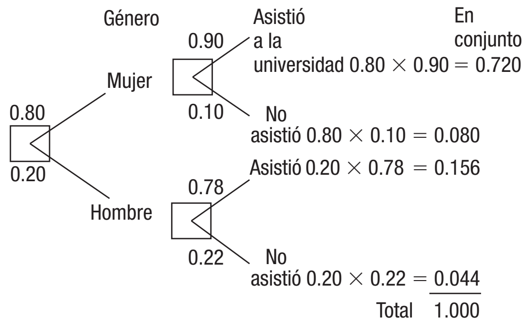
31. **a.** De los 545 estudiantes, 171 prefieren el esquí; así que la probabilidad es 171/545, o 0.3138
b. De los 545 estudiantes, 155 están en una escuela técnica; así que la probabilidad es 155/545, o 0.2844
c. De los 210 estudiantes de la universidad, 70 prefieren el patinaje sobre hielo; así que la probabilidad es 70/210, o 0.3333
d. De los 211 estudiantes que prefieren el snowboarding, 68 están en una escuela técnica; así que la probabilidad es 68/211, o 0.3223
e. De los 180 estudiantes de posgrado, 74 prefieren el esquí y 47 prefieren el patinaje sobre hielo; así que la probabilidad es (74 + 47)/180 = 121/180, o 0.6722

33.
$$P(A_1|B_1) = \frac{P(A_1) \times P(B_1|\,A_1)}{P(A_1) \times P(B_1|\,A_1) \times P(A_2) \times P(B_1|A_2)}$$
$$= \frac{0.60 \times 0.05}{(0.60 \times 0.05) + (0.40 \times 0.10)} = 0.4286$$

35.
$$P(\text{noche}|\text{ganar}) = \frac{P(\text{noche})\,P(\text{ganar}|\text{noche})}{P(\text{noche})\,P(\text{ganar}|\text{noche}) + P(\text{día})\,P(\text{ganar}|\text{día})}$$
$$= \frac{(0.70)(0.50)}{[(0.70)(0.50)] + [(0.30)(0.90]} = 0.5645$$

37.
$$P(\text{efectivo}\,|>\$50)$$
$$= \frac{P(\text{efectivo})\,P(>\$50\,|\,\text{efectivo})}{P(\text{efectivo})\,P(>\$50\,|\,\text{efectivo}) + P(\text{crédito})\,P(>\$50\,|\,\text{crédito}) + P(\text{débito})\,P(>\$50\,|\,\text{débito})}$$
$$= \frac{(0.30)(0.20)}{(0.30)(0.20) + (0.30)(.90) + (0.40)(0.60)} = 0.1053$$

39. **a.** 78 960 960
b. 840, calculado por (7)(6)(5)(4); es decir, 7!/3!
c. 10, calculado por 5!/3!2!
41. 210, calculado por (10)(9)(8)(7)/(4)(3)(2)
43. 120, calculado por 5!
45. 10 879 286 400, determinado por
$_{15}P_{10}$ = (15)(14)(13)(12)(11)(10)(9)(8)(7)(6)
47. **a.** Pedir a los adolescentes que comparen sus reacciones ante un refresco recién creado
b. Las respuestas variarán; una probabilidad consiste en que a más de la mitad de los entrevistados les guste
49. Subjetivo
51. **a.** 4/9, calculado por (2/3) · (2/3)
b. 3/4, porque (3/4) · (2/3) = 0.5
53. **a.** 0.8145, calculado mediante $(0.95)^4$
b. Regla especial de la multiplicación
c. $P(A \text{ y } B \text{ y } C \text{ y } D) = P(A) \times P(B) \times P(C) \times P(D)$
55. **a.** 0.08, calculado mediante 0.80 × 0.10
b. No; 90% de las mujeres asistió a la universidad; 78% de los hombres lo hizo
c.



d. Sí, ya que todos los resultados posibles aparecen en el diagrama de árbol
57. **a.** 0.57, calculado con 57/100
b. 0.97, calculado con (57/100) + (40/100)
c. Sí, ya que un empleado no puede ser ambas cosas
d. 0.03, calculado con 1 − 0.97
59. **a.** 1/2, calculado por (2/3)(3/4)
b. 1/12, calculado por (1/3)(1/4)
c. 11/12, calculado por 1 − 1/12
61. **a.** 0.9039, calculado por $(0.98)^5$
b. 0.0961, calculado por 1 − 0.9039
63. **a.** 0.0333, calculado por (4/10)(3/9)(2/8)
b. 0.1667, calculado por (6/10)(5/9)(4/8)
c. 0.8333, calculado por 1 − 0.1667
d. Dependiente
65. **a.** 0.3818, calculado mediante (9/12)(8/11)(7/10)
b. 0.6182, calculado mediante 1 − 0.3818
67. **a.** $P(S) \cdot P(R|S) = 0.60(0.85) = 0.51$
b. $P(S) \cdot P(PR|S) = 0.60(1 - 0.85) = 0.09$
69. **a.** $P(\text{no perfecto}) = P(\text{sector malo}) + P(\text{defectuoso})$
$$= \frac{112}{1\,000} + \frac{31}{1\,000} = 0.143$$
b. $P(\text{defectuoso}|\text{no perfecto}) = \dfrac{0.031}{0.143} = 0.217$
71. $$P(\text{pobre}|\text{ganancia}) = \frac{0.10(0.20)}{0.10(0.20) + 0.60(.80) + 0.30(0.60)} = 0.294$$
73. **a.** 0.1 + 0.02 = 0.12
b. 1 − 0.12 = 0.88
c. $(0.88)^3 = 0.6815$
d. 1 − 0.6815 = 0.3185
75. Sí; 256 se calcula mediante 2^8
77. 0.9744, calculado por $1 - (0.40)^4$
79. **a.** 0.185, calculado por (0.15)(0.95) + (0.05)(0.85)
b. 0.0075, calculado por (0.15)(0.05)
81. **a.** $P(\text{F y} > 60) = 0.25$, determinado por la regla general de la multiplicación:
$P(\text{F}) \cdot P(>60|\text{F}) = (0.5)(0.5)$
b. 0
c. 0.3333, calculado por 1/3
83. $26^4 = 456\,976$
85. 1/3, 628 800
87. **a.** $P(D) = 0.20(0.03) + 0.30(0.04) + 0.25(0.07) + 0.25(0.065) = 0.05175$
b. $$P(\text{Tyson}|\text{defectuoso}) = \frac{0.20(0.03)}{[0.20(0.03) + 0.30(0.04) + 0.25(0.07) + 0.25(0.065)]} = 0.1159$$

Proveedor	Conjunta	Revisada
Tyson	0.00600	0.1159
Fuji	0.01200	0.2319
Kirkpatricks	0.01750	0.3382
Parts	0.01625	0.3140
	0.05175	1.0000

89. 0.512, calculado por $(0.8)^3$
91. 0.525, calculado por $1 - (0.78)^3$
93. **a.**

Temporada de victorias	Asistencia Baja	Media	Alta	Total
No	4	7	2	13
Sí	3	7	7	17
Total	11	10	9	30

1. 0.5667, calculado por 17/30
2. 0.5667, calculado por 17/30 + 9/30 − 9/30 = 17/30
3. 0.7778, calculado por 7/9
4. 0.1000, calculado por 3/30

b.

	Temporada de derrotas	Temporada de victorias	Total
Nueva	7	8	15
Antigua	6	9	15
Total	13	17	30

1. 0.5667, calculado por 17/30
2. 0.2667, calculado por 8/30
3. 0.8000, calculado por 17/30 + 15/30 − 8/30 = 24/30

CAPÍTULO 6

1. Media = 1.3, varianza = 0.81, calculadas según:
$$\mu = 0(0.20) + 1(0.40) + 2(0.30) + 3(0.10) = 1.3$$
$$\sigma^2 = (0 - 1.3)^2(0.2) + (1 - 1.3)^2(0.4) + (2 - 1.3)^2(0.3) + (3 - 1.3)^2(0.1) = 0.81$$
3. Media = 14.5, varianza = 27.25, calculadas por:
$$\mu = 5(0.1) + 10(0.3) + 15(0.2) + 20(0.4) = 14.5$$
$$\sigma^2 = (5 - 14.5)^2(0.1) + (10 - 14.5)^2(0.3) + (15 - 14.5)^2(0.2) + (20 - 14.5)^2(0.4) = 27.25$$
5. **a.**

Llamadas, x	Frecuencia	$P(x)$	$xP(x)$	$(x - \mu)^2 P(x)$
0	8	0.16	0	0.4624
1	10	0.20	0.20	0.0980
2	22	0.44	0.88	0.0396
3	9	0.18	0.54	0.3042
4	1	0.02	0.08	0.1058
	50		1.70	1.0100

b. Distribución discreta, ya que solo son posibles ciertos resultados
c. $\mu = \Sigma x \cdot P(x) = 1.70$
d. $\sigma = \sqrt{1.01} = 1.005$
7.

Cantidad	$P(x)$	$xP(x)$	$(x - \mu)^2 P(x)$
10	0.50	5	60.50
25	0.40	10	6.40
50	0.08	4	67.28
100	0.02	2	124.82
		21	259.00

a. $\mu = \Sigma xP(x) = 21$
b. $\sigma^2 = \Sigma(x - \mu)^2 P(x) = 259$
$\sigma = \sqrt{259} = 16.093$

9. **a.** $P(2) = \frac{4!}{2!(4-2)!}(0.25)^2(0.75)^{4-2} = 0.2109$
b. $P(3) = \frac{4!}{3!(4-3)!}(0.25)^3(0.75)^{4-3} = 0.0469$

11. **a.**

x	$P(x)$
0	0.064
1	0.288
2	0.432
3	0.216

b. $\mu = 1.8$
$\sigma^2 = 0.72$
$\sigma = \sqrt{0.72} = 0.8485$

13. **a.** 0.2668, calculado por $P(2) = \frac{9!}{(9-2)!2!}(0.3)^2(0.7)^7$
b. 0.1715, calculado por $P(4) = \frac{9!}{(9-4)!4!}(0.3)^4(0.7)^5$
c. 0.0404, calculado por $P(0) = \frac{9!}{(9-0)!0!}(0.3)^0(0.7)^9$

15. **a.** 0.2824, calculado por $P(0) = \frac{12!}{(12-0)!0!}(0.1)^0(0.9)^{12}$
b. 0.3766, calculado por $P(1) = \frac{12!}{(12-1)!1!}(0.1)^1(0.9)^{11}$
c. 0.2301, calculado por $P(2) = \frac{12!}{(12-2)!2!}(0.1)^2(0.9)^{10}$
d. $\mu = 1.2$, calculado por 12(0.1)
$\sigma = 1.0392$, calculado con $\sqrt{1.08}$

17. **a.** 0.1858, calculado por $\frac{15!}{2!13!}(0.23)^2(0.77)^{13}$
b. 0.1416, calculado por $\frac{15!}{5!10!}(0.23)^5(0.77)^{10}$
c. 3.45, calculado por (0.23)(15)

19. **a.** 0.296, determinado utilizando el apéndice B.1, con n de 8; π de 0.30 y x de 2
b. $P(x \leq 2) = 0.058 + 0.198 + 0.296 = 0.552$
c. 0.448, determinado por $P(x \geq 3) = 1 - P(x \leq 2) = 1 - 0.552$

21. **a.** 0.387, determinado utilizando el apéndice B.1, con n de 9; π de 0.90 y x de 9
b. $P(x < 5) = 0.001$
c. 0.992, determinado por $1 - 0.008$
d. 0.947, determinado por $1 - 0.053$

23. **a.** $\mu = 10.5$, determinado por 15(0.7) y $\sqrt{15(0.7)(0.3)} = 1.7748$
b. 0.2061, determinado por $\frac{15!}{10!5!}(0.7)^{10}(0.3)^5$
c. 0.4247, determinado por $0.2061 + 0.2186$
d. 0.5154, determinado por
$0.2186 + 0.1700 + 0.0916 + 0.0305 + 0.0047$

25. $P(2) = \frac{[{}_6C_2][{}_4C_1]}{{}_{10}C_3} = \frac{15(4)}{120} = 0.50$

27. N es 10 (el número de préstamos en la población); S es 3 (el número de préstamos bajo el agua en la población); x es 0 (el número de préstamos bajo el agua seleccionados en la muestra); y n es 2 (el tamaño de la muestra). Utilice la fórmula [6.6] para encontrar:
$P(0) = \frac{[{}_7C_2][{}_3C_0]}{{}_{10}C_2} = \frac{21(1)}{45} = 0.4667$

29. $P(2) = \frac{[{}_9C_3][{}_6C_2]}{[{}_{15}C_5]} = \frac{84(15)}{3\,003} = 0.4196$

31. **a.** 0.6703
b. 0.3297

33. **a.** 0.0613
b. 0.0803

35. $\mu = 6$
$P(x \geq 5) = 1 - (0.0025 + 0.0149 + 0.0446 + 0.0892 + 0.1339)$
$= 0.7149$

37. Una variable aleatoria es un resultado cuantitativo o cualitativo que se deriva de un experimento aleatorio. Una distribución de probabilidad también incluye la posibilidad de cada resultado

39. $\mu = \$1\,000(0.25) + \$2\,000(0.60) + \$5\,000(0.15) = \$2\,200$
$\sigma^2 = (1\,000 - 2\,200)^2 0.25 + (2\,000 - 2\,200)^2 0.60 +$
$(5\,000 - 2\,200)^2 0.15$
$= 1\,560\,000$

41. $\mu = 12(0.25) + \cdots + 15(0.1) = 13.2$
$\sigma = (12 - 13.2)^2 0.25 + \cdots + (15 - 13.2)^2 0.10 = 0.86$
$\sigma = \sqrt{0.86 = 0.927}$

43. **a.** $\mu = 10(0.35) = 3.5$
b. $P(x = 4) = {}_{10}C_4(0.35)^4(0.65)^6 = 210(0.0150)(0.0754) = 0.2375$
c. $P(x \geq 4) = {}_{10}C_x(0.35)^x(0.65)^{10-x}$
$= 0.2375 + 0.1536 + \cdots + 0.0000 = 0.4862$

45. **a.** 6, calculado por 0.4×15
b. 0.0245, calculado por $\frac{15!}{10!5!}(0.4)^{10}(0.6)^5$
c. 0.0338, calculado por:
$0.0245 + 0.0074 + 0.0016 + 0.0003 + 0.0000$
d. 0.0093, calculado por $0.0338 - 0.0245$

47. **a.** $\mu = 20(0.075) = 1.5$
$\sigma = \sqrt{20(0.075)(0.925)} = 1.1779$
b. 0.02103, determinado por $\frac{20!}{0!20!}(0.075)^0(0.925)^{20}$
c. 0.07897, determinado por $1 - 0.2103$

49. **a.** 0.1311, calculado por $\frac{16!}{4!12!}(0.15)^4(0.85)^{12}$
b. 2.4, determinado por (0.15)(16)
c. 0.2100, determinado por
$1 - 0.0743 - 0.2097 - 0.2775 - 0.2285$

51. 0.2784, determinado por $0.1472 + 0.0811 + 0.0348 + 0.0116 + 0.0030 + 0.0006 + 0.0001 + 0.0000$

53. **a.**

0	0.0002	7	0.2075
1	0.0019	8	0.1405
2	0.0116	9	0.0676
3	0.0418	10	0.0220
4	0.1020	11	0.0043
5	0.1768	12	0.0004
6	0.2234		

b. $\mu = 12(0.52) = 6.24$ $\sigma = \sqrt{12(0.52)(0.48)} = 1.7307$
c. 0.1768
d. 0.3343, calculado por
$0.0002 + 0.0019 + 0.0116 + 0.0418 + 0.1020 + 0.1768$

55. **a.** $P(1) = \frac{[{}_7C_3][{}_3C_1]}{[{}_{10}C_3]} = \frac{(21)(3)}{120} = 0.5250$
b. $P(0) = \frac{[{}_7C_3][{}_3C_0]}{[{}_{10}C_3]} = \frac{(35)(1)}{120} = 0.2917$
$P(x \geq 1) = 1 - P(0) = 1 - 0.2917 = 0.7083$

57. $P(x = 0) = \frac{[{}_8C_4][{}_4C_0]}{[{}_{12}C_4]} = \frac{70}{495} = 0.141$

59. **a.** 0.0498
b. 0.7746, determinado por $(1 - 0.0498)^5$

61. $\mu = 4.0$, del apéndice B.2
a. 0.0183 **c.** 0.6289
b. 0.1954 **d.** 0.5665

63. **a.** 0.1733, determinado por $\frac{(3.1)^4 e^{-3.1}}{4!}$
b. 0.0450, determinado por $\frac{(3.1)^0 e^{-3.1}}{0!}$
c. 0.9550, determinado por $1 - 0.0450$

65. $\mu = n\pi = 23\left(\frac{2}{113}\right) = 0.407$
$P(2) = \frac{(0.407)^2 e^{-0.407}}{2!} = 0.0551$
$P(0) = \frac{(0.407)^0 e^{-0.407}}{0!} = 0.6656$

67. Sea $\mu = n\pi = 155(1/3\,709) = 0.042$
$P(4) = \frac{0.042^4 e^{-0.042}}{4!} = 0.00000012$
¡Muy poco probable!

69. **a.** $\mu = n\pi = 15(0.67) = 10.05$
$\sigma = \sqrt{n\pi(1-\pi)} = \sqrt{15(0.67)(0.33)} = 1.8211$
b. $P(8) = {}_{15}C_8(0.67)^8(0.33)^7 = 6435(0.0406)(0.000426) = 0.1114$
c. $P(x \geq 8) = 0.1114 + 0.1759 + \cdots + 0.0025 = 0.9163$

71. El número medio de cuadrangulares por partido es 2.0305, determinado por 4934/[(15)(162)]
a. $P(0) = \dfrac{2.0305^0 e^{-2.0305}}{0!} = 0.1313$
b. $P(2) = \dfrac{2.0305^2 e^{-2.0305}}{2!} = 0.2706$
c. $P(x \geq 4) = 0.1484$, calculado por
$1 - (0.1313 + 0.2665 + 0.2706 + 0.1832)$

CAPÍTULO 7

1. **a.** $b = 10, a = 6$
b. $\mu = \dfrac{6 + 10}{2} = 8$
c. $\sigma = \sqrt{\dfrac{(10-6)^2}{12}} = 1.1547$
d. $\text{Área} = \dfrac{1}{(10-6)} \cdot \dfrac{(10-6)}{1} = 1$
e. $P(x > 7) = \dfrac{1}{(10-6)} \cdot \dfrac{10-7}{1} = \dfrac{3}{4} = 0.75$
f. $P(7 \leq x \leq 9) = \dfrac{1}{(10-6)} \cdot \dfrac{(9-7)}{1} = \dfrac{2}{4} = 0.50$

3. **a.** 0.30, calculado por $(30 - 27)/(30 - 20)$
b. 0.40, calculado por $(24 - 20)/(30 - 20)$

5. **a.** $a = 0.5, b = 3.00$
b. $\mu = \dfrac{0.5\ 1\ 3.00}{2} = 1.75$
$\sigma = \sqrt{\dfrac{(3.00-0.50)^2}{12}} = 0.72$
c. $P(x > 1) = \dfrac{1}{(3.0-0.5)} \cdot \dfrac{1-0.5}{1} = \dfrac{0.5}{2.5} = 0.2$
d. 0, calculado por $\dfrac{1}{(3.0-0.5)} \cdot \dfrac{(1.0-1.0)}{1}$
e. $P(x > 1.5) = \dfrac{1}{(3.0-0.5)} \cdot \dfrac{3.0-1.5}{1} = \dfrac{1.5}{2.5} = 0.6$

7. La forma real de una distribución normal depende de su media y de su desviación estándar; por lo tanto, existe una distribución normal y una curva normal que la acompaña para una media de 7 y una desviación estándar de 2. Hay otra curva normal para una media de $25 000 y una desviación estándar de $1 742, etcétera

9. **a.** 490 y 510, determinado por $500 \pm 1(10)$
b. 480 y 520, determinado por $500 \pm 2(10)$
c. 470 y 530, determinado por $500 \pm 3(10)$

11. $z_{Rob} = \dfrac{\$50\,000 - \$60\,000}{\$5\,000} = -2$
$z_{Rachel} = \dfrac{\$50\,000 - \$35\,000}{\$8\,000} = 1.875$
Con el ajuste correspondiente a sus industrias, Rob está muy por debajo del promedio y Rachel está muy por encima

13. **a.** 1.25, determinado por $z = \dfrac{25-20}{4.0} = 1.25$
b. 0.3944, localizado en el apéndice B.3
c. 0.3085, determinado por $z = \dfrac{18-20}{2.5} = -0.5$
Encuentre 0.1915 en el apéndice B.3 para $z = -0.5$, enseguida $0.5000 - 0.1915 = 0.3085$

15. **a.** 0.3413, determinado por $z = \dfrac{\$24 - \$20.50}{\$3.50} = 1.00$, enseguida encuentre 0.3413 en al apéndice B.3 para $z = 1$
b. 0.1587, determinado por $0.5000 - 0.3413 = 0.1587$
c. 0.3336, determinado por $z = \dfrac{\$19.00 - \$20.50}{\$3.50} = -0.43$
Encuentre 0.1664 en el apéndice B.3 para $z = -0.43$, después $0.5000 - 0.1664 = 0.3336$

17. **a.** 0.8267; primero encuentre $z = -1.5$, calculado según $(44 - 50)/4$ y $z = 1.25 = (55 - 50)/4$. El área entre -1.5 y 0 es 0.4232, y el área entre 0 y 1.25 es 0.3944, las dos de acuerdo con el apéndice B.3; enseguida, al sumar ambas áreas, se encuentra que $0.4332 + 0.3944 = 0.8276$
b. 0.1056, determinado por $0.5000 - 0.3944$, donde $z = 1.25$
c. 0.2029; recuerde que el área para $z = 1.25$ es 0.3944, y el área para $z = 0.5$, calculada mediante $(52 - 50)/4$, es de 0.1915; enseguida reste $0.3944 - 0.1915$ para determinar 0.2029

19. **a.** 0.4129; inicie usando la fórmula [7.5] para encontrar el valor z para $ 3 100, que es $(3\,100 - 3\,000)/450$, o 0.22. Después consulte el apéndice B.3 para encontrar el área entre 0 y 0.22, que es 0.0871; finalmente, y dado que el área de interés está más allá de 0.22, reste esa probabilidad de 0.5000. El resultado es $0.555 - 0.0871$, o 0.4129
b. 0.1929; utilice la fórmula [7.5] para encontrar el valor z de 3 500, que es $(3\,500 - 3\,000)/450$, o 1.11. Después consulte el apéndice B.3 para encontrar el área bajo la curva normal estándar (la probabilidad es 0.3665); como ambos puntos (1.11 y 0.22) están en el mismo lado de la media, reste la probabilidad más pequeña de la mayor. El resultado es $0.3665 - 0.1736 = 0.1929$
c. 0.8190; utilice la fórmula [7.5] para encontrar el valor z de $2 250, que es -1.67, calculado por $(2\,250 - 3\,000)/450$. El área correspondiente es 0.4525, como 1.11 y -1.67 están en lados distintos de la media, sume las probabilidades correspondientes; así, se determina $0.4525 + 0.3665 = 0.8190$

21. **a.** 0.0764, calculado con $z = (20 - 15)/3.5 = 1.43$; enseguida $0.5000 - 0.4236 = 0.0764$
b. 0.9236, calculado según $0.5000 + 0.4236$, donde $z = 1.43$
c. 0.1185, calculado con $z = (12 - 15)/3.5 = -0.86$
El área bajo la curva es de 0.3051; entonces $z = (10 - 15)/3.5 = -1.43$. El área es 0.4236; finalmente, $0.4236 - 0.3051 = 0.1185$

23. $x = 56.60$, que se calcula sumando 0.5000 (el área a la izquierda de la media), y enseguida se determina un valor z que obliga a que 45% de los datos queden dentro de la curva. Al despejar x: $1.65 = (x - 50)/4$, así que $x = 56.60$

25. $1 630, que se determina mediante $2 100 − 1.88($250)

27. **a.** 214.8 horas: se determina un valor z para el que 0.4900 del área se localice entre 0 y z. Dicho valor es $z = 2.33$; enseguida se despeja $2.33 = (x - 195)/8.5$; así que $x = 214.8$ horas
b. 270.2 horas: se determina un valor z para el que 0.4900 del área se localice entre 0 y $(-z)$. Dicho valor es $z = -2.33$; enseguida se despeja x: $-2.33 = (x - 290)/8.5$; así que $x = 270.2$ horas

29. 41.7%, calculado por $12 + 1.65(18)$

31. **a.** $\mu = n\pi = 50(0.25) = 12.5$
$\sigma^2 = n\pi(1 - \pi) = 12.5(1 - 0.25) = 9.375$
$\sigma = \sqrt{9.375} = 3.0619$
b. 0.2578, determinado por $(14.5 - 12.5)/3.0619 = 0.65$. El área es 0.2422; entonces $0.5000 - 0.2422 = 0.2578$
c. 0.2578, determinado por $(10.5 - 12.5)/3.0619 = -0.65$. El área es 0.2422; entonces $0.5000 - 0.2422 = 0.2578$

33. **a.** $\mu = n\pi = 80(0.07) = 5.6 \qquad \sigma = \sqrt{5.208} = 2.2821$
0.3483, determinado por $z = (6.5 - 5.6)/2.2821 = 0.39$ con el área correspondiente de 0.1517, entonces $0.5000 - 0.1517 = 0.3483$
b. 0.5160, calculado por $z = (5.5 - 5.6)/2.2821 = -0.04$, con el área correspondiente de 0.0160, entonces $0.5000 + 0.0160 = 0.5160$
c. 0.1677, calculado por $0.5160 - 0.3483$

35. **a.** Sí. 1) Hay dos resultados mutuamente excluyentes: sobrepeso y no sobrepeso. 2) El resultado de contar el número de éxitos (miembros con sobrepeso). 3) Cada ensayo es independiente. 4) La probabilidad de 0.30 sigue siendo igual en cada ensayo
b. 0.0084, calculado por
$\mu = 500(0.30) = 150$
$\sigma^2 = 500(0.30)(0.70) = 105$
$\sigma = \sqrt{105} = 10.24695$
$z = \dfrac{x - \mu}{\sigma} = \dfrac{174.5 - 150}{10.24695} = -2.39$
El área bajo la curva para $z = 2.39$ es 0.4916
Entonces $0.5000 - 0.4916 = 0.0084$

c. 0.8461, calculado mediante $z = \frac{139.5 - 150}{10.24695} = -1.02$

El área entre 139.5 y 150 es 0.3461. Sumando, 0.3461 + 0.5000 = 0.8461

37. a. 0.3935, calculado por $1 - e^{[(-1/60)(30)]}$

b. 0.1353, calculado por $e^{[(-1/60)(120)]}$

c. 0.1859, calculado por $e^{[(-1/60)(45)]} - e^{[(-1/60)(75)]}$

d. 41.59 segundos, determinado por $-60 \ln(0.5)$

39. a. 0.5654, determinado por $1 - e^{[(-1/18)(15)]}$ y 0.2212, determinado por $1 - e^{[(-1/60)(15)]}$

b. 0.0013, calculado por $e^{[(-1/18)(120)]}$ y 0.1353, calculado por $e^{[(-1/60)(120)]}$

c. 0.1821, calculado por $e^{[(-1/18)(30)]} - e^{[(-1/18)(90)]}$ y 0.3834, calculado por $e^{[(-1/60)(30)]} - e^{[(-1/60)(90)]}$

d. 4 minutos, determinado por $-18 \ln(0.8)$ y 13.4 minutos, determinado por $-60 \ln(0.8)$

41. a. $\mu = \frac{11.96 + 12.05}{2} = 12.005$

b. $\sigma = \sqrt{\frac{(12.05 - 11.96)^2}{12}} = 0.0260$

c. $P(x < 12) = \frac{1}{(12.05 - 11.96)} \frac{12.05 - 11.96}{1} = \frac{0.04}{0.09} = 0.44$

d. $P(x > 11.98) = \frac{1}{(12.05 - 11.96)}\left(\frac{12.05 - 11.96}{1}\right)$

$= \frac{0.07}{0.09} = 0.78$

e. Todas las latas pueden tener más de 11.00 onzas, así que la probabilidad es de 100%

43. a. $\mu = \frac{4 + 10}{2} = 7$

b. $\sigma = \sqrt{\frac{(10 - 4)^2}{12}} = 1.732$

c. $P(x < 6) = \frac{1}{(10 - 4)} \cdot \left(\frac{6 - 4}{1}\right) = \frac{2}{6} = 0.33$

d. $P(x > 5) = \frac{1}{(10 - 4)} \cdot \left(\frac{10 - 5}{1}\right) = \frac{5}{6} = 0.83$

45. a. −0.4 de las ventas netas, calculado según (170 − 180)/25. 2.92 de los empleados, determinado por (1 850 − 1 500)/120

b. Las ventas netas se encuentran a 0.4 desviaciones estándar por debajo de la media. Los empleados se encuentran a 2.92 desviaciones estándar sobre la media

c. 65.64% de los fabricantes de aluminio tienen ventas netas más altas en comparación con Clarion, calculadas de acuerdo con 0.1554 + 0.5000. Solo 0.18% tienen más empleados que Clarion, calculados según 0.5000 − 0.4982

47. a. 0.5000, ya que $z = \frac{30 - 490}{90} = -5.11$

b. 0.2514, calculado por 0.5000 − 0.2486

c. 0.6374, calculado por 0.2486 + 0.3888

d. 0.3450, calculado por 0.3888 − 0.0438

49. a. 0.3015, calculado por 0.5000 − 0.1985

b. 0.2579, calculado por 0.4564 − 0.1985

c. 0.0011, calculado por 0.5000 − 0.4989

d. 1 818, calculado por 1 280 + 1.28(420)

51. a. 90.82%: primero se determina $z = 1.33$ mediante (40 − 34)/4.5. El área entre 0 y 1.33 es 0.4082 horas/semana para mujeres; enseguida sume 0.5000 y 0.4082 y encuentre 0.9082 o 90.82%

b. 78.23%: primero se determina $z = -0.78$ mediante (25 − 29)/5.1. El área entre 0 y 1.33 es (−0.78) es 0.2823; enseguida sume 0.5000 y 0.2823 y encuentre 0.7823 o 78.23%

c. Encuentre un valor z donde 0.4900 del área se encuentre entre 0 y z. El valor es 2.33; enseguida se despeja x: 2.33 = (x − 34)/4.5, así que x = 44.5 horas/semana

40.9 horas/semana en el caso de los hombres: 2.33 = (x − 29)/5.1, así que x = 40.9 horas/semana

53. Alrededor de 4 099 unidades, calculadas al despejar x. 1.65 = (x − 4 000)/60

55. a. 15.39%, calculado por (8 − 10.3)/2.25 = −1.02, Entonces 0.5000 − 0.3461 = 0.1539

b. 17.31%, calculado por:

$z = (12 - 10.3)/2.25 = 0.76$. El área es de 0.2764

$z = (14 - 10.3)/2.25 = 1.64$. El área es de 0.4495

El área entre 12 y 14 es de 0.1731, determinado por 0.4495 − 0.2764

c. En 99.73% de los días, las devoluciones son entre 3.55 y 17.05, calculadas mediante 10.3 ± 3(2.25); por consiguiente, la probabilidad de menos de 3.55 devoluciones es más bien remota

57. a. 0.9678, calculado por:

$\mu = 60(0.64) = 38.4$

$\sigma^2 = 60(0.64)(0.36) = 13.824$

$\sigma = \sqrt{13.824} = 3.72$

Entonces, (31.5 − 38.4)/3.72 = −1.85, para el cual el área es de 0.4678

Así, 0.5000 + 0.4678 = 0.9678

b. 0.0853, calculado por (43.5 − 38.4)/3.72 = 1.37, donde el área es de 0.4147; entonces, 0.5000 − 0.4147 = 0.0853

c. 0.8084, calculado por 0.4441 + 0.3643

d. 0.0348, calculado por 0.4495 − 0.4147

59. a. 0.0968, determinado mediante

$\mu = 50(0.40) = 20$

$\sigma^2 = 50(0.40)(0.60) = 12$

$\sigma = \sqrt{12} = 3.46$

$z = (24.5 - 20)/3.46 = 1.30$

El área es 0.4032; entonces, para 25 o más, 0.5000 − 0.4032 = 0.0968

61. a. $1.65 = (45 - \mu)/5 \quad \mu = 36.75$

b. $1.65 = (45 - \mu)/10 \quad \mu = 28.5$

c. $z = (30 - 28.5)/10 = 0.15$, entonces 0.5000 + 0.0596 = 0.5596

63. a. 21.19%, calculado mediante $z = (9.00 - 9.20)/0.25 = -0.80$; entonces 0.5000 − 0.2881 = 0.2119

b. Incremente la media, $z = (9.00 - 9.25)/0.25 = -1.00$, $P = 0.5000 - 0.3413 = 0.1587$

Reduzca la desviación estándar. $\sigma = (9.00 - 9.20)/0.15 = -1.33$; $P = 0.5000 - 0.4082 = 0.0918$

Reducir la desviación estándar es mejor porque un porcentaje menor de jamones estará por debajo del límite

65. a. $z = (60 - 52)/5 = 1.60$, así que 0.5000 − 0.4452 = 0.0548

b. Sea $z = 0.67$, entonces $0.67 = (x - 52)/5$ y $x = 55.35$, ajuste el millaje a 55 350

c. $z = (45 - 52)/5 = -1.40$, entonces 0.5000 − 0.4192 = 0.0808

67. $\frac{470 - \mu}{\sigma} = 0.25 \quad \frac{500 - \mu}{\sigma} = 1.28 \quad \sigma = 29\,126$ y $\mu = 462\,718$

69. $\mu = 150(0.15) = 22.5 \quad \sigma = \sqrt{150(0.15)(0.85)} = 4.37$

$z = (29.5 - 22.5)/4.37 = 1.60$

$P(z > 1.60) = 0.05000 - 0.4452 = 0.0548$

71. a. 0.4262, calculado por $1 - e^{[(-1/27)(15)]}$

b. 0.1084, calculado por $e^{[(-1/27)(60)]}$

c. 0.1403, calculado por $e^{[(-1/27)(30)]} - e^{[(-1/27)(45)]}$

d. 2.84 segundos, calculado por $-27 \ln(0.9)$

73. a. 0.2835, calculado por $1 - e^{[(-1/300\,000)(100\,000)]}$

b. 0.1889, calculado por $e^{[(-1/300\,000)(500\,000)]}$

c. 0.2020, calculado por $e^{[(-1/300\,000)(200\,000)]} - e^{[(-1/300\,000)(350\,000)]}$

d. Tanto la media como la desviación estándar son 300 000 horas.

75. a. 0.0582, calculado por 0.5000 − 0.4418, con z = (3.500 − 2.495)/0.642 = 1.57; así se determinan 1.71 equipos, calculado por 30(0.0582), en realidad, dos equipos tuvieron una asistencia de más de 3.5 millones, así que la estimación es exacta

b. 0.9032, calculado por 0.5000 + 0.4032, con z = (50 − 98.02)/36.83 = −1.30; así se establecen 27.1 equipos, calculado por 30(0.9032). Los 30 equipos tuvieron en realidad salarios superiores a $50 millones, así que la estimación no es realmente exacta

CAPÍTULO 8

1. a. 303 Lousiana, 5155 S. Main, 3501 Monroe, 2652 W. Central

b. Las respuestas variarán

c. 630 Dixie Hwy, 835 S. McCord Rd, 4624 Woodville Rd

d. Las respuestas variarán

3. a. Bob Schmidt Chevrolet

Great Lakes Ford Nissan

Grogan Towne Chrysler

Southside Lincoln Mercury
Rouen Chrysler Jeep Eagle

b. Las respuestas variarán

c. Yark Automotive
Thayer Chevrolet Geo Toyota
Franklin Park Lincoln Mercury
Mathews Ford Oregon, Inc.
Valiton Chrysler

5. a.

Muestra	Valores	Suma	Media
1	12, 12	24	12
2	12, 14	26	13
3	12, 16	28	14
4	12, 14	26	13
5	12, 16	28	14
6	14, 16	30	15

b. $\mu_{\bar{x}} = (12 + 13 + 14 + 13 + 14 + 15)/6 = 13.5$

$\mu = (12 + 12 + 14 + 16)/4 = 13.5$

c. Mayor dispersión con los datos de la población, si se compara con las medias muestrales, que varían de 12 a 15, mientras que la población varía de 12 a 16

7. a.

Muestra	Valores	Suma	Mediana
1	12, 12, 14	38	12.66
2	12, 12, 15	39	13.00
3	12, 12, 20	44	14.66
4	14, 15, 20	49	16.33
5	12, 14, 15	41	13.66
6	12, 14, 15	41	13.66
7	12, 15, 20	47	15.66
8	12, 15, 20	47	15.66
9	12, 14, 20	46	15.33
10	12, 14, 20	46	15.33

b. $\mu_{\bar{x}} = \dfrac{(12.66 + \cdots + 15.33 + 15.33)}{10} = 14.6$

$\mu = (12 + 12 + 14 + 15 + 20)/5 = 14.6$

c. La dispersión de la población es mayor que la de las medias muestrales. Las medias muestrales varían de 12.66 a 16.33, mientras que la población varía de 12 a 20

9. a. 20, calculado mediante ${}_6C_3$

b.

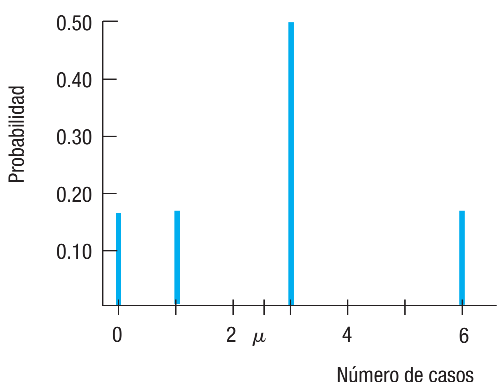
Muestra	Casos	Suma	Media
Ruud, Wu, Sass	3, 6, 3	12	4.00
Ruud, Sass, Flores	3, 3, 3	9	3.00
⋮	⋮	⋮	⋮
Sass, Flores, Schueller	3, 3, 1	7	2.33

c. $\mu_{\bar{x}} = 2.67$, calculado mediante $\dfrac{53.33}{20}$

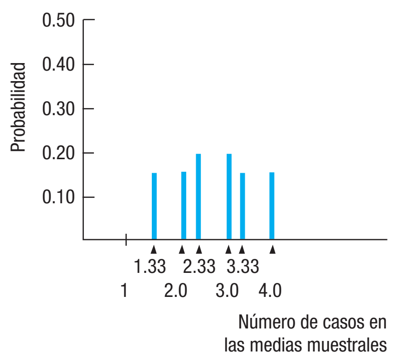
$\mu = 2.67$, calculado por $(3 + 6 + 3 + 3 + 0 + 1)/6$

Son iguales

d. Población



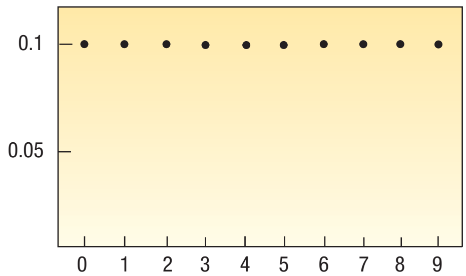
Distribución de las medias muestrales



Media de la muestra	Número de medias	Probabilidad
1.33	3	0.1500
2.00	3	0.1500
2.33	4	0.2000
3.00	4	0.2000
3.33	3	0.1500
4.00	3	0.1500
	20	1.0000

La población tiene mayor dispersión que las medias muestrales; las medias de la muestra varían de 1.33 a 4.0, y la población, de 0 a 6

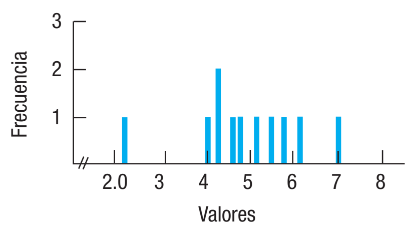
11. a.



$$\mu = \frac{0 + 1 + \cdots + 9}{10} = 4.5$$

b.

Muestra	Suma	$\bar{x}$	Muestra	Suma	$\bar{x}$
1	11	2.2	6	20	4.0
2	31	6.2	7	23	4.6
3	21	4.2	8	29	5.8
4	24	4.8	9	35	7.0
5	21	4.2	10	27	5.4



La media de las 10 medias muestrales es de 4.84, que se aproxima a la media de la población de 4.5. Las medias muestrales varían de 2.2 a 7.0, mientras que los valores de la población varían de 0 a 9; de acuerdo con lo que se muestra en la gráfica anterior, las medias muestrales tienden a agruparse entre 4 y 5

13. a.-c. Las respuestas variarán dependiendo de las monedas que tenga

15. a. $z = \dfrac{63 - 60}{12/\sqrt{9}} = 0.75$

$P = 0.2266$, calculado con $0.5000 - 0.2734$

b. $z = \dfrac{56 - 60}{12/\sqrt{9}} = -1.00$

$p = 0.1587$, calculado con $0.5000 = 0.3413$

c. $p = 0.6147$, calculado con $0.3413 + 0.2734$

17. $z = \dfrac{1950 - 2\,200}{250/\sqrt{50}} = -7.07 \qquad p = 1,$ o virtualmente cierta

19. **a.** Formal Man, Summit Stationers, Bootleggers, Leather Ltd, Petries

b. Las respuestas pueden variar

c. Elder-Beerman, Frederick's Hollywood, Summit Stationers, Lion Store, Leather Ltd., Things Remembered, County Seat, Coach House Gifts, Regis Hairstylists

21. **a.**

Muestras	Media	Desviación de la media	Cuadrado de la desviación
1, 1	1.0	−1.0	1.0
1, 2	1.5	−0.5	0.25
1, 3	2.0	0.0	0.0
2, 1	1.5	−0.5	0.25
2, 2	2.0	0.0	0.0
2, 3	2.5	0.5	0.25
3, 1	2.0	0.0	0.0
3, 2	2.5	0.5	0.25
3, 3	3.0	1.0	1.0

b. La media de las medias muestrales es $(1.0 + 1.5 + 2.0 + \cdots + 3.0)/9 = 18/9 = 2.0$. La media poblacional es $(1 + 2 + 3)/3 = 6/3 = 2$. Son el mismo valor

c. La varianza de las medias muestrales es $(1.0 + 0.25 + 0.0 + \cdots + 3.0)/9 = 3/9 = 1/3$; la varianza de los valores poblacionales es $(1 + 0 + 1)/3 = 2/3$; la varianza de la población es dos veces más grande que la de las medias muestrales

d. Las medias poblacionales siguen un pico triangular a 2; la población es uniforme entre 1 y 3

23. Muestras mayores proporcionan estimaciones más precisas de una media poblacional; así que la compañía con 200 clientes encuestados puede ofrecer estimaciones más precisas. Además, se trata de clientes selectos familiarizados con las computadoras portátiles, que pueden estar mejor calificados para evaluar la nueva computadora

25. **a.** Seleccione 60, 104, 75, 72 y 48. Las respuestas variarán

b. Seleccione la tercera observación; de modo que la muestra consiste en 75, 72, 68, 82 y 48. Las respuestas varían

c. El número de los primeros 20 moteles de 00 a 19. Seleccione tres números al azar; enseguida enumere los cinco últimos de 20 a 24. Seleccione al azar dos números de ese grupo

27. **a.** 15, calculado mediante ${}_6C_2$

b.

Muestra	Valor	Suma	Media
1	79, 64	143	71.5
2	79, 84	163	81.5
⋮	⋮	⋮	⋮
15	92, 77	169	84.5
			1 195.0

c. $\mu_{\bar{x}} = 79.67$, calculado por $1\,195/15$

$\mu = 79.67$, calculado por $478/6$

Son iguales

d. No. El estudiante no obtiene calificaciones en toda la información disponible; con base en la muestra, es tan probable que obtenga una calificación más baja como una calificación alta

29. **a.** 10, calculado con ${}_5C_2$

b.

Número de cortes	Media	Número de cortes	Media
4, 3	3.5	3, 3	3.0
4, 5	4.5	3, 2	2.5
4, 3	3.5	5, 3	4.0
4, 2	3.0	5, 2	3.5
3, 5	4.0	3, 2	2.5

Media muestral	Frecuencia	Probabilidad
2.5	2	0.20
3.0	2	0.20
3.5	3	0.30
4.0	2	0.20
4.5	1	0.10
	10	1.00

c. $\mu_{\bar{x}} = (3.5 + 4.5 + \cdots + 2.5)/10 = 3.4$

$\mu = (4 + 3 + 5 + 3 + 2)/5 = 3.4$

Ambas medias son iguales

d. La forma de los valores de la población es relativamente uniforme. La distribución de la muestra tiende a la normalidad

31. **a.** La distribución será normal

b. $\sigma_{\bar{x}} = \dfrac{5.5}{\sqrt{25}} = 1.1$

c. $z = \dfrac{36 - 35}{5.5/\sqrt{25}} = 0.91$

$p = 0.1814$, calculado por $0.5000 - 0.3186$

d. $z = \dfrac{34.5 - 35}{5.5/\sqrt{25}} = -0.45$

$p = 0.6736$, calculado por $0.5000 + 0.1736$

e. 0.4922, calculado por $0.3186 + 0.1736$

33. $z = \dfrac{\$335 - \$350}{\$45/\sqrt{40}} = -2.11$

$p = 0.9826$, calculado por $0.5000 + 0.4826$

35. $z = \dfrac{21.5 - 24.8}{2.5/\sqrt{60}} = 0.93$

$p = 0.8238$, calculado por $0.5000 + 0.3238$

37. Entre 5 954 y 6 046, calculado por $6\,000 \pm 1.96(150/\sqrt{40})$

39. $z = \dfrac{900 - 947}{205/\sqrt{60}} = -1.78$

$p = 0.0375$, calculado por $0.5000 - 0.4625$

41. **a.** Alaska, Connecticut, Georgia, Kansas, Nebraska, Carolina del Sur, Virginia, Utah

b. Arizona, Florida, Iowa, Massachusetts, Nebraska, Carolina del Norte, Rhode Island, Vermont

43. **a.** $z = \dfrac{600 - 510}{14.28/\sqrt{10}} = 19.9$, $P = 0.00$, o virtualmente nunca

b. $z = \dfrac{500 - 510}{14.28/\sqrt{10}} = -2.21$

$p = 0.4864 + 0.5000 = 0.9864$

c. $z = \dfrac{500 - 510}{14.28/\sqrt{10}} = -2.21$

$p = 0.5000 - 0.4864 = 0.0136$

45. **a.** $\sigma_{\bar{x}} = \dfrac{2.1}{\sqrt{81}} = 0.23$

b. $z = \dfrac{7.0 - 6.5}{2.1/\sqrt{81}} = 2.14, \quad z = \dfrac{6.0 - 6.5}{2.1/\sqrt{81}} = -2.14$

$p = 0.4838 + 0.4838 = 0.9676$

c. $z = \dfrac{6.75 - 6.5}{2.1/\sqrt{81}} = 1.07, \quad z = \dfrac{6.25 - 6.5}{2.1/\sqrt{81}} = -1.07$

$p = 0.3577 + 0.3577 = 0.7154$

d. 0.0162, calculado por $0.5000 - 0.4838$

47. La asistencia media de 2012 es de 2.495 millones. La probabilidad de una media muestral de este tamaño o mayor es 0.0274, calculado por $0.5000 - 0.4726$

donde $z = \dfrac{2.495 - 2.25}{\dfrac{0.70}{\sqrt{30}}} = 1.92$

CAPÍTULO 9

1. 51.314 y 58.686, que se determina por $55 \pm 2.58(10/\sqrt{49})$

3. **a.** 1.581, calculado por $\sigma_{\bar{x}} = 25/\sqrt{250}$

b. La población tiene una distribución normal y se conoce la varianza de la población

c. 16.901 y 23.099, que se determinan mediante 20 ± 3.099

5. **a.** \$20; es nuestra mejor estimación de la media de la población
b. \$18.60 y \$21.40, que se determinan por medio de $\$20 \pm 1.96(\$5/\sqrt{49})$; casi 95% de los intervalos construidos de manera similar incluirán la media de la población

7. **a.** 8.60 galones
b. 7.83 y 9.37, que se determinan por medio de $8.60 \pm 2.58(2.30/\sqrt{60})$
c. Si se determinan los 100 intervalos, la media de la población se incluirá en 99 intervalos

9. **a.** 2.201
b. 1.729
c. 3.499

11. **a.** Se desconoce la media, pero la mejor estimación es 20, la media de la muestra
b. Utilice la distribución *t*, ya que no se conoce la desviación estándar; sin embargo, suponga que la población tiene distribución normal
c. 2.093
d. Entre 19.06 y 20.94, que se determinan mediante $20 \pm 2.093(2/\sqrt{20})$
e. Ningún valor es razonable, porque no se localiza dentro del intervalo

13. Entre 95.39 y 101.81, que se determinan por $98.6 \pm 1.833(5.54/\sqrt{10})$

15. **a.** 0.8, que se determina por 80/100
b. Entre 0.72 y 0.88, que se calcula por $0.8 \pm 1.96\left(\sqrt{\frac{0.8(1-0.8)}{100}}\right)$
c. Hay seguridad razonable de que la proporción de la población se encuentra entre 72 y 88%

17. **a.** 0.625, que se determina mediante 250/400
b. Entre 0.563 y 0.687, que se determina mediante $0.625 \pm 2.58\left(\sqrt{\frac{0.625(1-0.625)}{400}}\right)$
c. Hay seguridad razonable de que la proporción de la población se encuentra entre 56 y 69%

19. 97, determinado por $n = \left(\frac{1.96 \times 10}{2}\right)^2 = 96.04$

21. 196, determinado por $n = 0.15(0.85)\left(\frac{1.96}{0.05}\right)^2 = 195.9216$

23. 554, determinado por $n = \left(\frac{1.96 \times 3}{0.25}\right)^2 = 553.19$

25. **a.** 577, que se determina mediante $n = 0.60(0.40)\left(\frac{1.96}{0.04}\right)^2 = 576.24$
b. 601, que se determina mediante $n = 0.50(0.50)\left(\frac{1.96}{0.04}\right)^2 = 600.25$

27. 33.41 y 36.59, determinados mediante $35 \pm 2.030\left(\frac{5}{\sqrt{36}}\right)\sqrt{\frac{300-36}{300-1}}$

29. 1.683 y 2.037, determinados por $1.86 \pm 2.680\left(\frac{0.5}{\sqrt{50}}\right)\sqrt{\frac{400-50}{400-1}}$

31. 6.13 a 6.87 años, que se determina por medio de $6.5 \pm 1.989(1.7/\sqrt{85})$

33. **a.** Entre \$313.41 y \$332.59, que se calcula mediante $323 \pm 2.426\left(\frac{25}{\sqrt{40}}\right)$
b. \$350 no es razonable, porque se encuentra fuera del intervalo de confianza

35. **a.** Se desconoce la población media
b. Entre 7.50 y 9.14, que se determina mediante $8.32 \pm 1.685(3.07/\sqrt{40})$
c. 10 no es razonable porque se encuentra fuera del intervalo de confianza

37. **a.** 65.49 a 71.71 horas, que se determina mediante $68.6 \pm 2.680(8.2/\sqrt{50})$
b. El valor sugerido por la NCAA se incluye en el intervalo de confianza; por lo tanto, es razonable
c. Cambiar el intervalo de confianza a 95 disminuiría la variación del intervalo. El valor de 2.680 cambiaría a 2.010

39. 61.47, redondeado a 62; determinado despejando *n* en la ecuación: $1.96(16/\sqrt{n}) = 4$

41. Entre \$13 734 y \$15 028, que se encuentra por medio de $14\,381 \pm 1.711(1\,892/\sqrt{25})$. 15 000 resulta razonable porque se encuentra dentro del intervalo de confianza

43. **a.** \$62.583, que se determina por medio de \$751/12
b. Entre \$60.54 y \$64.63, que se determina mediante $62.583 \pm 1.796(3.94/\sqrt{12})$
c. \$60 no es razonable porque se encuentra fuera del intervalo de confianza

45. **a.** 89.4667, que se determina mediante 1 342/15
b. Entre 84.99 y 93.94, que se determina por medio de $89.4667 \pm 2.145(8.08/\sqrt{15})$
c. Sí, porque inclusive el límite inferior del intervalo de confianza se encuentra por arriba de 80

47. El intervalo de confianza está entre 0.011 y 0.059, calculado por $0.035 \pm 2.576\left(\sqrt{\frac{0.035(1-0.035)}{400}}\right)$. No sería razonable concluir que menos de 5% de los empleados fallan en la prueba porque 0.05 está dentro del intervalo de confianza

49. Entre 0.648 y 0.752, calculado por

$$0.70 \pm 2.576\left(\sqrt{\frac{0.70(1-0.70)}{500}}\right)\left(\sqrt{\frac{20\,000-500}{20\,000-1}}\right)$$

Sí, porque incluso el límite inferior del intervalo de confianza está por encima de 0.500

51. \$52.51 y \$55.49, que se determina por

$$\$54.00 \pm 2.032\,\frac{\$450}{\sqrt{35}}\sqrt{\frac{(500-35)}{500-1}}$$

53. 369, que se encuentra por medio de $n = 0.60(1 - 0.60)(1.96/0.05)^2$

55. 97, que se determina mediante $[(1.96 \times 500)/100]^2$

57. **a.** Entre 7 849 y 8 151, calculado por $8\,000 \pm 2.756(300/\sqrt{30})$
b. 554, determinado por $n = \left(\frac{(1.96)(300)}{25}\right)^2$

59. **a.** Entre 75.44 y 80.56, que se determina por $78 \pm 2.010(9/\sqrt{50})$
b. 220, que se encuentra por $n = \left(\frac{(1.645)(9)}{1.0}\right)^2$

61. **a.** 30, calculado por $180/\sqrt{36}$
b. \$355.10 y \$476.90, calculado por $\$416 \pm 2.030\left(\frac{\$180}{\sqrt{36}}\right)$
c. Alrededor de 1 245, determinado por $\left(\frac{1.96(180)}{10}\right)^2$

63. **a.** 708.13, redondeado a 709, que se determina por $0.21(1 - 0.21)(1.96/0.03)^2$
b. 1 068, calculado por $0.50(0.50)(1.96/0.03)^2$

65. **a.** Entre 0.156 y 0.184, calculado por $0.17 \pm 1.96\sqrt{\frac{(0.17)(1-0.17)}{2\,700}}$
b. Sí, porque 18% está dentro del intervalo de confianza
c. 21 682; determinado por $0.17(1 - 0.17)[1.96/0.005]^2$

67. Entre 12.69 y 14.11, que se determina mediante $13.4 \pm 1.96(6.8/\sqrt{352})$

69. **a.** Para el precio de venta de 211.99 hasta 230.22, determinado por $221.1 \pm (1.983)(47.11/\sqrt{105}) = 221.1 \pm 9.12$
b. Para la distancia: 13.685 hasta 15.572, que se determina mediante $14.629 \pm (1.983)(4.874/\sqrt{105}) = 14.629 \pm 0.943$
c. Para la cochera: 0.5867 hasta 0.7657, que se determina por $0.6762 \pm (1.96)\sqrt{\frac{0.6762(1-0.6762)}{105}} = 0.6762 \pm 0.0895$
d. Las respuestas variarán

71. **a.** Entre \$438.34 y \$462.24, calculado por $450.29 \pm 1.99\left(\frac{53.69}{\sqrt{80}}\right)$
b. Entre 820.72 y 839.50, calculado por $830.11 \pm 1.99\left(\frac{42.19}{\sqrt{80}}\right)$
c. Las respuestas variarán

CAPÍTULO 10

1. **a.** De dos colas
b. Rechace H_0 y acepte H_1 cuando *z* no caiga en la región de −1.96 a 1.96

c. -1.2, que se calcula por medio de $z = (49 - 50)/(5/\sqrt{36}) = -1.2$
d. No se rechaza H_0
e. $p = 0.2302$, que se determina mediante $2(0.5000 - 0.3849)$. Una probabilidad de 23.02% de encontrar un valor z de este tamaño cuando H_0 es verdadera

3. **a.** Una cola
b. Rechace H_0 cuando $z > 1.65$
c. 1.2, que se determina mediante $z = (21 - 20)/(5/\sqrt{36})$
d. No se rechaza H_0 con el nivel de significancia 0.05
e. $p = 0.1151$, calculado por $0.5000 - 0.3849$. Una probabilidad de 11.51% de encontrar un valor z de ese tamaño o más grande

5. **a.** H_0: $\mu = 60\,000$ $\quad H_1$: $\mu \neq 60\,000$
b. Rechace H_0 si $z < -1.96$ o $z > 1.96$
c. -0.69, que se encuentra por $z = \dfrac{59\,500 - 60\,000}{(5\,000/\sqrt{48})}$
d. No rechace H_0
e. $p = 0.4902$ que se encuentra por $2(0.5000 - 0.2549)$. La experiencia de Crosset no es diferente de lo que afirma el fabricante. Si H_0 es verdadera, la probabilidad de encontrar un valor más grande que este es 0.4902

7. **a.** H_0: $\mu \geq 6.8$ $\quad H_1$: $\mu < 6.8$
b. Rechace H_0 si $z < -1.65$
c. $z = \dfrac{6.2 - 6.8}{1.8/\sqrt{36}} = -2.0$
d. Se rechaza H_0
e. $p = 0.0228$. El número medio de DVD que se observó es menor a 6.8 al mes. Si H_0 es verdadera, usted obtendrá un estadístico igual de pequeño en menos de una vez cada 40 pruebas

9. **a.** Se rechaza H_0 si $t > 1.833$
b. $t = \dfrac{12 - 10}{(3/\sqrt{10})} = 2.108$
c. Se rechaza H_0 porque la media es mayor que 10

11. H_0: $\mu \leq 40$ $\quad H_1$: $\mu > 40$ $\quad$ Rechace H_0 si $t > 1.703$

$$t = \frac{42 - 40}{(2.1/\sqrt{28})} = 5.040$$

Rechace H_0 y concluya que la cantidad media de llamadas es superior a 40 por semana

13. H_0: $\mu \leq 40\,000$ $\quad H_1$: $\mu > 40\,000$ $\quad$ Rechace H_0 si $t > 1.833$

$$t = \frac{50\,000 - 40\,000}{10\,000/\sqrt{10}} = 3.16$$

Rechace H_0 y concluya que el ingreso medio en Wilmington es mayor a \$40 000

15. **a.** Rechace H_0 si $t < -3.747$
b. $\bar{x} = 17$ y $s = \sqrt{\dfrac{50}{5-1}} = 3.536$

$$t = \frac{17 - 20}{(3.536/\sqrt{5})} = -1.90$$

c. No rechace H_0; no es posible concluir que la media de la población es menor a 20
d. Entre 0.05 y 0.10, cerca de 0.065

17. H_0: $\mu \leq 1.4$ $\quad H_1$: $\mu > 1.4$
Rechace H_0 si $t > 2.821$

$$t = \frac{1.6 - 1.4}{0.216/\sqrt{10}} = 2.93$$

Rechace H_0 y concluya que el consumo de agua ha aumentado. El valor p está entre 0.01 y 0.005, y hay una ligera probabilidad (entre 1 en 100 y una en 200) de que este aumento pueda haber sido casual

19. H_0: $\mu \leq 50$ $\quad H_1$: $\mu > 50$
Rechace H_0 si $t > 1.796$

$$t = \frac{82.5 - 50}{59.5/\sqrt{12}} = 1.89$$

Rechace H_0 y concluya que el número medio de mensajes de texto es mayor a 50. El valor p es menor a 0.05, y hay una ligera probabilidad (menos de 1 en 20) de que esto pueda haber sido casual

21. 1.05, que se determina por $z = (9\,992 - 9\,880)/(400/\sqrt{100})$; entonces $0.5000 - 0.3531 = 0.1469$, que es la probabilidad de cometer un error tipo II

23. H_0: $\mu = \$45\,000$ $\quad H_1$: $\mu \neq \$45\,000$
Rechace H_0 si $z < -1.65$

$$z = \frac{45\,500 - 45\,000}{\$3\,000/\sqrt{120}} = 1.83$$

Rechace H_0; puede concluir que el salario medio no es de \$45 000. Valor p de 0.0672, determinado mediante $2(0.5000 - 0.4664)$

25. H_0: $\mu \geq 10$ $\quad H_1$: $\mu < 10$
Rechace H_0 si $z < -1.65$

$$z = \frac{9.0 - 10.0}{2.8/\sqrt{50}} = -2.53$$

Rechace H_0; la pérdida media de peso es menor a 10 libras. Valor $p = 0.5000 - 0.4943 = 0.0057$

27. H_0: $\mu \geq 7.0$ $\quad H_1$: $\mu < 7.0$
Suponiendo 5% de nivel de significancia, rechace H_0 si $t < -1.677$

$$t = \frac{6.8 - 7.0}{0.9/\sqrt{50}} = -1.57$$

No se rechaza H_0; los estudiantes de West Virginia no duermen menos de 6 horas. El valor p se encuentra entre 0.05 y 0.10

29. H_0: $\mu \geq 3.13$ $\quad H_1$: $\mu < 3.13$
Rechace H_0 si $t < -1.711$

$$t = \frac{2.86 - 3.13}{1.20/\sqrt{25}} = -1.13$$

Rechace H_0 y concluya que el número medio de residentes no necesariamente es menor a 3.13

31. H_0: $\mu \leq 14$ $\quad H_1$: $\mu > 14$
Rechace H_0 si $t > 2.821$ $\quad \bar{x} = 15.66$ $\quad s = 1.544$

$$t = \frac{15.66 - 14.00}{1.544/\sqrt{10}} = 3.400$$

Rechace H_0; la tasa promedio es superior a 14%

33. H_0: $\mu = 3.1$ $\quad H_1$: $\mu \neq 3.1$. Suponga una población normal
Rechace H_0 si $t < -2.201$ o $t > 2.201$

$$\bar{x} = \frac{41.1}{12} = 3.425$$

$$s = \sqrt{\frac{4.0625}{12-1}} = 0.6077$$

$$t = \frac{3.425 - 3.1}{0.6077/\sqrt{12}} = 1.853$$

No rechace H_0; no se puede mostrar una diferencia entre los ciudadanos de la tercera edad y el promedio nacional. El valor p se encuentra cerca de 0.09

35. H_0: $\mu \geq 6.5$ $\quad H_1$: $\mu < 6.5$. Suponga una población normal
Rechace H_0 si $t < -2.718$
$\bar{x} = 5.1667$ $\quad s = 3.1575$

$$t = \frac{5.1667 - 6.5}{3.1575/\sqrt{12}} = -1.463$$

No rechace H_0; el valor p es mayor que 0.05

37. H_0: $\mu = 0$ $\quad H_1$: $\mu \neq 0$
Rechace H_0 si $t < -2.110$ o $t > 2.110$
$\bar{x} = -0.2322$ $\quad s = 0.3120$

$$t = \frac{-0.2322 - 0}{0.3120/\sqrt{18}} = -3.158$$

Rechace H_0; la media gana o pierde pero no es igual a 0. El valor p es menor que 0.01, aunque mayor que 0.001

39. H_0: $\mu \leq 100$ $\quad H_1$: $\mu > 100$. Suponga una población normal
Rechace H_0 si $t > 1.761$

$$\bar{x} = \frac{1\,641}{15} = 109.4$$

$$s = \sqrt{\frac{1389.6}{15-1}} = 9.9628$$

$$t = \frac{109.4 - 100}{9.9628/\sqrt{15}} = 3.654$$

Rechace H_0; el número medio con el escáner es mayor a 100. El valor p es 0.001

41. H_0: $\mu = 1.5$ $\quad H_1$: $\mu \neq 1.5$ $\quad$ Rechace H_0 si $t > 3.250$ o $t < -3.250$

$$t = \frac{1.3 - 1.5}{0.9/\sqrt{10}} = -0.703$$

No se rechaza H_0

43. H_0: $\mu \geq 10$ $\quad H_1$: $\mu < 10$ $\quad$ Rechace H_0 si $t < -1.895$

$$\bar{x} = \frac{78.3}{8} = 9.7875 \qquad s = \sqrt{\frac{5.889}{8-1}} = 0.9172$$

$t = \dfrac{9.7875 - 10}{0.9172/\sqrt{8}} = -0.655$ No rechace H_0; el costo no es menor a \$10 000

45. **a.** $9.00 \pm 1.645(1/\sqrt{36}) = 9.00 \pm 0.274$
De modo que los límites son 8.725 y 9.275

b. $z = \dfrac{8.726 - 8.6}{1/\sqrt{36}} = -0.756$
$P(z < 0.756) = 0.5000 + 0.2764 = 0.7764$

c. $z = \dfrac{9.274 - 9.6}{1/\sqrt{36}} = -1.956$
$P(z > -1.96) = 0.4750 + 0.5000 = 0.9750$

47. $50 + 2.33\dfrac{10}{\sqrt{n}} = 55 - 0.525\dfrac{10}{\sqrt{n}}$ $n = (5.71)^2 = 32.6$
Sea $n = 33$

49. H_0: $\mu \geq 8$ H_1: $\mu < 8$ Se rechaza H_0 si $t < -1.714$
$$t = \frac{7.5 - 8}{3.2/\sqrt{24}} = -0.77$$
No se rechaza la hipótesis nula porque el tiempo no es menor

51. **a.** H_0: $\mu = 80$ H_1: $\mu \neq 80$
Rechace H_0 si t no está entre -2.045 y 2.045
$t = \dfrac{98.02 - 80}{36.83/\sqrt{30}} = 2.68$ Rechace la hipótesis nula
El salario medio podría ser de \$80 millones

b. H_0: $\mu \leq 2\,000\,000$ H_1: $\mu > 2\,000\,000$
Rechace H_0 si t es > 1.699
$$t = \frac{2\,495\,000 - 2\,000\,000}{642\,000/\sqrt{30}} = 4.22$$
Rechace la hipótesis nula porque la asistencia media fue mayor a 2 000 000

CAPÍTULO 11

1. **a.** Prueba de dos colas
b. Rechace H_0 si $z < -2.05$ o $z > 2.05$
c. $z = \dfrac{102 - 99}{\sqrt{\dfrac{5^2}{40} + \dfrac{6^2}{50}}} = 2.59$
d. Rechace H_0
e. Valor $p = 0.0096$, determinado por $2(0.5000 - 0.4952)$

3. **Paso 1** H_0: $\mu_1 \geq \mu_2$ H_1: $\mu_1 < \mu_2$
Paso 2 Se eligió el nivel de significancia 0.05
Paso 3 Rechace H_0 si $z < -1.65$
Paso 4 -0.94, determinado mediante:
$$z = \frac{7.6 - 8.1}{\sqrt{\dfrac{(2.3)^2}{40} + \dfrac{(2.9)^2}{55}}} = -0.94$$
Paso 5 No se rechaza H_0
Paso 6 Los bebés que usaron la marca Gibbs no ganaron menos peso. Valor $p = 0.1736$, determinado mediante $0.5000 - 0.3264$

5. **Paso 1** H_0: $\mu_1 \leq \mu_2$ H_1: $\mu_1 > \mu_2$
Paso 2 Se eligió el nivel de significancia 0.05
Paso 3 Si $z > 1.645$, rechace H_0
Paso 4 $z = \dfrac{61.4 - 60.6}{\sqrt{\dfrac{(1.2)^2}{45} + \dfrac{(1.1)^2}{39}}} = 3.187$
Paso 5 Rechace la hipótesis nula
Paso 6 Es razonable concluir que quienes tuvieron una operación cesárea son más pequeños
El valor p es virtualmente cero; esa diferencia casi nunca se debe a un error de muestreo

7. **a.** Rechace H_0 si $t > 2.120$ o $t < -2.120$; $gl = 10 + 8 - 2 = 16$
b. $s_p^2 = \dfrac{(10 - 1)(4)^2 + (8 - 1)(5)^2}{10 + 8 - 2} = 19.9375$
c. $t = \dfrac{23 - 26}{\sqrt{19.9375\left(\dfrac{1}{10} + \dfrac{1}{8}\right)}} = -1.416$
d. No rechace H_0
El valor p es mayor que 0.10 y menor que 0.20

9. **Paso 1** H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
Paso 2 Se eligió el nivel de significancia 0.01
Paso 3 $gl = 12 + 13 - 2 = 23$. Rechace H_0 si t no está entre -2.807 y 2.807
Paso 4 $s_p^2 = \dfrac{(12 - 1)(7\,469)^2 + (13 - 1)(7\,474)^2}{12 + 13 - 2} = 55\,825\,000$
$$t = \frac{6\,590 - 5\,308}{\sqrt{55\,825\,000\left(\dfrac{1}{12} + \dfrac{1}{13}\right)}} = 0.429$$
Paso 5 No rechace H_0
Paso 6 No hay diferencia entre los salarios medios

11. **Paso 1** H_0: $\mu_s \leq \mu_a$ H_1: $\mu_s > \mu_a$
Paso 2 Se eligió el nivel de significancia 0.10
Paso 3 $gl = 6 + 7 - 2 = 11$ Rechace H_0 si $t > 1.363$
Paso 4 $s_p^2 = \dfrac{(6 - 1)(12.2)^2 + (7 - 1)(15.8)^2}{6 + 7 - 2} = 203.82$
$$t = \frac{142.5 - 130.3}{\sqrt{203.82\left(\dfrac{1}{6} + \dfrac{1}{7}\right)}} = 1.536$$
Paso 5 Rechace H_0
Paso 6 Los gastos medios diarios del personal de ventas son mayores; el valor p se encuentra entre 0.05 y 0.10

13. **a.** $gl = \dfrac{\left(\dfrac{25}{15} + \dfrac{225}{12}\right)^2}{\dfrac{\left(\dfrac{25}{15}\right)^2}{15 - 1} + \dfrac{\left(\dfrac{225}{12}\right)^2}{12 - 1}} = \dfrac{416.84}{0.1984 + 31.9602}$
$= 12.96 \rightarrow 12\ gl$
b. H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
Rechace H_0 si $t > 2.179$ o $t < -2.179$
c. $t = \dfrac{50 - 46}{\sqrt{\dfrac{25}{15} + \dfrac{225}{12}}} = 0.8852$
d. No rechace la hipótesis nula

15. **a.** $gl = \dfrac{\left(\dfrac{697\,225}{16} + \dfrac{2\,387\,025}{18}\right)^2}{\dfrac{\left(\dfrac{697\,225}{16}\right)^2}{16 - 1} + \dfrac{\left(\dfrac{2\,387\,025}{18}\right)^2}{18 - 1}} = 26.7 \rightarrow 26\ gl$
b. H_0: $\mu_{\text{Privada}} \leq \mu_{\text{Pública}}$ H_1: $\mu_{\text{Privada}} > \mu_{\text{Pública}}$
Rechace H_0 si $t > 1.706$
c. $t = \dfrac{12\,840 - 11\,045}{\sqrt{\dfrac{2\,387\,025}{18} + \dfrac{697\,225}{16}}} = 4.276$
d. Rechace la hipótesis nula porque el costo medio de adopción en una agencia privada es mayor que el costo medio de adopción en una agencia pública

17. **a.** Rechace H_0 si $t > 2.353$
b. $\bar{d} = \dfrac{12}{4} = 3.00$ $s_d = \sqrt{\dfrac{2}{3}} = 0.816$
c. $t = \dfrac{3.00}{0.816/\sqrt{4}} = 7.35$
d. Rechace H_0; hay más partes defectuosas producidas en el turno matutino
e. El valor p es menor que 0.005, pero mayor que 0.0005

19. H_0: $\mu_d \leq 0$ H_1: $\mu_d > 0$
$\bar{d} = 25.917$
$s_d = 40.791$
Rechace H_0 si $t > 1.796$
$$t = \frac{25.917}{40.791/\sqrt{12}} = 2.20$$
Rechace H_0; el plan de incentivos resultó en un aumento del ingreso diario. El valor p es aproximadamente 0.025

21. H_0: $\mu_M = \mu_W$ H_1: $\mu_M \neq \mu_W$
Rechace H_0 si $t < -2.645$ o $t > 2.645$ ($gl = 35 + 40 - 2$)
$$s_p^2 = \frac{(35 - 1)(4.48)^2 + (40 - 1)(3.86)^2}{35 + 40 - 2} = 17.3097$$

$$t = \frac{24.51 - 22.69}{\sqrt{17.3079\left(\frac{1}{35} + \frac{1}{40}\right)}} = 1.890$$

No rechace H_0; no hay diferencia entre los números de veces que los hombres y las mujeres compran comida para llevar en un mes. El valor p se encuentra entre 0.05 y 0.10

23. H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
Rechace H_0 si $z < -1.96$ o $z > 1.96$

$$z = \frac{4.77 - 5.02}{\sqrt{\frac{(1.05)^2}{40} + \frac{(1.23)^2}{50}}} = -1.04$$

No rechace H_0; no hay una diferencia entre los números medios de llamadas. El valor $p = 2(0.5000 - 0.3508) = 0.2984$

25. H_0: $\mu_B \leq \mu_A$ H_1: $\mu_B > \mu_A$ Rechace H_0 si $t > 1.668$

$$t = \frac{\$61\,000 - \$57\,000}{\sqrt{\frac{(\$7\,100)^2}{30} + \frac{(\$9\,200)^2}{40}}} = \frac{\$4\,000.00}{\$1\,948.42} = 2.05$$

Rechace H_0; el ingreso medio del plan B es mayor. El valor $p = 0.5000 - 0.4798 = 0.0202$. El sesgo no importa debido a los tamaños de las muestras

27. **a.** $$gl = \frac{\left(\frac{0.3136}{12} + \frac{0.0900}{12}\right)^2}{\frac{\left(\frac{0.3136}{12}\right)^2}{12-1} + \frac{\left(\frac{0.0900}{12}\right)^2}{12-1}}$$

$$= \frac{0.0011}{0.000062 + 0.0000051} = 16.37 \rightarrow 16\ gl$$

b. H_0: $\mu_a = \mu_w$ H_1: $\mu_a \neq \mu_w$
Rechace H_0 si $t > 2.120$ o $t < -2.120$

c. $$t = \frac{1.65 - 2.20}{\sqrt{\frac{0.3136}{12} + \frac{0.0900}{12}}} = -3.00$$

d. Rechace la hipótesis nula porque hay una diferencia

29. Asuma que las desviaciones estándar poblacionales son iguales
H_0: $\mu_n = \mu_s$ H_1: $\mu_n \neq \mu_s$ Rechace H_0 si $t < -2.086$ o $t > 2.086$

$$s_p^2 = \frac{(10 - 1)(10.5)^2 + (12 - 1)(14.5)^2}{10 + 12 - 2} = 161.2969$$

$$t = \frac{83.55 - 78.8}{\sqrt{161.2969\left(\frac{1}{10} + \frac{1}{12}\right)}} = 0.874$$

Valor $p > 0.10$. No rechace H_0; no hay diferencia entre los números medios de hamburguesas vendidas en ambas locaciones

31. Asuma que las desviaciones estándar poblacionales son iguales
H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$ Rechace H_0 si $t > 2.819$ o $t < -2.819$

$$s_p^2 = \frac{(10 - 1)(2.33)^2 + (14 - 1)(2.55)^2}{10 + 12 - 2} = 6.06$$

$$t = \frac{15.87 - 18.29}{\sqrt{6.06\left(\frac{1}{10} + \frac{1}{14}\right)}} = -2.374$$

No rechace H_0; no hay diferencia entre las cantidades medias compradas

33. Asuma que las desviaciones estándar poblacionales son iguales
H_0: $\mu_1 \leq \mu_2$ H_1: $\mu_1 > \mu_2$ Rechace H_0 si $t > 2.567$

$$s_p^2 = \frac{(8 - 1)(2.2638)^2 + (11 - 1)(2.4606)^2}{8 + 11 - 2} = 5.672$$

$$t = \frac{10.375 - 5.636}{\sqrt{5.672\left(\frac{1}{8} + \frac{1}{11}\right)}} = 4.28$$

Rechace H_0; el número medio de transacciones de los adultos jóvenes es mayor que el de los adultos mayores

35. H_0: $\mu_1 \leq \mu_2$ H_1: $\mu_1 > \mu_2$ Rechace H_0 si $t > 2.650$
$\bar{x}_1 = 125.125$ $s_1 = 15.094$
$\bar{x}_2 = 117.714$ $s_2 = 19.914$

$$s_p^2 = \frac{(8 - 1)(15.094)^2 + (7 - 1)(19.914)^2}{8 + 7 - 2} = 305.708$$

$$t = \frac{125.125 - 117.714}{\sqrt{305.708\left(\frac{1}{8} + \frac{1}{7}\right)}} = 0.819$$

No se rechaza H_0; no hay diferencia entre el número medio vendido al precio regular y el número medio vendido al precio reducido

37. H_0: $\mu_d \leq 0$ H_1: $\mu_d > 0$ Rechace H_0 si $t > 1.895$
$\bar{d} = 1.75$ $s_d = 2.9155$

$$t = \frac{1.75}{2.9155/\sqrt{8}} = 1.698$$

No rechace H_0; no hay diferencia entre los números medios de ausencias. El valor p es mayor que 0.05 pero menor que 0.10

39. H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
Rechace H_0 si $t < -2.024$ o $t > 2.204$

$$s_p^2 = \frac{(15 - 1)(40)^2 + (25 - 1)(30)^2}{15 + 25 - 2} = 1\,157.89$$

$$t = \frac{150 - 180}{\sqrt{1\,157.89\left(\frac{1}{15} + \frac{1}{25}\right)}} = -2.699$$

Rechace la hipótesis nula porque las medias de las poblaciones son distintas

41. H_0: $\mu_d \leq 0$ H_1: $\mu_d > 0$ Rechace H_0 si $t > 1.895$
$\bar{d} = 3.11$ $s_d = 2.91$

$$t = \frac{3.11}{2.91/\sqrt{8}} = 3.02$$

Rechace H_0; la media es menor

43. H_0: $\mu_O = \mu_R$ H_1: $\mu_O \neq \mu_R$ $gl = 25 + 28 - 2 = 51$
Rechace H_0 si $t < -2.008$ o $t > 2.008$
$\bar{x}_O = 86.24$, $s_o = 23.43$
$x_R = 92.04$, $s_R = 24.12$

$$s_p^2 = \frac{(25 - 1)(23.43)^2 + (28 - 1)(24.12)^2}{25 + 28 - 2} = 566.335$$

$$t = \frac{86.24 - 92.04}{\sqrt{566.335\left(\frac{1}{25} + \frac{1}{28}\right)}} = -0.886$$

No rechace H_0; no hay diferencia entre los números medios de automóviles vendidos en ambas concesionarias

45. H_0: $\mu_d \leq 0$ H_1: $\mu_d > 0$ Rechace H_0 si $t > 1.711$
$\bar{d} = 2.8$ $s_d = 6.59$

$$t = \frac{2.8}{6.59/\sqrt{25}} = 2.124$$

Rechace H_0; en promedio, hay más carros en el estacionamiento de US 17

47. **a.** μ_1 = sin alberca μ_2 = con alberca
H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
Rechace H_0 si $t > 2.000$ o $t < -2.000$
$\bar{x}_1 = 202.8$ $s_1 = 33.7$ $n_1 = 38$
$\bar{x}_2 = 231.5$ $s_2 = 50.46$ $n_2 = 67$

$$s_p^2 = \frac{(38 - 1)(33.7)^2 + (67 - 1)(50.46)^2}{38 + 67 - 2} = 2\,041.05$$

$$t = \frac{202.8 - 231.5}{\sqrt{2\,041.05\left(\frac{1}{38} + \frac{1}{67}\right)}} = -3.12$$

Rechace H_0; no hay diferencia entre los precios medios de venta de las casas con y sin alberca

b. μ_1 = sin cochera μ_2 = con cochera
H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
Rechace H_0 si $t > 2.000$ o $t < -2.000$
$\alpha = 0.05$ $gl = 34 + 71 - 2 = 103$
$\bar{x}_1 = 185.45$ $s_1 = 28.00$
$\bar{x}_2 = 238.18$ $s_2 = 44.88$

$$s_p^2 = \frac{(34 - 1)(28.00)^2 + (71 - 1)(44.88)^2}{103} = 1\,620.07$$

$$t = \frac{185.45 - 238.18}{\sqrt{1620.07\left(\frac{1}{34} + \frac{1}{71}\right)}} = -6.28$$

Rechace H_0; hay diferencia entre los precios medios de venta de las casas con y sin cochera

c. H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$
Rechace H_0 si $t > 2.036$ o $t < -2.036$
$\bar{x}_1 = 196.91 \quad s_1 = 35.78 \quad n_1 = 15$
$\bar{x}_2 = 227.45 \quad s_2 = 44.19 \quad n_2 = 20$

$$s_p^2 = \frac{(15-1)(35.78)^2 + (20-1)(44.19)^2}{15+20-2} = 1\,667.43$$

$$t = \frac{196.91 - 227.45}{\sqrt{1667.43\left(\frac{1}{15}+\frac{1}{20}\right)}} = -2.19$$

Rechace H_0; hay diferencia entre los precios medios de venta de las casas en los municipios 1 y 2

49. H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$
Rechace H_0 si t no está entre -1.991 y 1.991

$$s_p^2 = \frac{(53-1)(52.9)^2 + (27-1)(55.1)^2}{53+27-2} = 2\,878$$

$$t = \frac{454.8 - 441.5}{\sqrt{2\,878\left(\frac{1}{53}+\frac{1}{27}\right)}} = 1.05$$

No rechace H_0; puede no haber diferencia en el costo medio de mantenimiento entre ambos tipos de autobuses

CAPÍTULO 12

1. 9.01, del apéndice B.6

3. Rechace H_0 si $F > 10.5$, donde los grados de libertad en el numerador son 7 y, en el denominador, 5. $F = 2.04$, calculada mediante:

$$F = \frac{s_1^2}{s_2^2} = \frac{(10)^2}{(7)^2} = 2.04$$

No rechace H_0; no hay diferencia entre las variaciones de ambas poblaciones

5. H_0: $\sigma_1^2 = \sigma_2^2$ $\quad H_1$: $\sigma_1^2 \neq \sigma_2^2$
Rechace H_0 donde $F > 3.10$ (3.10 se encuentra casi a la mitad, entre 3.14 y 3.07). $F = 1.44$, calculada mediante:

$$F = \frac{(12)^2}{(10)^2} = 1.44$$

No rechace H_0; no hay diferencia entre las variaciones de ambas poblaciones

7. **a.** H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: no todas las medias de tratamiento son iguales
b. Rechace H_0 si $F > 4.26$
c. y d.

Fuente	SS	gl	MS	F
Tratamiento	62.17	2	31.08	21.94
Error	12.75	9	1.42	
Total	74.92	11		

e. Rechace H_0; no todas las medias de tratamiento son iguales

9. H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: no todas las medias de tratamiento son iguales. Rechace H_0 si $F > 4.26$

Fuente	SS	gl	MS	F
Tratamiento	276.50	2	138.25	14.18
Error	87.75	9	9.75	

Rechace H_0; no todas las medias de tratamiento son iguales

11. **a.** H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: no todas las medias de tratamiento son iguales
b. Rechace H_0 si $F > 4.26$
c. SST = 107.20, SSE = 9.47, SS total = 116.67
d.

Fuente	SS	gl	MS	F
Tratamiento	107.20	2	53.600	50.96
Error	9.47	9	1.052	
Total	116.67	11		

e. Como $50.96 > 4.26$, se rechaza H_0 porque al menos una de las medias difiere
f. $(\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\text{MSE}(1/n_1 + 1/n_2)}$
$= (9.667 - 2.20) \pm 2.262\sqrt{1.052(1/3 + 1/5)}$
$= 7.467 \pm 1.69$
$= [5.77, 9.157]$

Sí, puede concluir que los tratamientos 1 y 2 tienen medias diferentes

13. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$; H_1: no todas las medias son iguales
Se rechaza H_0 si $F > 3.71$

Fuente	SS	gl	MS	F
Tratamiento	32.33	3	10.77	2.36
Error	45.67	10	4.567	
Total	78.00	13		

Como 2.36 es menor que 3.71, no se rechaza H_0 porque no hay diferencia entre los números medios de semanas

15. **a.** H_0: $\mu_1 = \mu_2$; H_1: no todas las medias de tratamiento son iguales.
b. Rechace H_0 si $F > 18.5$
c. H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: no todas las medias de bloqueo son iguales. Se rechaza H_0 si $F > 19.0$
d. SS total $= (46.0 - 36.5)^2 + \cdots + (35 - 36.5)^2 = 289.5$
SST $= 3(42.33 - 36.5)^2 + 3(30.67 - 36.5)^2 = 204.167$
SSB $= 2(38.5 - 36.5)^2 + 2(31.5\ 2\ 36.5)^2 +$
$2(39.5 - 36.5)^2 = 8 + 50 + 18 = 76$
SSE $= 289.50 - 204.1667 - 76 = 9.3333$
e.

Fuente	SS	gl	MS	F
Tratamiento	204.167	1	204.167	43.75
Bloques	76.000	2	38.000	8.14
Error	9.333	2	4.667	
Total	289.5000	5		

f. $43.75 > 18.5$, rechace H_0 porque hay una diferencia entre los tratamientos. Como $8.14 < 19.0$, no rechace H_0 para los bloques porque no hay diferencia entre estos

17. Para tratamiento:
H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: no todas las medias son iguales
Rechace si $F > 4.46$

Para bloques:
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
H_1: no todas las medias son iguales
Rechace si $F > 3.84$

Fuente	SS	gl	MS	F
Tratamiento	62.53	2	31.2650	5.75
Bloques	33.73	4	8.4325	1.55
Error	43.47	8	5.4338	
Total	139.73			

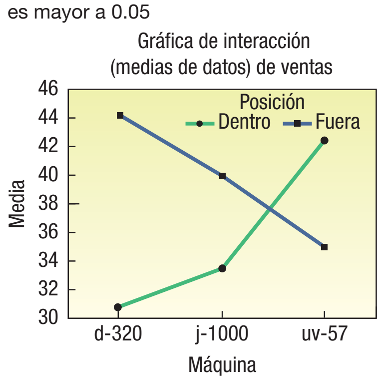
Hay una diferencia entre los turnos, no entre empleados

19.

Fuente	SS	gl	MS	F	P
Tamaño	156.333	2	78.1667	1.98	0.180
Peso	98.000	1	98.000	2.48	0.141
Interacción	36.333	2	18.1667	0.46	0.642
Error	473.333	12	39.444		
Total	764.000	17			

a. Como el valor p (0.18) es mayor a 0.05, no hay diferencia entre las medias del tamaño
b. El valor p de "Peso" (0.141) también es mayor que 0.05; por lo tanto, no hay diferencia entre esas medias
c. No existe una interacción significativa porque el valor p (0.642) es mayor a 0.05

21. **a.**



Sí, parece haber un efecto de interacción. Las ventas son diferentes con base en la posición de la máquina, ya sea en la posición dentro o fuera

b.

ANOVA de dos vías: ventas contra posición, máquina					
Fuente	***df***	**SS**	**MS**	***F***	**P**
Posición	1	104.167	104.167	9.12	0.007
Máquina	2	16.333	8.167	0.72	0.502
Interacción	2	457.333	228.667	20.03	0.000
Error	18	205.500	11.417		
Total	23	783.333			

La posición y la interacción de la posición y los efectos de la máquina son relevantes; pero el efecto de la máquina en las ventas no es importante

c.

ANOVA de una vía: ventas contra posición D-320					
Fuente	***gl***	**SS**	**MS**	***F***	**P**
Posición	1	364.50	364.50	40.88	0.001
Error	6	53.50	8.92		
Total	7	418.00			

ANOVA de una vía: ventas contra posición J-1000					
Fuente	***gl***	**SS**	**MS**	***F***	**P**
Posición	1	84.5	84.5	5.83	0.052
Error	6	87.0	14.5		
Total	7	171.5			

ANOVA de una vía: ventas contra posición UV-57					
Fuente	***gl***	**SS**	**MS**	***F***	**P**
Posición	1	112.5	112.5	10.38	0.018
Error	6	65.0	10.8		
Total	7	177.5			

Recomendaciones al utilizar los resultados estadísticos y las ventas medias graficadas en el inciso a): posicione la máquina D-320 fuera. De manera estadística, la posición de J-1000 no importa; posicione la máquina UV-57 dentro

23. H_0: $\sigma_1^2 \leq \sigma_2^2$; H_1: $\sigma_1^2 > \sigma_2^2$; $gl_1 = 21 - 1 = 20$; $gl_2 = 18 - 1 = 17$; se rechaza H_0 si $F > 3.16$

$$F = \frac{(45\,600)^2}{(21\,330)^2} = 4.57$$

Rechace H_0; hay más variación entre los precios de venta de las casas con frente al mar

25. Sharkey: $n = 7$ $s_s = 14.79$
White: $n = 8$ $s_w = 22.95$
H_0: $\sigma_w^2 \leq \sigma_s^2$; H_1: $\sigma_w^2 > \sigma_s^2$ $gl_s = 7 - 1 = 6$; $gl_w = 8 - 1 = 7$
Rechace H_0 si $F > 8.26$

$$F = \frac{(22.95)^2}{(14.79)^2} = 2.41$$

No puede rechazar H_0 porque no hay diferencia entre las variaciones de las ventas mensuales

27. **a.** H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
H_1: no todas las medias de tratamiento son iguales
b. $\alpha = 0.05$ Rechace H_0 si $F > 3.10$
c.

Fuente	**SS**	***gl***	**MS**	***F***
Tratamiento	50	$4 - 1 = 3$	50/3	1.67
Error	200	$24 - 4 = 20$	10	
Total	250	$24 - 1 = 23$		

b. No rechace H_0

29. H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: no todas las medias de tratamiento son iguales. Se rechaza H_0 si $F > 3.89$

Fuente	**SS**	***gl***	**MS**	***F***
Tratamiento	63.33	2	31.667	13.38
Error	28.40	12	2.367	
Total	91.73	14		

Se rechaza H_0 porque hay una diferencia entre las medias de tratamiento

31. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$; H_1: no todas las medias son iguales. Se rechaza H_0 si $F > 3.10$

Fuente	**SS**	***gl***	**MS**	***F***
Factor	87.79	3	29.26	9.12
Error	64.17	20	3.21	
Total	151.96	23		

Como la F calculada de $9.12 > 3.10$, se rechaza la hipótesis nula de que no hay diferencia con el nivel de 0.05

33. **a.** H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$. Valor crítico de $F = 4.75$

Fuente	**SS**	***gl***	**MS**	***F***
Tratamiento	219.43	1	219.43	23.10
Error	114.00	12	9.5	
Total	333.43	13		

b. $t = \dfrac{19 - 27}{\sqrt{9.5\left(\dfrac{1}{6} + \dfrac{1}{8}\right)}} = -4.806$

Entonces $t^2 = F$; es decir $(-4.806)^2 = 23.10$

c. Se rechaza H_0 porque hay una diferencia entre las calificaciones medias

35. Se rechaza la hipótesis nula debido a que el estadístico F (8.26) es mayor que el valor crítico (5.61) al nivel de significancia 0.01. El valor p (0.0019) también es menor que el nivel de significancia; los rendimientos medios en millas no son iguales

37. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ H_1: al menos una media es diferente
Rechace H_0 si $F > 2.7395$. Como $11.90 > 2.74$ se rechaza H_0; también puede ver esta conclusión a partir del valor p de $0.000 < 0.05$. El correo de prioridad express es más rápido que las otras tres clases, y el correo de prioridad es más rápido que el de primera clase o el estándar; sin embargo, el correo de primera clase y el estándar pueden ser iguales

39. Para el color, el valor crítico de F es 4.76; para el tamaño, es 5.14

Fuente	**SS**	***gl***	**MS**	***F***
Tratamiento	25.0	3	8.3333	5.88
Bloques	21.5	2	10.75	7.59
Error	8.5	6	1.4167	
Total	55.0	11		

Las H_0 del tratamiento y los bloques (color y tamaño) se rechazan. Al menos una media del color difiere y al menos una media del tamaño

41. **a.** El valor crítico de F es 3.49; la F calculada es 0.688.
No rechace H_0
b. El valor crítico de F es 3.26; el valor calculado de F es 100.204
Rechace H_0 para las medias de los bloques
Hay una diferencia entre las casas pero no entre los asesores

43. En el caso de la gasolina:
H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: el millaje medio no es el mismo
Rechace H_0 si $F > 3.89$.
En el caso del automóvil:
H_0: $\mu_1 = \mu_2 = \cdots = \mu_7$; H_1: el millaje medio no es el mismo
Rechace H_0 si $F > 3.00$

Tabla ANOVA				
Fuente	**SS**	***gl***	**MS**	***F***
Gasolina	44.095	2	22.048	26.71
Autos	77.238	6	12.873	15.60
Error	9.905	12	0.825	
Total	131.238	20		

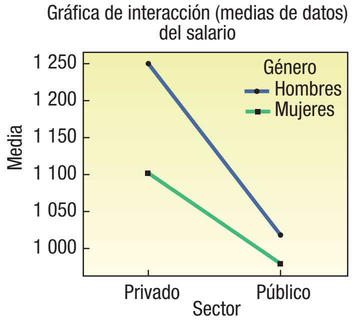
Hay una diferencia tanto entre los automóviles como entre la gasolina

45. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$; H_1: las medias de tratamiento no son iguales; rechace H_0 si $F > 2.37$

Fuente	**SS**	***gl***	**MS**	***F***
Tratamiento	0.03478	5	0.00696	3.86
Error	0.10439	58	0.0018	
Total	0.13917	63		

Se rechaza H_0 porque hay una diferencia entre las ponderaciones medias de los colores

47. **a.**



b. ANOVA de dos vías: salario contra género, sector

```
Fuente        GL      SS      MS      F      P
Género         1   44086   44086  11.44  0.004
Sector         1  156468  156468  40.61  0.000
Interacción    1   14851   14851   3.85  0.067
Error         16   61640    3853
Total         19  277046
```

No hay efecto de interacción del género y el sector en los salarios; sin embargo, hay diferencias relevantes entre los salarios medios con base en el género y diferencias significativas entre los salarios medios con base en el sector

c. ANOVA de una vía: salario frente a sector

```
Fuente  GL      SS      MS      F      P
Sector   1  156468  156468  23.36  0.000
Error   18  120578    6699
Total   19  277046

s = 81.85   R cuadrada = 56.48%
R cuadrada ajustada = 54.06%
```

ANOVA de dos vías: salario frente al género

```
Fuente  GL      SS      MS     F      P
Género   1   44086   44086  3.41  0.081
Error   18  232960   12942
Total   19  277046

s = 113.8  R cuadrada = 15.91%
R cuadrada ajustada = 11.24%
```

d. Los resultados estadísticos muestran que solo el sector (público o privado) tiene un efecto relevante en los salarios de los contadores

49. **a.** H_0: $\sigma_{np}^2 = \sigma_p^2$ $\quad H_1$: $\sigma_{np}^2 \neq \sigma_p^2$
Rechace H_0 si $F > 2.05$ (estimado)
$gl_1 = 67 - 1 = 66$; $gl_2 = 38 - 1 = 37$

$$F = \frac{(50.57)^2}{(33.71)^2} = 2.25$$

Rechace H_0; hay una diferencia entre las varianzas de los precios de venta

b. H_0: $\sigma_g^2 = \sigma_{ng}^2$ $\quad H_1$: $\sigma_g^2 \neq \sigma_{ng}^2$
Rechace H_0 si $F > 2.21$ (estimado)

$$F = \frac{(44.88)^2}{(28.00)^2} = 2.27$$

Rechace H_0; hay una diferencia entre las varianzas de ambos precios de venta

c.

Fuente	SS	*gl*	MS	*F*
Municipio	13 263	4	3 316	1.52
Error	217 505	100	2 175	
Total	230 768	104		

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$; H_1: no todas las medias de tratamiento son iguales. Rechace H_0 si $F > 2.46$

No rechace H_0; no hay una diferencia entre los precios de venta medios en los cinco municipios

51. **a.** H_0: $\mu_1 = \mu_2 = \mu_3$; H_1: las medias de tratamiento no son iguales. Rechace H_0 si $F > 4.89$

Fuente	SS	*gl*	MS	*F*
Tratamiento	28 996	2	14 498	5.62
Error	198 696	77	2 580	
Total	227 692	79		

Rechace H_0; las millas medias recorridas son diferentes

b. H_0: $\mu_1 = \mu_2 = \mu_3$ $\quad H_1$: las medias de tratamiento no son iguales. Rechace H_0 si $F > 3.12$

Fuente	SS	*gl*	MS	*F*
Tratamiento	5 095	2	2 547	1.45
Error	135 513	77	1 760	
Total	140 608	79		

No rechace H_0; las millas medias recorridas no son diferentes

c. $(441.81 - 506.75) \pm 1.991\sqrt{2\,580\left(\frac{1}{47} + \frac{1}{8}\right)}$

Esto se reduce a -64.94 ± 38.68, así que la diferencia está entre -103.62 y -26.26; en otras palabras, Bluebird es menos costoso que Thompson por una cantidad entre \$26.26 y \$103.62

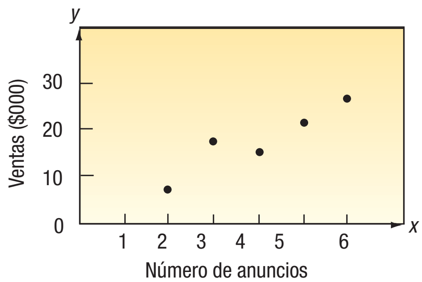
CAPÍTULO 13

1. $\Sigma(x - \bar{x})(y - \bar{y}) = 10.6$, $s_x = 2.7$, $s_y = 1.3$

$$r = \frac{10.6}{(5 - 1)(2.709)(1.38)} = 0.75$$

3. **a.** Ventas

b.



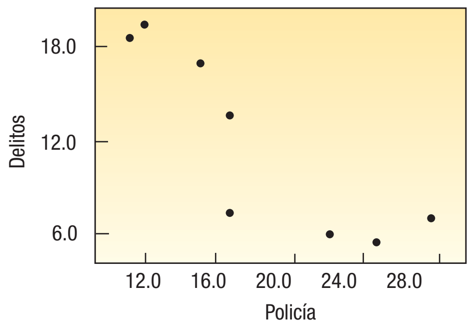
c. $\Sigma(x - \bar{x})(y - \bar{y}) = 36$, $n = 5$, $s_x = 1.5811$, $s_y = 6.1237$

$$r = \frac{36}{(5 - 1)(1.5811)(6.1237)} = 0.9295$$

d. Hay una fuerte asociación positiva entre las variables

5. **a.** Cualquier variable podía ser independiente; en el diagrama de dispersión, "policía" es la variable independiente

b.



c. $n - 8$, $\Sigma(x - \bar{x})(y - \bar{y}) = -231.75$
$s_x = 5.8737$, $s_y = 6.4462$

$$r = \frac{-231.75}{(8 - 1)(5.8737)(6.4462)} = -0.8744$$

d. Relación inversa fuerte. Conforme aumenta el número de policías disminuyen los delitos; o, a medida que los delitos aumentan, decrece el número de policías

7. Rechace H_0 si $t > 1.812$

$$t = \frac{0.32\sqrt{12-2}}{\sqrt{1-(0.32)^2}} = 1.068$$

No rechace H_0

9. H_0: $\rho \le 0$ H_1: $\rho > 0$. Rechace H_0 si $t > 2.552$ $gl = 18$

$$t = \frac{0.78\sqrt{20-2}}{\sqrt{1-(0.78)^2}} = 5.288$$

Rechace H_0; hay una correlación positiva entre los galones vendidos y el precio

11. H_0: $\rho \le 0$ H_1: $\rho > 0$
Rechace H_0 si $t > 2.650$ con $gl = 13$

$$t = \frac{0.667\sqrt{15-2}}{\sqrt{1-0.667^2}} = 3.228$$

Rechace H_0; hay una correlación positiva entre el número de pasajeros y el peso del avión

13. **a.** $\hat{y} = 3.7671 + 0.3630x$

$$b = 0.7522\left(\frac{1.3038}{2.7019}\right) = 0.3630$$

$$a = 5.8 - 0.3630(5.6) = 3.7671$$

b. 6.3081, determinado por $\hat{y} = 3.7671 + 0.3630(7)$

15. **a.** $\Sigma(x - \bar{x})(y - \bar{y}) = 44.6$, $s_x = 2.726$, $s_y = 2.011$

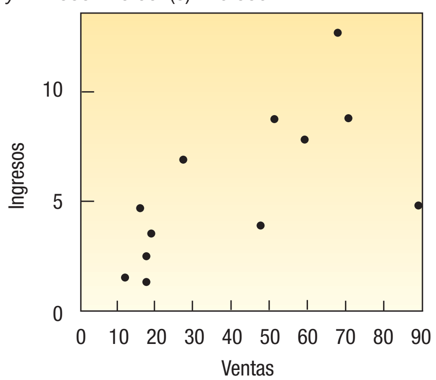
$$r = \frac{44.6}{(10-1)(2.726)(2.011)} = 0.904$$

$$b = 0.904\left(\frac{2.011}{2.726}\right) = 0.667$$

$$a = 7.4 - 0.677(9.1) = 1.333$$

b. $\hat{y} = 1.333 + 0.667(6) = 5.335$

17. **a.**



b. $\Sigma(x - \bar{x})(y - \bar{y}) = 629.64$, $s_x = 26.17$, $s_y = 3.248$

$$r = \frac{629.64}{(12-1)(26.17)(3.248)} = 0.6734$$

c. $b = 0.6734\left(\frac{3.248}{26.170}\right) = 0.0836$

$$a = \frac{64.1}{12} - 0.0836\left(\frac{501.10}{12}\right) = 1.8507$$

d. $\hat{y} = 1.8507 + 0.0836(50.0) = 6.0307$ (millones de dólares)

19. **a.** $b = 0.8744\left(\frac{6.4462}{5.8737}\right) = -0.9596$

$$a = \frac{95}{8} - (-0.9596)\left(\frac{146}{8}\right) = 29.3877$$

b. 10.1957, determinado mediante $29.3877 - 0.9596(20)$

c. Por cada policía adicional, los delitos disminuyen en casi uno.

21. H_0: $\beta \ge 0$ H_1: $\beta < 0$ $gl = n - 2 = 8 - 2 = 6$
Rechace H_0 si $t < -1.943$

$$t = -0.96/0.22 = -4.364$$

Rechace H_0 y concluya que la pendiente es distinta a cero

23. H_0: $\beta = 0$ H_1: $\beta \ne 0$ $gl = n - 2 = 12 - 2 = 10$
Rechace H_0 si t no está entre -2.228 y 2.228

$$t = 0.08/0.03 = 2.667$$

Rechace H_0 y concluya que la pendiente es distinta a cero

25. El error estándar de estimación es 3.378, calculado por $\sqrt{\frac{68.4814}{8-2}}$

El coeficiente de determinación es 0.76, calculado por $(-0.874)^2$

76% de la variación en los delitos puede explicarse por la variación en los policías

27. El error estándar de estimación es 0.913, calculado por $\sqrt{\frac{6.667}{10-2}}$

El coeficiente de determinación es 0.82, calculado por 29.733/36.4. 82% de la variación en las horas-kilovatio puede explicarse por la variación en el número de habitaciones

29. **a.** $r^2 = \frac{1\,000}{1\,500} = 0.6667$

b. $r = \sqrt{0.6667} = 0.8165$

c. $s_{y\cdot x} = \sqrt{\frac{500}{13}} = 6.2017$

31. **a.** $6.308 \pm (3.182)(0.993)\sqrt{0.2 + \frac{(7-5.6)^2}{29.2}}$
$= 6.308 \pm 1.633$
$= [4.675, 7.941]$

b. $6.608 \pm (3.182)(0.993)\sqrt{1 + 1/5 + 0.0671}$
$= [2.751, 9.865]$

33. **a.** 4.2939, 6.3721

b. 2.9854, 7.6806

35. La correlación entre ambas variables es 0.298. Elevando al cuadrado x, la correlación aumenta a 0.998

37. H_0: $\rho \le 0$ H_1: $\rho > 0$. Rechace H_0 si $t > 1.714$

$$t = \frac{0.94\sqrt{25-2}}{\sqrt{1-(0.94)^2}} = 13.213$$

Rechace H_0; hay una correlación positiva entre pasajeros y el peso del equipaje

39. H_0: $\rho \le 0$ H_1: $\rho > 0$. Rechace H_0 si $t > 2.764$

$$t = \frac{0.47\sqrt{12-2}}{\sqrt{1-(0.47)^2}} = 1.684$$

No rechace H_0; no hay una correlación positiva entre el tamaño del motor y el desempeño. El valor p es mayor que 0.05, pero menor que 0.10

41. **a.** El volumen de ventas está inversamente relacionado con la participación en el mercado

b. $r = \frac{-299.792}{(13-1)(3.686)(9.808)} = -0.691$

c. H_0: $\rho \ge 0$ H_1: $\rho < 0$ Rechace H_0 si $t < -2.718$ $gl = 11$

$t = \frac{-0.691\sqrt{13-2}}{\sqrt{1-(-0.691)^2}} = -3.17$ Rechace H_0; hay una correlación negativa entre los carros vendidos y la participación en el mercado

d. 47.7%, calculado por $(-0.691)^2$, de la variación en la participación del mercado está representado por la variación en los autos vendidos

43. **a.** $r = -0.393$

b. El coeficiente de determinación es 0.1544, calculado elevando al cuadrado (-0.393)

c. H_0: $\rho \ge 0$ H_1: $\rho < 0$ Rechace H_0 si $t < -1.697$

$$t = \frac{-0.393\sqrt{32-2}}{\sqrt{1-(-0.393)^2}} = -2.34$$

Se rechaza H_0 porque hay una correlación negativa entre los puntos registrados y los permitidos

d. Para la Conferencia Nacional (NFC), H_0: $\rho \ge 0$ H_1: $\rho < 0$
Rechace H_0 si $t < -1.761$

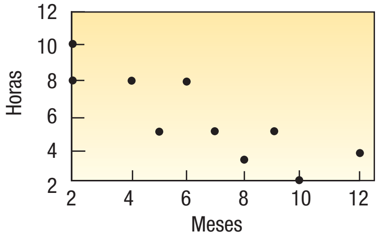
$$t = \frac{-0.139\sqrt{16-2}}{\sqrt{1-(-0.139)^2}} = -0.53$$

No rechace H_0; *no* se puede decir que existe una correlación negativa entre los puntos registrados y los permitidos en la NFC. Para la Conferencia Americana (AFC), H_0: $\rho \ge 0$ H_1: $\rho < 0$.
Rechace H_0 si $t < -1.761$

$$t = \frac{-0.576\sqrt{16-2}}{\sqrt{1-(-0.576)^2}} = -2.64$$

Rechace H_0; se puede decir que existe una correlación negativa entre los puntos registrados y los permitidos en la AFC

45. **a.**



Hay una relación inversa entre las variables. Conforme aumentan los meses de posesión, el número de horas de ejercicio disminuye

b. $r = -0.827$

c. H_0: $\rho \geq 0$ H_1: $\rho < 0$. Rechace H_0 si $t < -2.896$

$$t = \frac{-0.827\sqrt{10-2}}{\sqrt{1-(-0.827)^2}} = -4.16$$

Rechace H_0; hay una asociación negativa entre los meses en posesión y las horas ejercitadas

47. **a.** La edad mediana y la población están directamente relacionadas

b. $r = \dfrac{11.93418}{(10 - 1)(2.207)(1.330)} = 0.452$

c. La pendiente de 0.272 indica que por cada incremento de un millón en la población, la edad mediana aumenta 0.272 años en promedio

d. La edad mediana es 32.08 años, calculado por $31.4 + 0.272(2.5)$

e. El valor p (0.190) de la variable población es mayor que, digamos, 0.05. No se puede rechazar una prueba de significancia de dicho coeficiente; en otras palabras, es posible que el coeficiente de la población sea cero

f. H_0: $\rho = 0$ H_1: $\rho \neq 0$. Rechace H_0 si t no está entre -1.86 y 1.86

$gl = 8$ $\quad t = \dfrac{0.452\sqrt{10-2}}{\sqrt{1-(0.452)^2}} = 1.433$ No rechace H_0; quizá no haya relación entre la edad y la población

49. **a.** $b = -0.4667$, $a = 11.2358$

b. $\hat{y} = 11.2358 - 0.4667(7.0) = 7.9689$

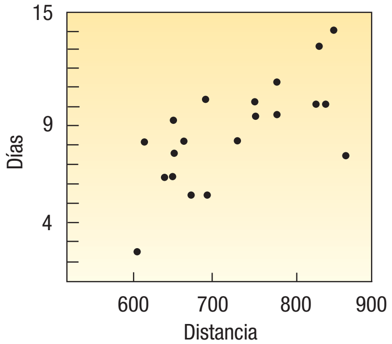
c. $7.9689 \pm (2.160)(1.114)\sqrt{1 + \dfrac{1}{15} + \dfrac{(7-7.1333)^2}{73.7333}}$

$= 7.9689 \pm 2.4854$

$= [5.4835, 10.4543]$

d. $R^2 = 0.499$. Casi 50% de la variación en la cantidad de la licitación se explica por el número de los licitadores

51. **a.**



Parece haber una relación entre ambas variables: conforme aumenta la distancia, también lo hace el tiempo de embarque

b. $r = 0.692$

H_0: $\rho \leq 0$ H_1: $\rho > 0$ Rechace H_0 si $t > 1.734$

$$t = \frac{0.692\sqrt{20-2}}{\sqrt{1-(0.692)^2}} = 4.067$$

Se rechaza H_0 porque hay una asociación positiva entre la distancia de embarque y el tiempo de envío

c. $R^2 = 0.479$. Casi la mitad de la variación en el tiempo de envío se explica por la distancia de embarque

d. $S_{y \cdot x} = 2.004$

53. **a.** $b = 2.41$

$a = 26.8$

La ecuación de regresión es: Precio $= 26.8 + 2.41 \times$ Dividendo. Por cada dólar adicional de "dividendo", "precio" aumenta \$2.41

b. Para probar la significancia de la pendiente, se usa $n - 2$, o $30 - 2 = 28$ grados de libertad. Con el nivel de significancia 0.05, los valores críticos son -2.048 y 2.048. El estadístico de prueba t es $t = \dfrac{b - 0}{s_b} = \dfrac{2.408}{0.328} = 7.34$. Se rechaza la hipótesis nula (que la pendiente es igual a cero)

c. $R^2 = \dfrac{5\,057.6}{7\,682.7} = 0.658$ Por lo tanto, 65.8% de la variación del precio se explica por el dividendo

d. $r = \sqrt{0.658} = 0.811$ $\quad H_0$: $\rho \leq 0$ $\quad H_1$: $\rho > 0$.

Con un nivel de significancia de 5%, rechace H_0 cuando $t > 1.701$

$$t = \frac{0.811\sqrt{30-2}}{\sqrt{1-(0.811)^2}} = 7.34$$

Se rechaza H_0 porque la correlación de la población es positiva

55. **a.** 35

b. $s_{y \cdot x} = \sqrt{29\,778\,406} = 5\,456.96$

c. $r^2 = \dfrac{13\,548\,662\,082}{14\,531\,349\,747} = 0.932$

d. $r = \sqrt{0.932} = 0.966$

e. H_0: $\rho \leq 0$ $\quad H_1$: $\rho > 0$; rechace H_0 cuando $t > 1.692$

$$t = \frac{0.966\sqrt{35-2}}{\sqrt{1-(0.966)^2}} = 21.46$$

Rechace H_0; hay una relación directa entre el tamaño de la casa y su valor de mercado

57. **a.** La ecuación de regresión es Precio $= -773 + 1\,048$ Velocidad

b. La segunda computadora portátil (1.6, 992) tiene un residuo de -557.60, es decir, cuesta \$557.60 por debajo del precio pronosticado; es una "oferta" relevante

c. La correlación de Velocidad y Precio es 0.835

H_0: $\rho \leq 0$ $\quad H_1$: $\rho > 0$ $\quad$ Rechace H_0 si $t > 1.8125$

$$t = \frac{0.835\sqrt{12-2}}{\sqrt{1-(0.835)^2}} = 4.799$$

Rechace H_0; es razonable decir que la correlación de la población es positiva

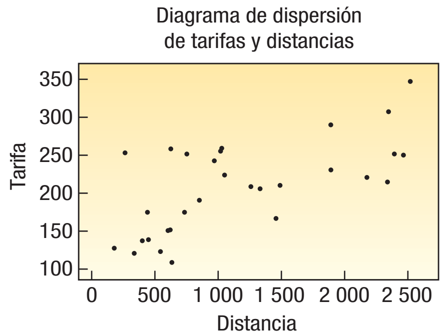
59. **a.** $r = 0.987$ $\quad H_0$: $\rho \leq 0$ $\quad H_1$: $\rho > 0$ $\quad$ Rechace H_0 si $t > 1.746$

$$t = \frac{0.987\sqrt{18-2}}{\sqrt{1-(0.987)^2}} = 24.564$$

b. $\hat{y} = -29.7 + 22.93x$; una taza adicional aumenta el peso del perro casi 23 libras

c. El perro número 4 come demasiado

61. **a.**



La relación es directa: las tarifas aumentan en los vuelos más largos

b. La correlación de "Distancia" y "Tarifa" es 0.656

H_0: $\rho \leq 0$ $\quad H_1$: $\rho > 0$ $\quad$ Rechace H_0 si $t > 1.701$ $\quad gl = 28$

$$t = \frac{0.656\sqrt{30-2}}{\sqrt{1-(0.656)^2}} = 4.599$$

Rechace H_0; hay una correlación positiva significativa entre las tarifas y las distancias

c. 43%, calculado por $(0.656)^2$, de la variación en las tarifas se explica por la variación en la distancia

d. La ecuación de regresión es Tarifa = 147.08 + 0.05265 Distancia. Cada milla adicional añade \$0.05265 a la tarifa. Un vuelo de 1 500 millas costaría \$226.06, calculado por \$147.08 + 0.05265(1 500)

e. Un vuelo de 4 218 millas está fuera del rango de los datos muestreados, así que la ecuación de regresión puede no ser útil

63. a. Parece haber una relación directa entre las variables

b. La asistencia esperada con un salario de 80 millones es 2.2529 millones, calculado por 1.17855 + 0.01343(80)

c. Incrementar el salario en 30 millones aumentará la asistencia en 0.4029 en promedio, calculado por 0.01343(30); esta es también la diferencia entre la asistencia esperada con un salario de 110 millones, y la asistencia esperada de 80 millones

d. A continuación se presenta la salida de regresión de Excel

Resumen

Estadísticas de regresión	
R múltiple	0.770317814
R cuadrada	0.593389535
R cuadrada ajustada	0.578867732
Error estándar	0.416717999
Observaciones	30

ANOVA

	gl	SS	MS	*F*	*F* de significancia
Regresión	1	7.095841062	7.095841062	40.86197572	6.41022E-07
Residuo	28	4.862308938	0.173653891		
Total	29	11.95815			

	Coeficientes	Error estándar	Estadístico *t*	*Valor P*
Intercesión	1.178554138	0.21954557	5.368152674	1.01455E-05
Salario 2012	0.013429923	0.002100941	6.392337265	6.41022E-07

H_0: $\beta \leq 0$ $\quad H_1$: $\beta > 0$ $\quad gl = n - 2 = 30 - 2 = 28$
Rechace H_0 si $t > 1.701$ $\quad t = 0.013429923/0.002100941 = 6.39233$. Rechace H_0 y concluya que la pendiente es positiva

e. 0.593389 o 59.3389% de la variación se explica por la variación en el salario

f. La correlación entre “asistencia” y “promedio de bateo” es 0.6538
H_0: $\rho \leq 0$ $\quad H_1$: $\rho > 0$ $\quad$ En un nivel de 5%, rechace H_0 si $t >$ 1.701

$$t = \frac{0.6538\sqrt{30-2}}{\sqrt{1-(0.6538)^2}} = 4.573 \qquad \text{Rechace } H_0$$

El promedio de bateo y la asistencia están correlacionados positivamente. La correlación entre asistencia y ERA es −0.0548. La correlación entre asistencia y promedio de bateo es más fuerte que la que existe entre la asistencia y ERA
H_0: $\rho \geq 0$ $\quad H_1$: $\rho < 0$ $\quad$ En un nivel de 5%, rechace H_0 si t −1.701

$$t = \frac{-0.0548\sqrt{30-2}}{\sqrt{1-(-0.0548)^2}} = 0.291 \qquad \text{No rechace } H_0$$

ERA y asistencia podrían no tener una correlación

CAPÍTULO 14

1. a. Ecuación de regresión múltiple

b. La intersección *y*

c. $\hat{y} = 64\,100 + 0.394(796\,000) + 9.6(6\,940) - 11\,600(6.0) = \$374\,748$

3. a. 497.736 determinado por
$\hat{y} = 16.24 + 0.017(18) + 0.0028(26\,500) + 42(3) + 0.0012(156\,000) + 0.19(141) + 26.8(2.5)$

b. Dos actividades sociales más. El ingreso solo agregó 28 al índice; las actividades sociales agregaron 53.6

5. a. $s_{Y\cdot 12} = \sqrt{\frac{SSE}{n-(k+1)}} = \sqrt{\frac{583.693}{65-(2+1)}}$
$= \sqrt{9.414} = 3.068$
95% de los residuos estarán entre ±6.136, determinado mediante 2(3.068)

b. $R^2 = \frac{SSR}{SS\ total} = \frac{77.907}{661.6} = 0.118$
Las variables independientes explican 11.8% de la variación

c. $$R^2_{adj} = 1 - \frac{\frac{SSE}{n-(k+1)}}{\frac{SS\ total}{n-1}} = 1 - \frac{\frac{583.693}{65-(2+1)}}{\frac{661.6}{65-1}}$$
$$= 1 - \frac{9.414}{10.3375} = 1 - 0.911 = 0.089$$

7. a. $\hat{y} = 84.998 + 2.391x_1 - 0.4086x_2$

b. 90.0674, determinado por $\hat{y} = 84.998 + 2.391(4) - 0.4086(11)$

c. $n = 65$ y $k = 2$

d. H_0: $\beta_1 = \beta_2 = 0$ $\quad H_1$: no todas las β son cero
Rechace H_0 si $F > 3.15$
$F = 4.14$, rechace H_0; no todos los coeficientes de regresión netos son iguales a cero

e.

Para x_1	Para x_2
H_0: $\beta_1 = 0$	H_0: $\beta_2 = 0$
H_1: $\beta_1 \neq 0$	H_1: $\beta_2 \neq 0$
$t = 1.99$	$t = -2.38$

Rechace H_0 si $t > 2.0$ o bien $t < -2.0$
Elimine la variable 1 y mantenga la 2

f. El análisis de regresión se debe repetir solo con x_2 como variable independiente

9. a. La ecuación de regresión es: Desempeño = 29.3 + 5.22 Aptitud + 22.1 Sindicato

Predictor	Coef	Error estándar del coeficiente	T	P
Constante	29.28	12.77	2.29	0.041
Aptitud	5.222	1.702	3.07	0.010
Sindicato	22.135	8.852	2.50	0.028

S = 16.9166 R cuadrada = 53.3%
R cuadrada ajustada = 45.5%

Análisis de varianza

Fuente	GL	SS	MS	F	P
Regresión	2	3919.3	1959.6	6.85	0.010
Error residual	12	3434.0	286.2		
Total	14	7353.3			

b. Estas variables son eficaces para predecir el desempeño. Explican 53.3% de la variación en el desempeño; en particular, los miembros de un sindicato aumentan 22.1 el desempeño típico

c. H_0: $\beta_2 = 0$ $\quad H_1$: $\beta_2 \neq 0$
Rechace H_0 si $t < -2.179$ o bien $t > 2.179$
Como 2.50 es mayor que 2.179, rechace la hipótesis nula y concluya que la membresía del sindicato es relevante y se debe incluir

d. Cuando usted considera la variable interacción, la ecuación de regresión es Desempeño = 38.7 + 3.80 Aptitud − 0.1 Sindicato + 3.61 x_1x_2

Predictor	Coef	Error estándar del coeficiente	T	P
Constante	38.69	15.62	2.48	0.031
Aptitud	3.802	2.179	1.74	0.109
Sindicato	−0.10	23.14	−0.00	0.997
X_1X_2	3.610	3.473	1.04	0.321

El valor correspondiente al término interacción es 1.04; esto no es relevante, por lo tanto, concluya que no hay interacción entre aptitud y membresía en sindicato cuando se predice el desempeño laboral

11. **a.** La ecuación de regresión es
Precio $= 3\,080 - 54.2$ Licitadores $+ 16.3$ Edad

Predictor	Coef	Error estándar del coeficiente	T	P
Constante	3080.1	343.9	8.96	0.000
Licitadores	−54.19	12.28	−4.41	0.000
Edad	16.289	3.784	4.30	0.000

El precio disminuye 54.2 conforme participa un licitador adicional; en tanto que el precio aumenta 16.3 conforme la pintura envejece. ¡Aunque uno podría esperar que las pinturas antiguas valgan más, es inesperado que el precio disminuya conforme participen más licitadores!

b. La ecuación de regresión es
Precio $= 3\,972 - 185$ Licitadores $+ 6.35$ Edad $+ 1.46\, x_1x_2$

Predictor	Coef	Error estándar del coeficiente	T	P
Constante	3971.7	850.2	4.67	0.000
Licitadores	−185.0	114.9	−1.61	0.122
Edad	6.353	9.455	0.67	0.509
X_1X_2	1.462	1.277	1.15	0.265

El valor t correspondiente al término interacción es 1.15; esto no es relevante, por lo tanto, concluya que no hay interacción

c. En el procedimiento por pasos, el número de licitadores ingresa primero a la ecuación; luego ingresa el término interacción. La variable "edad" no se debe incluir ya que no es significativa. La respuesta es "Precio" en 3 factores de predicción, con $N = 25$

Paso	1	2
Constante	4,507	4,540
Licitadores	−57	−256
Valor T	−3.53	−5.59
Valor P	0.002	0.000
X_1X_2		2.25
Valor T		4.49
Valor P		0.000
S	295	218
R cuadrada	35.11	66.14
R cuadrada ajustada	32.29	63.06

13. **a.** $n = 40$ **c.** $R^2 = \dfrac{750}{1\,250} = 0.60$

b. 4 **d.** $s_{y\cdot 1234} = \sqrt{500/35} = 3.7796$

e. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
H_1: no todas las β son iguales a cero
H_0 se rechaza si $F > 2.65$

$$F = \frac{750/4}{500/35} = 13.125$$

Se rechaza H_0 porque al menos una β_i no es igual a cero

15. **a.** $n = 26$
b. $R^2 = 100/140 = 0.7143$
c. 1.4142, calculado por $\sqrt{2}$
d. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
H_1: no todas las β son 0
H_0: se rechaza si $F > 2.71$
$F = 10.0$ calculada. Rechace H_0; al menos un coeficiente de regresión no es cero
e. H_0 se rechaza en cada caso si $t < -2.086$ o bien $t > 2.086$. Se deben eliminar x_1 y x_5

17. **a.** $28 000
b. $R^2 = \dfrac{SSR}{SS\ total} = \dfrac{3\,050}{5\,250} = 0.5809$
c. 9.199, determinado mediante $\sqrt{84.62}$
d. Se rechaza H_0 si $F > 2.97$ (aproximadamente)

$$F \text{ calculada} = \frac{1\,016.67}{84.62} = 12.01$$

Se rechaza H_0 porque al menos un coeficiente de regresión no es cero
e. Si la t calculada está a la derecha de -2.056 o a la derecha de 2.056, se rechaza la hipótesis nula en cada uno de estos casos. La t calculada para x_2 y x_3 sobrepasa el valor crítico; por lo tanto, "población" y "gastos en publicidad" se deben conservar y eliminar "número de competidores", x_1

19. **a.** La correlación más fuerte es entre el promedio de calificaciones en el bachillerato y el paralegal; no hay problema con multicolinealidad
b. $R^2 = \dfrac{4.3595}{5.0631} = 0.8610$
c. Se rechaza H_0 si $F > 5.41$

$$F = \frac{1.4532}{0.1407} = 10.328$$

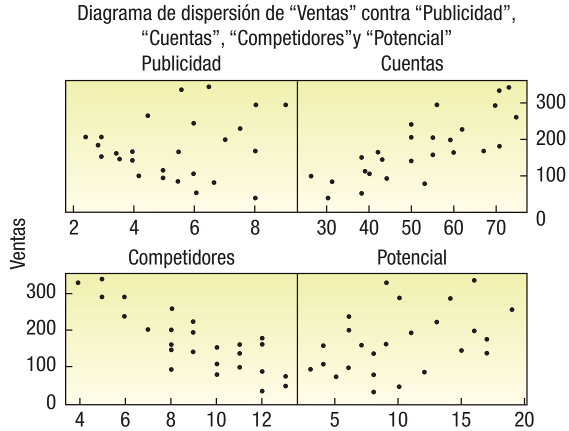
Al menos un coeficiente no es cero
d. Se rechaza cualquier H_0 si $t < -2.571$ o bien $t > 2.571$; parece que solo el promedio del bachillerato es relevante. Se pueden eliminar "Verbal" y "Matemáticas"
e. $R^2 = \dfrac{4.2061}{5.0631} = 0.8307$
R^2 solo se ha reducido 0.0303
f. Los residuos parecen ligeramente sesgados (positivos), pero aceptables
g. No parece haber un problema con la gráfica

21. **a.** La matriz de correlación de "Pantalla" y "Precio" es 0.893; así que no parece haber una relación lineal entre ambas
b. "Precio" es la variable *dependiente*
c. La ecuación de regresión es Precio $= -2\,484 + 101 \times$ Pantalla. Por cada pulgada de aumento de tamaño de la pantalla, el precio se eleva $101 en promedio
d. Usando variables indicadoras "ficticias" para Sharp y Sony, la ecuación de regresión es Precio $= -2\,308 + 94.1$ Pantalla $+ 15$ fabricante Sharp $+ 381$ fabricante Sony. Sharp puede obtener, en promedio, $15 más que Samsung, y Sony puede obtener una ganancia adicional de $381 más que Samsung
e. A continuación se muestra una parte de la salida:

Predictor	Coeficiente	Error estándar de coeficiente	T	P
Constante	-2308.2	492.0	-4.69	0.000
Pantalla	94.12	10.83	8.69	0.000
Sharp	15.1	171.6	0.09	0.931
Sony	381.4	168.8	2.26	0.036

El valor p de Sharp es relativamente grande. No puede rechazarse una prueba de su coeficiente; eso significa que puede no tener una ventaja real sobre Samsung. Por otra parte, el valor p del coeficiente de Sony es bastante pequeño. Eso indica que no ocurrió por azar, y que existe cierta ventaja real de Sony sobre Samsung
f. Un histograma de los residuos indica que siguen una distribución normal
g. La variación residual puede estar aumentando para valores ajustados más grandes

23. **a.**



Las ventas parecen disminuir con el número de competidores y aumentan con el número de cuentas y el potencial

b. Correlaciones de Pearson

	Ventas	Publicidad	Cuentas	Competidores
Publicidad	0.159			
Cuentas	0.783	0.173		
Competidores	−0.833	−0.038	−0.324	
Potencial	0.407	−0.071	0.468	−0.202

El número de cuentas y el potencial de mercado están moderadamente correlacionados

c. La ecuación de regresión es:
Ventas = 178 + 1.81 Publicidad + 3.32 Cuentas − 21.2 Competidores + 0.325 Potencial

Predictor	Coeficiente	Error estándar del coeficiente	T	P
Constante	178.32	12.96	13.76	0.000
Publicidad	1.807	1.081	1.67	0.109
Cuentas	3.3178	0.1629	20.37	0.000
Competidores	−21.1850	0.7879	−26.89	0.000
Potencial	0.3245	0.4678	0.69	0.495

S = 9.60441 *R* cuadrada = 98.9%
R cuadrada ajustada = 98.7%

Analisis de la varianza

Fuente	GL	SS	MS	F	P
Regresión	4	176777	44194	479.10	0.000
Error residual	21	1937	92		
Total	25	178714			

El valor *F* calculado es muy grande; por lo tanto, puede rechazar la hipótesis nula que todos los coeficientes de regresión son cero. Concluya que algunas de las variables independientes son eficaces en explicar las ventas

d. El potencial de "mercado" y "publicidad" tienen valores *p* grandes (0.495 y 0.109, respectivamente); probablemente deba omitirlas

e. Si omite el potencial, la ecuación de regresión es:
Ventas = 180 + 1.68 Publicidad + 3.37 Cuentas − 21.1 Competidores

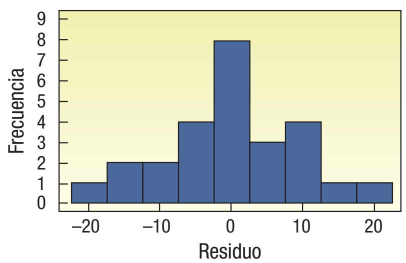
Predictor	Coeficiente	Error estándar del coeficiente	T	P
Constante	179.84	12.62	14.25	0.000
Publicidad	1.677	1.052	1.59	0.125
Cuentas	3.3694	0.1432	23.52	0.000
Competidores	−21.2165	0.7773	−27.30	0.000

Ahora la publicidad no es importante; esto también conduciría a dejar fuera la variable "publicidad" y reportar que la ecuación de regresión pulida es:
Ventas = 187 + 3.41 Cuentas − 21.2 Competidores

Predictor	Coeficiente	Error estándar del coeficiente	T	P
Constante	186.69	12.26	15.23	0.000
Publicidad	3.4081	0.1458	23.37	0.000
Competidores	−21.1930	0.8028	−26.40	0.000

f.

Histograma de residuos
(la respuesta es "ventas")



El histograma parece ser normal. No hay problemas indicados en esta gráfica

g. El factor de inflación de la varianza de ambas variables es 1.1. Son menores que 10; no hay problemas ya que este valor indica que las variables independientes no están fuertemente correlacionadas entre ellas

25. He aquí la captura de pantalla:

Predictor	*Coeficiente*	*Desv. est.*	*Valor t*	*p*
Constante	651.9	345.3	1.89	0.071
Servicio	13.422	5.125	2.62	0.015
Edad	−6.710	6.349	−1.06	0.301
Género	205.65	90.27	2.28	0.032
Trabajo	−33.45	89.55	−0.37	0.712

Análisis de la varianza

FUENTE	*GL*	*SS*	*MS*	*F*	*p*
Regresión	4	1066830	266708	4.77	0.005
Error	25	1398651	55946		
Total	29	2465481			

a. $\hat{y} = 651.9 + 13.422x_1 - 6.710x_2 + 205.65x_3 - 33.45x_4$

b. $R^2 = 0.433$, que es un poco bajo para este tipo de estudio.

c. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$; H_1: no todas las β son iguales a cero
Rechace H_0 si $F > 2.76$

$$F = \frac{1\,066\,830/4}{1\,398\,651/25} = 4.77$$

Se rechaza H_0 porque no todas las β son iguales a cero

d. Usando el nivel de significancia 0.05, rechace la hipótesis de que el coeficiente de regresión es 0 si $t < -2.060$ o $t > 2.060$. "Servicio" y "género" deben permanecer en el análisis; "edad" y "empleo" pueden eliminarse

e. A continuación se presenta la imagen de la captura de pantalla usando las variables independientes servicio y género:

Predictor	*Coeficiente*	*Desv. est.*	*Valor t*	*p*
Constante	784.2	316.8	2.48	0.020
Servicio	9.021	3.106	2.90	0.007
Género	224.41	87.35	2.57	0.016

Análisis de varianza

FUENTE	*GL*	*SS*	*MS*	*F*	*p*
Regresión	2	998779	499389	9.19	0.001
Error	27	1466703	54322		
Total	29	2465481			

Un hombre gana $224 más al mes que una mujer. La diferencia entre empleos técnicos y administrativos no es relevante

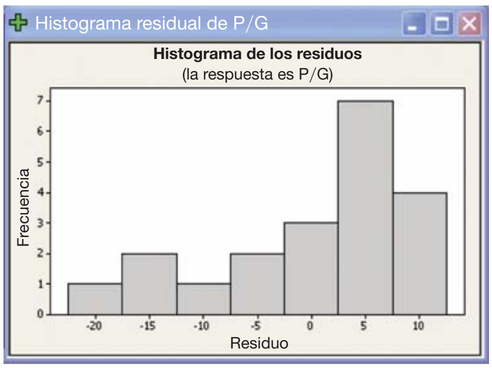
27. **a.** $\hat{y} = 29.913 - 5.324x_1 + 1.449x_2$

b. EPS es ($t = -3.26$, valor $p = 0.005$). Producción no es ($t = 0.81$, valor $p = 0.431$)

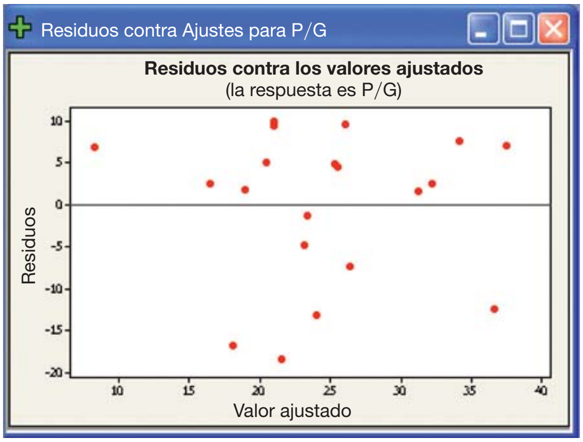
c. Un aumento de 1 en GPA genera una disminución de 5.324 en P/G

d. El número 2 de acciones está devaluada

e. A continuación se muestra una gráfica residual que *no* parece seguir la distribución normal:



f. No parece haber problema con la gráfica de los residuos contra los valores ajustados



g. La correlación entre producción y GPA no es un problema. No hay problema con la multicolinealidad

	P/E	EPS
EPS	−0.602	
Producción	0.054	0.162

29. a. La ecuación de regresión es
Ventas (000) = 1.02 + 0.0829 Infomerciales

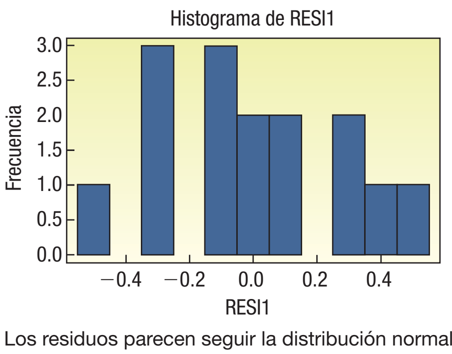
Predictor	Coeficiente	Error estándar del coeficiente	T	P
Constante	1.0188	0.3105	3.28	0.006
Informerciales	0.08291	0.01680	4.94	0.000

Análisis de la varianza

FUENTE	GL	SS	MS	F	P
Regresión	1	2.3214	2.3214	24.36	0.000
Error residual	13	1.2386	0.0953		
Total	14	3.5600			

La prueba global demuestra que hay una relación entre las ventas y el número de informerciales

b.



Los residuos parecen seguir la distribución normal

31. a. La ecuación de regresión es
Precio en la subasta = −118 929 + 1.63 Préstamo + 2.1 Pago mensual + 50 Pagos realizados

Análisis de varianza

FUENTE	*GL*	SS	MS	F	P
Regresión	3	5966725061	1988908354	39.83	0.000
Residuo Error	16	798944439	49934027		
Total	19	6765669500			

La *F* calculada es 39.83; mucho mayor que el valor crítico 3.24. Asimismo, el valor *p* es muy pequeño; por lo tanto, la hipótesis nula que todos los coeficientes de regresión son cero se puede rechazar. Al menos uno de los coeficientes de regresión múltiples es diferente a cero

b.

Predictor	Coeficiente	Error estándar del coeficiente	T	P
Constante	−118929	19734	−6.03	0.000
Préstamo	1.6268	0.1809	8.99	0.000
Pago mensual	2.06	14.95	0.14	0.892
Pagos realizados	50.3	134.9	0.37	0.714

La hipótesis nula es que el coeficiente es cero en la prueba individual; se debería rechazar si *t* es menor que −2.120 o mayor que 2.120. En este caso, el valor *t* de la variable préstamo es mayor que el valor crítico; por lo tanto, no se debe eliminar, sin embargo, las variables pago mensual y pagos realizados es probable que se eliminen

c. La ecuación de regresión revisada es: Precio en la subasta = −119 893 + 1.67 Préstamo

33. He aquí la captura de pantalla:

Predictor	*Coeficiente*	*Error estándar del coeficiente*	*T*	*P*
Constante	38.71	39.02	.99	.324
Habitaciones	7.118	2.551	2.79	0.006
Tamaño	0.03800	0.01468	2.59	0.011
Alberca	18.321	6.999	2.62	0.010
Distancia	−0.9295	0.7279	−1.28	0.205
Garaje	35.810	7.638	4.69	0.000
Baños	23.315	9.025	2.58	0.011

S = 33.21 *R* cuadrada = 53.2%
R cuadrada ajustada = 50.3%

Análisis de la varianza

FUENTE	*GL*	*SS*	*MS*	*F*	*P*
Regresión	6	122676	20446	18.54	0.000
Error residual	98	108092	1103		
Total	104	230768			

a. Cada recámara adicional agrega $7 000 al precio de venta, cada pie cuadrado agrega $38, una alberca reduce el valor en $18 300, una cochera aumenta $35 800 el valor y cada milla que la casa está alejada del centro de la ciudad reduce $929 al precio de venta

b. El valor de *R* cuadrada es 0.532

c. La matriz de correlación es como sigue:

	Precio	Recámaras	Tamaño	Alberca	Distancia	Garaje
Recámaras	0.467					
Tamaño	0.371	0.383				
Alberca	0.294	0.005	0.201			
Distancia	−0.347	−0.153	−0.117	−0.139		
Garaje	0.526	0.234	0.083	0.114	−0.359	
Baños	0.382	0.329	0.024	0.055	−0.195	0.221

La variable independiente “garaje” tiene la correlación más fuerte con el precio. La distancia está inversamente relacionada, como se esperaba, y parece haber un problema con la correlación entre las variables independientes

d. Los resultados de la prueba global sugieren que algunas de las variables independientes tienen coeficientes de regresión netos diferentes a cero

e. Se puede eliminar “distancia”

f. He aquí la captura de pantalla de la nueva regresión:

Predictor	*Coeficiente*	*Error estándar del coeficiente*	*T*	*P*
Constante	17.01	35.24	.48	.630
Habitaciones	7.169	2.559	2.80	0.006
Tamaño	0.03919	0.01470	−2.67	0.009
Alberca	19.110	6.994	2.73	0.007
Garaje	38.847	7.281	5.34	0.000
Baños	24.624	8.995	2.74	0.007

(*continúa*)

(*continuación*)

```
S = 33.32  R cuadrada = 52.4%
R cuadrada ajustada = 50.0%

Análisis de la varianza
FUENTE            GL      SS      MS       F       P
Regresión          5  120877   24175   21.78   0.000
Error residual    99  109890    1110
Total            104  230768
```

Al revisar los valores p de los diversos coeficientes de regresión, todos son menores que 0.05. Deje todas las variables independientes

g. y **h.** El análisis de los residuos, que no se muestra, indica que la suposición de normalidad es razonable; además, no hay un patrón en las gráficas de los residuos y los valores ajustados de Y

35. **a.** La ecuación de regresión es

Mantenimiento = 102 + 5.94 Edad + 0.374 Millas − 11.8 indicador de gasolina

Cada año adicional de edad agrega $5.94 al costo de mantenimiento. Los autobuses de gasolina son más baratos de mantener que los de diésel por $11.80 por año

b. El coeficiente de determinación es 0.286, calculado por 65 135/227 692; 29% de la variación del costo de mantenimiento se explica por estas variables

c. La matriz de correlación es:

	Mantenimiento	Edad	Millas
Edad	0.465		
Millas	0.450	0.522	
Indicador de gasolina	−0.118	−0.068	0.025

Edad y millas tienen correlaciones moderadamente fuertes con el costo de mantenimiento. La correlación más alta entre las variables independientes es 0.522, entre edad y millas; esta es menos que 0.70, así que puede no ser un problema de multicolinealidad

d.

```
Análisis de varianza
FUENTE          GL      SS     MS      F      P
Regresión        3   65135  21712  10.15  0.000
Error residual  76  162558   2139
Total           79  227692
```

El valor p es cero; rechace la hipótesis nula de que todos los coeficientes son cero y afirme que al menos uno es importante

e.

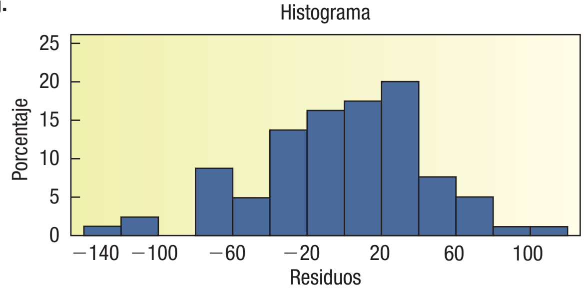
```
                                   Error
                              estándar del
Predictor      Coeficiente     coeficiente      T       P
Constante            102.3           112.9   0.91   0.368
Edad                 5.939           2.227   2.67   0.009
Millas              0.3740          0.1450   2.58   0.012
Indicador de
  gasolina          −11.80           10.99  −1.07   0.286
```

El valor p del indicador de gasolina es mayor a 0.10; considere eliminarlo

f. La ecuación de regresión condensada es

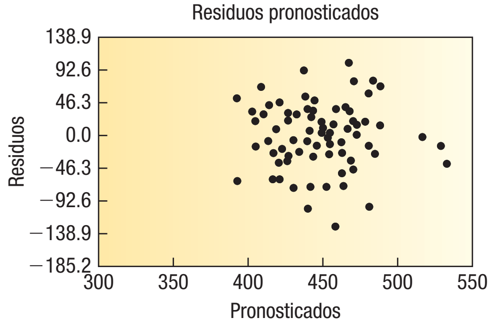
Mantenimiento = 106 + 6.17 Edad + 0.363 Millas

g.



La conjetura de normalidad parece ser realista

h.



Este diagrama parece ser aleatorio, y tener una varianza constante

CAPÍTULO 15

1. **a.** Se rechaza H_0 si $z > 1.65$

b. 1.09, calculado por $z = (0.75 - 0.70)/\sqrt{(0.70 \times 0.30)/100}$

c. No se rechaza H_0

3. **a.** $H_0: \pi \le 0.52$ $\quad H_1: \pi > 0.52$

b. H_0 se rechaza si $z > 2.33$

c. 1.62, determinado por $z = (0.5667 - 0.52)/\sqrt{(0.52 \times 0.48)/300}$

d. H_0 no se rechaza; no puede concluir que la proporción de hombres que manejan en Ohio Tumpike es mayor a 0.52

5. **a.** $H_0: \pi \ge 0.90$ $\quad H_1: \pi < 0.90$

b. H_0 se rechaza si $z < -1.28$

c. −2.67, que se determina por medio de $z = (0.82 - 0.90)/\sqrt{(0.90 \times 0.10)/100}$

d. Se rechaza H_0 porque menos de 90% de los clientes recibieron sus órdenes en menos de 10 minutos

7. **a.** Rechace H_0 si $z > 1.65$

b. 0.64, determinado por $p_c = \dfrac{70 + 90}{100 + 150}$

c. 1.61, determinado por

$$z = \frac{0.70 - 0.60}{\sqrt{[(0.64 \times 0.36)/100] + [(0.64 \times 0.36)/150]}}$$

d. No rechace H_0

9. **a.** $H_0: \pi_1 = \pi_2$ $\quad H_1: \pi_1 \ne \pi_2$

b. Rechace H_0 si $z < -1.96$ o bien $z > 1.96$

c. $p_c = \dfrac{24 + 40}{400 + 400} = 0.08$

d. −2.09, determinado por

$$z = \frac{0.06 - 0.10}{\sqrt{[(0.08 \times 0.92)/400] + [(0.08 \times 0.92)/400]}}$$

e. Rechace H_0; la proporción infestada no es la misma en ambos campos

11. $H_0: \pi_d \le \pi_r$ $\quad H_1: \pi_d > \pi_r$

Rechace H_0 si $z > 2.05$

$$p_c = \frac{168 + 200}{800 + 1\,000} = 0.2044$$

$$z = \frac{0.21 - 0.20}{\sqrt{\dfrac{(0.2044)(0.7956)}{800} + \dfrac{(0.2044)(0.7956)}{1\,000}}} = 0.52$$

No rechace H_0; no hay diferencia entre las proporciones de demócratas y republicanos que favorecen los estándares. Valor p = 0.3015

13. **a.** 3

b. 7.815

15. **a.** Rechace H_0 si $\chi^2 > 5.991$

b. $\chi^2 = \dfrac{(10 - 20)^2}{20} + \dfrac{(20 - 20)^2}{20} + \dfrac{(30 - 20)^2}{20} = 10.0$

c. Rechace H_0; las proporciones no son iguales

17. H_0: los resultados son iguales; H_1: los resultados no son iguales

Rechace H_0 si $\chi^2 > 9.236$

$$\chi^2 = \frac{(3 - 5)^2}{5} + \cdots + \frac{(7 - 5)^2}{5} = 7.60$$

No rechace H_0; no se puede rechazar que los resultados son iguales

19. H_0: no hay una diferencia entre las proporciones
H_1: hay una diferencia entre las proporciones
Rechace H_0 si $\chi^2 > 15.086$

$$\chi^2 = \frac{(47 - 40)^2}{40} + \cdots + \frac{(34 - 40)^2}{40} = 3.400$$

No rechace H_0; no hay diferencia entre las proporciones

21. **a.** Rechace H_0 si $\chi^2 > 9.210$

b. $\chi^2 = \frac{(30 - 24)^2}{24} + \frac{(20 - 24)^2}{24} + \frac{(10 - 12)^2}{12} = 2.50$

c. No rechace H_0

23. H_0: las proporciones son como se indicaron; H_1: las proporciones no son como se indicaron. Rechace H_0 si $\chi^2 > 11.345$

$$\chi^2 = \frac{(50 - 25)^2}{25} + \cdots + \frac{(160 - 275)^2}{275} = 115.22$$

Rechace H_0; las proporciones no son como se indicaron

25.

Número de clientes	Valores z	Área	Calculada por	f_e
Menos de 30	Menos de −1.58	0.0571	0.5000 − 0.4429	2.855
30 hasta 40	−1.58 hasta −0.51	0.2479	0.4429 − 0.1950	12.395
40 hasta 50	−0.51 hasta 0.55	0.4038	0.1950 + 0.2088	20.19
50 hasta 60	0.55 hasta 1.62	0.2386	0.4474 − 0.2088	11.93
60 o más	1.62 o más	0.0526	0.5000 − 0.4474	2.63

La primera y la última clase tienen frecuencias esperadas menores a 5; están combinadas con las clases adyacentes
H_0: la población de clientes sigue una distribución normal
H_1: la población de clientes no sigue una distribución normal
Rechace la hipótesis nula si *ji* cuadrada es mayor a 5.991

Número de clientes	Área	f_e	f_o	$f_e - f_o$	$(f_o - f_e)^2$	$[(f_o - f_e)^2]/f_e$
Menos de 40	0.3050	15.25	16	−0.75	0.5625	0.0369
40 hasta 50	0.4038	20.19	22	−1.81	3.2761	0.1623
50 o más	0.2912	14.56	12	2.56	6.5536	0.4501
Total	1.0000	50.00	50	0		0.6493

Como 0.6493 no es mayor a 5.991, no se rechaza la hipótesis nula; estos datos podrían provenir de una distribución normal

27. H_0: no hay relación entre los tamaños de la comunidad y la sección leída. H_1: hay una relación. Rechace H_0 si $\chi^2 > 9.488$

$$\chi^2 = \frac{(170 - 157.50)^2}{157.50} + \cdots + \frac{(88 - 83.62)^2}{83.62} = 7.340$$

No rechace H_0; no hay relación entre el tamaño de la comunidad y la sección leída

29. H_0: no hay relación entre las tasas de error y el tipo de artículo
H_1: hay una relación entre las tasas de error y el tipo de artículo
Rechace H_0 si $\chi^2 > 9.21$

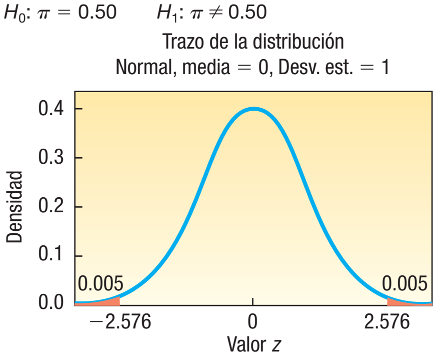
$$\chi^2 = \frac{(20 - 14.1)^2}{14.1} + \cdots + \frac{(225 - 225.25)^2}{225.25} = 8.033$$

No rechace H_0; no hay relación entre las tasas de error y el tipo de artículo

31. **a.** Esta es una situación binomial, en donde tanto el número medio de éxitos como el de fracasos son iguales a 23.5, calculado por 0.5(47)

b. H_0: $\pi = 0.50$ $\quad H_1$: $\pi \neq 0.50$

c.



Rechace H_0 si z no está entre −2.576 y 2.576

d. $z = \dfrac{\frac{31}{47} - 0.5}{\sqrt{0.5(1-0.5)/47}} = 2.19$. No se rechaza la hipótesis nula; estos datos no prueban que el giro de la moneda tenga un sesgo

e. El valor p es 0.0286, calculado por 2(0.5000 − 0.4857). Un valor así de extremo ocurrirá aproximadamente solo una vez en 35 ensayos con una moneda

33. H_0: $\pi \leq 0.60$ $\quad H_1$: $\pi > 0.60$
Rechace H_0 si $z > 2.33$

$$z = \frac{0.70 - 0.60}{\sqrt{\dfrac{0.60(0.40)}{200}}} = 2.89$$

Se rechaza H_0 porque la señorita Dennis está en lo correcto: más de 60% de las cuentas tiene una antigüedad superior a 3 meses.

35. H_0: $\pi \leq 0.44$ $\quad H_1$: $\pi > 0.44$
H_0 se rechaza si $z > 1.65$

$$z = \frac{0.480 - 0.44}{\sqrt{(0.44 \times 0.56)/1.000}} = 2.55$$

Se rechaza H_0 porque ha aumentado la proporción de personas que quieren ir a Europa

37. H_0: $\pi \leq 0.20$ $\quad H_1$: $\pi > 0.20$
Se rechaza H_0 si $z > 2.33$

$$z = \frac{(56/200) - 0.20}{\sqrt{(0.20 \times 0.80)/200}} = 2.83$$

Se rechaza H_0 porque más de 20% de los propietarios se muda durante un año en particular. Valor p = 0.5000 − 0.4977 = 0.0023

39. H_0: $\pi \geq 0.0008$ $\quad H_1$: $\pi < 0.0008$
Se rechaza H_0 si $z < -1.645$

$$z = \frac{0.0006 - 0.0008}{\sqrt{\dfrac{0.0008\,(0.9992)}{10\,000}}} = -0.707 \qquad \text{No se rechaza } H_0$$

Estos datos no prueban que haya una reducción en el rango de fatalidades

41. H_0: $\pi_1 \leq \pi_2$ $\quad H_1$: $\pi_1 > \pi_2$
Si $z > 2.33$, rechace H_0

$$p_c = \frac{990 + 970}{1\,500 + 1\,600} = 0.63$$

$$z = \frac{0.6600 - 0.60625}{\sqrt{\dfrac{0.63(0.37)}{1\,500} + \dfrac{0.63(0.37)}{1\,600}}} = 3.10$$

Rechace la hipótesis nula; no se puede concluir que es mayor la proporción de hombres que considera que la división es justa

43. H_0: $\pi_1 \leq \pi_2$ $\quad H_1$: $\pi_1 > \pi_2$ $\quad$ Rechace H_0 si $z > 1.65$

$$p_c = \frac{0.091 + 0.085}{2} = 0.88$$

$$z = \frac{0.091 - 0.085}{\sqrt{\dfrac{(0.088)(0.912)}{5\,000} + \dfrac{(0.088)(0.912)}{5\,000}}} = 1.059$$

No rechace la hipótesis nula; no ha existido un aumento en la proporción de condiciones de llamadas "buenas". El valor p 0.1446, calculado por 0.5000 − 0.3554. El incremento de los porcentajes ocurrirá por azar en uno de cada siete casos

45. H_0: $\pi_1 = \pi_2$ $\quad H_1$: $\pi_1 \neq \pi_2$
Rechace H_0 si z no está entre −1.96 y 1.96

$$p_c = \frac{100 + 36}{300 + 200} = 0.272$$

$$z = \frac{\dfrac{100}{300} - \dfrac{36}{200}}{\sqrt{\dfrac{(0.272)(0.728)}{300} + \dfrac{(0.272)(0.728)}{200}}} = 3.775$$

Se rechaza la hipótesis nula; hay diferencias entre las respuestas de los sexos

47. H_0: $\pi_s = 0.50$, $\pi_r = \pi_e = 0.25$
H_1: la distribución no es como se dio antes

Vuelta	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2/f_e$
Derecho	112	100	12	1.44
Derecha	48	50	−2	0.08
Izquierda	40	50	−10	2.00
Total	200	200		3.52

No se rechaza H_0 porque las proporciones son como se dieron en la hipótesis nula

49. H_0: no hay preferencia con respecto a las estaciones de TV
H_1: hay preferencia con respecto a las estaciones de TV
$gl = 3 - 1 = 2$. Se rechaza H_0 si $\chi^2 > 5.991$

Estación TV	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
WNAE	53	50	3	9	0.18
WRRN	64	50	14	196	3.92
WSPD	33	50	−17	289	5.78
	150	150	0		9.88

Se rechaza H_0 porque hay una preferencia por las estaciones de televisión

51. H_0: $\pi_n = 0.21$, $\pi_m = 0.24$, $\pi_s = 0.35$, $\pi_w = 0.20$
H_1: la distribución no es como se dio
Rechace H_0 si $\chi^2 > 11.345$

Región	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2/f_e$
Noreste	68	84	−16	3.0476
Oeste medio	104	96	8	0.6667
Sur	155	140	15	1.6071
Oeste	73	80	−7	0.6125
Total	400	400	0	5.9339

No se rechaza H_0 porque la distribución del orden de los destinos refleja la población

53. H_0: las proporciones son las mismas
H_1: las proporciones no son las mismas
Rechace H_0 si $\chi^2 > 16.919$

f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
44	28	16	256	9.143
32	28	4	16	0.571
23	28	−5	25	0.893
27	28	−1	1	0.036
23	28	−5	25	0.893
24	28	−4	16	0.571
31	28	3	9	0.321
27	28	−1	1	0.036
28	28	0	0	0.000
21	28	−7	49	1.750
				14.214

No rechace H_0; los dígitos siguen una distribución uniforme

55.

Salario por hora	f	M	fM	$M - x$	$(M - x)^2$	$f(M - x)^2$
$5.50 hasta 6.50	20	6	120	−2.222	4.938	98.8
6.50 hasta 7.50	24	7	168	−1.222	1.494	35.9
7.50 hasta 8.50	130	8	1 040	−0.222	0.049	6.4
8.50 hasta 9.50	68	9	612	0.778	0.605	41.1
9.50 hasta 10.50	28	10	280	1.778	3.161	88.5
Total	270		2 220			270.7

La media muestral es 8.222, calculada por 2 220/270. La desviación estándar de la muestra es 1.003, calculada como la raíz cuadrada de 270.7/269
H_0: la población de salarios sigue una distribución normal
H_1: la población de salarios no sigue una distribución normal
Rechace la nula si $\chi^2 > 4.605$

Salario	Valores z	Área	Calculada por	f_e	f_o	$f_e - f_o$	$(f_o - f_e)^2$	$[(f_o - f_e)^2]/f_e$
Menor a $6.50	Menor a −1.72	0.0427	0.5000 − 0.4573	11.529	20	−8.471	71.7578	6.2241
6.50 hasta 7.50	−1.72 hasta −0.72	0.1931	0.4573 − 0.2642	52.137	24	28.137	791.6908	15.1848
7.50 hasta 8.50	−0.72 hasta 0.28	0.3745	0.2642 + 0.1103	101.115	130	−28.885	834.3432	8.2514
8.50 hasta 9.50	0.28 hasta 1.27	0.2877	0.3980 − 0.1103	77.679	68	9.679	93.6830	1.2060
9.50 o más	1.27 o más	0.1020	0.5000 − 0.3980	27.54	28	−0.46	0.2116	0.0077
Total		1.0000		270	270	0		30.874

Como 30.874 es mayor a 4.605, se rechaza la hipótesis nula, es decir, no hay una distribución normal

57. H_0: el género y la actitud hacia el déficit no están relacionados
H_1: el género y la actitud hacia el déficit están relacionados
Rechace H_0 si $\chi^2 > 5.991$

$$\chi^2 = \frac{(244 - 292.41)^2}{292.41} + \frac{(194 - 164.05)^2}{164.05} + \frac{(68 - 49.53)^2}{49.53} + \frac{(305 - 256.59)^2}{256.59} + \frac{(114 - 143.95)^2}{143.95} + \frac{(25 - 43.47)^2}{43.47} = 43.578$$

Como $43.578 > 5.991$, rechace H_0; la posición de una persona respecto al déficit está influenciada por su género

59. H_0: si se hace un reclamo y la edad no están relacionados
H_1: si se hace un reclamo y la edad están relacionados
Rechace H_0 si $\chi^2 > 7.815$

$$\chi^2 = \frac{(170 - 203.33)^2}{203.33} + \cdots + \frac{(24 - 35.67)^2}{35.67} = 53.639$$

Rechace H_0; la edad está relacionada con hacer un reclamo

61. H_0: $\pi_{BL} = \pi_O = 0.23$, $\pi_Y = \pi_G = 0.15$, $\pi_{BR} = \pi_R = 0.12$. H_1: Las proporciones no son como se dieron; rechace H_0 si $\chi^2 > 15.086$

Color	f_o	f_e	$(f_o - f_e)^2/f_e$
Azul	12	16.56	1.256
Café	14	8.64	3.325
Amarillo	13	10.80	0.448
Rojo	14	8.64	3.325
Naranja	7	16.56	5.519
Verde	12	10.80	0.133
Total	72		14.006

No rechace H_0; la distribución del color concuerda con la información del fabricante

63. **a.** H_0: el salario y las victorias no están relacionados
H_1: el salario y las victorias están relacionados
Rechace H_0 si $\chi^2 > 3.84$

	Salario		
Victoria	Mitad baja	Mitad alta	Total
No	8	5	13
Sí	7	10	17
Total	15	15	

$$\chi^2 = \frac{(8 - 6.5)^2}{6.5} + \frac{(5 - 6.5)^2}{6.5} + \frac{(7 - 8.5)^2}{8.5} + \frac{(10 - 8.5)^2}{8.5} = 1.22$$

No rechace H_0; concluya que el salario y las victorias pueden no estar relacionados

CAPÍTULO 16

1. **a.** Si el número de pulsos (éxitos) en la muestra es 9 o mayor, rechace H_0
b. Rechace H_0 debido a que la probabilidad acumulada asociada con nueve o más éxitos (0.073) no sobrepasa el nivel de significancia (0.10)

3. **a.** H_0: $\pi \leq 0.50$; H_1: $\pi > 0.50$; $n = 10$

b. Se rechaza H_0 si hay nueve o más signos de más. Un "+" representa una pérdida

c. Rechace H_0; es un programa eficaz, ya que nueve personas bajaron de peso

5. a. H_0: $\pi \leq 0.50$ (no hay cambio de peso)
H_1: $\pi > 0.50$ (hay una pérdida de peso)

b. Rechace H_0 si $z > 1.65$

c. $z = \dfrac{(32 - 0.50) - 0.50(45)}{0.50\sqrt{45}} = 2.68$

d. Rechace H_0; el programa de pérdida de peso es eficaz

7. H_0: $\pi \leq 0.50$, H_1: $\pi > 0.50$. Se rechaza H_0 si $z > 2.05$

$$z = \frac{42.5 - 40.5}{4.5} = 0.44$$

Como $0.44 < 2.05$, no rechace H_0; no hay preferencia

9. a. H_0: mediana $\leq$ \$81 500; H_1: mediana $>$ \$81 500

b. Se rechaza H_0 si $z > 1.65$

c. $z = \dfrac{170 - 0.50 - 100}{7.07} = 9.83$

Se rechaza H_0 porque el ingreso mediano es mayor que \$81 500

11.

Pareja	Diferencia	Rango
1	550	7
2	190	5
3	250	6
4	−120	3
5	−70	1
6	130	4
7	90	2

Sumas: −4, +24; por lo tanto, $T = 4$ (la menor de las dos sumas). Del apéndice B.8, nivel de significancia 0.05, $n = 7$, el valor crítico es 3; como T de $4 > 3$, no rechace H_0 (prueba de una cola). No hay diferencia entre los pies cuadrados; las parejas de profesionales no viven en casas más grandes

13. a. H_0: la producción de ambos sistemas es la misma
H_1: la producción utilizando el método de Mump es mayor

b. Se rechaza H_0 si $T \leq 21$, $n = 13$

c. Los cálculos de los primeros tres empleados son:

Empleado	Edad	Mump	d	Rango	R^+	R^-
A	60	64	4	6	6	
B	40	52	12	12.5	12.5	
C	59	58	−1	2		2

La suma de los rangos negativos es 6.5 (menor que 21), entonces se rechaza H_0 porque la producción empleando el método de Mump es mayor

15. H_0: las distribuciones son iguales. H_1: las distribuciones no son iguales. Rechace H_0 si $z < -1.96$ o bien $z > 1.96$

A		B	
Calificación	Rango	Calificación	Rango
38	4	26	1
45	6	31	2
56	9	35	3
57	10.5	42	5
61	12	51	7
69	14	52	8
70	15	57	10.5
79	16	62	13
	86.5		49.5

$$z = \frac{86.5 - \dfrac{8(8+8+1)}{2}}{\sqrt{\dfrac{8(8)(8+8+1)}{2}}} = 1.943$$

No se rechaza H_0; no hay diferencia en ambas poblaciones

17. H_0: las distribuciones son iguales. H_1: la distribución del campus es a la derecha. Rechace H_0 si $z > 1.65$

Campus		En línea	
Edad	Rango	Edad	Rango
26	6	28	8
42	16.5	16	1
65	22	42	16.5
38	13	29	9.5
29	9.5	31	11
32	12	22	3
59	21	50	20
42	16.5	42	16.5
27	7	23	4
41	14	25	5
46	19		94.5
18	2		
	158.5		

$$z = \frac{158.5 - \dfrac{12(12+10+1)}{2}}{\sqrt{\dfrac{12(10)(12+10+1)}{12}}} = 1.35$$

No se rechaza H_0 porque no hay diferencia en las distribuciones

19. ANOVA requiere que tenga dos o más poblaciones. Los datos están a nivel de intervalo o de razón, las poblaciones están normalmente distribuidas, y las desviaciones estándar de las poblaciones son iguales. Kruskal-Wallis solo requiere datos a nivel ordinal, y no se hacen suposiciones respecto a la forma de las poblaciones

21. a. H_0: las tres distribuciones de la población son iguales. H_1: no todas las distribuciones son iguales

b. Rechace H_0 si $H > 5.991$

c.

Muestra 1 Rango	Muestra 2 Rango	Muestra 3 Rango
8	5	1
11	6.5	2
14.5	6.5	3
14.5	10	4
16	12	9
64	13	19
	53	

$$H = \frac{12}{16(16+1)}\left[\frac{(64)^2}{5} + \frac{(53)^2}{6} + \frac{(19)^2}{5}\right] - 3(16 + 1)$$
$$= 59.98 - 51 = 8.98$$

d. Rechace H_0 debido a que $8.98 > 5.991$. Las tres distribuciones no son iguales

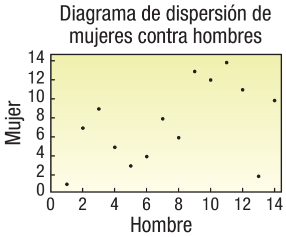
23. H_0: las distribuciones de las duraciones de vida son iguales
H_1: las distribuciones de las duraciones de vida no son iguales
Se rechaza H_0 si $H > 9.210$

Sal		Dulce		Otros	
Horas	Rango	Horas	Rango	Horas	Rango
167.3	3	160.6	1	182.7	13
189.6	15	177.6	11	165.4	2
177.2	10	185.3	14	172.9	7
169.4	6	168.6	4	169.2	5
180.3	12	176.6	9	174.7	8
	46		39		35

$$H = \frac{12}{15(16)}\left[\frac{(46)^2}{5} + \frac{(39)^2}{5} + \frac{(35)^2}{5}\right] - 3(16) = 0.62$$

No se rechaza H_0 porque no hay diferencia entre las tres distribuciones

25. a.



b.

Hombre	Mujer	d	d^2
4	5	−1	1
6	4	2	4
7	8	−1	1
2	7	−5	25
12	11	1	1
8	6	2	4
5	3	2	4
3	9	−6	36
13	2	11	121
14	10	4	16
1	1	0	0
9	13	−4	16
10	12	−2	4
11	14	−3	9
		Total	242

$$r_s = 1 - \frac{6(242)}{14(14^2 - 1)} = 0.47$$

c. H_0: no hay correlación entre los rangos
H_1: hay una correlación positiva entre los rangos
Rechace H_0 si $t > 1.782$

$$t = 0.47\sqrt{\frac{14-2}{1-(0.47)^2}} = 1.84$$

Se rechaza H_0 porque la correlación de rangos entre la población es positiva; a los maridos y las esposas en general les gustan los mismos programas

27.

Representante	Ventas	Rango	Rango de entrenamiento	d	d^2
1	319	8	8	0	0
2	150	1	2	1	1
3	175	2	5	3	9
4	460	10	10	0	0
5	348	9	7	−2	4
6	300	6.5	1	5.5	30.25
7	280	5	6	1	1
8	200	4	9	5	25
9	190	3	4	1	1
10	300	6.5	3	−3.5	12.25
					83.50

a. $r_s = 1 - \frac{6(83.5)}{10(10^2 - 1)} = 0.494$

Una correlación positiva moderada

b. H_0: no hay correlación entre los rangos. H_1: hay correlación positiva entre los rangos. Rechace H_0 si $t > 1.860$

$$t = 0.494\sqrt{\frac{10-2}{1-(0.494)^2}} = 1.607$$

No se rechaza H_0 porque la correlación de los rangos entre la población podría ser 0

29. H_0: $\pi = 0.50$. H_1: $\pi \neq 0.50$. Utilice un paquete de software para desarrollar la distribución de probabilidad normal para $n = 19$ y $\pi = 0.50$. Se rechaza H_0 si hay 5 o menos signos "+" o bien 14 o más. El total de 12 signos "+" cae en la región de aceptación; no se rechaza H_0; no hay preferencia entre ambos programas

31. H_0: $\pi = 0.50$ H_1: $\pi \neq 0.50$
Se rechaza H_0 si hay 12 o más, o 3 o menos signos de menos. Como solo hay 8 signos de más, no se rechaza H_0 porque no hay preferencia con respecto a las dos marcas de componentes

33. H_0: $\pi = 0.50$; H_1: $\pi \neq 0.50$. Rechace H_0 si $z > 1.96$ o bien $z < -1.96$

$$z = \frac{159.5 - 100}{7.071} = 8.415$$

Rechace H_0; hay una diferencia entre las preferencias por ambos tipos de jugo de naranja

35. H_0: las tasas son iguales; H_1: las tasas no son iguales
Se rechaza H_0 si $H > 5.991$. $H = 0.082$; no rechace H_0

37. H_0: las poblaciones son las mismas. H_1: las poblaciones difieren
Rechace H_0 si $H > 7.815$. $H = 14.30$; rechace H_0

39. $r_s = 1 - \frac{6(78)}{12(12^2 - 1)} = 0.727$

H_0: no hay correlación entre los rangos de los entrenadores y de los cronistas deportivos
H_1: hay una correlación positiva entre los rangos de los entrenadores y de los cronistas deportivos
Rechace H_0 si $t > 1.812$

$$t = 0.727\sqrt{\frac{12-2}{1-(0.727)^2}} = 3.348$$

Se rechaza H_0 porque hay una correlación positiva entre los escritores deportivos y los entrenadores

41. a. H_0: no hay diferencia entre las distribuciones de los precios de venta en los cinco municipios. H_1: hay una diferencia entre las distribuciones de los precios de venta de los cinco municipios; se rechaza H_0 si H es mayor que 9.488. El valor calculado de H es 4.70, por lo que se rechaza la hipótesis nula; los datos de la muestra no sugieren una diferencia entre las distribuciones de los precios de venta

b. H_0: no hay diferencia entre las distribuciones de los precios de venta dependiendo del número de recámaras. H_1: hay una diferencia entre las distribuciones de los precios de venta dependiendo del número de recámaras; se rechaza H_0 si H es mayor que 9.448. El valor calculado de H es 16.34, por lo que se rechaza la hipótesis nula; los datos de la muestra indican que hay una diferencia entre las distribuciones de los precios de venta con base en el número de recámaras. Nota: Combine seis o más en un solo grupo

c. H_0: no hay diferencia entre las distribuciones de las distancias desde el centro de la ciudad dependiendo de si la casa tiene alberca o no. H_1: hay una diferencia entre las distribuciones de las distancias desde el centro de la ciudad dependiendo de si la casa tiene una alberca o no; se rechaza H_0 si H es mayor que 3.84. El valor calculado de H es 3.37, por lo que no se rechaza la hipótesis nula; los datos de la muestra no sugieren una diferencia entre las distribuciones de las distancias

43. a. H_0: las distribuciones de los costos de mantenimiento de todos los fabricantes son las mismas
H_1: las distribuciones de los costos no son iguales
Rechace H_0 si $H > 5.991$

$$H = \frac{12}{80(81)}\left[\frac{(1\,765)^2}{47} + \frac{(972)^2}{25} + \frac{(503)^2}{8}\right] - 3(81) = 8.29$$

Se rechaza H_0 porque hay diferencia entre el costo de mantenimiento de los tres fabricantes de autobuses

b. H_0: las distribuciones de los costos de mantenimiento son iguales para las capacidades de los autobuses
H_1: las distribuciones de los costos no son iguales
Rechace H_0 si $H > 7.815$

$$H = \frac{12}{80(81)}\left[\frac{(96.5)^2}{4} + \frac{(332.5)^2}{7} + \frac{(388.5)^2}{9} + \frac{(2\,422.5)^2}{60}\right] - 3(81) = 2.74$$

No se rechaza H_0 porque no hay diferencia entre los costos de mantenimiento de las cuatro distintas capacidades

c. H_0: las distribuciones son iguales
H_1: las distribuciones son diferentes
Rechace H_0 si $z < -1.96$ o $z > 1.96$

$$W = \frac{2\,252 - \frac{53(53+27+1)}{2}}{\sqrt{\frac{(53)(27)(53+27+1)}{12}}} = 1.07$$

No rechace H_0; las distribuciones podrían ser iguales

CAPÍTULO 17

1. 114.6, calculado por ($19 989/$17 446)(100)
123.1, calculado por ($21 468/$17 446)(100)
124.3, calculado por ($21 685/$17 446)(100)
91.3, calculado por ($15 922/$17 446)(100)
105.3, calculado por ($18 375/$17 446)(100)
314.2, calculado por ($54 818/$17 446)(100)
475.4, calculado por ($83 040/$17 446)(100)

3. Las ventas medias por los primeros tres años son $(486.6 + 506.8 + 522.2)/3, o $505.2
2011: 90.4, calculado por (456.6/505.2)(100)
2012: 85.8, calculado por (433.3/505.2)(100)
Las ventas netas descendieron en 9.5 y 14.2%, respectivamente.

5. **a.** $P_t = \frac{3.35}{2.49}(100) = 134.54 \qquad P_s = \frac{4.49}{3.49}(100) = 136.47$

$P_c = \frac{4.19}{1.59}(100) = 263.52 \qquad P_a = \frac{2.49}{1.79}(100) = 139.11$

b. $P = \frac{14.52}{9.16}(100) = 158.52$

c. $P = \frac{\$3.35(6) + 4.49(4) + 4.19(2) + 2.49(3)}{\$2.49(6) + 3.29(4) + 1.59(2) + 1.79(3)}(100) = 147.1$

d. $P = \frac{\$3.35(6) + 4.49(5) + 4.19(3) + 2.49(4)}{\$2.49(6) + 3.29(5) + 1.59(3) + 1.79(4)}(100) = 150.2$

e. $I = \sqrt{(147.1)(150.2)} = 148.64$

7. **a.** $P_W = \frac{0.10}{0.07}(100) = 142.9 \qquad P_C = \frac{0.03}{0.04}(100) = 75.0$

$P_S = \frac{0.15}{0.15}(100) = 100 \qquad P_H = \frac{0.10}{0.08}(100) = 125.0$

b. $P = \frac{0.38}{0.34}(100) = 111.8$

c.
$$P = \frac{0.10(17\,000) + 0.03(125\,000) + 0.15(40\,000) + 0.10(62\,000)}{0.07(17\,000) + 0.04(125\,000) + 0.15(40\,000) + 0.08(62\,000)} \times (100) = 102.92$$

d.
$$P = \frac{0.10(20\,000) + 0.03(130\,000) + 0.15(42\,000) + 0.10(65\,000)}{0.07(20\,000) + 0.04(130\,000) + 0.15(42\,000) + 0.08(65\,000)} \times (100) = 103.32$$

e. $I = \sqrt{102.92(103.32)} = 103.12$

9. $V = \frac{\$3.8(64) + 7.9(2) + 6.9(10\,780) + 6.4(220)}{\$1.81(116) + 3.56(2) + 2.32(8\,967) + 2.72(227)}(100) = 351.46$

11. **a.** $I = \frac{6.8}{5.3}(0.20) + \frac{362.26}{265.88}(0.40) + \frac{125.0}{109.6}(0.25) + \frac{622\,864}{529\,917}(0.15) = 1.263$

El índice es 126.3

b. La actividad bursátil aumentó 26.3% de 2000 a 2014

13. x = ($89 673)/2.32531 = $38 564
El salario “real” aumentó $38 564 − $19 800 = $18 764

15.

Año	Tinora	Índice Tinora	Índice nacional
2000	$28 650	100.0	100
2009	33 972	118.6	122.5
2013	37 382	130.5	136.9

Los maestros de Tinora recibieron aumentos menores que el promedio nacional

17. El índice (2000 = 100) de años seleccionados es:

Año	Índice	Año	Índice	Año	Índice	Año	Índice
2001	114.5	2004	160.4	2007	187.4	2010	170.0
2002	129.7	2005	163.9	2008	186.6	2011	166.9
2003	146.0	2006	172.0	2009	178.4	2012	172.3

Las ventas domésticas se incrementaron casi 90% entre 2000 y 2007 y después se redujeron cerca de 10% en los siguientes cinco años

19. El índice (2000 = 100) de años seleccionados es:

Año	Índice	Año	Índice	Año	Índice	Año	Índice
2001	105.4	2004	165.1	2007	241.7	2010	271.0
2002	116.8	2005	186.7	2008	265.2	2011	304.5
2003	139.9	2006	198.6	2009	261.5	2012	315.2

Las ventas internacionales casi se triplicaron entre 2000 y 2012

21. El índice (2000 = 100) de años seleccionados es:

Año	Índice	Año	Índice	Año	Índice	Año	Índice
2001	100.9	2004	108.9	2007	118.1	2010	113.0
2002	107.3	2005	114.6	2008	117.6	2011	116.8
2003	109.6	2006	121.1	2009	114.5	2012	117.6

El número de empleados aumentó casi 15% entre 2000 y 2005, permaneciendo en el mismo nivel desde 2012

23. El índice (2004 = 100) de años seleccionados es:

Año	2005	2006	2007	2008	2009	2010	2011	2012
Índice	113.4	117.2	125.4	132.1	136.6	111.6	109.9	109.9

El ingreso aumentó casi 37%, y volvió casi a los niveles de 2005

25. El índice (2004 = 100) de años seleccionados es:

Año	2005	2006	2007	2008	2009	2010	2011	2012
Índice	94.5	97.2	98.2	100.6	99.4	93.5	88.3	92.6

El número de empleados disminuyó alrededor de 7% entre 2008 y 2012

27. $P_{ma} = \frac{2.00}{0.81}(100) = 246.91 \qquad P_{sh} = \frac{1.88}{0.84}(100) = 223.81$

$P_{mi} = \frac{2.89}{1.44}(100) = 200.69 \qquad P_{po} = \frac{3.99}{2.91}(100) = 137.11$

29. $P = \frac{\$2.00(18) + 1.88(5) + 2.89(70) + 3.99(27)}{\$0.81(18) + 0.84(5) + 1.44(70) + 2.91(27)}(100) = 179.37$

31. $I = \sqrt{179.37(178.23)} = 178.80$

33. $P_R = \frac{0.60}{0.50}(100) = 120 \qquad P_S = \frac{0.90}{1.20}(100) = 75.0$

$P_W = \frac{1.00}{0.85}(100) = 117.65$

35. $P = \frac{0.60(320) + 0.90(110) + 1.00(230)}{0.50(320) + 1.20(110) + 0.85(230)}(100) = 106.87$

37. $P = \sqrt{(106.87)(106.04)} = 106.45$

39. $P_C = \frac{0.05}{0.06}(100) = 83.33 \qquad P_C = \frac{0.12}{0.10}(100) = 120$

$P_P = \frac{0.18}{0.20}(100) = 90 \qquad P_E = \frac{0.015}{0.15}(100) = 100$

41. $P = \frac{0.05(2\,000) + 0.12(200) + 0.18(400) + 0.15(100)}{0.06(2\,000) + 0.10(200) + 0.20(400) + 0.15(100)}(100) = 89.79$

43. $I = \sqrt{(89.79)(91.25)} = 90.52$

45. $P_A = \frac{0.76}{0.287}(100) = 264.81 \qquad P_N = \frac{2.50}{0.17}(100) = 1\,470.59$

$P_P = \frac{26.00}{3.18}(100) = 817.61 \qquad P_P = \frac{490}{133}(100) = 368.42$

47.
$$P = \frac{0.76(1\,000) + 2.50(5\,000) + 26(60\,000) + 490(500)}{0.287(1\,000) + 0.17(5\,000) + 3.18(60\,000) + 133(500)} \times (100) = 703.56$$

49. $I = \sqrt{(703.56)(686.58)} = 659.02$

51. $I = 100\left[\frac{1\,971.0}{1\,159.0}(0.20) + \frac{91}{87}(0.10) + \frac{114.7}{110.6}(0.40) + \frac{1\,501}{1\,214}(0.30)\right] = 123.05$

La economía aumentó 23.05% de 1996 a 2013

53. Febrero: $I = 100\left[\frac{6.8}{8.0}(0.40) + \frac{23}{20}(0.35) + \frac{303}{300}(0.25)\right] = 99.50$

Marzo: $I = 100\left[\frac{6.4}{8.0}(0.40) + \frac{21}{20}(0.35) + \frac{297}{300}(0.25)\right]$
$= 93.50$

55. En 1995: $1 876 466, calculado por $2 400 000/1.279
En 2012: $1 799 486, calculado por $3 500 000/1.945

CAPÍTULO 18

1. Los promedios móviles ponderados son: 31 584.8, 33 088.9, 34 205.4, 34 899.8, 35 155.0, 34 887.1

3. La ecuación de regresión es: $\hat{y} = 8\,842 - 88.1273t$.
Para 2014, $t = 12$ y $\hat{y} = 8\,842 - 88.1273(12) = 7\,784.47$

5. $\hat{y} = 1.30 + 0.90t$
$\hat{y} = 1.30 + 0.90(7) = 7.6$

7. **a.** $b = \dfrac{5.274318 - (1.390087)(15)/5}{55 - (15)^2/5}$

$= \dfrac{1.104057}{10} = 0.1104057$

$a = \dfrac{1.390087}{5} - 0.1104057\left(\dfrac{15}{5}\right) = -0.0531997$

b. 28.95%, determinado por 1.28945 − 1.0
c. $\hat{y} = -0.0531997 + 0.1104057t$ para 2010, $t = 8$
$\hat{y} = -0.0531997 + 0.1104057(8) = 0.8300459$
Antilogaritmo de 0.8300459 = 6.76

9.

Timestre	Componente promedio del índice estacional	Índice estacional
1	0.6859	0.6911
2	1.6557	1.6682
3	1.1616	1.1704
4	0.4732	0.4768

11.

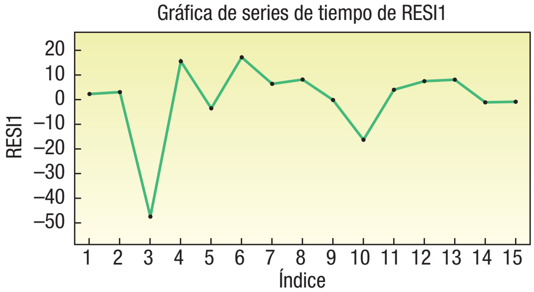
t	Pares estimados (millones)	Índice estacional	Predicción trimestral (millones)
21	40.05	110.0	44.055
22	41.80	120.0	50.160
23	43.55	80.0	34.840
24	45.30	90.0	40.770

13. $\hat{y} = 5.1658 + 0.37805t$. A continuación se muestran algunas estimaciones de ventas:

Estimación	Índice	Ajustado estacional
10.080	0.6911	6.966
10.458	1.6682	17.446
10.837	1.1704	12.684
11.215	0.4768	5.343

15. **a.** Los residuos ordenados son: 2.61, 2.83, −48.50, 15.50, −3.72, 17.17, 6.39, 7.72, −0.41, −16.86, 3.81, 7.25, 8.03, −1.08 y −0.75

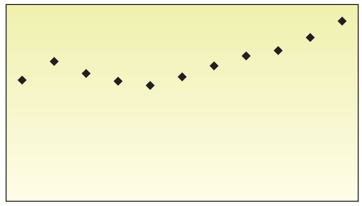
Gráfica de series de tiempo de RESI1



b. Hay 2 variables independientes (k) y el tamaño de la muestra (n) es 15. Con el nivel de significancia 0.05, el valor superior es 1.54; como el valor calculado del estadístico de Durbin-Watson es 2.48, que está arriba del límite superior, no se rechaza la hipótesis nula. No hay autocorrelación entre estos residuos

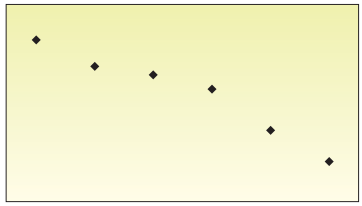
17. **a.** $\hat{y} = 18\,000 - 400t$, asumiendo que la recta inicia en 18 000 en 1990 y disminuye a 10 000 en 2010
b. 400
c. 8 000, calculado por 18 000 = 400(25)

19. **a.**



b. $\hat{y} = 1.00455 + 0.04409t$, utilizando $t = 1$ para 2005
c. En 2007, $\hat{y} = 1.18091$ y para 2012, $\hat{y} = 1.40136$
d. En 2019, $\hat{y} = 1.70999$
e. Cada activo cambió 0.044 veces

21. **a.**

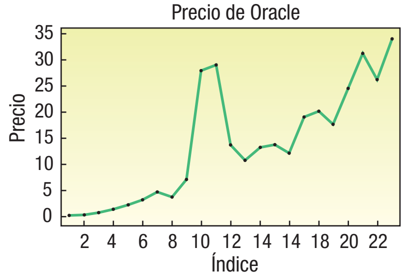


b. $\hat{y} = 49.140 - 2.9829t$
c. En 2011, $\hat{y} = 40.1913$. Para 2014, $\hat{y} = 31.2426$
d. En 2017, $\hat{y} = 22.2939$
e. El número de empleados disminuyó a una tasa de 2 983 por año

23. **a.** $\text{Log}\,\hat{y} = 0.790231 + 0.113669t$
b. $\text{Log}\,\hat{y} = 0.9039$, calculado por 0.790231 + 0.113669(1), el antilogaritmo es 8.015
$\text{Log}\,\hat{y} = 1.699583$, calculado por 0.790231 + 0.113699(8), el antilogaritmo es 50.071
c. 29.92, que es el antilogaritmo de 0.113669 menos 1
d. $\text{Log}\,\hat{y} = 2.154258$, el antilogaritmo es 142.65

25. **a.**

Precio de Oracle



b. Las ecuaciones son $\hat{y} = -2.58 + 1.37t$ y/o $\log \hat{y} = -0.133 + 0.083t$.
Utilizando el logaritmo, la ecuación parece mejor porque R^2 es mayor
c. $\text{Log}\,\hat{y} = -0.133 + 0.0831(4) = 0.1994$, el antilogaritmo es 1.5827
$\text{Log}\,\hat{y} = -0.133 + 0.0831(9) = 0.6149$, el antilogaritmo es 4.12
d. $\text{Log}\,\hat{y} = -0.133 + 0.0831(26) = 2.0276$, el antilogaritmo es 106.56. ¡Razonable si el precio se eleva a nivel histórico!
e. La tasa anual de incremento es 21%, calculado por el antilogaritmo de 0.0831 menos 1

27. **a.** Julio 87.5; agosto 92.9; septiembre 99.3; octubre 109.1
b.

Mes	Total	Media	Corregida
Jul.	348.9	87.225	86.777
Ago.	368.1	92.025	91.552
Sep.	395.0	98.750	98.242
Oct.	420.4	105.100	104.560
Nov.	496.2	124.050	123.412
Dic.	572.3	143.075	142.340
Ene.	333.5	83.375	82.946
Feb.	297.5	74.375	73.993
Mar.	347.3	86.825	86.379
Abr.	481.3	120.325	119.707
May.	396.2	99.050	98.541
Jun.	368.1	92.025	91.552
		1 206.200	

Corrección = 1 200/1 206.2 = 0.99486

c. Abril, noviembre y diciembre son periodos de ventas altas, en tanto que las ventas de febrero son las más bajas

Nota: La solución de los ejercicios 29 a 33 puede variar debido al redondeo y al paquete de software empleado

29. a.

Índice estacional por trimestre

Trimestre	Componente promedio del IE	Índice estacional
1	0.5014	0.5027
2	1.0909	1.0936
3	1.7709	1.7753
4	0.6354	0.6370

b. La producción es mayor en el tercer trimestre, pues es 77.5% superior a la del trimestre promedio. El segundo trimestre también está por encima del promedio, el primero y el cuarto trimestres están muy abajo del promedio, con el primer trimestre en casi 50% de un trimestre típico

31. a. A continuación se muestran los índices estacionales de un juego en paquete; recuerde que el periodo 1 en realidad es julio, ya que los datos inician ese mes:

Periodo	Índice	Periodo	Índice
1	0.19792	7	0.26874
2	0.25663	8	0.63189
3	0.87840	9	1.67943
4	2.10481	10	2.73547
5	0.77747	11	1.67903
6	0.18388	12	0.60633

Observe que el periodo 4 (octubre) y el periodo 10 (abril) son superiores al doble del promedio

b. Los índices estacionales del juego sin paquete son:

Periodo	Índice	Periodo	Índice
1	1.73270	7	0.23673
2	1.53389	8	0.69732
3	0.94145	9	1.00695
4	1.29183	10	1.13226
5	0.66928	11	0.98282
6	0.52991	12	1.24486

Estos índices son más constantes. Observe los valores muy bajos en los periodos 6 (diciembre) y 7 (enero)

c. Los índices estacionales del juego total son:

Periodo	Índice	Periodo	Índice
1	0.63371	7	0.25908
2	0.61870	8	0.65069
3	0.89655	9	1.49028
4	1.86415	10	2.28041
5	0.74353	11	1.48235
6	0.29180	12	0.78876

Estos índices muestran tanto los picos en octubre (periodo 4) y abril (periodo 10) como los valles en diciembre (periodo 6) y enero (periodo 7)

d. El juego en paquete es relativamente más alto en abril. El juego que no está en paquete es relativamente alto en julio; como 70% del juego total proviene del juego en paquete, el juego total es muy similar al juego en paquete

33.

Índice estacional por trimestre

Trimestre	Componente promedio del IE	Índice estacional
1	1.1962	1.2053
2	1.0135	1.0212
3	0.6253	0.6301
4	1.1371	1.1457

La ecuación de regresión es: $\hat{y} = 43.611 + 7.21153t$

Periodo	Visitantes	Índice	Predicción
29	252.86	1.2053	304.77
30	260.07	1.0212	265.58
31	267.29	0.6301	168.42
32	274.50	1.1457	314.50

En 2013 hubo un total de 928 visitantes. Un aumento de 10% en 2014 significa que habrá 1 021 visitantes; las estimaciones trimestrales son $1\,021/4 = 255.25$ visitantes por trimestre

Periodo	Visitantes	Índice	Predicción
Invierno	255.25	1.2053	307.65
Primavera	255.25	1.0212	260.66
Verano	255.25	0.6301	160.83
Otoño	255.25	1.1457	292.44

La aproximación de regresión es probablemente superior debido a que se considera la tendencia

35. La línea de tendencia lineal es $\hat{y} = 0.469 + 0.114t$ y la línea de tendencia logarítmica es $\log \hat{y} = -0.0854 + 0.0255t$

Utilizando el logaritmo, la ecuación parece mejor, porque R^2 es mayor; los años 2013 y 2014 se codificarán $t = 24$ y $t = 25$, respectivamente

Para 2013: $\log \hat{y} = -0.0854 + 0.0255(24) = 0.5266$, y el antilogaritmo es 3.36

Para 2014: $\log \hat{y} = -0.0854 + 0.0255(25) = 0.5521$, y el antilogaritmo es 3.57

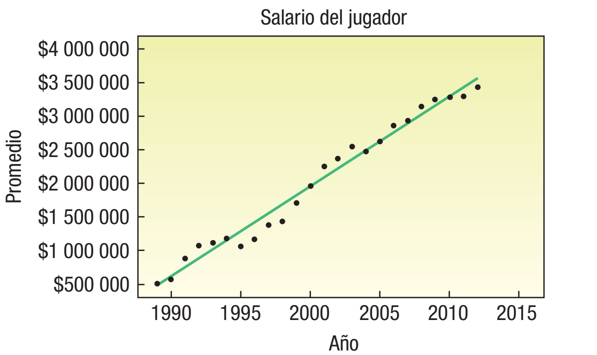
37. La línea de tendencia lineal para la cantidad media por transacción (expresada en dólares constantes) es $\hat{y} = 41.58 - 0.238528t$

El gasto de los clientes está declinando definitivamente en cerca de \$0.24 por año. A continuación se muestra una salida de computadora para la regresión de las transacciones con el tiempo (observe el valor p para el tiempo):

La ecuación de regresión es CPI = 41.6 − 0.239 Tiempo

```
                                 Error
                               estándar
                                    del
Predictor  Coeficiente  coeficiente      T      P
Constante       41.583        1.078  38.56  0.000
Tiempo        −0.23853      0.09003  −2.65  0.016
```

39.

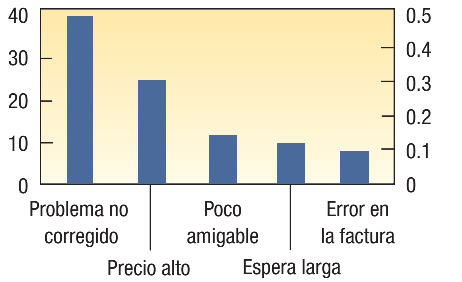


Con 1988 como año base, la ecuación de regresión es:

$\hat{y} = 352\,624 + 134\,221t$. El salario se incrementó a un rango de \$134 221 por año durante el periodo

CAPÍTULO 19

1.



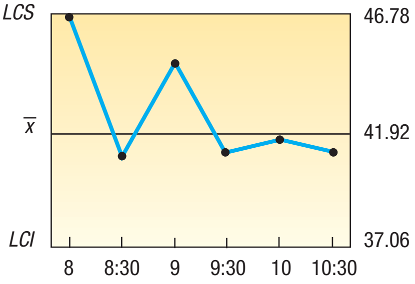
Conteo	38	23	12	10	8
Porcentaje	42	25	13	11	9
Porcentaje acumulado	42	67	80	91	100

Casi 67% de las quejas se refieren al problema que no está siendo corregido y a que el precio es demasiado alto

3. La variación casual es de naturaleza aleatoria; como la causa es una variedad de factores, no se puede eliminar por completo. La variación asignable no es aleatoria; en general, se debe a una causa específica y se puede eliminar

5. **a.** El factor A_2 es 0.729
b. El valor de D_3 es 0, y para D_4 es 2.282

7. **a.**



Hora	$\bar{x}$ Medias aritméticas	*R*, Rango
8:00 a.m.	46	16
8:30 a.m.	40.5	6
9:00 a.m.	44	6
9:30 a.m.	40	2
10:00 a.m.	41.5	9
10:30 a.m.	39.5	1
	251.5	40

$\bar{\bar{x}} = \dfrac{251.5}{6} = 41.92 \qquad \bar{R} = \dfrac{40}{6} = 6.67$

$LCS = 41.92 + 0.729(6.67) = 46.78$
$LCI = 41.92 - 0.729(6.67) = 37.06$

b. Interpretando, la lectura media fue 341.92 grados Fahrenheit; si el horno continúa operando según la evidencia de las primeras seis lecturas por hora, casi 99.7% de las lecturas medias se encontrarán entre 337.06 grados y 346.78 grados

9. **a.** La fracción defectuosa es 0.0507. El límite de control superior es 0.0801 y el límite de control inferior es 0.0213
b. Sí, las muestras 7 y 9 indican que el proceso está fuera de control
c. El proceso parece permanecer igual

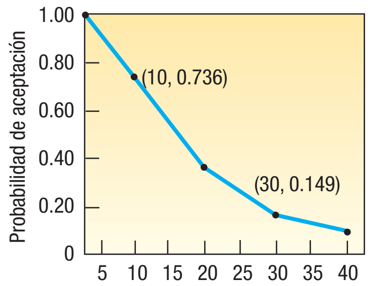
11. $\bar{c} = \dfrac{37}{14} = 2.64$
$2.64 \pm 3\sqrt{2.64}$
Los límites de control son 0 y 7.5; el proceso está fuera de control en el séptimo día

13. $\bar{c} = \dfrac{6}{14} = 0.545$
$0.545 \pm 3\sqrt{0.545} = 0.545 \pm 2.215$
Los límites de control van de 0 a 2.760, por lo que no hay recibos fuera de control

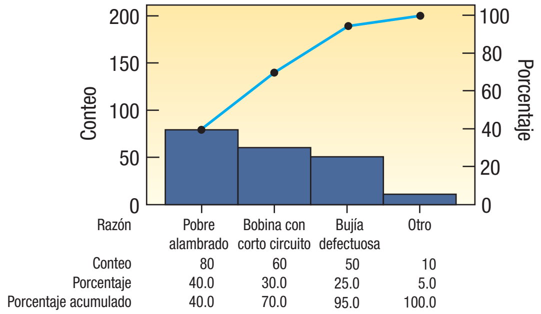
15.

Porcentaje defectuoso	Probabilidad de aceptar el lote
10	0.889
20	0.558
30	0.253
40	0.083

17. $P(x \le 1 | n = 10, \pi = 0.10) = 0.736$
$P(x \le 1 | n = 10, \pi = 0.20) = 0.375$
$P(x \le 1 | n = 10, \pi = 0.30) = 0.149$
$P(x \le 1 | n = 10, \pi = 0.40) = 0.046$



19.

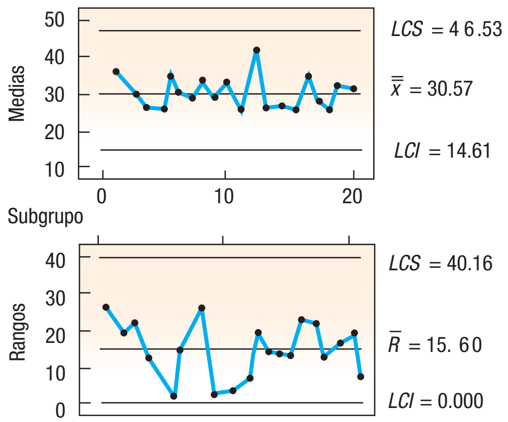


21. **a.** Media: $LCS = 10.0 + 0.577(0.25) = 10.0 + 0.14425$
$= 10.14425$
$LCS = 10.0 - 0.577(0.25) = 10.0 - 0.14425$
$= 9.85575$
Rango: $LCI = 2.115(0.25) = 0.52875$
$LCS = 0(0.25) = 0$

b. La media es 10.16, que está arriba del límite de control superior y fuera de control; hay demasiada cola en las bebidas gaseosas. La variación del proceso está bajo control; es necesario un ajuste

23. **a.** $\bar{\bar{x}} = \dfrac{611.3333}{20} = 30.57$
$\bar{R} = \dfrac{312}{20} = 15.6$
Media: $LCS = 30.5665 + (1.023)(15.6) = 46.53$
$LCI = 30.5665 - (1.023)(15.6) = 14.61$
Rango: $LCS = 2.575(15.6) = 40.17$

b.



c. Todos los puntos parecen estar dentro de los límites de control; no es necesario hacer ajustes

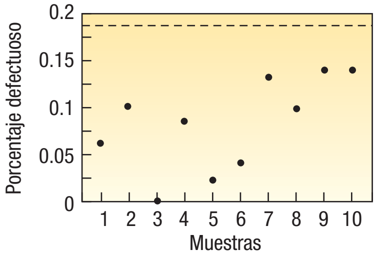
25. $\bar{\bar{x}} = \dfrac{4\,183}{10} = 418.3 \qquad \bar{R} = \dfrac{162}{10} = 16.2$

Media: $LCS = 418.3 + (0.577)(16.2) = 427.65$
$LCI = 418.3 - (0.577)(16.2) = 408.95$
Rango: $LCS = 2.115(16.2) = 34.26$
Todos los puntos están en control, tanto en el caso de la media como del rango

27. **a.** $P = \dfrac{40}{10(50)} = 0.08 \qquad 3\sqrt{\dfrac{0.08(0.92)}{50}} = 0.115$

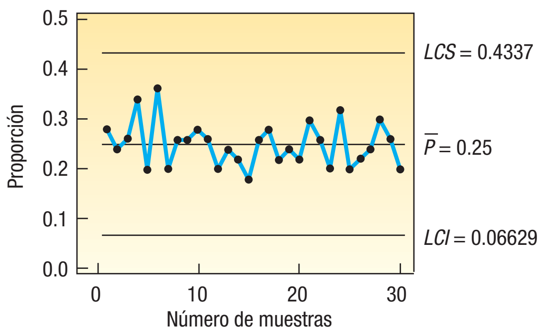
$LCS = 0.08 + 0.115 = 0.195$
$LCI = 0.08 - 0.115 = 0$

b.



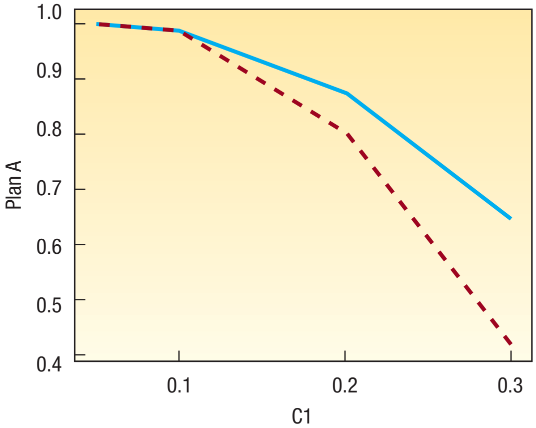
c. No hay puntos que sobrepasen los límites

29. Gráfica P de C1



Estos resultados muestrales indican que las probabilidades de aumento son mucho menores que 50-50; el porcentaje de acciones que aumentan está "en control" alrededor de 0.25 o 25%. Los límites de control son 0.06629 y 0.4337

31. $P(x \leq 3|n = 10, \pi = 0.05) = 0.999$
$P(x \leq 3|n = 10, \pi = 0.10) = 0.987$
$P(x \leq 3|n = 10, \pi = 0.20) = 0.878$
$P(x \leq 3|n = 10, \pi = 0.30) = 0.649$
$P(x \leq 5|n = 20, \pi = 0.05) = 0.999$
$P(x \leq 5|n = 20, \pi = 0.10) = 0.989$
$P(x \leq 5|n = 20, \pi = 0.20) = 0.805$
$P(x \leq 5|n = 20, \pi = 0.30) = 0.417$

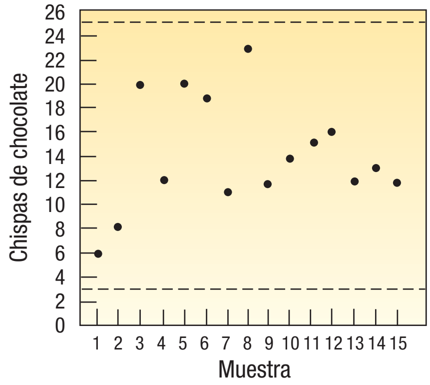


La línea continua es la curva característica de operación del primer plan y la línea discontinua del segundo. El proveedor debería preferir el primero debido a que la probabilidad de aceptación es más alta (arriba); sin embargo, si está completamente seguro de su calidad, el segundo plan parece más alto en el rango muy bajo de porcentajes defectuosos y se podría preferir

33. **a.** $\bar{c} = \frac{213}{15} = 14.2; 3\sqrt{14.2} = 11.30$

$LCS = 14.2 + 11.3 = 25.5$
$LCI = 14.2 - 11.3 = 2.9$

b.

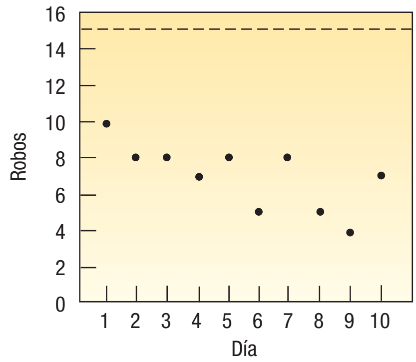


c. Todos los puntos están en control

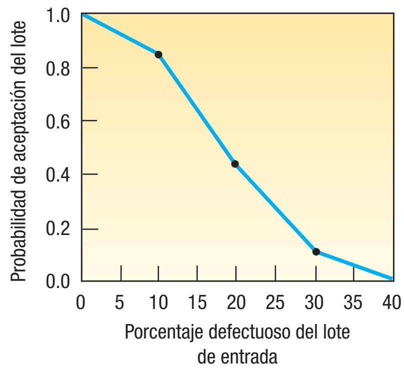
35. **a.** $\bar{c} = \frac{70}{10} = 7.0$

$LCL = 7.0 + 3\sqrt{7} = 14.9$
$LCI = 7.0 - 3\sqrt{7} = 0$

b.



37. $P(x \leq 3|n = 20, \pi = 0.10) = 0.867$
$P(x \leq 3|n = 20, \pi = 0.20) = 0.412$
$P(x \leq 3|n = 20, \pi = 0.30) = 0.108$



APÉNDICE D: RESPUESTAS

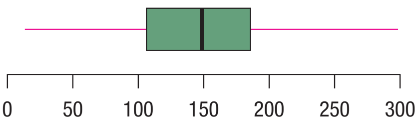
Respuestas a los ejercicios impares de repaso

REPASO DE LOS CAPÍTULOS 1 A 4
PROBLEMAS

1. **a.** La media es 147.9; la mediana es 148.5; la desviación estándar es 69.24

b. El primer cuartil es 106; el tercer cuartil es 186.25

c.



No hay datos atípicos; la distribución es simétrica; los bigotes y las cajas son más o menos iguales en ambos lados

d. $2^6 = 64$, use 6 clases; $i = \dfrac{299 - 14}{6} = 47.5$, use $i = 50$

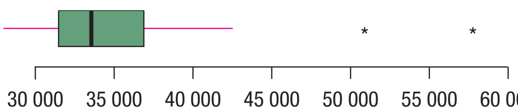
Cantidad	Frecuencia
\$ 0 hasta \$ 50	3
50 hasta 100	8
100 hasta 150	15
150 hasta 200	13
200 hasta 250	7
250 hasta 300	7
Total	50

e. Las respuestas variarán, pero se debe incluir toda la información anterior

3. **a.** La media es \$35 768; la mediana es \$34 405; la desviación estándar es \$5 992

b. El primer cuartil es \$32 030; el tercer cuartil es \$38 994

c.



Hay dos datos atípicos por encima de \$50 000; la distribución tiene un sesgo positivo; los bigotes y las cajas de la derecha son mucho más grandes que los de la izquierda

d.

Cantidades	Frecuencia
\$24 000 hasta 30 000	8
30 000 hasta 36 000	22
36 000 hasta 42 000	15
42 000 hasta 48 000	4
48 000 hasta 54 000	1
54 000 hasta 60 000	1
Total	51

e. Las respuestas variarán, pero se debe incluir toda la información anterior

5. **a.** Diagrama de caja

b. La mediana es 48, el primer cuartil es 24, y el tercero es 84

c. Con sesgo positivo, con la cola larga a la derecha

d. No es posible determinar el número de observaciones

REPASO DE LOS CAPÍTULOS 5 A 7
PROBLEMAS

1. **a.** 0.035 **b.** 0.018 **c.** 0.648

3. **a.** 0.0401 **b.** 0.6147 **c.** 7 440

5. **a.** $\mu = 1.10$

$\sigma = 1.18$

b. Cerca de 550

c. $\mu = 1.833$

REPASO DE LOS CAPÍTULOS 8 Y 9
PROBLEMAS

1. $z = \dfrac{8.8 - 8.6}{2.0/\sqrt{35}} = 0.59$, $0.5000 - 0.2224 = 0.2776$

3. $160 \pm 2.426 \dfrac{20}{\sqrt{40}}$, 152.33 hasta 167.67

5. $985.5 \pm 2.571 \dfrac{115.5}{\sqrt{6}}$, 864.27 hasta 1 106.73

7. $240 \pm 2.131 \dfrac{35}{\sqrt{16}}$, 221.35 hasta 258.65

Porque 250 está en el intervalo, la evidencia *no* indica un aumento de la producción

9. $n = \left[\dfrac{1.96(25)}{4}\right]^2 = 150$

11. $n = 0.08(0.92)\left(\dfrac{2.33}{0.02}\right)^2 = 999$

13. $n = 0.4(0.6)\left(\dfrac{2.33}{0.03}\right)^2 = 1\,448$

REPASO DE LOS CAPÍTULOS 10 A 12
PROBLEMAS

1. H_0: $\mu \geq 36$ $\quad H_1$: $\mu < 36$. Rechace H_0 si $t < -1.683$

$$t = \frac{35.5 - 36.0}{0.9/\sqrt{42}} = -3.60$$

Rechace H_0; la altura media es menor a 36 pulgadas

3. H_0: $\mu_1 = \mu_2$ $\quad H_1$: $\mu_1 \neq \mu_2$

Rechace H_0 si $t < -2.845$ o $t > 2.845$

$$s_p^2 = \frac{(12 - 1)(5)^2 + (10 - 1)(8)^2}{12 + 10 + 2} = 42.55$$

$$t = \frac{250 - 252}{\sqrt{42.55\left(\dfrac{1}{12} + \dfrac{1}{10}\right)}} = -0.716$$

H_0 no se rechaza; no hay diferencia en la fuerza media de ambos pegamentos

5. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ $\quad H_1$: las medias no son iguales

Rechace H_0 si $F > 3.29$

Fuente	SS	*gl*	MS	*F*
Tratamientos	20.736	3	6.91	1.04
Error	100.00	15	6.67	
Total	120.736	18		

H_0 no se rechaza; no existe diferencia en las ventas medias

7. **a.** De la gráfica, los salarios de marketing pueden estar actuando diferente

b. H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: cuando menos una media es distinta (para cuatro especialidades)

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: cuando menos una media es distinta (para tres años)

H_0: no hay interacción

H_1: hay interacción

c. El valor *p* (0.482) es alto; no rechace la hipótesis de no interacción

d. El valor *p* para las especialidades es pequeño ($0.034 < 0.05$), así que hay diferencias entre los salarios mínimos por especialidad; no hay diferencia de un año al siguiente en los salarios medios ($0.894 > 0.05$)

REPASO DE LOS CAPÍTULOS 13 Y 14
PROBLEMAS

1. **a.** Utilidad
   b. $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$
   c. \$163 200
   d. Cerca de 86% de la variación de la utilidad neta se explica por las cuatro variables
   e. Cerca de 68% de las utilidades netas estarían dentro de \$3 000 de los estimados, cerca de 95% estaría dentro de 2(\$3 000), o \$6 000, de los estimados; y virtualmente todas estarían dentro de 3(\$3 000), o \$9 000, de las estimaciones
3. **a.** 0.9261
   b. 2.0469, calculado por $\sqrt{83.8/20}$
   c. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
   H_1: no todos los coeficientes son 0
   Rechace si $F > 2.87$, calculado $F = 62.697$, determinado por 162.70/4.19
   d. Podría eliminar x_2 porque la razón de t (1.29) es menor que el valor crítico de t de 2.086. De otro modo, rechace H_0 para x_1, x_3 y x_4 porque todas las razones de t son mayores que 2.086

REPASO DE LOS CAPÍTULOS 15 Y 16
PROBLEMAS

1. H_0: mediana ≤ 60
   H_1: mediana > 60
   $\mu = 20(0.5) = 10$
   $\sigma = \sqrt{20(0.5)(0.5)} = 2.2361$
   H_0 se rechaza si $z > 1.65$. Hay 16 observaciones mayores a 60
   $$z = \frac{15.5 - 10.0}{2.2361} = 2.46$$
   Rechace H_0; la media de ventas por día es mayor a 60
3. H_0: la longitud de población es la misma
   H_1: la longitud de población no es la misma
   H_0 se rechaza si H es > 5.991
   $$H = \frac{12}{24(24+1)}\left[\frac{(104.5)^2}{7} + \frac{(125.5)^2}{9} + \frac{(70)^2}{8}\right] - 3(24 + 1)$$
   $$= 78.451 - 75 = 3.451$$
   No rechace H_0; la longitud de población es la misma

REPASO DE LOS CAPÍTULOS 17 Y 18
PROBLEMAS

1. **a.** 137.8, calculado por (14 883/10 799)100
   b. 134.6, calculado por (14 883/11 056.7)100. Observe que 11 056.7 es el promedio para el periodo 2008 a 2010
   c. 8 941 + 1 112t y 16 725, calculado por 8 941 + 1 112(7)
3. 55.44, calculado por 1.20[3.5 + (0.7) (61)] y 44.73, calculado por 0.90[3.5 + (0.7)(66)]

Soluciones a los test de práctica

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 4)

Parte 1

1. estadísticas
2. estadísticas descriptivas
3. población
4. cuantitativo y cualitativo
5. discreta
6. nominal
7. nominal
8. cero
9. siete
10. 50
11. varianza
12. nunca
13. mediana

Parte 2

1. $\sqrt[3]{(1.18)(1.04)(1.02)} = 1.0777$, o 7.77%
2. **a.** 30 mil dólares
   b. 105
   c. 52
   d. 0.19, calculado por 20/105
   e. 165
   f. 120 y 330
3. **a.** 70
   b. 71.5
   c. 67.8
   d. 28
   e. 9.34
4. \$44.20, calculado por [(200)\$36 + (300)\$40 + (500)\$50]/1 000
5. **a.** gráfica de pastel
   b. 11.1
   c. tres veces
   d. 65%

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 7)

Parte 1

1. nunca
2. experimento
3. evento
4. conjunta
5. **a.** permutación
   b. combinación
6. uno
7. tres o más resultados
8. infinitas
9. una
10. 0.2764
11. 0.0475
12. independiente
13. mutuamente excluyentes
14. solo dos resultados
15. en forma de campana

Parte 2

1. **a.** 0.526, calculado por (5/20)(4/19)
   b. 0.4474, calculado por 1 − (15/20)(14/19)
2. **a.** 0.2097, calculado por $16(0.15)(0.85)^{15}$
   b. 0.9257, calculado por $1 - (0.85)^{16}$
3. 720, calculado por $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2$
4. **a.** 2.2, calculado por 0.2(1) + 0.5(2) + 0.2(3) + 0.1(4)
   b. 0.76, calculado por 0.2(1.44) + 0.5(0.04) + 0.2(0.64) + 0.1(3.24)
5. **a.** 0.1808. El valor z para \$2 000 es 0.47, calculado por (2 000 − 1 600)/850
   b. 0.4747, calculado por 0.2939 + 0.1808
   c. 0.0301, calculado por 0.5000 − 0.4699
6. **a.** tabla de contingencia
   b. 0.625, calculado por 50/80
   c. 0.75, calculado por 60/80
   d. 0.40, calculado por 20/50
   e. 0.125, calculado por 10/80
7. **a.** 0.0498, calculado por $\dfrac{3^0e^{-3}}{0!}$
   b. 0.2240, calculado por $\dfrac{3^3e^{-3}}{3!}$
   c. 0.1847, calculado por 1 − [0.0498 + 0.1494 + 0.2240 + 0.2240 + 0.1680]
   d. 0.0025

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 9)

Parte 1

1. muestra aleatoria
2. error de muestreo
3. error estándar
4. se reducirá
5. estimación de puntos
6. intervalo de confianza
7. tamaño de la población
8. proporción
9. sesgo positivo
10. 0.5

Parte 2

1. 0.0351, calculado por 0.5000 − 0.4649. El valor z correspondiente es: $z = \dfrac{11 - 12.2}{2.3/\sqrt{12}} = -1.81$
2. **a.** Se desconoce la media de la población
   b. 9.3 años, que es la media muestral
   c. 0.3922, calculado por $2/\sqrt{26}$
   d. El intervalo de confianza es de 8.63 a 9.97, calculado por
   $$9.3 \pm 1.708\left(\frac{2}{\sqrt{26}}\right)$$
3. 2 675, calculado por $0.27(1 - 0.27)\left(\dfrac{2.33}{0.02}\right)^2$
4. El intervalo de confianza es de 0.5459 a 0.7341, calculado por
   $$0.64 \pm 1.96\sqrt{\frac{0.64(1-0.64)}{100}}$$

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 12)

Parte 1

1. hipótesis nula
2. nivel de significancia
3. valor p
4. desviación estándar
5. normalidad
6. estadístico de prueba
7. repartido uniformemente entre ambas colas
8. va de infinito negativo a infinito positivo
9. independiente
10. tres y 20

Parte 2

1. H_0: $\mu \leq 90$ H_1: $\mu > 90$ si $t > 2.567$, rechace H_0.
   $$t = \frac{96 - 90}{12/\sqrt{18}} = 2.12$$
   No rechace la nula; el tiempo medio en el parque podría ser de 90 minutos

2. H_0: $\mu_1 = \mu_2$ H_1: $\mu_1 \neq \mu_2$
$gl = 14 + 12 - 2 = 24$
Si $t < -2.064$ o $t > 2.064$, entonces rechace H_0

$$s_p^2 = \frac{(14-1)(30)^2 + (12-1)(40)^2}{14+12-2} = 1\,220.83$$

$$t = \frac{837 - 797}{\sqrt{1220.83\left(\frac{1}{14}+\frac{1}{12}\right)}} = \frac{40.0}{13.7455} = 2.910$$

Rechace la hipótesis nula; hay diferencia entre las medias de las millas recorridas

3. a. tres, porque hay 2 *gl* entre los grupos
   b. 21, calculado por los grados totales de libertad más 1
   c. Si el nivel de significancia es 0.05, el valor crítico es 3.55
   d. H_0: $\mu_1 = \mu_2 = \mu_3$ H_1: las medias de tratamiento no son iguales
   e. A un nivel de significancia de 5%, se rechaza la hipótesis nula
   f. A un nivel de significancia de 5%, se puede concluir que las medias de tratamiento difieren

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 14)

Parte 1

1. vertical
2. intervalo
3. cero
4. −0.77
5. nunca
6. 7
7. disminución de 0.5
8. −0.9
9. cero
10. ilimitado
11. lineal
12. residual
13. dos
14. matriz de correlación
15. distribución normal

Parte 2

1. a. 30
   b. La ecuación de regresión es $\hat{y} = 90.619x - 0.9401$. Si *x* es cero, la línea cruza el eje vertical en −0.9401; a medida que aumenta la variable independiente en una unidad, la variable independiente aumenta en 90.619 unidades
   c. 905.2499
   d. 0.3412, calculado por 129.7275/380.1667; 34% de la variación en la variable dependiente se explica por la variable independiente
   e. 0.5842, calculado por $\sqrt{0.3412}$ H_0: $p \geq 0$ H_1: $p < 0$
   Con el nivel de significancia 0.01, rechace H_0 si $t > 2.467$

$$t = \frac{0.5842\sqrt{30-2}}{\sqrt{1-(0.5842)^2}} = 3.81$$

   Rechace H_0; existe una correlación negativa entre las variables
2. a. 30
   b. 4
   c. 0.5974, calculado por 227.0928/380.1667
   d. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ H_1: no todas las β son 0
   Rechace H_0 si $F > 4.18$ (con un nivel de significancia de 1%)
   Como el valor calculado de *F* es 9.27, rechace H_0; no todos los coeficientes de regresión son cero
   e. Rechace H_0 si $t > 2.787$ o $t < -2.787$ (con un nivel de significancia de 1%). Descarte inicialmente la variable 2 y recalcule; tal vez también se descarten las variables 1 y 4

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 16)

Parte 1

1. nominal
2. al menos 30 observaciones
3. dos
4. 6
5. número de categorías
6. dependiente
7. binomial
8. comparando dos o más muestras independientes
9. nunca
10. poblaciones normales, desviaciones estándar iguales

Parte 2

1. H_0: las proporciones son como se estableció
H_1: las proporciones no son como se estableció
Con el nivel de significancia 0.05, rechace H_0 si $\chi^2 > 7.815$.

$$\chi^2 = \frac{(120-130)^2}{130} + \frac{(40-40)^2}{40} + \frac{(30-20)^2}{20} + \frac{(10-10)^2}{10} = 5.769$$

No rechace H_0; las proporciones podrían ser como se estableció

2. H_0: no hay relación entre el género y el tipo de libro
H_1: hay relación entre el género y el tipo de libro
Con el nivel de significancia 0.01, rechace H_0 si $\chi^2 > 9.21$

$$\chi^2 = \frac{(250-197.3)^2}{197.3} + \cdots + \frac{(200-187.5)^2}{187.5} = 54.84$$

Rechace H_0; hay una relación entre el género y el tipo de libro

3. H_0: las distribuciones son iguales
H_1: las distribuciones no son iguales
Rechace H_0 si $H > 5.99$

8:00 a.m.	Rangos	10:00 a.m.	Rangos	1:30 p.m.	Rangos
68	6	59	1.5	67	5
84	20	59	1.5	69	7
75	10.5	63	4	75	10.5
78	15.5	62	3	76	12.5
70	8	78	15.5	79	17
77	14	76	12.5	83	19
88	24	80	18	86	21.5
71	9			86	21.5
				87	23
Sumas	107		56		137
Conteo	8		7		9

$$H = \frac{12}{24(25)}\left[\frac{107^2}{8} + \frac{56^2}{7} + \frac{137^2}{9}\right] - 3(25) = 4.29$$

No rechace H_0; no hay diferencia en las tres distribuciones

4. H_0: $\pi \leq 1/3$ H_1: $\pi > 1/3$
Con el nivel de significancia 0.01, la regla de decisión es rechazar H_0 si $z > 2.326$

$$z = \frac{\left[\frac{210}{500} - \frac{1}{3}\right]}{\sqrt{\frac{\left(\frac{1}{3}\right)\left(1-\frac{1}{3}\right)}{500}}} = \frac{0.08667}{0.02108} = 4.11$$

Rechace la hipótesis nula
La proporción real de niños de Louisiana que son obesos o padecen sobrepeso es más de uno en tres

TEST DE PRÁCTICA (DESPUÉS DEL CAPÍTULO 18)

Parte 1

1. denominador
2. índice
3. cantidad
4. periodo base
5. 1982-1984
6. tendencia
7. promedio móvil
8. autocorrelación
9. residuo
10. el mismo

Parte 2

1. a. 111.54, calculado por (145 000/130 000) × 100 para 2009
   92.31, calculado por (120 000/130 000) × 100 para 2010
   130.77, calculado por (170 000/130 000) × 100 para 2011
   146.15, calculado por (190 000/130 000) × 100 para 2012
   b. 87.27, calculado por (120 000/137 500) × 100 para 2010
   126.64, calculado por (170 000/137 500) × 100 para 2011
   138.18, calculado por (190 000/137 500) × 100 para 2012
2. a. 108.91, calculado por (1 100/1 010) × 100
   b. 111.18, calculado por (4 525/4 070) × 100
   c. 110.20, calculado por (5 400/4 900) × 100
   d. 110.69, calculado por la raíz cuadrada de (111.18) × (110.20)
3. Para enero del quinto año, el pronóstico estacional ajustado es 70.0875, calculado por 1.05 × [5.50 + 1.25(49)]
Para febrero del quinto año, el pronóstico estacional ajustado es 66.844, calculado por 0.983 × [5.50 + 1.25(50)]

APÉNDICE E: RESPUESTAS A LAS AUTOEVALUACIONES

CAPÍTULO 1

1.1. **a.** Con base en la muestra de 1 960 consumidores, se estima que, si lo comercializa, 60% de ellos comprará el platillo de pollo (1 176/1 960) $\times$ 100 $=$ 60%

b. Estadística inferencial, ya que se empleó una muestra para llegar a una conclusión relativa a la reacción de los consumidores de la población en caso de que se comercializara el platillo de pollo

1.2. **a.** La edad es una variable de escala de razón. Una persona de 40 años tiene el doble de edad que una de 20

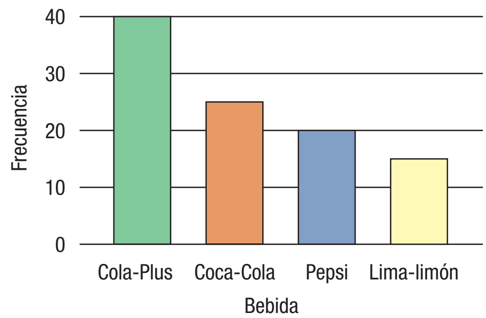
b. Escala nominal; podría ordenar indistintamente los estados

CAPÍTULO 2

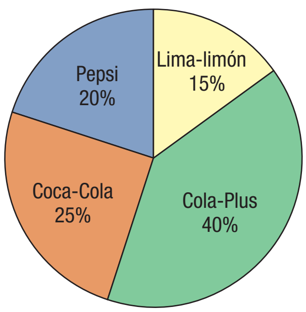
2.1. **a.** Datos cualitativos, ya que la respuesta de los consumidores a la prueba de degustación es el nombre de una bebida

b. Tabla de frecuencias; en esta se muestra el número de personas que prefiere cada una de las bebidas

c.



d.



2.2. **a.** Los datos brutos (o no agrupados)

b.

Comisión	Número de vendedores
\$1 400 hasta \$1 500	2
1 500 hasta 1 600	5
1 600 hasta 1 700	3
1 700 hasta 1 800	1
Total	11

c. Frecuencias de clase

d. La concentración más grande de comisiones se encuentra entre \$1 500 y \$1 600. La comisión más pequeña es de aproximadamente \$1 400 y la más grande, de casi \$1 800. La cantidad típica que se obtuvo fue de \$1 500

2.3. **a.** $2^6 = 64 < 73 < 128 = 2^7$; así que se recomiendan 7 clases

b. La variación del intervalo debería ser de por lo menos $(488 - 320)/7 = 24$. Los intervalos de clase de 25 o 30 pies son razonables

c. Asumiendo un intervalo de clase de 25 y se comienza con un límite inferior de 300, serían necesarias ocho clases. Un intervalo de clase de 30 que comience con un límite inferior de 300 también es razonable; esta alternativa requiere solo siete clases

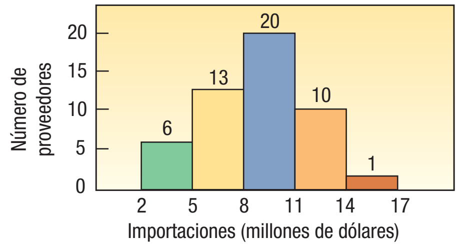
Clases de distancia	Frecuencia	Porcentaje
300 hasta 330	2	2.7
330 hasta 360	2	2.7
360 hasta 390	17	23.3
390 hasta 420	27	37.0
420 hasta 450	22	30.1
450 hasta 480	1	1.4
480 hasta 510	2	2.7
Gran total	73	100.00

d. 17

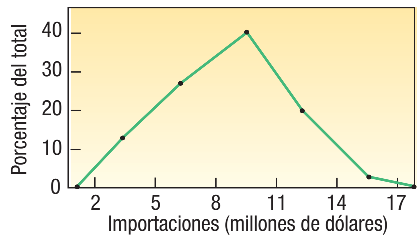
e. 23.3%, calculado por 17/23

f. 71.2%, calculado por (27 + 22 + 1 + 2)/73

2.4. **a.**



b.



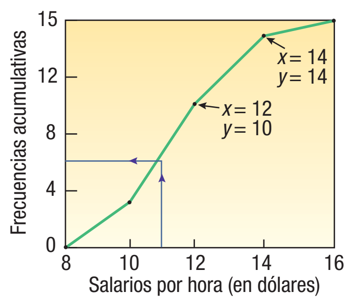
Las puntos son: (3.5, 12), (6.5, 26), (9.5, 40), (12.5, 20) y (15.5, 2)

c. El mínimo volumen anual de importaciones por parte de un proveedor es de aproximadamente \$2 millones, el máximo, de \$17 millones. La frecuencia más alta se encuentra entre \$8 millones y \$11 millones

2.5. **a.** Una distribución de frecuencias

b.

Salarios por hora	Número acumulador
Menos de \$8	0
Menos de \$10	3
Menos de \$12	10
Menos de \$14	14
Menos de \$16	15



c. Alrededor de siete empleados ganan \$11.00 o menos

CAPÍTULO 3

3.1. **1.** **a.** $\bar{x} = \dfrac{\Sigma x}{n}$

b. $\bar{x} = \dfrac{\$267\,100}{4} = \$66\,775$

c. Estadístico, pues se trata de un valor muestral

d. $66 775. La media de la muestra constituye nuestra mejor aproximación de la media poblacional

2. **a.** $\mu = \dfrac{\Sigma x}{n}$

b. $\mu = \dfrac{498}{6} = 83$

c. Parámetro, porque se calculó empleando todos los valores de la población

3.2. **1.** **a.** $878

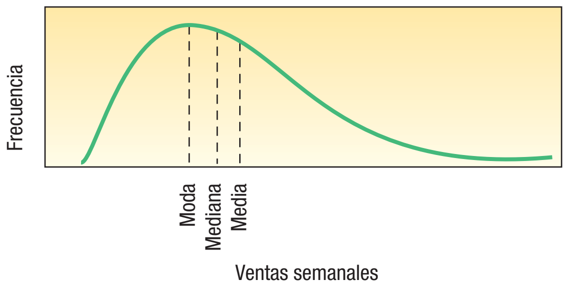
b. 3, 3

2. **a.** 7, que se calcula por $(6 + 8)/2 = 7$

b. 3, 3

c. 0

3.3. **a.**



b. Con sesgo positivo, ya que la media es el promedio más grande y la moda es el más pequeño

3.4. **a.** $237; se calcula de la siguiente manera:

$$\frac{(95 \times \$400) + (126 \times \$200) + (79 \times \$100)}{95 + 126 + 79} = \$237.00$$

b. La ganancia por traje es de $12, que se determina mediante la operación $237 − costo de $200 − $25 de comisión. La ganancia total que generaron los 300 trajes es de $3 600, la cual se calcula multiplicando 300 × $12

3.5. **1.** **a.** Alrededor de 9.9%, que se obtiene con la raíz $\sqrt[4]{1.458602236}$ entonces $1.099 - 1.00 = 0.099$

b. Alrededor de 10.095%

c. Mayor que, porque $10.095 > 9.9$

2. 8.63%, que se determina mediante la operación $\sqrt[20]{\dfrac{120\,520}{23\,000}} - 1 = 1.0863 - 1$

3.6. **a.** 22 mil libras, que se determina restando $112 - 90$

b. $\bar{x} = \dfrac{824}{8} = 103$ mil libras

c.

x	$(x - x)$	$(x - \bar{x})^2$
95	−8	64
103	0	0
105	2	4
110	7	49
104	1	2
105	2	4
112	9	81
90	−13	169
	Total	373

$$\text{Varianza} = \frac{373}{8} = 46.625$$

3.7. **a.** $\mu = \dfrac{\$16\,900}{5} = \$3\,380$

b.
$$\sigma^2 = \frac{(3\,536 - 3\,380)^2 + \cdots + (3\,622 - 3\,380)^2}{5}$$
$$= \frac{(156)^2 + (-207)^2 + (68)^2 + (-259)^2 + (242)^2}{5}$$
$$= \frac{197\,454}{5} = 39\,490.8$$

c. $\sigma = \sqrt{39\,490.8} = 198.72$

d. Hay más variación en la oficina de Pittsburgh, ya que la desviación estándar y la media son mayores

3.8. 2.33, que se calcula por:

$$\bar{x} = \frac{\Sigma x}{n} = \frac{28}{7} = 4$$

x	$x - \bar{x}$	$(x - \bar{x})^2$
4	0	0
2	−2	4
5	1	1
4	0	0
5	1	1
2	−2	4
6	2	4
28	0	14

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n - 1}$$
$$= \frac{14}{7 - 1}$$
$$= 2.33$$
$$s = \sqrt{2.33} = 1.53$$

3.9. **a.** $k = \dfrac{14.15 - 14.00}{0.10} = 1.5$

$k = \dfrac{13.85 - 14.00}{0.10} = -1.5$

$1 = \dfrac{1}{(1.5)^2} = -1 - 0.44 = 0.56$

b. 13.8 y 14.2

3.10. **a.** Distribución de frecuencias

b.

f	M	fM	$(M - \bar{x})$	$f(M - \bar{x})^2$
1	4	4	−8.2	67.24
4	8	32	−4.2	70.56
10	12	120	−0.2	0.40
3	16	48	3.8	43.32
2	20	40	7.8	121.68
20		244		303.20

$$\bar{x} = \frac{\Sigma fM}{M} = \frac{\$244}{20} = \$12.20$$

c. $s = \sqrt{\dfrac{303.20}{20 - 1}} = \$3.99$

CAPÍTULO 4

4.1. **1.** **a.** 79, 105

b. 15

c. De 88 a 97; 75% de las tiendas se encuentran en este rango

2.

7	7
8	0013488
9	1256689
10	1248
11	26

a. 8

b. 10.1, 10.2, 10.4, 10.8

c. 9.5

d. 11.6, 7.7

4.2. **a.** 7.9

b. $Q_1 = 7.76$, $Q_3 = 8.015$

4.3. El valor más bajo es 10 y el más alto, 85; el primer cuartil es 25 y el tercero, 60; alrededor de 50% de los valores se encuentran entre 25 y 60 (el de la mediana es de 40), y la distribución es positivamente sesgada

4.4. **a.** $\bar{x} = \dfrac{407}{5} = 81.4$, mediana $= 84$

$s = \sqrt{\dfrac{923.2}{5 - 1}} = 15.19$

b. $sk = \dfrac{3(81.4 - 84.0)}{15.19} = -0.51$

c.

x	$\dfrac{x - \bar{x}}{s}$	$\left[\dfrac{x - \bar{x}}{s}\right]^3$
73	−0.5530	−0.1691
98	1.0928	1.3051
60	−1.4088	−2.7962
92	0.6978	0.3398
84	0.1712	0.0050
		−1.3154

$$sk = \frac{5}{(4)(3)}[-1.3154]$$
$$= -0.5481$$

d. La distribución es de alguna forma negativamente sesgada

4.5. **a.** Diagrama de dispersión
b. 16
c. \$7 500
d. Fuerte y directa

CAPÍTULO 5

5.1. **a.** Cuente el número que piensa que el nuevo juego es operable
b. A 73 jugadores les gustó el juego; hay muchas otras respuestas probables
c. No, la probabilidad no puede ser mayor que 1; la probabilidad de que el juego sea un éxito si se comercializa es de 65/80, o 0.8125
d. No puede ser menor que 0; tal vez se trate de un error aritmético
e. A más de la mitad de los jugadores que probaron el juego les gustó (por supuesto, hay otras posibles respuestas)

5.2. **1.** $\dfrac{\text{4 reinas en una baraja}}{\text{52 cartas en total}} = \dfrac{4}{52} = -0.0769$ Clásico

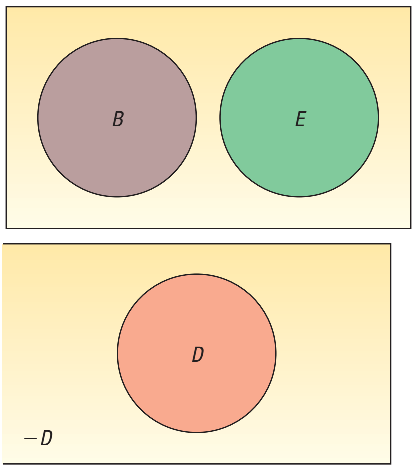
2. $\dfrac{182}{539} = 0.338$ Empírico

3. La posibilidad del resultado se estima aplicando el enfoque subjetivo a la estimación de la probabilidad; si piensa que es probable ahorrar \$1 millón, entonces su probabilidad debería estar entre 0.5 y 0.10

5.3. **a.** **i** $\dfrac{(50 + 68)}{2\,000} = 0.059$

ii. $1 - \dfrac{302}{2\,000} = 0.849$

b.

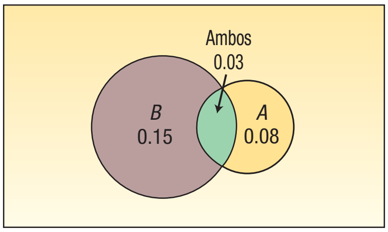


c. No son complementarios, pero son mutuamente excluyentes

5.4. **a.** El evento *A* se refiere a la necesidad de zapatos ortopédicos; el evento *B* se refiere a la necesidad de un tratamiento dental

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$$
$$= 0.08 + 0.15 - 0.03$$
$$= 0.20$$

b. Una probabilidad es:



5.5. $(0.95)(0.95)(0.95)(0.95) = 0.8145$

5.6. **a.** 0.002, que se determina por:

$$\left(\frac{4}{12}\right)\left(\frac{3}{11}\right)\left(\frac{2}{10}\right)\left(\frac{1}{9}\right) = \frac{24}{11\,880} = 0.002$$

b. 0.14, que se determina por:

$$\left(\frac{8}{12}\right)\left(\frac{7}{11}\right)\left(\frac{6}{10}\right)\left(\frac{5}{9}\right) = \frac{1\,680}{11\,880} = 0.1414$$

c. No, porque existen otras probabilidades, como tres mujeres y un hombre

5.7. **a.** $P(B_2) = \dfrac{225}{500} = 0.45$

b. Ambos eventos son mutuamente excluyentes, así que aplique la regla especial de la adición

$$P(B_1 \text{ o } B_2) = P(B_1) + P(B_2) = \frac{100}{500} + \frac{225}{500} = 0.65$$

c. Ambos eventos no son mutuamente excluyentes, así que aplique la regla general de la adición

$$P(B_1 \text{ o } A_1) = P(B_1) + P(A_1) - P(B_1 \text{ y } A_1)$$
$$= \frac{100}{500} + \frac{75}{500} + \frac{15}{500} = 0.32$$

d. Como se muestra en el ejemplo/solución, la atención de los cines por mes y edad no son independientes; por lo tanto, aplique la regla general de la multiplicación

$$P(B_1 \text{ y } A_1) = P(B_1)\,P(A_1|B_1)$$
$$= \left(\frac{100}{500}\right)\left(\frac{15}{500}\right) = 0.03$$

5.8. **a.** $P(\text{visitan a menudo}) = \dfrac{80}{195} = 0.41$

b. *P*(visitaron una tienda en un centro comercial cerrado)

$$= \frac{90}{195} = 0.46$$

c. Ambos eventos no son mutuamente excluyentes, así que aplique la regla general de la adición

P(visitan a menudo o visitaron un Sears en un centro comercial cerrado)
= *P*(a menudo) + *P*(centro comercial cerrado)
− *P*(a menudo y en un centro comercial cerrado)

$$= \frac{80}{195} + \frac{90}{195} - \frac{60}{195} = 0.56$$

d. *P*(visitan a menudo|fueron a un Sears en un centro comercial cerrado) $= \dfrac{60}{90} = 0.67$

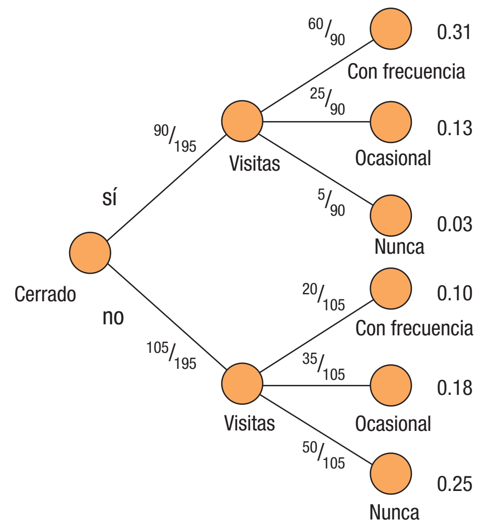
e. La independencia requiere que $P(A|B) = P(A)$. Una probabilidad es: *P*(visitas frecuentes|sí, centro comercial cerrado) = *P*(visitas frecuentes), ¿60/90 = 80/195? No, ambas variables *no* son independientes; por consiguiente, cualquier probabilidad conjunta que se encuentre en la tabla debe calcularse aplicando la regla general de la multiplicación

f. Como se muestra en el punto e, las visitas a menudo y en un centro comercial cerrado no son independientes, así que aplique la regla general de la multiplicación

P(a menudo y centro comercial cerrado)
= *P*(a menudo)*P*(cerrado|a menudo)

$$= \left(\frac{80}{195}\right)\left(\frac{60}{80}\right) = 0.31$$

g.



5.9. **a.** $P(A_3 \text{ y } B_2) = \dfrac{P(A_3)\,P(B_2|A_3)}{P(A_1)\,P(B_2|A_1) + P(A_2)\,P(B_2|A_2) + P(A_3)\,P(B_2|A_3)}$

b. $= \dfrac{(0.50)(0.96)}{(0.30)(0.97) + (0.20)(0.95) + (0.50)(0.96)}$

$= \dfrac{0.480}{0.961} = 0.499$

5.10. **1.** $(5)(4) = 20$

2. $(3)(2)(4)(3) = 72$

5.11. **1.** **a.** 60, que se calcula multiplicando (5)(4)(3)

b. 60, que se calcula por:

$$\frac{5!}{(10-4)!} = \frac{5 \cdot 4 \cdot 3 \cdot \cancel{2 \cdot 1}}{\cancel{2 \cdot 1}}$$

2. 5 040, que se calcula por:

$$\frac{10!}{(10-4)!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot \cancel{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}{\cancel{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}}$$

3. **a.** 35 es correcto, el cual se calcula:

$${}_7C_3 = \frac{n!}{r!(n-r)!} = \frac{7!}{3!(7-3)!} = 35$$

b. Sí, hay 21 combinaciones, que se calculan:

$${}_7C_5 = \frac{n!}{r!(n-r)!} = \frac{7!}{5!(7-5)!} = 21$$

4. **a.** ${}_{50}P_3 = \dfrac{50!}{(50-3)!} = 117\ 600$

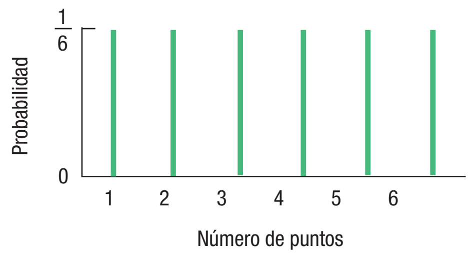
b. ${}_{50}C_3 = \dfrac{50!}{3!(50-3)!} = 19\ 600$

CAPÍTULO 6

6.1. **a.**

Número de puntos	Probabilidad
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$
Total	$\frac{6}{6} = 1.00$

b.



c. $\frac{6}{6}$, o 1

6.2. **a.** Discreta, pues los valores \$0.80, \$0.90 y \$1.20 se encuentran claramente separados entre sí; asimismo, la suma de las probabilidades es 1.00 y los resultados son mutuamente excluyentes

b.

x	$P(x)$	$xP(x)$
\$ .80	0.30	0.24
0.90	0.50	0.45
1.20	0.20	0.24
		0.93

La media es de 93 centavos

c.

x	$P(x)$	$(x-\mu)$	$(x-\mu)^2P(x)$
\$0.80	0.30	−0.13	0.00507
0.90	0.50	−0.03	0.00045
1.20	0.20	0.27	0.01458
			0.02010

La varianza es de 0.02010, y la desviación estándar, de 14 centavos

6.3. **a.** Es razonable, porque a cada empleado se le hace un depósito directo o no se le hace; los empleados son independientes; la probabilidad de que se hagan depósitos directos es de 0.95 en el caso de todos, y se cuentan los empleados de entre siete que se benefician del servicio

b. $P(7) = {}_7C_7(0.95)^7(0.05)^0 = 0.6983$

c. $P(4) = {}_7C_4(0.95)^4(0.05)^3 = 0.0036$

d. Las respuestas concuerdan

6.4. **a.** $n = 8,\ \pi = 0.40$

b. $P(x = 3) = 0.2787$

c. $P(x > 0) = 1 - P(x = 0) = 1 - 0.0168 = 0.9832$

6.5. $P(3) = \dfrac{{}_8C_3\ {}_4C_2}{{}_{12}C_5} = \dfrac{\left(\frac{8!}{3!5!}\right)\left(\frac{4!}{2!2!}\right)}{\frac{12!}{5!7!}}$

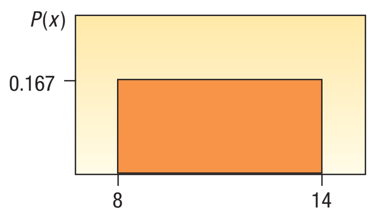
$= \dfrac{(56)(6)}{792} = 0.424$

6.6. $\mu = 4\ 000(0.0002) = 0.8$

$P(1) = \dfrac{0.8^1 e^{-0.8}}{1!} = 0.3595$

CAPÍTULO 7

7.1. **a.**



b. $P(x) = (\text{altura})(\text{base})$

$= \left(\dfrac{1}{14-8}\right)(14-8)$

$= \left(\dfrac{1}{6}\right)(6) = 1.00$

c. $\mu = \dfrac{a+b}{2} = \dfrac{14+8}{2} = \dfrac{22}{2} = 11$

$\sigma = \sqrt{\dfrac{(a-b)^2}{12}} = \sqrt{\dfrac{(14-8)^2}{12}} = \sqrt{\dfrac{36}{12}} = \sqrt{3}$
$= 1.73$

d. $P(10 < x < 14) = (\text{altura})(\text{base})$
$= \left(\dfrac{1}{14-8}\right)(14-10)$
$= \dfrac{1}{6}(4)$
$= 0.667$

e. $P(x < 9) = (\text{altura})(\text{base})$
$= \left(\dfrac{1}{14-8}\right)(9-8)$
$= 0.167$

7.2. **a.** $z = (64 - 48)/12.8 = 1.25$. La diferencia de esta persona de 16 onzas más que el promedio es de 1.25 desviaciones estándar por encima de este
b. $z = (32 - 48)/12.8 = -1.25$. La diferencia de esta persona de 16 onzas menos que el promedio es de 1.25 desviaciones estándar por debajo de este

7.3. **a.** \$46 400 y \$48 000, que se obtienen mediante el cálculo de \$47 200 ± 1(\$800)
b. \$45 600 y \$48 800, que se obtienen mediante el cálculo de \$47 200 ± 2(\$800)
c. \$44 800 y \$49 600, que se obtienen mediante el cálculo de \$47 200 ± 3(\$800)
d. \$47 200. La media, la mediana y la moda son iguales para una distribución normal
e. Sí; una distribución normal es simétrica

7.4. **a.** Cálculo de z:

$$z = \frac{154 - 150}{5} = 0.80$$

De acuerdo con el apéndice B.3, el área es de 0.2881; así que $P(150 < \text{temp} < 154) = 0.2881$

b. Cálculo de z:

$$z = \frac{164 - 150}{5} = 2.80$$

De acuerdo con el apéndice B.3, el área es de 0.4974; así que $P(164 > \text{temp}) = 0.5000 - 0.4974 = 0.0026$

7.5. **a.** Cálculo de los valores z:

$$z = \frac{146 - 150}{5} = -0.80 \quad \text{y} \quad z = \frac{156 - 150}{5} = 1.20$$

$$P(146 < \text{temp} < 156) = P(-0.80 < z < 1.20) = 0.2881 + 0.3849 = 0.6730$$

b. Cálculo de los valores z:

$$z = \frac{162 - 150}{5} = 2.40 \quad \text{y} \quad z = \frac{156 - 150}{5} = 1.20$$

$$P(156 < \text{temp} < 162) = P(1.20 < z < 2.40) = 0.4918 - 0.3849 = 0.1069$$

7.6. 85.24 (sin duda, el profesor lo convertirá en 85). El área más próxima a 0.4000 es de 0.3997; z equivale a 1.28, por consiguiente:

$$1.28 = \frac{x - 75}{8}$$
$$10.24 = x - 75$$
$$x = 85.24$$

7.7. **a.** 0.0465, que se calcula mediante $\mu = n\pi = 200(0.80) = 160$, y $\sigma^2 = n\pi(1 - \pi) = 200(0.80)(1 - 0.80) = 32$; entonces,

$$\sigma = \sqrt{32} = 5.66$$
$$z = \frac{169.5 - 160}{5.66} = 1.68$$

De acuerdo con el apéndice B.3, el área es de 0.4535; al restar de 0.500, se obtiene 0.0465

b. 0.9686, que se calcula mediante 0.4686 + 0.5000; primero se calcula z:

$$z = \frac{149.5 - 160}{5.66} = 1.86$$

De acuerdo con el apéndice B.3, el área es de 0.4686

7.8. **a.** 0.7769, que se calcula mediante:

$$P(\text{Llegada} < 15) = 1 - e^{-\frac{1}{10}(15)} = 1 - 0.2231 = 0.7769$$

b. 0.0821, que se calcula mediante:

$$P(\text{Llegada} > 25) = e^{-\frac{1}{10}(25)} = 0.0821$$

c. 0.1410, que se calcula mediante:

$$P(15 < x < 25) = P(\text{Llegada} < 25) - P(\text{Llegada} < 15) = 0.9179 - 0.7769 = 0.1410$$

d. 16.09 minutos, que se calcula mediante:

$$0.80 = 1 - e^{-\frac{1}{10}(x)}$$
$$-\ln 0.20 = \frac{1}{10}x$$
$$x = -(-1.609)(10) = 1.609(10) = 16.09$$

CAPÍTULO 8

8.1. **a.** Los estudiantes seleccionados son Price, Detley y Molter
b. Las respuestas varían
c. Se salta y se desplaza al siguiente número aleatorio

8.2. Los estudiantes seleccionados son Berry, Francis, Kopp, Poteau y Swetye

8.3. **a.** 10, que se calcula por:

$$_5C_2 = \frac{5!}{2!(5-2)!}$$

b.

	Servicio	Media muestral
Snow, Tolson	20, 22	21
Snow, Kraft	20, 26	23
Snow, Irwin	20, 24	22
Snow, Jones	20, 28	24
Tolson, Kraft	22, 26	24
Tolson, Irwin	22, 24	23
Tolson, Jones	22, 28	25
Kraft, Irwin	26, 24	25
Kraft, Jones	26, 28	27
Irwin, Jones	24, 28	26

c.

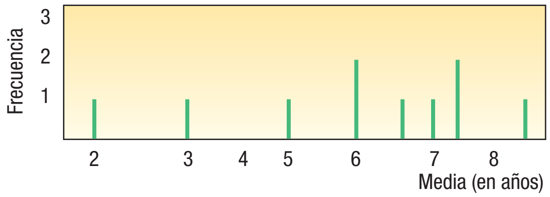
Media	Número	Probabilidad
21	1	0.10
22	1	0.10
23	2	0.20
24	2	0.20
25	2	0.20
26	1	0.10
27	1	0.10
	10	1.00

d. Idénticos: la media de población, μ, es 24, y la media de las medias de la muestra, $\mu_{\bar{x}}$, también es 24
e. Medias muestrales con rango de 21 a 27; valores de la población de 20 a 28
f. No, la población no está uniformemente distribuida
g. Sí

8.4. Las respuestas varían; a continuación se muestra una solución:

	Número de muestra									
	1	2	3	4	5	6	7	8	9	10
	8	2	2	19	3	4	0	4	1	2
	19	1	14	9	2	5	8	2	14	4
	8	3	4	2	4	4	1	14	4	1
	0	3	2	3	1	2	16	1	2	3
	2	1	7	2	19	18	18	16	3	7
Total	37	10	29	35	29	33	43	37	24	17
$\bar{x}$	7.4	2	5.8	7.0	5.8	6.6	8.6	7.4	4.8	3.4

La media de las 10 medias muestrales es 5.88



8.5. $z = \dfrac{31.8 - 31.20}{0.4/\sqrt{16}} = -1.20$

La probabilidad de que z sea mayor que -1.20 es $0.5000 + 0.3849 = 0.8849$; existe más de 88% de probabilidad de que la operación de llenado produzca botellas con al menos 31.08 onzas

CAPÍTULO 9

9.1. **a.** Desconocido; se trata del valor que se desea calcular

b. $20 000, estimador puntual

c. $\$20\,000 \pm 1.960 \dfrac{\$3\,000}{\sqrt{40}} = \$20\,000 \pm \$930$

d. Los puntos extremos del intervalo de confianza son $19 070 y $20 930; aproximadamente 95% de los intervalos construidos de forma similar incluirían la media poblacional

9.2. **a.** $\bar{x} = \dfrac{18}{10} = 1.8 \quad s = \sqrt{\dfrac{11.6}{10-1}} = 1.1353$

b. La media poblacional no se conoce; el mejor estimador es la media de la muestra, 1.8 días

c. $1.80 \pm 2.262 \dfrac{1.1353}{\sqrt{10}} = 1.80 \pm 0.81$

Los puntos extremos son 0.99 y 2.61

d. Se utiliza t porque no se conoce la desviación estándar

e. El valor de 0 no se encuentra en el intervalo; no es razonable concluir que la cantidad media de días de ausencias laborales sea de 0 por empleado

9.3. **a.** $p = \dfrac{420}{1\,400} = 0.30$

b. $0.30 \pm 2.576(.0122) = 0.30 \pm 0.03$

c. El intervalo se encuentra entre 0.27 y 0.33; alrededor de 99% de los intervalos construidos de forma similar incluirían la media poblacional

9.4. $n = \left(\dfrac{2.576(0.279)}{0.05}\right)^2 = 206.6$; la muestra debe redondearse a 207

9.5. $0.375 \pm 1.96 \sqrt{\dfrac{0.375(1-0.375)}{40}} \sqrt{\dfrac{250-40}{250-1}}$

$= 0.375 \pm 1.96(0.0765)(0.9184) = 0.375 \pm 0.138$

CAPÍTULO 10

10.1. **a.** H_0: $\mu = 16.0 \qquad H_1$: $\mu \neq 16.0$

b. 0.05

c. $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

d. Se rechaza H_0 si $z < -1.96$ o $z > 1.96$

e. $z = \dfrac{16.017 - 16.0}{0.15/\sqrt{50}} = \dfrac{0.0170}{0.0212} = 0.80$

f. No se rechaza H_0

g. No se puede concluir que la cantidad media gastada sea distinta a 16 onzas

10.2. **a.** H_0: $\mu \leq 16.0 \qquad H_1$: $\mu > 16.0$

b. Se rechaza H_0 si $z > 1.65$

c. $z = \dfrac{16.040 - 16.0}{0.15/\sqrt{50}} = \dfrac{0.400}{0.0212} = 1.89$

d. Se rechaza H_0

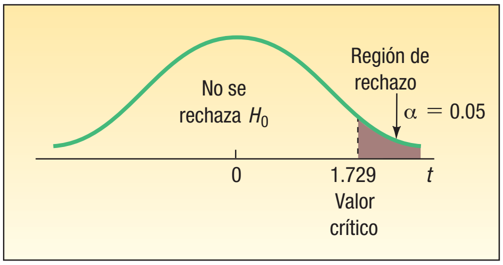
e. La cantidad media gastada es superior a 16.0 onzas

f. Valor $p = 0.5000 - 0.4706 = 0.0294$; el valor p es menor que $\alpha(0.05)$, así que se rechaza H_0 (es la misma conclusión que en la parte *d*)

10.3. **a.** H_0: $\mu \leq 305 \qquad H_1$: $\mu > 305$

b. $gl = n - 1 = 20 - 1 = 19$

La regla de decisión consiste en rechazar H_0 si $t > 1.729$



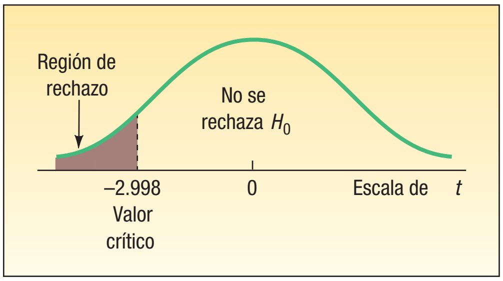
c. $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{311 - 305}{12/\sqrt{20}} = 2.236$

Se rechaza H_0 porque $2.236 > 1.729$; la modificación incrementa la vida media de las baterías a más de 305 días

10.4. **a.** H_0: $\mu \geq 9.0 \qquad H_1$: $\mu < 9.0$

b. 7, que se calcula mediante $n - 1 = 8 - 1 = 7$

c. Se rechaza H_0 si $t < -2.998$



d. $t = -2.494$, que se calcula:

$$s = \sqrt{\frac{0.36}{8-1}} = 0.2268$$

$$\bar{x} = \frac{70.4}{8} = 8.8$$

De esta manera,

$$t = \frac{8.8 - 9.0}{0.2268/\sqrt{8}} = -2.494$$

Como -2.494 se encuentra a la derecha de -2.998, no se rechaza H_0; no se demostró que la media es menor que 9.0

e. El valor p se localiza entre 0.025 y 0.010

10.5. 0.0054, que se encuentra al determinar el área bajo la curva entre 10 078 y 10 180

$$t = \frac{\bar{x}_c - \mu_1}{\sigma/\sqrt{n}}$$

$$= \frac{10\,078 - 10\,180}{400/\sqrt{100}} = -2.55$$

El área bajo la curva para un valor z de -2.55 es 0.4946 (apéndice B.3), y $0.5000 - 0.4946 = 0.0054$

CAPÍTULO 11

11.1. **a.** H_0: $\mu_W \leq \mu_M \qquad H_1$: $\mu_W > \mu_M$

El subíndice W se refiere a las mujeres, y M, a los hombres

b. Se rechaza H_0 si $z > 1.645$

c. $z = \dfrac{\$1\,500 - \$1\,400}{\sqrt{\dfrac{(\$250)^2}{50} + \dfrac{(\$200)^2}{40}}} = 2.11$

d. Se rechaza la hipótesis nula

e. Valor $p = 0.5000 - 0.4826 = 0.0174$

f. La cantidad media vendida por día es mayor para las mujeres

11.2. **a.** H_0: $\mu_d = \mu_a \qquad H_1$: $\mu_d \neq \mu_a$

b. $gl = 6 + 8 - 2 = 12$

Se rechaza H_0 si t es menor que -2.179 o si t es mayor que 2.179

c. $\bar{x}_1 = \dfrac{42}{6} = 7.00 \quad s_1 = \sqrt{\dfrac{10}{6-1}} = 1.4142$

$\bar{x}_2 = \dfrac{80}{8} = 10.00 \quad s_2 = \sqrt{\dfrac{36}{8-1}} = 2.2678$

$$s_p^2 = \frac{(6-1)(1.4142)^2 + (8-1)(2.2678)^2}{6+8-2}$$
$$= 3.8333$$
$$t = \frac{7.00 - 10.00}{\sqrt{3.8333\left(\frac{1}{6}+\frac{1}{8}\right)}} = -2.837$$

d. Se rechaza H_0 porque -2.837 es menor que el valor crítico
e. El valor p es menor que 0.02
f. El número medio de defectos no es el mismo en ambos turnos
g. Poblaciones independientes, las poblaciones siguen la distribución normal, las poblaciones tienen desviaciones estándar iguales

11.3. a. H_0: $\mu_c \geq \mu_a$ $\quad H_1$: $\mu_c < \mu_a$

b. $gl = \dfrac{[(356^2/10)+(857^2/8)]^2}{\dfrac{(356^2/10)^2}{10-1}+\dfrac{(857^2/8)^2}{8-1}} = 8.93$

así que $gl = 8$

c. Se rechaza H_0 si $t < -1.860$

d. $t = \dfrac{\$1\,568 - \$1\,967}{\sqrt{\dfrac{356^2}{10}+\dfrac{857^2}{8}}} = \dfrac{-399.00}{323.23} = -1.234$

e. No se rechaza H_0
f. No hay diferencia entre los saldos medios de la cuenta de los que solicitaron la tarjeta de crédito o fueron contactados por teléfono por un agente

11.4. a. H_0: $\mu_d \leq 0$ $\quad H_1$: $\mu_d > 0$
b. Se rechaza H_0 si $t > 2.998$
c.

Nombre	Antes	Después	d	$(d-\bar{d})$	$(d-d)^2$
Hunter	155	154	1	−7.875	62.0156
Cashman	228	207	21	12.125	147.0156
Mervine	141	147	−6	−14.875	221.2656
Massa	162	157	5	−3.875	15.0156
Creola	211	196	15	6.125	37.5156
Peterson	164	150	14	5.125	26.2656
Redding	184	170	14	5.125	26.2656
Poust	172	165	7	−1.875	3.5156
			71		538.8750

$$\bar{d} = \frac{71}{8} = 8.875$$
$$s_d = \sqrt{\frac{538.857}{8-1}} = 8.774$$
$$t = \frac{8.875}{8.774/\sqrt{8}} = 2.861$$

d. No se rechaza H_0 porque no se puede concluir que los estudiantes bajaron de peso; el valor p es menor que 0.025 pero mayor que 0.01
e. La distribución de las diferencias debe seguir una distribución normal

CAPÍTULO 12

12.1. Suponga que los ensamblados de Mark son la población 1; entonces, H_0: $\sigma_1^2 \leq \sigma_2^2$; H_1: $\sigma_1^2 > \sigma_2^2$; $gl_1 = 10 - 1 = 9$; y gl_2 también es igual a 9. H_0 se rechaza si $F > 3.18$

$$F = \frac{(2.0)^2}{(1.5)^2} = 1.78$$

H_0 no se rechaza; la variación es la misma para ambos empleados

12.2. a. H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: al menos una media de tratamiento es diferente
b. Rechace H_0 si $F > 4.26$
c. $\bar{x} = \dfrac{240}{12} = 20$

$$\text{SS total} = (18-20)^2 + \cdots + (32-20)^2$$
$$= 578$$
$$\text{SSE} = (18-17)^2 + (14-17)^2 + \cdots + (32-29)^2$$
$$= 74$$
$$\text{SST} = 578 - 74 = 504$$

d.

Fuente	Suma de cuadrados	Grados de libertad	Media cuadrática	F
Tratamiento	504	2	252	30.65
Error	74	9	8.22	
Total	578	11		

e. H_0 se rechaza; hay una diferencia entre los números medios de botellas vendidas en las distintas ubicaciones

12.3. a. H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: no todas las medias son iguales
b. H_0 se rechaza si $F > 3.98$
c. $\bar{x}_G = 8.86$, $\bar{x}_1 = 11$, $\bar{x}_2 = 8.75$, $\bar{x}_3 = 6.8$
SS total = 53.71
SST = 44.16
SSE = 9.55

Fuente	Suma de cuadrados	gl	Media cuadrática	F
Tratamiento	44.16	2	22.08	25.43
Error	9.55	11	0.8682	
Total	53.71	13		

d. H_0 se rechaza; las medias de tratamiento difieren
e. $(11.0 - 6.8) \pm 2.201\sqrt{0.8682\left(\frac{1}{5}+\frac{1}{5}\right)} = 4.2 \pm 1.30 = 2.90$ y 5.50
Estas medias de tratamiento difieren debido a que ambos puntos extremos del intervalo de confianza tienen signo igual, que en este problema es positivo

12.4. Para los tipos:
H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: las medias de tratamiento no son iguales
Rechace H_0 si $F > 4.46$
Para los meses:
H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
H_1: las medias de bloqueo no son iguales
Rechace H_0 si $F > 3.84$
He aquí el análisis de la tabla de la varianza:

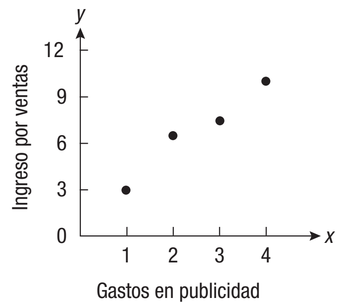
Fuente:	gl	SS	MS	F
Tipos	2	3.60	1.80	0.39
Meses	4	31.73	7.93	1.71
Error	8	37.07	4.63	
Total	14	72.40		

La hipótesis nula no se puede rechazar para cualquier tipo o mes. No hay diferencia entre las ventas medias entre tipos o meses

12.5 a. Hay cuatro niveles del factor A, y el valor p es menor que 0.05, por lo cual las medias del factor A difieren
b. Hay tres niveles del factor B, y el valor p es menor que 0.05, por lo cual las medias del factor B difieren
c. Hay tres observaciones en cada celda, y una interacción entre las medias del factor A y del factor B, debido a que el valor p es menor que 0.05

CAPÍTULO 13

13.1 a. "Gastos en publicidad" es la variable independiente, e "ingreso por ventas", la dependiente
b.



c.

x	y	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
2	7	−0.5	.25	0	0	0
1	3	−1.5	2.25	−4	16	6
3	8	0.5	.25	1	1	0.5
4	10	1.5	2.25	3	9	4.5
10	28		5.00		26	11.0

$$\bar{x} = \frac{10}{4} = 2.5 \quad \bar{y} = \frac{28}{4} = 7$$

$$s_x = \sqrt{\frac{5}{3}} = 1.2910$$

$$s_y = \sqrt{\frac{26}{3}} = 2.9439$$

$$r = \frac{\Sigma(X-\bar{X})(y-\bar{y})}{(n-1)s_x s_y} = \frac{11}{(4-1)(1.2910)(2.9439)}$$
$$= 0.9648$$

d. Hay una correlación fuerte entre los gastos de publicidad y las ventas

13.2. H_0: $\rho \le 0$ H_1: $\rho > 0$. H_0 se rechaza si $t > 1.714$

$$t = \frac{0.43\sqrt{25-2}}{\sqrt{1-(0.43)^2}} = 2.284$$

H_0 se rechaza; hay una correlación positiva entre el porcentaje de los votos recibidos y la cantidad que se gastó en la campaña

13.3. a. Vea los cálculos en la autoevaluación 13.1, inciso c

$$b = \frac{rs_y}{s_x} = \frac{(0.9648)(2.9439)}{1.2910} = 2.2$$

$$a = \frac{28}{4} - 2.2\left(\frac{10}{4}\right) = 7 - 5.5 = 1.5$$

b. La pendiente es 2.2, lo cual indica que un aumento de \$1 millón en publicidad generará un aumento de \$2.2 millones en las ventas. La intersección es 1.5; si no hubiera gastos en publicidad, las ventas serían \$1.5 millones

c. $\hat{Y} = 1.5 + 2.2(3) = 8.1$

13.4. H_0: $\beta_1 \le 0$ H_1: $\beta > 0$. H_0 se rechaza si $t > 3.182$

$$t = \frac{2.2 - 0}{0.4243} = 5.1850$$

Rechace H_0; la recta de la pendiente es mayor a cero

13.5. a.

y	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$
7	5.9	1.1	1.21
3	3.7	−0.7	.49
8	8.1	−0.1	.01
10	10.3	−0.3	.09
			1.80

$$s_{y \cdot x} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$$

$$= \sqrt{\frac{1.80}{4-2}} = 0.9487$$

b. $r^2 = (0.9648)^2 = 0.9308$

c. Los gastos de publicidad representan 90% de la variación de las ventas

13.6. 6.58 y 9.62, dado que $\hat{y}$ para una x de 3 es 8.1, calculado por: $\hat{y} = 1.5 + 2.2(3) = 8.1$, entonces $\bar{x} = 2.5$ y $\Sigma(x-\bar{x})^2 = 5$

t del apéndice B.5 para $4 - 2 = 2$ grados de libertad con el nivel 0.10 es 2.920

$$\hat{y} \pm t(s_{y \cdot x})\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\Sigma(x-\bar{x})^2}}$$

$$= 8.1 \pm 2.920(0.9487)\sqrt{\frac{1}{4} + \frac{(3-2.5)^2}{5}}$$

$$= 8.1 \pm 2.920(0.9487)(0.5477)$$

$$= 6.58 \text{ y } 9.62 \text{ (en millones de dólares)}$$

CAPÍTULO 14

14.1. a. \$389 500, o bien, 389.5 (en miles de dólares); determinado por $2.5 + 3(40) + 4(72) - 3(10) + 0.2(20) + 1(5) = 3\,895$

b. La b_2 de 4 indica que la ganancia aumentará hasta \$4 000 por cada hora extra que abra el restaurante (si no cambia ninguna otra variable). La b_3 de -3 implica que la ganancia disminuirá \$3 000 por cada milla adicional desde el área central (si no cambia ninguna otra variable)

14.2. a. Los grados totales de libertad ($n - 1$) son 25; por lo tanto, el tamaño muestral es 26

b. Hay cinco variables independientes

c. Solo hay una variable dependiente (ganancia)

d. $S_{Y.12345} = 1.414$, determinada por $\sqrt{2}$; 95% de los residuos estará entre -2.828 y 2.828, determinado por $\pm 2(1.414)$

e. $R^2 = 0.714$, determinado por 100/140. De la desviación de la ganancia, 71.4% se debe a estas cinco variables

f. $R^2_{\text{ajust}} = 0.643$, determinado por

$$1 - \left[\frac{40}{(26-(5+1))}\right] \Big/ \left[\frac{140}{(26-1)}\right]$$

14.3. a. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_1: no todas las β son cero

La regla de decisión es rechazar H_0 si $F > 2.71$. El valor calculado de F es 10, determinado por 20/2; por lo tanto, se rechaza H_0, lo que indica que al menos uno de los coeficientes de regresión es diferente de cero

Basados en los valores p, la regla de decisión es rechazar la hipótesis nula si el valor p es menor a 0.05. El valor calculado de F es 10, determinado por 20/2, y tiene un valor p de 0.000; así, se rechaza la hipótesis nula, es decir, cuando menos uno de los coeficientes de regresión es distinto a cero

b. En el caso de la variable 1: H_0: $\beta_1 = 0$ y H_1: $\beta_1 \ne 0$. La regla de decisión es: rechazar H_0 si $t < -2.086$, o si $t > 2.086$. Como 2.000 no sobrepasa estos límites, no se rechaza la hipótesis nula; este coeficiente de regresión puede ser cero. Puede considerar eliminar esta variable; por lógica paralela, se rechaza la hipótesis nula para las variables 3 y 4

Para la variable 1, la regla de decisión es rechazar H_0: $\beta_1 = 0$ si el valor p es menor a 0.05. Como el valor p es 0.056, no se puede rechazar la hipótesis nula. Este coeficiente de regresión podría ser cero; por lo tanto, se puede prescindir de esta variable. Por lógica paralela, se rechazan las hipótesis nulas para las variables 3 y 4

c. Se debe considerar la eliminación de las variables 1, 2 y 5. La variable 5 tiene un valor absoluto menor de t; por lo tanto, elimínela primero y vuelva a elaborar el análisis de regresión

14.4. a. $\hat{y} = 15.7625 + 0.4415x_1 + 3.8598x_2$

$\hat{y} = 15.7625 + 0.4415(30) + 3.8598(1)$

$= 32.87$

b. Las agentes ganan \$3 860 más que los agentes

c. H_0: $\beta_3 = 0$

H_1: $\beta_3 \ne 0$

$gl = 17$, rechace H_0 si $t < -2.110$, o si $t > 2.110$

$$t = \frac{3.8598 - 0}{1.4724} = 2.621$$

El estadístico t excede el valor crítico de 2.110; también, el valor $p = 0.0179$ es menor que 0.05. Rechace H_0; se debería incluir al género en la ecuación de regresión

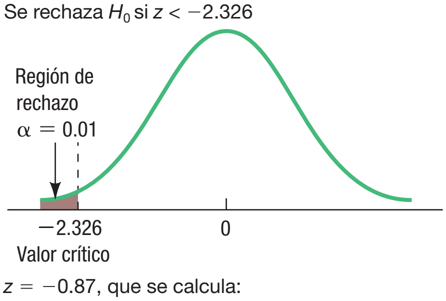
CAPÍTULO 15

15.1. a. Sí, porque tanto $n\pi$ como $n(1 - \pi)$ son mayores a 5: $n\pi = 200(0.40) = 80$ y $n(1 - \pi) = 200(0.60) = 120$

b. H_0: $\pi \ge 0.40$

H_1: $\pi < 0.40$

c. Se rechaza H_0 si $z < -2.326$



d. $z = -0.87$, que se calcula:

$$z = \frac{0.37 - 0.40}{\sqrt{\frac{0.40(1-0.40)}{200}}} = \frac{-0.03}{\sqrt{0.0012}} = -0.87$$

No se rechaza H_0

e. El valor p es de 0.1922, que se calcula mediante 0.5000 − 0.3078

15.2. a. H_0: $\pi_a = \pi_{ch}$
H_1: $\pi_a \neq \pi_{ch}$

b. 0.10

c. Dos colas

d. Se rechaza H_0 si $z < -1.645$ o $z > 1.645$

e. $$p_c = \frac{87 + 123}{150 + 200} = \frac{210}{350} = 0.60$$

$$p_a = \frac{87}{150} = 0.58 \qquad p_{ch} = \frac{123}{200} = 0.615$$

$$z = \frac{0.58 - 0.615}{\sqrt{\dfrac{0.60(0.40)}{150} + \dfrac{0.60(0.40)}{200}}} = -0.66$$

f. No se rechaza H_0

g. Valor p = 2(0.5000 − 0.2454) = 0.5092
No hay diferencia en la proporción de adultos y niños a quienes les gustó el sabor propuesto

15.3. a. Frecuencias observadas

b. Seis (seis días de la semana)

c. 10; total de las frecuencias observadas ÷ 6 = 60/6 = 10

d. 5; $k - 1 = 6 - 1 = 5$

e. 15.086 (de la tabla *ji* cuadrada en el apéndice B.7)

f. $$\chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = \frac{(12 - 10)^2}{10} + \cdots + \frac{(9 - 10)^2}{10} = 0.8$$

g. No se rechaza H_0

h. La evidencia no puede mostrar una diferencia en la proporción de ausencias por día de la semana

15.4. H_0: $P_C = 0.60$, $P_L = 0.30$ y $P_U = 0.10$
H_1: la distribución no se parece a la anterior
Se rechaza H_0 si $\chi^2 > 5.991$

Categoría	f_o	f_e	$\frac{(f_o - f_e)^2}{f_e}$
Actuales	320	300	1.33
Atrasadas	120	150	6.00
Irrecuperables	60	50	2.00
	500	500	9.33

Se rechaza H_0; los datos de las cuentas por cobrar no reflejan el promedio nacional

15.5. a. Tabla de contingencia

b. H_0: no hay relación entre el ingreso y jugar a la lotería
H_1: hay relación entre el ingreso y jugar a la lotería

c. Se rechaza H_0 si χ^2 es mayor que 5.991

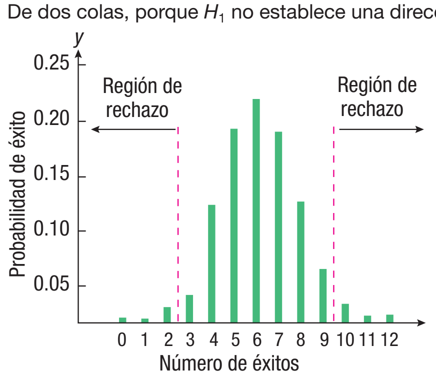
d. $$\chi^2 = \frac{(46 - 40.71)^2}{40.71} + \frac{(28 - 27.14)^2}{27.14} + \frac{(21 - 27.14)^2}{27.14} + \frac{(14 - 19.29)^2}{19.29} + \frac{(12 - 12.86)^2}{12.86} + \frac{(19 - 12.86)^2}{12.86}$$
$$= 6.544$$

e. Se rechaza H_0; hay relación entre el nivel de ingreso y jugar a la lotería

CAPÍTULO 16

16.1. a. De dos colas, porque H_1 no establece una dirección

b.



Al sumar hacia abajo, 0.000 + 0.003 + 0.016 = 0.019; esta es la probabilidad acumulada mayor hasta 0.050 (pero sin excederlo), que es la mitad del nivel de significancia. La regla de decisión es rechazar H_0 si el número de signos más es 2 o menor, o 10 o mayor

c. Rechace H_0; acepte H_1; sí existe una preferencia

16.2. a. H_0: $\pi \leq 0.50$ $\qquad H_1$: $\pi > 0.50$

b. Rechace H_0 si $z > 1.645$

c. Como 80 es mayor que $n/2 = 100/2 = 50$, se emplea:
$$z = \frac{(80 - 0.50) - 0.50(100)}{0.50\sqrt{100}} = \frac{29.5}{5} = 5.9$$

d. H_0 se rechaza

e. la supervisión fue eficaz

16.3. H_0: la mediana ≤ \$2 505 H_1: la mediana es mayor que \$2 505. La regla de decisión es rechazar H_0 si $z > 1.645$
$$z = \frac{(42 - 0.50) - 32}{0.50\sqrt{64}} = \frac{9.5}{4} = 2.38$$
Se rechaza H_0, debido a que 2.38 es mayor que 1.645. La mediana de la cantidad gastada es mayor a \$2 505

16.4. a. $n = 10$; debido a que no hubo cambio para A.A.

b.

Antes	Después	Diferencia	Diferencia absoluta	Rango	R^-	R^+
17	18	−1	1	1.5	1.5	
21	23	−2	2	3.0	3.0	
25	22	3	3	5.0		5.0
15	25	−10	10	8.0	8.0	
10	28	−18	18	10.0	10.0	
16	16	—	—	—	—	—
10	22	−12	12	9.0	9.0	
20	19	1	1	1.5		1.5
17	20	−3	3	5.0	5.0	
24	30	−6	6	7.0	7.0	
23	26	−3	3	5.0	5.0	
					48.5	6.5

H_0: la producción es la misma
H_1: la producción aumentó
La suma de los rangos con signos positivos es 6.5; la suma negativa es 48.5. Del apéndice B.8, prueba de una cola, $n = 10$, el valor crítico es 10. Como 6.5 es menor que 10, se rechaza la hipótesis nula y se acepta la hipótesis alternativa. Los nuevos procedimientos aumentaron la producción

c. No es necesaria una suposición respecto de la forma de la distribución

16.5. H_0: no hay diferencia entre las distancias recorridas por XL-5000 y D2
H_1: hay una diferencia entre las distancias recorridas por XL-5000 y D2
No rechace H_0 si el valor calculado z aparece entre 1.96 y −1.96 (del apéndice B.3); de lo contrario, rechace H_0 y acepte H_1; $n_1 = 8$, el número de observaciones en la primera muestra

XL-5000		D2	
Distancia	Rango	Distancia	Rango
252	4	262	9
263	10	242	2
279	15	256	5
273	14	260	8
271	13	258	7
265	11.5	243	3
257	6	239	1
280	16	265	11.5
Total	89.5		46.5

$W = 89.5$

$$z = \frac{89.5 - \dfrac{8(8+8+1)}{2}}{\sqrt{\dfrac{(8)(8)(8+8+1)}{12}}}$$

$$= \frac{21.5}{9.52} = 2.26$$

Rechace H_0; acepte H_1. Hay evidencia de una diferencia en las distancias recorridas por ambas pelotas de golf

16.6.

Rangos			
Englewood	**West Side**	**Great Northern**	**Sylvania**
17	5	19	7
20	1	9.5	11
16	3	21	15
13	5	22	9.5
5	2	14	8
18			12

$\Sigma R_1 = 89$ $\Sigma R_2 = 16$ $\Sigma R_3 = 85.5$ $\Sigma R_4 = 62.5$

$n_1 = 6$ $n_2 = 5$ $n_3 = 5$ $n_4 = 6$

H_0: las distribuciones de las poblaciones son idénticas

H_1: las distribuciones de las poblaciones no son idénticas

$$H = \frac{12}{22(22+1)}\left[\frac{(89)^2}{6} + \frac{(16)^2}{5} + \frac{(85.5)^2}{5} + \frac{(62.5)^2}{6}\right] - 3(22+1)$$

$$= 13.635$$

El valor crítico de *ji* cuadrada para $k - 1 = 4 - 1 = 3$ grados de libertad es 11.345. Como el valor calculado de 13.635 es mayor que 11.345, se rechaza la hipótesis nula; en conclusión, los índices de movimientos no son iguales

16.7. a.

		Rango			
x	y	x	y	d	d^2
805	23	5.5	1	4.5	20.25
777	62	3.0	9	−6.0	36.00
820	60	8.5	8	0.5	0.25
682	40	1.0	4	−3.0	9.00
777	70	3.0	10	−7.0	49.00
810	28	7.0	2	5.0	25.00
805	30	5.5	3	2.5	6.25
840	42	10.0	5	5.0	25.00
777	55	3.0	7	−4.0	16.00
820	51	8.5	6	2.5	6.25
				0	193.00

$$r_s = 1 - \frac{6(193)}{10(99)} = -0.170$$

b. H_0: $\rho = 0$; H_1: $\rho \neq 0$. Rechace H_0 si $t < -2.306$ o bien $t > 2.306$

$$t = -0.170\sqrt{\frac{10-2}{1-(-0.170)^2}} = -0.488$$

H_0 no se rechaza; no se demostró una relación entre ambas pruebas

CAPÍTULO 17

17.1. 1.

Nación	Cantidad (millones de toneladas)	Índice
China	716.5	807.8
Japón	107.2	120.9
Estados Unidos	88.7	100.0
India	77.6	87.5
Rusia	70.4	79.4

China produjo 707.8% más acero que Estados Unidos

2.

Año	Ingreso promedio por hora	(a) Índice	(b) Índice
1995	$11.65	100.0	90.8
2000	14.02	120.3	109.2
2005	16.13	138.5	125.7
2013	19.97	171.4	155.6

a. El ingreso promedio por hora aumentó 71.4% de 1995 a 2013

b. Promediando el ingreso por hora para los años 1995 y 2000, y utilizando esto como base, se muestra que los ingresos aumentaron 55.6%

17.2. a. $P_1 = (\$85/\$75)(100) = 113.3$

$P_2 = (\$45/\$40)(100) = 112.5$

$P = (113.3 + 112.5)/2 = 112.9$

b. $P = (\$130/\$115)(100) = 113.0$

c. $$P = \frac{\$85(500) + \$45(1\,200)}{\$75(500) + \$40(1\,200)}(100)$$

$$= \frac{\$96\,500}{\$85\,500}(100) = 112.9$$

d. $$P = \frac{\$85(520) + \$45(1\,300)}{\$75(520) + \$40(1\,300)}(100)$$

$$= \frac{\$102\,700}{\$91\,000}(100) = 112.9$$

e. $P = \sqrt{(112.9)(122.9)} = 112.9$

17.3. a. $$P = \frac{\$4(9\,000) + \$5(200) + \$8(5\,000)}{\$3(10\,000) + \$1(600) + \$10(3\,000)}(100)$$

$$= \frac{\$77\,000}{\$60\,600}(100) = 127.1$$

b. El valor de las ventas aumentó 27.1% de 2001 a 2013

17.4. a.

En 2009	
Artículo	**Ponderación**
Algodón	($0.25/$0.20)(100)(0.10) = 12.50
Automóviles	(1 200/1 000)(100)(0.30) = 36.00
Cambio de dinero	(90/80)(100)(0.60) = 67.50
Total	116.00

En 2014	
Artículo	**Ponderación**
Algodón	($0.50/$0.20)(100)(0.10) = 25.00
Automóviles	(900/1 000)(100)(0.30) = 27.00
Cambio de dinero	(75/80)(100)(0.60) = 56.25
Total	108.25

b. La actividad comercial aumentó 16% de 2004 a 2009, y se incrementó 8.25% de 2004 a 2014

17.5. a. $14 637, determinado por ($25 000/170.8)(100)

b. $17 944, determinado por ($41 200/229.6)(100)

c. En términos del periodo base, el salario de Jon fue de $14 637 en 2000 y de $17 944 en 2013; esto indica que su ingreso neto aumentó con una tasa mayor que el precio de alimentos, transporte, etcétera

17.6. $0.51, determinado por ($1.00/195.4)(100); el poder de compra disminuyó $0.49

17.7. Con 2010 como periodo base de ambas series:

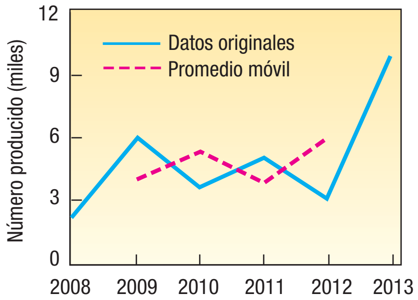
Año	IPI	IPP
2007	111.6	89.9
2008	114.3	96.6
2009	99.5	95.7
2010	100.0	100.0
2011	105.3	103.6
2012	109.4	107.9
2013	111.5	109.4

El Índice de Producción Industrial (IPI) aumentó 11.5% de 2010 a 2013, y el Índice de Precios al Productor (IPP) aumentó 9.4%; el IPI disminuyó de 2007 a 2010 y para 2013 volvió a su valor de 2007. El IPP aumentó uniformemente de 2007 a 2013

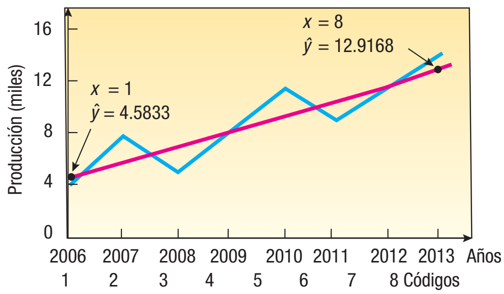
CAPÍTULO 18

18.1.

Año	Producción (miles)	Total móvil de tres años	Promedio móvil de tres años
2008	2	—	—
2009	6	12	4
2010	4	15	5
2011	5	12	4
2012	3	18	6
2013	10	—	—



18.2. a.



b. $\hat{y} = a + bt = 3.3928 + 1.1905t$ (en miles)

c. En 2006:

$$\hat{y} = 3.3928 + 1.1905(1) = 4.5833$$

en 2013:

$$\hat{y} = 3.3928 + 1.1905(8) = 12.9168$$

d. En 2016 $t = 11$, así que

$$\hat{y} = 3.3928 + 1.1905(11) = 16.4883$$

o 16 488 mecedoras

18.3. a.

Año	y	Códigos y	t
2009	2.13	0.3284	1
2010	18.10	1.2577	2
2011	39.80	1.5999	3
2012	81.40	1.9106	4
2013	112.00	2.0492	5

$$b = 0.40945$$
$$a = 0.20081$$

b. Casi 156.7%. El antilogaritmo de 0.40945 es 2.567; al restar 1 se obtiene 1.567

c. Casi 454.5, determinado por $\hat{y} = 0.20081 + 0.40945(6) = 2.65751$. El antilogaritmo de 2.65751 es 454.5

18.4. a. He aquí los valores de un paquete de software; debido al redondeo, se podrían encontrar diferencias en sus cifras:

	Invierno	Primavera	Verano	Otoño
Media	119.35	81.66	125.31	74.24
Estacional habitual	119.18	81.55	125.13	74.13

El factor de corrección es 0.9986

b. Las ventas totales en Teton Village para la temporada de invierno en general están 19.18% arriba del promedio anual

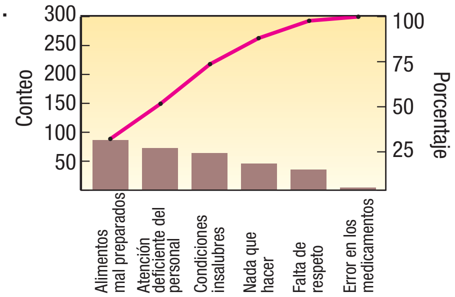
18.5. El valor proyectado para enero del sexto año es 34.9, determinado por:

$$\hat{y} = 4.4 + 0.5(61) = 34.9$$

Al ajustar estacionalmente la proyección, $34.9(120)/100 = 41.88$. Para febrero, $\hat{y} = 4.4 + 0.5(62) = 35.4$; así, $(35.4)(95)/100 = 33.63$

CAPÍTULO 19

19.1.



Porcentaje	84	71	63	45	35	2
acumulado	28	24	21	15	12	0
	28	52	73	88	100	100

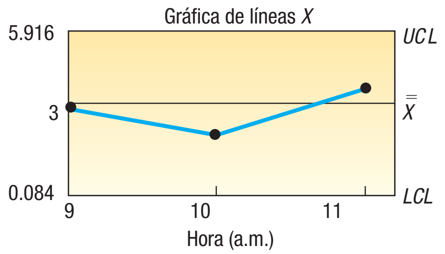
Del total de quejas, 73% se debe a alimentos malos, atención deficiente o condiciones insalubres; estos son los factores que el administrador debe corregir

19.2. a.

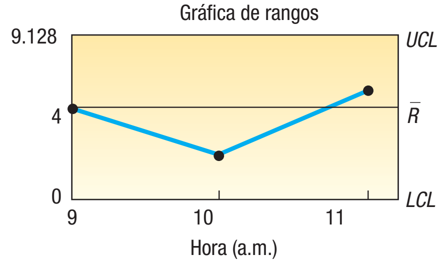
Veces de la muestra 1	2	3	4	Total	Promedio	Rango
1	4	5	2	12	3	4
2	3	2	1	8	2	2
1	7	3	5	16	4	6
					9	12

$$\bar{\bar{x}} = \frac{9}{3} = 3 \qquad \bar{R} = \frac{12}{3} = 4$$

$$LCS \text{ y } LCI = \bar{\bar{x}} \pm A_2\bar{R}$$
$$= 3 \pm 0.729(4)$$
$$LCS = 5.916 \qquad LCI = 0.084$$



$$LCI = D_3\bar{R} = 0(4) = 0$$
$$LCS = D_4\bar{R} = 2.282(4) = 9.128$$



b. Sí, tanto en la gráfica de la media como la gráfica del rango se indica que el proceso está bajo control

19.3. $\bar{c} = \dfrac{25}{12} = 2.083$

$LCS = 2.083 + 3\sqrt{2.083} = 6.413$
$LCI = 2.083 - 3\sqrt{2.083} = -2.247$

Como LCI es negativo, se establece $LCI = 0$; el turno con 7 defectos está fuera de control

19.4. $P(x \leq 2|\pi = 0.30 \text{ y } n = 20) = 0.036$

CAPÍTULO 20

20.1.

Suceso	Pago	Probabilidad del evento	Valor esperado
Mercado a la alza	$2 200	0.60	$1 320
Mercado a la baja	1 100	0.40	440
			$1 760

20.2. a. Suponga que el inversionista compró acciones de Rim Homes y su valor en un mercado a la baja disminuyó a $1 100, como se anticipó (tabla 20.1); en lugar de eso, si el inversionista hubiera comprado acciones de Texas Electronics y el mercado fuera a la baja, el valor de las acciones de Texas Electronics sería $1 150. La diferencia de $50, determinada mediante $1 150 − $1 100, representa el arrepentimiento del inversionista por comprar acciones de Rim Homes

b. Suponga que el inversionista compró acciones de Texas Electronics y después subió el mercado. Las acciones subieron a $1 900, como se anticipó (tabla 20.1); sin embargo, si el inversionista hubiera comprado acciones de Kayser Chemicals y el valor del mercado hubiera aumentado a $2 400 como se anticipó, la diferencia de $500 representaría la ganancia adicional que el inversionista hubiera obtenido al comprar acciones de Kayser Chemicals

20.3.

Evento	Pago	Probabilidad del evento	Valor esperado de la oportunidad
Mercado a la alza	$500	0.60	$300
Mercado a la baja	0	0.40	0
			$300

20.4. a.

Evento	Pago	Probabilidad del evento	Valor esperado
Mercado a la alza	$1 900	0.40	$ 760
Mercado a la baja	1 150	0.60	690
			$1 450

b.

Evento	Pago	Probabilidad del evento	Valor esperado
Mercado a la alza	$2 400	0.50	$1 200
Mercado a la baja	1 000	0.50	500
			$1 700

20.5. Con probabilidades de un mercado a la alza (o a la baja) a 0.333, las acciones de Kayser Chemicals proporcionarían el mayor pago esperado. Con probabilidades de 0.333 a 0.143, las acciones de Rim Homes representarían la mejor compra. Con probabilidades de 0.143 y menores, las acciones de Texas Electronics darían el mayor pago esperado. Las soluciones algebraicas son:

Kayser: $2\,400p + (1 - p)1\,000$
Rim: $\underline{2\,200p + (1 - p)1\,100}$
$1\,400p + 1\,000 = 1\,100p + 1\,100$
$p = 0.333$

Rim: $2\,200p + (1 - p)1\,100$
Texas: $\underline{1\,900p + (1 - p)1\,150}$
$1\,100p + 1\,100 = 750p + 1\,150$
$p = 0.143$

Glosario

Alpha. Probabilidad de un error tipo I o del nivel de significancia; se identifica con la letra griega α.

Análisis de correlación. Grupo de técnicas estadísticas que se emplean para medir la fuerza de la relación entre dos variables.

Análisis de la varianza (ANOVA). Técnica que se utiliza para probar de manera simultánea si las medias de varias poblaciones son iguales; se vale de la distribución *F* como la del estadístico de prueba.

Análisis de la varianza en una dirección de los rangos de Kruskal-Wallis. Prueba que se utiliza cuando no se pueden cumplir las suposiciones del análisis de la varianza (ANOVA) paramétrico y se quiere probar si varias poblaciones son iguales; los datos se deben encontrar al menos en escala ordinal.

Autocorrelación. Los residuos sucesivos en una serie de tiempo se correlacionan.

Bloque. En un ANOVA de dos vías, un bloque se considera la segunda fuente de variación, además de los tratamientos.

Clase. Intervalo en el que se recopilan los datos; por ejemplo, \$4 hasta \$7 constituye una clase; \$7 hasta \$11 es otra.

Coeficiente de correlación. Medida de la fuerza de asociación que se da entre dos variables.

Coeficiente de correlación por rangos de Spearman. Medida de la asociación que se da entre los rangos de dos variables; puede variar de -1.00 a 1.00. En un valor de -1.00 se indica una asociación negativa perfecta entre los rangos, y en un valor de 1.00, una asociación positiva perfecta; en un valor de 0 se indica que no hay asociación entre estos.

Coeficiente de determinación. Proporción de la variación total de la variable dependiente que se explica por la variable independiente; adopta cualquier valor entre 0 y $+1.00$, inclusive. Este coeficiente se estima al elevar al cuadrado el coeficiente de correlación, *r*.

Coeficiente de determinación múltiple. Porcentaje de variación que se encuentra en la variable dependiente *y*, explicado por el grupo de variables independientes, $x_1, x_2, x_3, \ldots, x_k$.

Coeficiente de sesgo1. Por medio de este se mide la falta de simetría de una distribución; como en el caso de una distribución simétrica no existe sesgo, su coeficiente se establece en cero; de lo contrario, puede ser positivo o negativo, con límites ± 3.0.

Colectivamente exhaustivo. Cuando menos uno de los eventos debe ocurrir cuando se realiza un experimento.

Concepto subjetivo de probabilidad. La oportunidad de que un evento se realice con base en estimaciones personales, como presentimientos, opiniones o rumores.

Cuartiles. Valores de un conjunto que se ordenan de manera ascendente y se dividen en cuatro intervalos.

Dato atípico. Se suele ubicar muy lejos de los otros. Una regla aceptada es clasificar una observación como dato atípico si el rango intercuartil está 1.5 veces por encima del tercer cuartil o por debajo del primero.

Deciles. Valores de un conjunto que se ordenan de manera ascendente y se dividen en diez intervalos de frecuencias iguales.

Desviación estándar. Raíz cuadrada de la varianza.

Diagrama de caja. Representación gráfica que muestra la forma general de la distribución de una variable; se basa en cinco estadísticos descriptivos: los valores máximo y mínimo, el primer y tercer cuartiles y la mediana.

Diagrama de dispersión. Técnica gráfica que se emplea para mostrar la relación entre dos variables medidas con escalas de intervalo o de razón.

Diagrama de puntos. Mediante este se resume la distribución de una variable apilando los puntos sobre una línea de puntos en la que se muestran todos los valores de la variable.

Diagrama de tallo y hojas. Método mediante el cual se representa la distribución de una variable utilizando todos los valores; estos se clasifican por medio de su dígito principal; por ejemplo, si un conjunto contiene valores entre 13 y 84, se utilizarían para los tallos ocho clases basadas en los dígitos de las decenas (las unidades corresponderían a las hojas).

Distribución de frecuencias. Agrupamiento de datos en clases en el que se muestra la cantidad de estos en cada una de las clases mutuamente excluyentes y colectivamente exhaustivas.

Distribución de frecuencias relativas. En esta se muestra la fracción o parte del total de observaciones de cada clase.

Distribución de Poisson. Se emplea con frecuencia para aproximar probabilidades binomiales cuando *n* es grande y π es pequeño.

Distribución de probabilidad. Se define como la lista de posibles resultados de un experimento y la probabilidad asociada con cada uno de ellos.

Distribución de probabilidad binomial. Se basa en una variable aleatoria discreta; se caracteriza por lo siguiente:

1. Cada resultado se puede clasificar en una o dos categorías mutuamente excluyentes.
2. La distribución se determina al contar el número de éxitos.
3. Cada prueba se realiza de manera independiente, lo que significa que la respuesta a la prueba 1 (correcta o incorrecta) no afecta la respuesta a la prueba 2.

Distribución de probabilidad normal. Se caracteriza por ser continua y en forma de campana; se describe completamente por su media y su desviación estándar. La mitad de las observaciones son menores que la media.

Distribución exponencial. Se caracteriza por ser continua y de sesgo positivo; se describe por el parámetro de rango, λ.

Distribución *F*. Se emplea como estadístico de prueba en los problemas ANOVA y de otro tipo; se caracteriza por lo siguiente:

1. Nunca es negativa.
2. Es una distribución continua que se aproxima al eje *X*, pero nunca lo toca.
3. Tiene sesgo positivo.
4. Se basa en dos conjuntos de grados de libertad.
5. Al igual que la distribución *t*, hay una familia de distribuciones *F*. Existe una distribución con 17 grados de libertad en

el numerador y 9 en el denominador, hay otra distribución F con 7 grados de libertad en el numerador y 12 en el denominador, y así sucesivamente.

Distribución *ji* cuadrada. Se caracteriza por lo siguiente:

1. Su valor solo puede ser positivo.
2. Hay una familia de distribuciones *ji* cuadrada, una diferente por cada grado de libertad distinto.
3. Las distribuciones tienen sesgo positivo, pero, a medida que aumenta el número de grados de libertad, la distribución se aproxima a la distribución normal.

Distribución hipergeométrica. Clasifica probabilidades con base en una variable aleatoria discreta. Se caracteriza, más que nada, por tener un número fijo de pruebas, solo dos posibles resultados y porque la probabilidad de éxito no es la misma en cada prueba.

Distribución muestral de medias de la muestra. Clasificación de probabilidad en la que se encuentran todas las posibles medias de muestras de tamaño determinado seleccionadas de la población.

Distribución *t*. William S. Gosset la investigó, reportó y publicó en 1908 con el seudónimo Student; se considera similar a la distribución normal estándar que se presentó en el capítulo 7. Se caracteriza por lo siguiente:

1. Es una distribución continua.
2. Puede adoptar valores entre menos infinito y más infinito.
3. Es simétrica respecto de su media de cero; sin embargo, está más dispersa y es más plana en el ápice que la distribución normal estándar.
4. Se aproxima a la distribución normal estándar cuando n aumenta.
5. Existe una familia de distribuciones t. Existe una distribución t en una muestra de 15 observaciones, otra en 25, y así sucesivamente.

Distribución uniforme. Se caracteriza por ser continua y de forma rectangular; se le describe de forma completa por sus valores mínimo y máximo.

Ecuación de regresión. Mediante esta se expresa la relación lineal entre dos variables.

Ecuación de regresión lineal. Por medio de esta se define la relación entre dos variables; tiene la forma $y = a + bx$. Se emplea para predecir y con base en un valor x seleccionado; y es la variable dependiente, y x, la independiente.

Ecuación de regresión múltiple. Es la relación que se da entre la forma de una ecuación matemática y diversas variables independientes y una variable dependiente; la forma general es $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_kx_k$. Se utiliza para estimar y con k variables independientes, x_i.

Error de muestreo. Se define como la diferencia entre un estadístico muestral y el parámetro poblacional correspondiente; por ejemplo: el ingreso medio muestral es de \$22 100; la media poblacional es de \$22 000. El error de muestreo es: \$22 100 − \$22 000 = \$100; este error se puede atribuir al muestreo, es decir, al azar.

Error estándar de estimación. Mediante este se mide la dispersión de los valores y reales respecto de la recta de regresión; se reporta en las mismas unidades que la variable dependiente.

Error tipo I. Se comete cuando se rechaza una H_0 verdadera.

Error tipo II. Se comete cuando se acepta una H_0 falsa.

Estadística. Ciencia encargada de recolectar, organizar, analizar e interpretar datos numéricos con el fin de que se tomen decisiones más efectivas.

Estadística descriptiva. Técnicas que se emplean para describir las características importantes de un conjunto de datos; estos pueden organizarse en una distribución de frecuencias, e incluir el cálculo de medidas de dispersión y sesgos.

Estadística inferencial (o, inferencia estadística). Se relaciona con el cálculo de un parámetro basado en la estadística de una muestra; por ejemplo, si una muestra de 10 calculadoras solares TI-36X reveló que 2 estaban defectuosas, se puede inferir que 20% de la producción es defectuosa.

Estadístico. Característica de una muestra.

Estadístico de prueba. Un valor, determinado por la información de la muestra, que se utiliza para determinar si se rechaza la hipótesis nula.

Estimado conjunto de la varianza de la población. Promedio ponderado de s_1^2 y s_2^2 para estimar la varianza común σ^2, cuando se utilizan muestras pequeñas para probar la diferencia entre dos medias poblacionales.

Estimador de intervalo. Con base en información de la muestra es probable que en este se localice un parámetro de población; por ejemplo, de acuerdo con los datos de la muestra, la media de la población se encuentra en el intervalo entre 0.9 y 2.0 libras.

Estimador puntual. Valor único que se determina a partir de una muestra para calcular un parámetro poblacional; por ejemplo: si la media de la muestra es de 1 020, este constituye el mejor estimador de la media de la población.

Evento. Conjunto de uno o más resultados de un experimento.

Exhaustivo. Cada observación se debe ubicar en alguna de las categorías.

Experimento. Actividad que se observa o se mide.

Factor de corrección de continuidad. Se utiliza para mejorar la exactitud de la aproximación de una distribución discreta por medio de una distribución continua.

Factor de corrección de una población finita (FCP). Cuando se lleva a cabo un muestreo sin reemplazo a partir de una población finita, se utiliza un término de corrección para reducir el error estándar de la media, de acuerdo con el tamaño relativo de la muestra respecto del tamaño de la población. El factor de corrección se aplica cuando la muestra constituye más de 5% de una población finita.

Factor de inflación de la varianza. Prueba mediante la cual se detecta la correlación entre variables independientes.

Fórmula de la multiplicación. Se emplea para contar el número de posibles resultados de un experimento; establece que si hay m formas de hacer algo y n formas de hacer otra cosa, hay $(m * n)$ formas de hacer ambas.

Fórmula de las combinaciones. Se utiliza para enumerar los posibles arreglos cuando el orden de los resultados no es importante; por ejemplo, el resultado {a, b, c} se considera el mismo que {c, b, a}.

Fórmula de las permutaciones. Con esta se cuenta el número de posibles resultados, cuando el orden de los resultados es importante; por ejemplo, el resultado {a, b, c} se considera distinto a {c, b, a}.

Frecuencia de clase. Se define como el número de observaciones de cada clase; si se realizan 16 de la clase de \$4 hasta \$6, 16 es la frecuencia de clase.

Grados de libertad. Número de elementos de una muestra que tienen libertad para variar. Suponga que hay dos elementos en una muestra y se conoce la media. Se tiene libertad de especificar solo uno de ambos valores, debido a que el otro se determina de manera automática (pues el total de los dos es el doble de la media). Ejemplo: si la media es \$6, se tiene libertad de elegir solo un valor. Si elige \$4, el otro valor es \$8, porque \$4 + \$8 = 2*(\$6); por lo tanto, hay 1 grado de libertad en este ejemplo. Se pueden determinar los grados de libertad mediante $n - 1 = 2 - 1 = 1$. Si n es 4, hay 3 grados de libertad, determinados por $n - 1 = 4 - 1 = 3$.

Gráfica de barras. En esta se muestran las clases cualitativas en el eje horizontal y las frecuencias de clase en el eje vertical. Las frecuencias de clase son proporcionales a las alturas de las barras.

Gráfica de pastel. En esta se muestra la proporción o porcentaje del número total de frecuencias que cada clase representa.

Gráficas. Formatos especiales de representación que se utilizan para mostrar una distribución de frecuencias, incluyendo histogramas, polígonos de frecuencias y polígonos de frecuencias acumulativas. Otros dispositivos gráficos que se emplean para representar datos son las gráficas de barras y las gráficas de pastel.

Hipótesis. Declaración o afirmación que se hace acerca del valor de un parámetro de la población. Ejemplos: 40.7% de todas las personas de 65 años o mayores viven solas. El número medio de personas en un automóvil es 1.33.

Hipótesis alternativa. Conclusión que se acepta cuando se demuestra que la hipótesis nula es falsa; también se denomina hipótesis de investigación.

Hipótesis nula. Declaración que se hace en relación con el valor del parámetro poblacional; se compara para probarla ante la evidencia numérica (se escribe H_0).

Histograma. Representación gráfica de una frecuencia o una distribución de frecuencias relativas; en el eje horizontal se muestran las clases. La frecuencia o frecuencia relativa de cada clase se muestra mediante la altura vertical de barras adyacentes.

Homoscedasticidad. El error estándar de estimación es el mismo para todos los valores ajustados de la variable dependiente.

Independencia. La probabilidad de un evento no influye en la probabilidad de que ocurra otro evento.

Índice de Precios al Consumidor. Se reporta de manera mensual mediante el U.S. Department of Labor; en este se describe el cambio de precios de una canasta básica de bienes y servicios del periodo base 1982-1984 al presente.

Índice ponderado. Los precios en el periodo base y el periodo dado se multiplican por cantidades (ponderaciones).

Índice simple. Se define como el valor en el periodo dado dividido entre el valor en el periodo base. En general, el resultado suele multiplicarse por 100 y reportarse como porcentaje.

Interacción. Caso en el cual una variable independiente (como x_2) afecta la relación que se da entre otra variable independiente (x_1) y la variable dependiente (y).

Intervalo de confianza. Rango de valores construidos a partir de otros más simples, de modo que es posible que el parámetro poblacional se presente dentro de dicho rango en una probabilidad especificada.

Ley de los grandes números. En un gran número de pruebas, la probabilidad empírica de un evento se aproximará a su probabilidad real.

Matriz de correlación. Enumeración de todos los coeficientes de correlación simples posibles; se compone de las correlaciones entre todas las variables.

Media aritmética. Se define como la suma de valores dividida entre el número de estos; el símbolo de la media de una muestra es μ, y el símbolo de una media poblacional es m.

Medida de dispersión. Valor en el cual se muestra la propagación de los datos; por ejemplo, el rango, la varianza y la desviación estándar.

Media geométrica. Se define como la e-ésima raíz del producto de los valores; es de particular utilidad para promediar razones de cambio y números indicadores. Minimiza la importancia de los valores extremos. Una segunda aplicación de la media geométrica se relaciona con determinar el cambio porcentual anual medio durante cierto periodo de tiempo.

Media ponderada. A cada valor se le asigna un peso relativo (a veces una frecuencia); cada uno se multiplica por el peso y después se suman los valores ponderados. La suma se divide entre la suma de los pesos.

Mediana. Valor de la observación media después de que un grupo de estas se ordenaron de menor a mayor; por ejemplo, si 6, 9 y 4 se ordenan (4, 6 y 9), la mediana es 6, el valor medio.

Medida de ubicación. Número mediante el cual se indica un solo valor típico de los datos; señala al centro de una distribución. La media aritmética, la media ponderada, la mediana, la moda y la media geométrica son medidas de ubicación.

Método de mínimos cuadrados. Técnica mediante la cual se llega a la ecuación de regresión minimizando la suma de los cuadrados de las distancias verticales entre los valores y actuales y los anticipados.

Moda. Valor que se presenta con mayor frecuencia en un conjunto de datos. En el caso de datos agrupados, es el punto medio de la clase que contiene el máximo número de estos.

Muestra. Porción, o subconjunto, de la población que se estudia.

Muestra probabilística. Conjunto de elementos o individuos elegidos de manera que cada miembro de la población cuente con la misma posibilidad de que se le incluya en la muestra.

Muestras dependientes. Se caracterizan por una medición, después algún tipo de intervención, seguida por otra medición. Las muestras pareadas también son dependientes debido a que el mismo individuo o elemento es un miembro de ambas muestras. Ejemplo: diez participantes en un maratón se pesaron antes y después de competir en la carrera, y se desea estudiar la cantidad media de pérdida de peso.

Muestras independientes. Se eligen al azar y no están relacionadas entre sí; por ejemplo, se desea estudiar la edad media de los empleados de Google y Yahoo; entonces, se selecciona una muestra de 28 empleados de Google y una muestra de 19 empleados de Yahoo. Como una persona no puede estar empleada en ambas compañías, las muestras son independientes, es decir, no se relacionan.

Muestreo aleatorio estratificado. Una población primero se divide en subgrupos denominados estratos; luego se elige una muestra de cada estrato. Si, por ejemplo, la población de interés consta de todos los estudiantes universitarios, el diseño de la muestra puede indicar que formen parte de la muestra 62 estudiantes de primer año, 51 de segundo, 40 de tercero y 39 del último grado.

Muestreo aleatorio simple. Esquema en el que cada miembro de la población posee la misma probabilidad de que se le seleccione como parte de la muestra.

Muestreo aleatorio sistemático. Si la población se ordena de cierta forma, ya sea alfabética, por estatura o en un archivero, se selecciona un punto de partida aleatorio; después, cada k-ésimo elemento se convierte en miembro de la muestra. Si el diseño de una muestra requiere que se entreviste a cada novena familia en Main Street comenzando con el 932 de esa calle, la muestra constaría de familias de los números 932, 941, 950 de Main, etcétera.

Muestreo por conglomerados. Método común que se emplea para reducir el costo del muestreo si la población se encuentra dispersa en un área geográfica amplia. El área se divide en pequeñas unidades (condados, distritos, manzanas, etc.), denominadas unidades primarias; después se eligen unas cuantas unidades primarias y se selecciona una muestra aleatoria de cada una.

Multicolinealidad. Condición que se presenta en el análisis de regresión múltiple si las variables independientes se correlacionan entre sí.

Mutuamente excluyente. La ocurrencia de un evento significa que ninguno de los otros eventos puede ocurrir al mismo tiempo.

Nivel de intervalo de medición. Si una observación es mayor que otra por cierta cantidad, y el punto cero es arbitrario, se considera que la medición está en una escala de intervalo; por ejemplo, la diferencia entre temperaturas de 70 grados y 80 grados es de 10 grados. Igualmente, una temperatura de 90 grados es 10 grados mayor que una temperatura de 80 grados, y así sucesivamente.

Nivel de razón de medición. Se presenta cuando las distancias entre números son de cierto tamaño constante conocido, *existe* un punto cero real y la razón entre dos valores es significativa, la medida es de escala de razón; por ejemplo, la distancia entre $200 y $300 es $100, y en el caso del dinero, existe un punto cero real. Si se tienen cero dólares, no hay dinero (no se tiene nada); asimismo, la razón entre $200 y $300 es significativa.

Nivel de significancia. Se considera como la probabilidad de rechazar la hipótesis nula cuando es verdadera.

Nivel nominal de medición. Grado "más bajo" de medición. Estos datos solo se clasifican en categorías, sin un orden particular de ellas; por ejemplo, no hay ninguna diferencia si las categorías "hombre" y "mujer" se listan en ese orden, o primero mujer y luego hombre. Las categorías son mutuamente excluyentes, lo que quiere decir, en esta ilustración, que una persona no puede ser un hombre y una mujer al mismo tiempo.

Nivel ordinal de medición. Los datos que se pueden clasificar se conocen como mediciones ordinales; por ejemplo, la respuesta del consumidor al sonido de una nueva bocina puede ser excelente, muy bueno, regular o malo.

Número de índice. Mediante este se expresa el cambio relativo en precio, cantidad o valor comparado con un periodo base.

Parámetro. Característica de una población.

Percentiles. Valores de un conjunto, ordenados de manera ascendente, que se dividen en cien intervalos.

Población. Colección o conjunto de individuos, objetos o medidas cuyas propiedades se estudian.

Probabilidad. Valor entre 0 y 1, inclusive, que indica hasta qué grado se podría dar un evento.

Probabilidad clásica. Se basa en el supuesto de que cada uno de los resultados tiene la misma oportunidad de suceder.

Probabilidad condicional. Se considera como la posibilidad de que un evento ocurra dado que haya ocurrido ya otro evento.

Probabilidad conjunta. Mediante esta se mide la posibilidad de que dos o más eventos ocurran conjuntamente.

Probabilidad empírica. Concepto probabilístico que se basa en la información histórica.

Probabilidad posterior. Esta se ha revisado con base en información adicional.

Probabilidad principal. Con base en el nivel de información actual, es la primera que se da.

Proporción. Fracción del porcentaje de una muestra o una población con una asimetría particular. Si a 5 de 50 en una muestra les gustó un nuevo cereal, la proporción es 5/50, o bien, 0.10.

Prueba de bondad de ajuste *ji* cuadrada. Su objetivo es determinar el ajuste de un conjunto observado de frecuencias a un conjunto esperado de frecuencias; se relaciona con una variable de escala nominal, como el color de un automóvil.

Prueba de dos colas. Se emplea cuando la hipótesis alternativa no indica una dirección, como H_1: $\mu \neq 75$, y se lee "la media poblacional no es igual a 75". Existe una región de rechazo en cada cola.

Prueba de una cola. Se emplea cuando la hipótesis alternativa indica una dirección, como H_1: $\mu > 40$, y se lee "la media poblacional es mayor que 40". En este caso la región de rechazo se encuentra solo en una cola (la derecha).

Prueba de hipótesis. Procedimiento estadístico con base en evidencia muestral y teoría de la probabilidad, para determinar si es razonable la declaración acerca del parámetro poblacional.

Prueba de los signos. Se utiliza en muestras dependientes para determinar si hay una preferencia por una marca entre dos productos o si es mejor el desempeño después de un experimento que antes de él; además, la prueba de los signos se utiliza para probar una hipótesis respecto de la mediana.

Prueba de Wilcoxon de los rangos. Es no paramétrica y para realizarla se requieren al menos datos de nivel ordinal (es decir, deben tener la posibilidad de clasificarse) y muestras dependientes. La prueba se utiliza cuando no se cumplen los supuestos de la prueba t de Student; su objetivo es descubrir si es posible considerar que dos muestras independientes provienen de la misma población. Se usa si no se cumplen las suposiciones que requiere la prueba t por pares.

Prueba global. Medainte esta se determina si alguna de las variables del conjunto de variables independientes tiene coeficientes de regresión diferentes de cero.

Pruebas no paramétricas o sin distribución. Análisis de hipótesis que comprenden datos de nivel nominal u ordinal. No es necesario hacer suposiciones acerca de la forma de la distribución de la población; es decir, no se supone que la población está normalmente distribuida.

Punto medio. Valor que divide a la clase en dos partes iguales. En las clases que van de $10 hasta $20 y de $20 hasta $30, los puntos medios se encuentran en $15 y $25, respectivamente.

Rango. Medida de dispersión que se encuentra al restar el valor mínimo del valor máximo.

Rango intercuartil. Valor absoluto de la diferencia numérica entre el primer y tercer cuartiles; 50% de los valores de una distribución se presenta en este rango.

Regla empírica. Para una distribución de frecuencias simétrica, con forma de campana, aproximadamente 68% de las observaciones residen dentro de ± 1 desviación estándar de la media; cerca de 95% de estas se encuentran dentro de ± 2 desviaciones estándar de la media; y prácticamente todas (99.7%) se ubican dentro de ± 3 desviaciones estándar de la media.

Regla especial de la adición. Se utiliza para encontrar la probabilidad de eventos constituidos por A o B cuando los eventos son mutuamente excluyentes.

Regla especial de la multiplicación. Se utiliza para encontrar las probabilidades de la ocurrencia conjunta de eventos independientes.

Regla general de la adición. Se utiliza para determinar las probabilidades de eventos complejos compuestos por *A* o *B* cuando los eventos no son mutuamente excluyentes.

Regla general de la multiplicación. Se utiliza para determinar probabilidades de eventos *A* y *B* que ocurren cuando los eventos no son independientes.

Regresión por pasos. Proceso gradual mediante el cual se determina la ecuación de regresión; solo las variables independientes con coeficientes de regresión distintos de cero están en la ecuación de regresión. Se agrega una variable independiente a la vez a la ecuación de regresión.

Residuo. Diferencia que se halla entre el valor real de la variable dependiente y el valor estimado de la variable dependiente, es decir, $y - \hat{y}$.

Resultado. Se define como la observación o medición particular de un experimento.

Sesgo. Posible consecuencia que se da al negar a determinados miembros de la población la oportunidad de ser seleccionados para la muestra; como resultado, la muestra puede no ser representativa de la población.

Tabla de contingencia. Se utiliza para clasificar observaciones de acuerdo con dos características. Si dos características, como el género y el grado más alto otorgado a una muestra de corredores de bolsa, se clasifican en forma cruzada en una tabla, el resultado se denomina tabla de contingencia.

Tabla de frecuencias. Agrupación de datos cualitativos en clases mutuamente excluyentes y colectivamente exhaustivas en la que se muestra el número de estos en cada clase.

Tendencia secular. Se define como la dirección de largo plazo suavizada de una serie de tiempo.

Teorema central del límite. A medida que el tamaño de la muestra aumenta, la distribución de la media muestral se aproximará a la distribución normal sin importar la forma de la población.

Teorema de Bayes. El reverendo Bayes lo formuló en el siglo VIII, se diseñó para determinar la probabilidad de que ocurra un evento *A,* dado que haya ocurrido otro evento *B*.

Teorema de Chebyshev. Para cualquier grupo de valores (muestra o población), la proporción de estos que reside dentro de desviaciones estándar *k* de la media es al menos $1 - 1/k^2$, donde *k* es cualquier valor mayor a 1.

Tratamiento. Fuente de variación; mediante esta se identifican las diversas poblaciones que se están examinando.

Valor crítico. Indica el punto de división entre la región donde la hipótesis nula no se rechaza y la región donde se rechaza.

Valor de la información perfecta. Diferencia que se halla entre el pago máximo bajo condiciones de certidumbre y el pago máximo en situación de incertidumbre.

Valor *p*. Probabilidad de calcular un valor del estadístico de prueba por lo menos tan extremo como el que se encuentra en los datos muestrales cuando la hipótesis nula es verdadera.

Valor *z*. Para una variable aleatoria que se encuentra casi normalmente distribuida, es la distancia entre un valor seleccionado y la media poblacional medida en unidades de desviación estándar.

Variable aleatoria. Se define por una distribución de probabilidades; puede tener diferentes valores y su posibilidad se determina por su distribución de probabilidades.

Variable aleatoria continua. Adopta una infinidad de valores dentro de un intervalo.

Variable aleatoria discreta. Adopta solo ciertos valores claramente separados.

Variable dependiente. Es la que se predice o estima.

Variable ficticia. Es cualitativa; asume solo uno de dos resultados posibles.

Variable independiente. Proporciona la base para la estimación.

Variables cualitativas. De escala nominal, se codifica para asumir solo uno de dos resultados posibles; por ejemplo, una persona se considera empleada o desempleada.

Variación aleatoria. La suma de las diferencias al cuadrado entre cada observación y su media de tratamiento.

Variación asignable. No es aleatoria; se puede eliminar o reducir cuando se investiga el problema y se encuentra la causa.

Variación casual. Es de naturaleza aleatoria, no se puede eliminar por completo a menos que ocurra un cambio mayor en las técnicas, tecnologías, métodos, equipo o materiales utilizados en el proceso.

Variación cíclica. Aumento y disminución de una serie de tiempo que se da durante periodos mayores de un año.

Variación de tratamiento. Suma de las diferencias al cuadrado que se presenta entre cada media de tratamiento y la gran media o media total.

Variación episódica. Es de naturaleza aleatoria, pero su causa se puede identificar.

Variación estacional. Patrones de cambio en una serie de tiempo en un año; estos se repiten cada año.

Variación irregular. Cambio de naturaleza aleatoria que se observa en una serie de tiempo y que no se repite regularmente.

Variación residual. Cambio de naturaleza aleatoria que no se puede identificar ni predecir.

Variación total. Se define como la suma de las diferencias al cuadrado entre cada observación y la media total.

Varianza. Se considera la media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética.

Créditos fotográficos

Capítulo 1
Página 1: Cortesía de Barnes & Noble; p. 2: Gregor Schuster/Getty Images; p. 7: Rachel Epstein/The Image Works.

Capítulo 2
Página 16: José Luis Peláez/Getty Images; p. 17: Justin Sullivan/Getty Images; p. 18: Steve Cole/Getty Images.

Capítulo 3
Página 45: Andy Lyons/Getty Images; p. 46: Digital Vision/Getty Images; p. 48: Bloomberg vía Getty Images; p. 62: Spencer Grant/Photo Edit.

Capítulo 4
Página 82: Randy Fairs/Corbis; p. 87: Somos/Veer/Getty Images; p. 90: Ramin Talaie/Getty Images; p. 99: Steve Mason/Getty Images.

Capítulo 5
Página 116: Karin Slade/Getty Images; p. 117: Franziska Tetzner/Corbis; p. 125: whiteboxmedia limited/Alamy; p. 127: Tony Arruza/Corbis; p. 139: The McGraw-Hill Companies, Inc./Marker Dieker, fotógrafo.

Capítulo 6
Página 154: JGI/Jamie Grill/Getty Images; p. 159: Thinkstock/Jupiterimages; p. 162: David Madison/Digital Vision/Getty Images RF; p. 171: Howard Berman/Getty Images.

Capítulo 7
Página 184: Ilene MacDonald/Alamy RF; p., 185: C. Sherbume/PhotoLink/Getty Images; p. 199: Jupiterimages/ Getty Images RF; p. 205: zumaamericasse-ven871062/Newscom.

Capítulo 8
Página 220: JB Reed/Landov; p. 222: David Epperson/Getty Images RF.

Capítulo 9
Página 249: Jack Hollingsworth/Photodisc/Getty Images RF; p. 251: Andersen Ross/Getty Images; p. 254: Cortesía de Del Monte Corporation; p. 261: The McGraw-Hill Companies, Inc./Andrew Resek, fotógrafo; p. 264: Rich Pedrocelli/AP Photo.

Capítulo 10
Página 281: Photo Source Hawaii/Alamy; p. 282: Russell Illig/Getty Images; p. 284: Jim Stern/Bloomberg vía Getty Images; p. 289: Robert Nicholas/Getty Images; p. 291: Gene J. Puskar/AP Photo.

Capítulo 11
Página 310: © JGI/Blend Images LLC RF; p. 311: John Lund/Drew Kelly/Blend Images LLC; p. 313: © Debbie Noda/Modesto Bee/ZUMAPRESS.com/Alamy; p. 325: Photodisc/Getty Images RF.

Capítulo 12
Página 338: Ilustración por Alexander Hassenstein/Getty Images; p. 340: The McGraw-Hill Companies, Inc./John Flournoy, fotógrafo; p. 340: Daniel Acker/Bloomberg/Getty Images; p. 355: John A. Rizzo/Getty Images RF.

Capítulo 13
Página 380: Ingram Publishing/Superstock RF; p. 392: Image Source/Getty Images RF.

Capítulo 14
Página 424: Image Source/Getty Images RF.

Capítulo 15
Página 474: Najiah Feanny/Corbis; p. 479: Digital Vision/PunchStock RF; p. 483: © Ian Dagnall/Alamy; p. 483: © Blend Images/Getty Images RF.

Capítulo 16
Página 505: ITAR-TASS/Landov; p. 506: Foto cortesía de Nestlé; p. 509: © Corbis, derechos reservados, RF; p. 514: Blend Images/Getty Images RF.

Capítulo 17
Página 539: Steve Cole/Photodisc/Getty Images RF; p. 540: Image Source/Getty Images RF; p. 554: Image Ideas Inc./Picture Quest.

Capítulo 18
Página 567: Bob Levey/Getty Images; p. 568: Flying Colours Ltd/Photodisc/Getty Images RF; p. 574: © First Light/Alamy; p. 581: Daniel Belenguer/Alamy; p. 592: Arthur Tilley/Getty Images.

Capítulo 19
Página 605: Jerry Lampen/Landov; p. 607: Cortesía de The National Institute of Standards and Technology, Office of Quality Programs, Gaithersburg, MD; p. 613: © Kevpix/Alamy; p. 625: Ingram Publishing.

Capítulo 20 (en el sitio: www.mhhe.com/uni/lind_ae16e)
Página 20.1: Mark Horn/Getty Images; p. 20.2: Gary C. Knapp/AP Photo.

Índice analítico

S

T

CAPÍTULO 3

- Media poblacional

$$\mu = \frac{\Sigma x}{n} \quad [3.1]$$

- Media muestral

$$\bar{x} = \frac{\Sigma x}{n} \quad [3.2]$$

- Media ponderada

$$\bar{x}_w = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} \quad [3.3]$$

- Media geométrica

$$GM = \sqrt[n]{(x_1)(x_2)(x_3)\cdots(x_n)} \quad [3.4]$$

- Tasa de incremento con el tiempo

$$GM = \sqrt[n]{\frac{\text{Valor al final del periodo}}{\text{Valor al inicio del periodo}}} - 1.0 \quad [3.5]$$

- Rango

$$\text{Rango} = \text{valor máximo} - \text{valor mínimo} \quad [3.6]$$

- Varianza de la población

$$\sigma^2 = \frac{\Sigma (x - \mu)^2}{N} \quad [3.7]$$

- Desviación estándar de la población

$$\sigma = \sqrt{\frac{\Sigma (x - \mu)^2}{N}} \quad [3.8]$$

- Varianza muestral

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1} \quad [3.9]$$

- Desviación estándar de la muestra

$$s = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n - 1}} \quad [3.10]$$

- Media aritmética de datos agrupados

$$\bar{x} = \frac{\Sigma fM}{n} \quad [3.11]$$

- Desviación estándar, datos agrupados

$$s = \sqrt{\frac{\Sigma f(M - \bar{x})^2}{n - 1}} \quad [3.12]$$

CAPÍTULO 4

- Localización de un percentil

$$L_p = (n + 1)\frac{P}{100} \quad [4.1]$$

- Coeficiente de sesgo de Pearson

$$sk = \frac{3(\bar{x} - \text{Mediana})}{s} \quad [4.2]$$

- Coeficiente de sesgo calculado con software

$$sk = \frac{n}{(n - 1)(n - 2)}\left[\Sigma\left(\frac{x - \bar{x}}{s}\right)^3\right] \quad [4.3]$$

CAPÍTULO 5

- Probabilidad clásica

$$\begin{array}{c}\text{Probabilidad}\\ \text{de un evento}\end{array} = \frac{\text{Número de resultados favorables}}{\text{Número total de posibles resultados}} \quad [5.1]$$

- Regla especial de la adición

$$P(A \text{ o } B) = P(A) + P(B) \quad [5.2]$$

- Regla del complemento

$$P(A) = 1 - P(\sim A) \quad [5.3]$$

- Regla general de la adición

$$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B) \quad [5.4]$$

- Regla especial de la multiplicación

$$P(A \text{ y } B) = P(A)P(B) \quad [5.5]$$

- Regla general de la multiplicación

$$P(A \text{ y } B) = P(A)P(B|A) \quad [5.6]$$

- Teorema de Bayes

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \quad [5.7]$$

- Fórmula de la multiplicación

$$\text{Número total de disposiciones} = (m)(n) \quad [5.8]$$

- Fórmula de las permutaciones

$${}_nP_r = \frac{n!}{(n - r)!} \quad [5.9]$$

- Fórmula de las combinaciones

$${}_nC_r = \frac{n!}{r!(n - r)!} \quad [5.10]$$

CAPÍTULO 6

- Media de una distribución de probabilidad

$$\mu = \Sigma[xP(x)] \quad [6.1]$$

- Varianza de una distribución de probabilidad

$$\sigma^2 = \Sigma[(x - \mu)^2 P(x)] \quad [6.2]$$

- Fórmula de la probabilidad binomial

$$P(x) = {}_nC_x\, \pi^x (1 - \pi)^{n - x} \quad [6.3]$$

- Media de una distribución binomial

$$\mu = n\pi \quad [6.4]$$

- Varianza de una distribución binomial

$$\sigma^2 = n\pi(1 - \pi) \quad [6.5]$$

- Distribución hipergeométrica

$$P(x) = \frac{({}_SC_x)({}_{N-S}C_{n-x})}{{}_NC_n} \quad [6.6]$$

- Distribución de Poisson

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \quad [6.7]$$

- Media de una distribución de Poisson

$$\mu = n\pi \quad [6.8]$$

CAPÍTULO 7

- Media de la distribución uniforme

$$\mu = \frac{a + b}{2} \quad [7.1]$$

- Desviación estándar de la distribución uniforme

$$\sigma = \sqrt{\frac{(b - a)^2}{12}} \quad [7.2]$$

- Distribución uniforme

$$P(x) = \frac{1}{b - a} \quad [7.3]$$

si $a \le x \le b$ y 0 en cualquier otro lugar

- Distribución de probabilidad normal

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} \quad \textbf{[7.4]}$$

- Valor normal estándar

$$z = \frac{x - \mu}{\sigma} \quad \textbf{[7.5]}$$

- Distribución exponencial

$$P(x) = \lambda e^{-\lambda x} \quad \textbf{[7.6]}$$

- Encontrar la probabilidad usando la distribución exponencial

$$P(\text{Tiempo de llegada} < x) = 1 - e^{-\lambda x} \quad \textbf{[7.7]}$$

CAPÍTULO 8

- Error estándar de la media

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \textbf{[8.1]}$$

- Cálculo del valor z de $\bar{x}$ cuando se conoce la desviación estándar de la población

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \textbf{[8.2]}$$

CAPÍTULO 9

- Intervalo de confianza de la media poblacional con σ conocida

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}} \quad \textbf{[9.1]}$$

- Intervalo de confianza de la media poblacional con σ desconocida

$$\bar{x} \pm t\frac{s}{\sqrt{n}} \quad \textbf{[9.2]}$$

- Proporción muestral

$$p = \frac{x}{n} \quad \textbf{[9.3]}$$

- Intervalo de confianza de la proporción de una población

$$p \pm z\sqrt{\frac{p(1-p)}{n}} \quad \textbf{[9.4]}$$

- Tamaño de la muestra para estimar la media de la población

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad \textbf{[9.5]}$$

- Tamaño de la muestra de la proporción de la población

$$n = \pi(1 - \pi)\left(\frac{z}{E}\right)^2 \quad \textbf{[9.6]}$$

CAPÍTULO 10

- Prueba de la media cuando se conoce σ

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad \textbf{[10.1]}$$

- Prueba de una media; σ desconocida

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \textbf{[10.2]}$$

- Error tipo II

$$z = \frac{\bar{x}_c - \mu_1}{\sigma/\sqrt{n}} \quad \textbf{[10.3]}$$

CAPÍTULO 11

- Varianza de la distribución de las diferencias entre medias

$$\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad \textbf{[11.1]}$$

- Prueba de dos medias de muestras σ conocida

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \quad \textbf{[11.2]}$$

- Varianza conjunta

$$s_p^2 = \frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2} \quad \textbf{[11.3]}$$

- Pruebas de medias de dos muestras σ^2s desconocidas

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \quad \textbf{[11.4]}$$

- Estadístico de prueba de medias sin diferencia, varianzas desiguales σ^2s

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} \quad \textbf{[11.5]}$$

- Grados de libertad para prueba con varianza desigual

$$gl = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} \quad \textbf{[11.6]}$$

- Prueba t apareada

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \quad \textbf{[11.7]}$$

CAPÍTULO 12

- Estadístico de prueba para comparar dos varianzas

$$F = \frac{s_1^2}{s_2^2} \quad \textbf{[12.1]}$$

- Suma de cuadrados, total (SS)

$$\text{SS total} = \Sigma(x - \bar{x}_G)^2 \quad \textbf{[12.2]}$$

- Suma de cuadrados, error (SSE)

$$\text{SSE} = \Sigma(x - \bar{x}_c)^2 \quad \textbf{[12.3]}$$

- Suma de cuadrados, tratamientos (SST)

$$\text{SST} = \text{SS total} - \text{SSE} \quad \textbf{[12.4]}$$

- Intervalo de confianza de la diferencia entre las medias de tratamiento

$$(\bar{x}_1 - \bar{x}_2) \pm t\frac{\Sigma(x - \bar{x})(y - \bar{y})}{(n - 1)\, s_x s_y} \quad \textbf{[12.5]}$$

- Suma de cuadrados, bloques

$$\text{SSB} = k\Sigma(\bar{x}_b - \bar{x}_G)^2 \quad \textbf{[12.6]}$$

- Suma de errores cuadráticos, dos vías ANOVA

$$\text{SSE} = \text{SS total} - \text{SST} - \text{SSB} \quad \textbf{[12.7]}$$

CAPÍTULO 13

- Coeficiente de correlación

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{(n - 1)\, s_x s_y} \quad \textbf{[13.1]}$$

- Prueba t del coeficiente de correlación

$$r = \frac{r\sqrt{n - 2}}{\sqrt{1 - r^2}} \quad \textbf{[13.2]}$$

- Forma general de la ecuación de regresión lineal

$$\hat{y} = a + bx \quad \textbf{[13.3]}$$

- Pendiente de la recta de regresión

$$b = r\frac{s_y}{s_x} \quad [13.4]$$

- Intersección con el eje Y

$$a = \bar{y} - b\bar{x} \quad [13.5]$$

- Prueba de la pendiente

$$t = \frac{b - 0}{s_b} \quad [13.6]$$

- Error estándar de estimación

$$s_{y \cdot x} = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} \quad [13.7]$$

- Coeficiente de determinación

$$r^2 = \frac{\text{SSR}}{\text{SS total}} = 1 - \frac{\text{SSE}}{\text{SS total}} \quad [13.8]$$

- Error estándar de estimación

$$s_{y \cdot x} = \sqrt{\frac{\text{SSE}}{n - 2}} \quad [13.9]$$

- Intervalo de confianza de la media de y, dada x

$$\hat{y} \pm ts_{y \cdot x}\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x - \bar{x})^2}} \quad [13.10]$$

- Intervalo de predicción de y, dada x

$$\hat{y} \pm ts_{y \cdot x}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\Sigma(x - \bar{x})^2}} \quad [13.11]$$

CAPÍTULO 14

- Ecuación general de regresión múltiple

$$\hat{y} = a + b_1x_1 + b_2x_2 + \cdots + b_kx_k \quad [14.1]$$

- Error estándar de estimación múltiple

$$s_{y \cdot 123\ldots k} = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - (k + 1)}} = \sqrt{\frac{\text{SSE}}{n - (k + 1)}} \quad [14.2]$$

- Coeficiente de determinación múltiple

$$R^2 = \frac{\text{SSR}}{\text{SS total}} \quad [14.3]$$

- Coeficiente de determinación ajustado

$$R^2_{\text{ajust}} = 1 - \frac{\dfrac{\text{SSE}}{n - (k + 1)}}{\dfrac{\text{SS total}}{n - 1}} \quad [14.4]$$

- Prueba global

$$F = \frac{\text{SSR}/k}{\text{SSE}/[n - (k + 1)]} \quad [14.5]$$

- Prueba de los coeficientes de regresión individuales

$$t = \frac{b_i - 0}{s_{b_i}} \quad [14.6]$$

- Factor de inflación de la varianza

$$VIF = \frac{1}{1 - R_j^2} \quad [14.7]$$

CAPÍTULO 15

- Prueba de hipótesis de una proporción

$$z = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \quad [15.1]$$

- Prueba de proporciones de dos muestras

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_c(1 - p_c)}{n_1} + \dfrac{p_c(1 - p_c)}{n_2}}} \quad [15.2]$$

- Proporción conjunta

$$p_c = \frac{x_1 + x_2}{n_1 + n_2} \quad [15.3]$$

- Estadístico de prueba *ji* cuadrada

$$\chi^2 = \sum\left[\frac{(f_o - f_e)^2}{f_e}\right] \quad [15.4]$$

- Frecuencia esperada

$$f_e = \frac{\text{(Total de filas)(Total de columnas)}}{\text{Gran total}} \quad [15.5]$$

CAPÍTULO 16

- Prueba de los signos, $n > 10$

$$z = \frac{(x \pm 0.50) - \mu}{\sigma} \quad [16.1]$$

- Prueba de Wilcoxon de la suma de rangos

$$z = \frac{W - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1n_2(n_1 + n_2 + 1)}{12}}} \quad [16.4]$$

- Prueba de Kruskal-Wallis

$$H = \frac{12}{n(n + 1)}\left[\frac{(\Sigma R_1)^2}{n_1} + \frac{(\Sigma R_2)^2}{n_2} + \cdots + \frac{(\Sigma R_k)^2}{n_k}\right] - 3(n + 1) \quad [16.5]$$

- Coeficiente de correlación por rangos de Spearman

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)} \quad [16.6]$$

- Prueba de hipótesis, correlación por rangos

$$t = r_s\sqrt{\frac{n - 2}{1 - r_s^2}} \quad [16.7]$$

CAPÍTULO 17

- Índice simple

$$P = \frac{p_t}{p_0}(100) \quad [17.1]$$

- Promedio simple de los precios relativos

$$P = \frac{\Sigma P_i}{n} \quad [17.2]$$

- Índice agregado simple

$$P = \frac{\Sigma p_t}{\Sigma p_0}(100) \quad [17.3]$$

- Índice de precios de Laspeyres

$$P = \frac{\Sigma p_tq_0}{\Sigma p_0q_0}(100) \quad [17.4]$$

- Índice de precios de Paasche

$$P = \frac{\Sigma p_tq_t}{\Sigma p_0q_t}(100) \quad [17.5]$$

- Índice ideal de Fisher

$$\sqrt{\text{(Índice de Laspeyres)(Índice de Paasche)}} \quad [17.6]$$

- Índice de valores

$$V = \frac{\Sigma p_t q_t}{\Sigma p_0 q_0}(100) \quad \textbf{[17.7]}$$

- Ingreso real

$$\text{Ingreso real} = \frac{\text{Ingreso monetario}}{\text{IPC}}(100) \quad \textbf{[17.8]}$$

- Uso de un índice como factor de deflación

$$\text{Ventas deflacionadas} = \frac{\text{Ventas reales}}{\text{Un índice apropiado}}(100) \quad \textbf{[17.9]}$$

- Poder de compra del dólar

$$\text{Poder de compra del dólar} = \frac{\$1}{\text{CPI}}(100) \quad \textbf{[17.10]}$$

CAPÍTULO 18

- Ecuación de tendencia lineal

$$\hat{y} = a + bt \quad \textbf{[18.1]}$$

- Ecuación de tendencia logarítmica

$$\log \hat{y} = \log a + \log b(t) \quad \textbf{[18.2]}$$

- Factor de corrección para ajustar medias trimestrales

$$\text{Factor de corrección} = \frac{4.00}{\text{Total de cuatro medias}} \quad \textbf{[18.3]}$$

- Estadístico de Durbin-Watson

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \quad \textbf{[18.4]}$$

CAPÍTULO 19

- Media total

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}}{k} \quad \textbf{[19.1]}$$

- Límites de control de la media

$$\text{LCS} = \bar{\bar{x}} + A_2\bar{R} \qquad \text{LCI} = \bar{\bar{x}} - A_2\bar{R} \quad \textbf{[19.4]}$$

- Diagrama de control de rangos

$$\text{LCS} = D_4\bar{R} \qquad \text{LCI} = D_3\bar{R} \quad \textbf{[19.5]}$$

- Proporción media de defectos

$$p = \frac{\text{Número total de defectos}}{\text{Número total de artículos de la muestra}} \quad \textbf{[19.6]}$$

- Límites de control de proporciones

$$\text{LCS y LCI} = p \pm 3\sqrt{\frac{p(1-p)}{n}} \quad \textbf{[19.8]}$$

- Límites de control del número de defectos por unidad

$$\text{LCS y LCI} = \bar{c} \pm 3\sqrt{\bar{c}} \quad \textbf{[19.9]}$$

CAPÍTULO 20 (en el sitio web del libro: www.mhhe.com/uni/lind_ae16e)

- Valor monetario esperado

$$\text{EMV}(A_i) = \Sigma[P(S_j) \cdot V(A_i, S_j)] \quad \textbf{[20.1]}$$

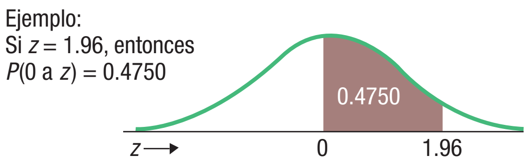
- Pérdida de oportunidad esperada

$$\text{EOL}(A_i) = \Sigma[P(S_j) \cdot R(A_i, S_j)] \quad \textbf{[20.2]}$$

- Valor esperado de la información perfecta

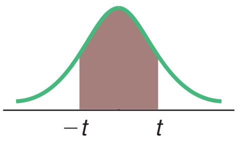
EVPI = Valor esperado en condiciones de certidumbre
− Valor esperado en condiciones
de incertidumbre **[20.3]**

Áreas bajo la curva normal

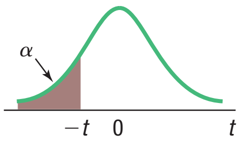


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	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

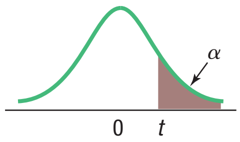
Distribución *t* de Student



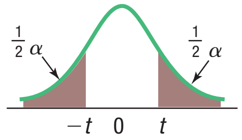
Intervalo de confianza



Prueba de cola izquierda



Prueba de cola derecha



Prueba de dos colas

	Intervalos de confianza, *c*					
***gl* (grados de libertad)**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.9%**
	Nivel de significancia de una prueba de una cola, α					
	0.10	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
	Nivel de significancia de una prueba de dos colas, α					
	0.20	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
31	1.309	1.696	2.040	2.453	2.744	3.633
32	1.309	1.694	2.037	2.449	2.738	3.622
33	1.308	1.692	2.035	2.445	2.733	3.611
34	1.307	1.691	2.032	2.441	2.728	3.601
35	1.306	1.690	2.030	2.438	2.724	3.591

(*continúa*)

(*continuación*)

	Intervalos de confianza, *c*					
***gl* (grados de libertad)**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.9%**
	Nivel de significancia de una prueba de una cola, α					
	0.10	**0.05**	**0.025**	**0.01**	**0.005**	**0.0005**
	Nivel de significancia de una prueba de dos colas, α					
	0.20	**0.10**	**0.05**	**0.02**	**0.01**	**0.001**
36	1.306	1.688	2.028	2.434	2.719	3.582
37	1.305	1.687	2.026	2.431	2.715	3.574
38	1.304	1.686	2.024	2.429	2.712	3.566
39	1.304	1.685	2.023	2.426	2.708	3.558
40	1.303	1.684	2.021	2.423	2.704	3.551
41	1.303	1.683	2.020	2.421	2.701	3.544
42	1.302	1.682	2.018	2.418	2.698	3.538
43	1.302	1.681	2.017	2.416	2.695	3.532
44	1.301	1.680	2.015	2.414	2.692	3.526
45	1.301	1.679	2.014	2.412	2.690	3.520
46	1.300	1.679	2.013	2.410	2.687	3.515
47	1.300	1.678	2.012	2.408	2.685	3.510
48	1.299	1.677	2.011	2.407	2.682	3.505
49	1.299	1.677	2.010	2.405	2.680	3.500
50	1.299	1.676	2.009	2.403	2.678	3.496
51	1.298	1.675	2.008	2.402	2.676	3.492
52	1.298	1.675	2.007	2.400	2.674	3.488
53	1.298	1.674	2.006	2.399	2.672	3.484
54	1.297	1.674	2.005	2.397	2.670	3.480
55	1.297	1.673	2.004	2.396	2.668	3.476
56	1.297	1.673	2.003	2.395	2.667	3.473
57	1.297	1.672	2.002	2.394	2.665	3.470
58	1.296	1.672	2.002	2.392	2.663	3.466
59	1.296	1.671	2.001	2.391	2.662	3.463
60	1.296	1.671	2.000	2.390	2.660	3.460
61	1.296	1.670	2.000	2.389	2.659	3.457
62	1.295	1.670	1.999	2.388	2.657	3.454
63	1.295	1.669	1.998	2.387	2.656	3.452
64	1.295	1.669	1.998	2.386	2.655	3.449
65	1.295	1.669	1.997	2.385	2.654	3.447
66	1.295	1.668	1.997	2.384	2.652	3.444
67	1.294	1.668	1.996	2.383	2.651	3.442
68	1.294	1.668	1.995	2.382	2.650	3.439
69	1.294	1.667	1.995	2.382	2.649	3.437
70	1.294	1.667	1.994	2.381	2.648	3.435

(*continúa*)

Distribución *t* de Student

(*continuación*)

gl (grados de libertad)	Intervalos de confianza, *c*					
	80%	90%	95%	98%	99%	99.9%
	Nivel de significancia de una prueba de una cola, α					
	0.10	0.05	0.025	0.01	0.005	0.0005
	Nivel de significancia de una prueba de dos colas, α					
	0.20	0.10	0.05	0.02	0.01	0.001
71	1.294	1.667	1.994	2.380	2.647	3.433
72	1.293	1.666	1.993	2.379	2.646	3.431
73	1.293	1.666	1.993	2.379	2.645	3.429
74	1.293	1.666	1.993	2.378	2.644	3.427
75	1.293	1.665	1.992	2.377	2.643	3.425
76	1.293	1.665	1.992	2.376	2.642	3.423
77	1.293	1.665	1.991	2.376	2.641	3.421
78	1.292	1.665	1.991	2.375	2.640	3.420
79	1.292	1.664	1.990	2.374	2.640	3.418
80	1.292	1.664	1.990	2.374	2.639	3.416
81	1.292	1.664	1.990	2.373	2.638	3.415
82	1.292	1.664	1.989	2.373	2.637	3.413
83	1.292	1.663	1.989	2.372	2.636	3.412
84	1.292	1.663	1.989	2.372	2.636	3.410
85	1.292	1.663	1.988	2.371	2.635	3.409
86	1.291	1.663	1.988	2.370	2.634	3.407
87	1.291	1.663	1.988	2.370	2.634	3.406
88	1.291	1.662	1.987	2.369	2.633	3.405
89	1.291	1.662	1.987	2.369	2.632	3.403
90	1.291	1.662	1.987	2.368	2.632	3.402
91	1.291	1.662	1.986	2.368	2.631	3.401
92	1.291	1.662	1.986	2.368	2.630	3.399
93	1.291	1.661	1.986	2.367	2.630	3.398
94	1.291	1.661	1.986	2.367	2.629	3.397
95	1.291	1.661	1.985	2.366	2.629	3.396
96	1.290	1.661	1.985	2.366	2.628	3.395
97	1.290	1.661	1.985	2.365	2.627	3.394
98	1.290	1.661	1.984	2.365	2.627	3.393
99	1.290	1.660	1.984	2.365	2.626	3.392
100	1.290	1.660	1.984	2.364	2.626	3.390
120	1.289	1.658	1.980	2.358	2.617	3.373
140	1.288	1.656	1.977	2.353	2.611	3.361
160	1.287	1.654	1.975	2.350	2.607	3.352
180	1.286	1.653	1.973	2.347	2.603	3.345
200	1.286	1.653	1.972	2.345	2.601	3.340
∞	1.282	1.645	1.960	2.326	2.576	3.291