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.

Estimadores puntuales e intervalos de confianza de una media

OA9-1 Calcular e interpretar un estimador puntual de la media poblacional.

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.

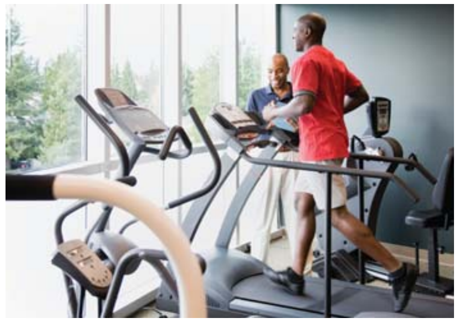
ESTIMADOR PUNTUAL Estadístico calculado a partir de información de la muestra para estimar el parámetro poblacional.

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.

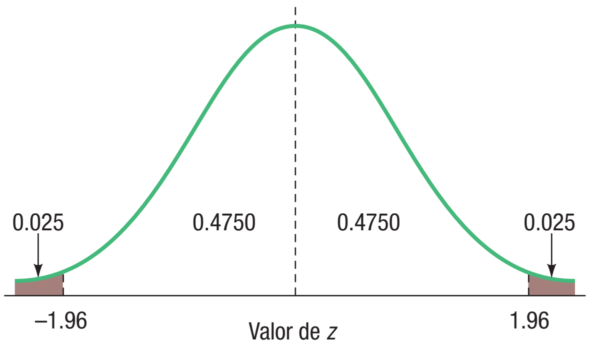
Recuerde que la distribución muestral de la media es la distribución de todas las medias muestrales, $\bar{x}$, con tamaño de la muestra, *n*, de una población, y que se conoce la desviación estándar de la población, σ. A partir de esta información, y del teorema central del límite, se sabe que la distribución 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

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \qquad [9.1]$$

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

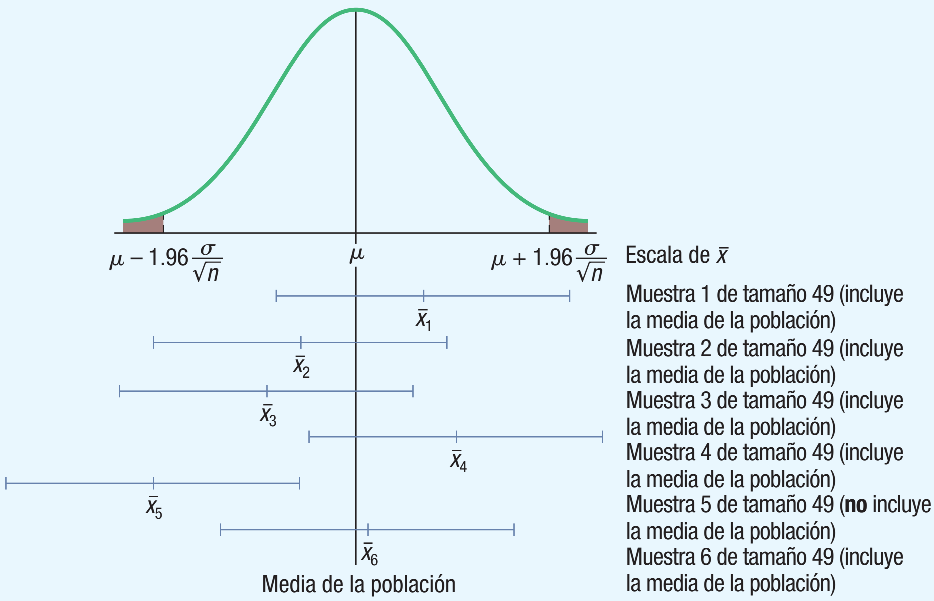
En general, las distribuciones de los salarios e ingresos tienen un sesgo positivo, pues unos cuantos individuos ganan considerablemente más que otros, lo cual sesga la distribución en dirección positiva. Por fortuna, el teorema central del límite estipula que, si se selecciona una muestra grande, la 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 ±$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.

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*.

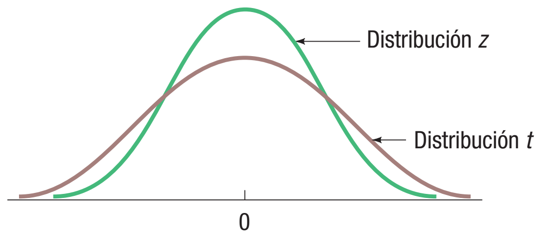
Por fortuna, se utiliza la desviación estándar de la muestra para estimar la desviación estándar poblacional. Es decir, se utiliza s (la desviación estándar de la muestra) para estimar σ (la desviación estándar de la población); no obstante, al hacerlo no es posible utilizar la fórmula [9.1]. Al desconocer 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:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

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.

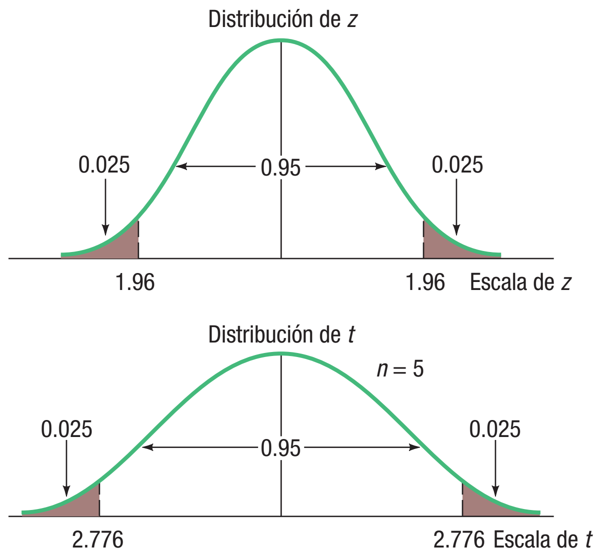


GRÁFICA 9.1 Distribución normal estándar y distribución *t* de Student

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.

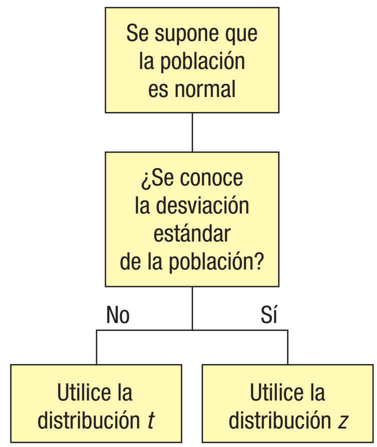
INTERVALO DE CONFIANZA DE LA MEDIA POBLACIONAL CON UNA σ DESCONOCIDA

$$\bar{x} \pm t\frac{s}{\sqrt{n}} \quad [9.2]$$

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	

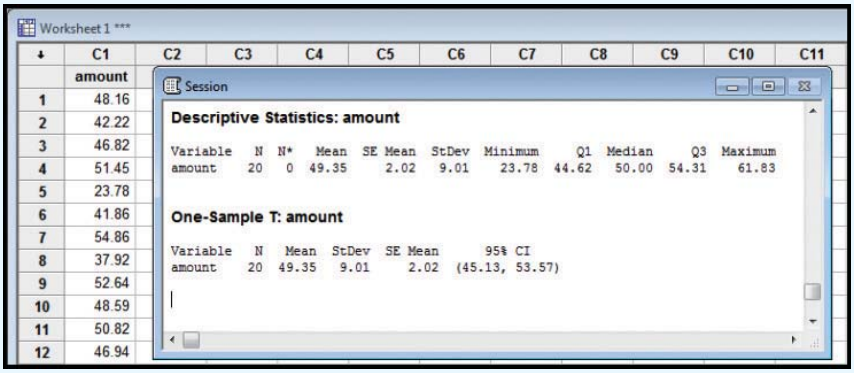
¿Cuál es la mejor estimación de la media poblacional? Determine un intervalo de confianza de 95% e interprete el resultado. ¿Concluiría de forma razonable que la media poblacional es de 50 o 60 dólares?

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



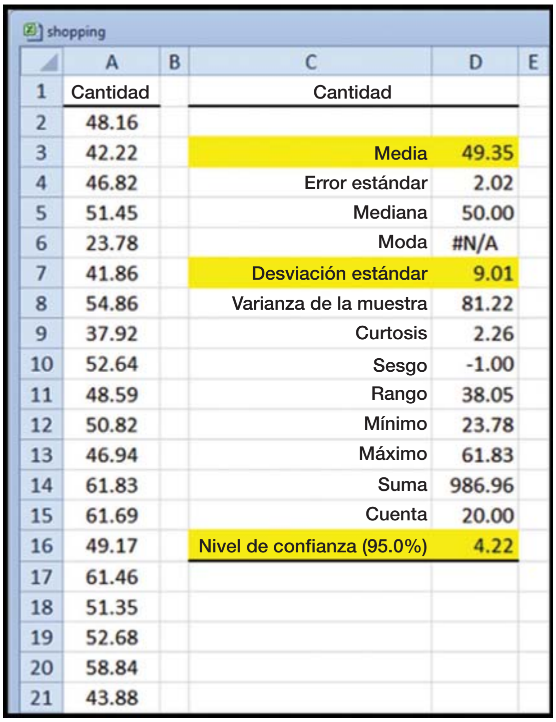
dólares, que constituye la mejor aproximación, la *estimación puntual,* de la media poblacional desconocida.

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.



	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.