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

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.

OA8-1
Explicar por qué se muestrean las poblaciones, describir cuatro métodos para seleccionar 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

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.

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.

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

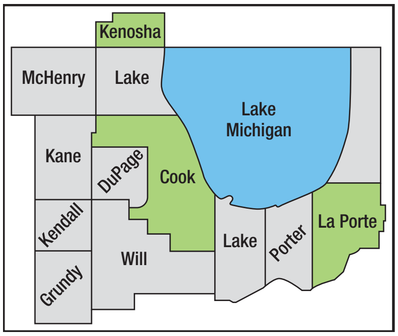
* 0.02 de 50 = 1, 0.10 de 50 = 5, etcétera.

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	Número de identificación	Distribuidor	Número de identificación	Distribuidor
00	Dave White Acura	11	Thayer Chevrolet/Toyota	23	Kistler Ford, Inc.
01	Autofair Nissan	12	Spurgeon Chevrolet Motor Sales, Inc.	24	Lexus of Toledo
02	Autofair Toyota-Suzuki	13	Dunn Chevrolet	25	Mathews Ford Oregon, Inc.
03	George Ball's Buick GMC Truck	14	Don Scott Chevrolet	26	Northtowne Chevrolet
04	Yark Automotive Group	15	Dave White Chevrolet Co.	27	Quality Ford Sales, Inc.
05	Bob Schmidt Chevrolet	16	Dick Wilson Infinity	28	Rouen Chrysler Jeep Eagle
06	Bowling Green Lincoln Mercury Jeep Eagle	17	Doyle Buick	29	Saturn of Toledo
		18	Franklin Park Lincoln Mercury	30	Ed Schmidt Jeep Eagle
07	Brondes Ford	19	Genoa Motors	31	Southside Lincoln Mercury
08	Brown Honda	20	Great Lakes Ford Nissan	32	Valiton Chrysler
09	Brown Mazda	21	Grogan Towne Chrysler	33	Vin Divers
10	Charlie's Dodge	22	Hatfield Motor Sales	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	12	**Joehlin Bob** 3358 Navarre Av	21	**Seibert John H** 201 S Main
03	**Denker Rollie** 7445 Airport Hwy	13	**Keisser David** 3030 W Sylvania Av	22	**Smithers Bob** 229 Superior St
04	**Farley Ron** 1837 W Alexis Rd	14	**Keisser Keith** 5902 Sylvania Av	23	**Smithers Jerry** 229 Superior St
05	**George Mark** 7247 W Central Av	15	**Lawrence Grant** 342 W Dussel Dr	24	**Wright Steve** 105 S Third St
06	**Gibellato Carlo** 6616 Monroe St	16	**Miller Ken** 2427 Woodville Rd	25	**Wood Tom** 112 Louisiana Av
07	**Glemser Cathy** 5602 Woodville Rd	17	**O'Donnell Jim** 7247 W Central Av	26	**Yoder Scott** 6 Willoughby Av
08	**Green Mike** 4149 Holland Sylvania Rd	18	**Priest Harvey** 5113 N Summit St		
09	**Harris Ev** 2026 Albon Rd				

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 muestro 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.

$$\mu_{\bar{x}} = \frac{\text{Suma de todas las medias muestrales}}{\text{Total de muestras}} = \frac{\$7.00 + \$7.50 + \$7.50 + \cdots + \$8.00 + \$8.50}{21}$$

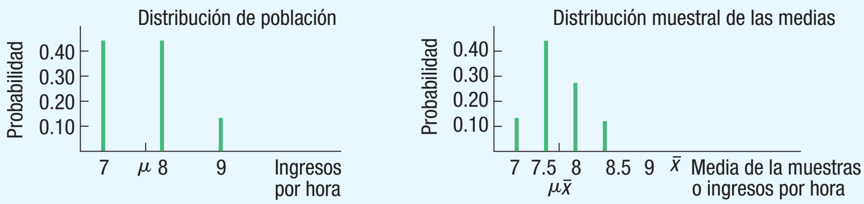
$$= \frac{\$162}{21} = \$7.71$$

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.

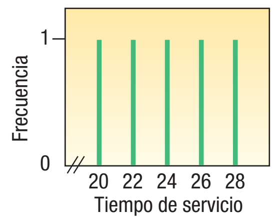
Dada una distribución de probabilidad normal (forma de campana), se aplican los conceptos del capítulo 7 para determinar la probabilidad de seleccionar una muestra con una media muestral específica. En la siguiente sección se resalta la importancia del tamaño de una muestra en relación con la distribución 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.

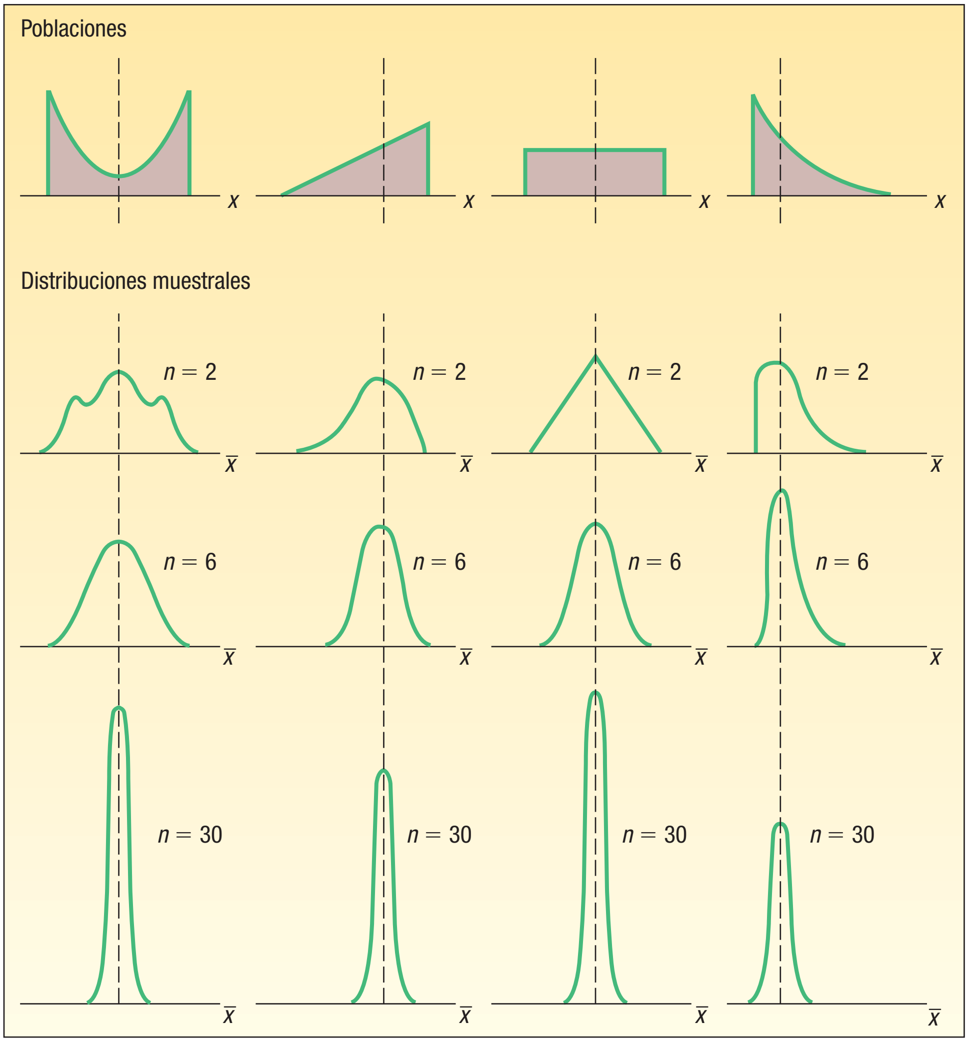
En esta sección se estudia el **teorema central del límite.** Su aplicación a la distribución muestral de medias se introdujo en la sección anterior; esta permite utilizar la distribución de probabilidad normal para crear intervalos de confianza de la media poblacional (vea el capítulo 9) y llevar a 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

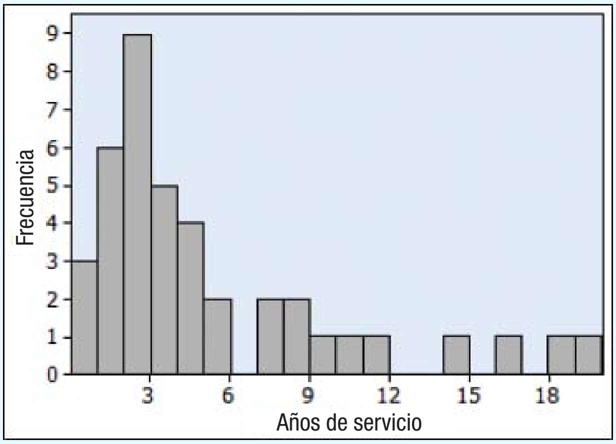
Ed Spence fundó su negocio de engranes hace 20 años; este creció a lo largo del tiempo y ahora cuenta con 40 empleados. Spence Sprockets, Inc., encara algunas decisiones importantes relacionadas con la atención médica de su personal. Antes de tomar una decisión definitiva sobre el programa de atención 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.

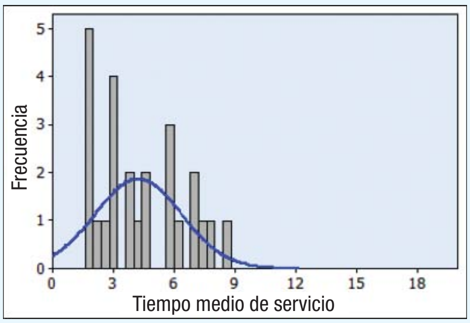
$$\mu = \frac{11 + 4 + 18 + \cdots + 2 + 3}{40} = 4.80$$

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

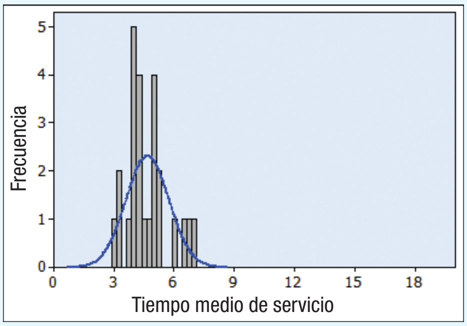
Cambiemos ahora el ejemplo aumentando el tamaño de la muestra de cinco empleados a 20. En la tabla 8.6 se indican los resultados de seleccionar 25 muestras de 20 empleados cada una y el cálculo de las medias muestrales. Estas medias muestrales se representan en la gráfica 8.6; compare la forma de esta 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 $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \textbf{[8.1]}$$

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.

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

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

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.

2	3	2	3	3	4	2	4	3	2	2	7	3	4	5	3	3	3	3	5

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

  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

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \qquad \textbf{[8.2]}$$

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

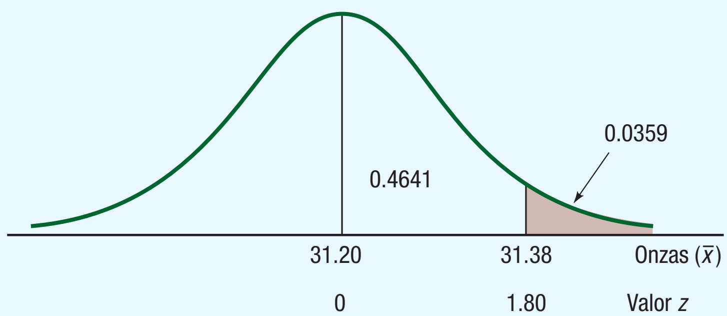
Utilice los resultados de la sección anterior para determinar la probabilidad de seleccionar una muestra de 16 (n) botellas de una población normal con una media de 31.2 (μ) onzas y una desviación estándar de la población de 0.4 (σ) onzas, y encontrar que la media muestral es de 31.38 ($\bar{x}$) o superior. 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?