

# CAPÍTULO 8



## Estimación por intervalo

### CONTENIDO

LA ESTADÍSTICA EN  
LA PRÁCTICA: FOOD LION

- 8.1** MEDIA POBLACIONAL:  
 $\sigma$  CONOCIDA  
Margen de error y estimación  
por intervalo  
Recomendación práctica
- 8.2** MEDIA POBLACIONAL:  
 $\sigma$  DESCONOCIDA  
Margen de error en estimación  
por intervalo

Recomendación práctica  
Uso de una muestra pequeña  
Resumen de los procedimientos  
de estimación por intervalo

- 8.3** DETERMINACIÓN DEL  
TAMAÑO DE LA MUESTRA
- 8.4** PROPORCIÓN  
POBLACIONAL  
Determinación del tamaño  
de la muestra



## LA ESTADÍSTICA *(en)* LA PRÁCTICA

### FOOD LION\*

*SALISBURY, CAROLINA DEL NORTE*

Fundada en 1957 como Food Town, Food Lion es una de las cadenas de supermercados más grande de Estados Unidos, con 1 200 tiendas en 11 estados. La empresa vende más de 24 000 productos diferentes y ofrece mercancías de marcas nacionales y regionales, así como una cantidad cada vez mayor de productos de gran calidad de marca propia fabricados especialmente. La empresa mantiene su liderazgo en precios bajos y asegura la calidad a partir de eficientes controles como formatos estándar de tienda, diseño innovador de los almacenes, instalaciones eficaces y sincronización de datos con los proveedores. Food Lion mira hacia un futuro de continua innovación, crecimiento, liderazgo de precio y servicio a sus clientes.

Puesto que es un negocio con inventario intensivo, Food Lion decidió adoptar como método de evaluación de inventario el método LIFO (*last-in, first-out*). Este método compara los costos corrientes con los ingresos corrientes. Lo cual minimiza los efectos de los cambios radicales de precios sobre los resultados de ganancia y pérdida. Además, el método LIFO reduce la ganancia neta disminuyendo con esto los impuestos sobre la renta durante los períodos de inflación.

Food Lion establece un índice LIFO para cada uno de los siete grupos de inventario: abarrotes, papel/artículos para el hogar, artículos para mascotas, artículos para la salud y la belleza, lácteos, cigarros/tabaco y cervezas/vinos. Por ejemplo, un índice LIFO de 1.008 para el grupo abarrotes indica que el valor de este inventario, a los costos corrientes, refleja un aumento de 0.8% debido a la inflación en el último periodo de un año.

Un índice LIFO para cada grupo de inventario requiere que el inventario de fin de año de cada producto se evalúe al costo actual de fin de año y al costo del año anterior.

\*Los autores agradecen a Keith Cunningham director de impuestos y a Bobby Harkey del equipo de contadores por proporcionar este artículo para *La estadística en la práctica*.



El almacén Food Lion en el centro comercial Cambridge, Charlotte, North Carolina. © Cortesía de Food Lion.

Para ahorrar tiempo y gastos excesivos por el conteo para el inventario en las 1200 tiendas, Food Lion selecciona una muestra aleatoria simple de 50 tiendas. El inventario físico de fin de año se realiza en cada una de las tiendas de la muestra. Para obtener el índice LIFO de cada uno de los grupos de inventario se usan los costos del año actual y los costos del año anterior.

En uno de los últimos años, la estimación muestral del índice LIFO para el inventario del grupo de productos para la salud y la belleza fue 1.015. Con un nivel de confianza de 95%, Food Lion calculó un margen de error de 0.006 para la estimación muestral. Por tanto, el intervalo de 1.009 a 1.021 proporciona una estimación por intervalo de confianza de 95% del índice LIFO poblacional. Este nivel de precisión se consideró muy bueno.

En ese capítulo aprenderá cómo calcular un margen de error para una estimación puntual. También verá cómo usar esta información para construir e interpretar estimaciones por intervalo para una media poblacional y para una proporción poblacional.

En el capítulo 7 se dijo que un estimador puntual es un estadístico muestral que se usa para estimar un parámetro poblacional. Por ejemplo, la media muestral  $\bar{x}$  es un estimador puntual de la media poblacional  $\mu$  y la proporción muestral  $\bar{p}$  es un estimador puntual de la proporción poblacional  $p$ . Como no se puede esperar que un estimador puntual suministre el valor exacto del parámetro poblacional, se suele calcular una **estimación por intervalo** al sumar y restar al estimador puntual una cantidad llamada **margen de error**. La fórmula general de una estimación por intervalo es

Estimación puntual  $\pm$  Margen de error

El objetivo de la estimación por intervalo es aportar información de qué tan cerca se encuentra la estimación puntual, obtenida de la muestra, del valor del parámetro poblacional.

En este capítulo se muestra cómo obtener una estimación por intervalo para la media poblacional  $\mu$  y para la proporción poblacional  $p$ . La fórmula general para obtener una estimación por intervalo para la media poblacional es

$$\bar{x} \pm \text{Margen de error}$$

De manera similar, la fórmula general para obtener una estimación por intervalo para la proporción poblacional es

$$\bar{p} \pm \text{Margen de error}$$

Las distribuciones muestrales de  $\bar{x}$  y de  $\bar{p}$  son clave para calcular estas estimaciones por intervalo.

## 8.1

## Media poblacional: $\sigma$ conocida

Con objeto de obtener una estimación por intervalo para la media poblacional, se necesita la desviación estándar poblacional  $\sigma$  o la desviación estándar muestral  $s$  para calcular el margen de error. En la mayor parte de los casos, no se conoce  $\sigma$  y para calcular el margen de error se emplea  $s$ . Sin embargo, en algunas ocasiones, se cuenta con una gran cantidad de datos anteriores (históricos) que se pueden usar para calcular la desviación estándar poblacional antes de tomar la muestra. También, en aplicaciones al control de calidad, en las que se supone que el proceso se desarrolla correctamente “en control”, se considera que se conoce la desviación estándar. A tales casos se les conoce como **casos de  $\sigma$  conocida**. En esta sección se presenta un caso en el que es razonable considerar que se conoce  $\sigma$  y se muestra cómo obtener una estimación por intervalo.

Cada semana, la empresa Lloyd's Department Store selecciona una muestra aleatoria simple de 100 clientes con objeto de conseguir información acerca de la cantidad que gastan en cada visita a la tienda. Si  $x$  representa la cantidad gastada en cada visita a la tienda, la media muestral  $\bar{x}$  es una estimación puntual de  $\mu$ , la cantidad media gastada en cada visita a la tienda por la población formada por los clientes de Lloyd's Department Store. Lloyd's ha estado realizando estos estudios semanales durante varios años. Con base en sus datos anteriores, Lloyd's supone que el valor conocido de la desviación estándar poblacional es  $\sigma = \$20$ . Los datos anteriores (históricos) indican también que la población tiene una distribución normal.

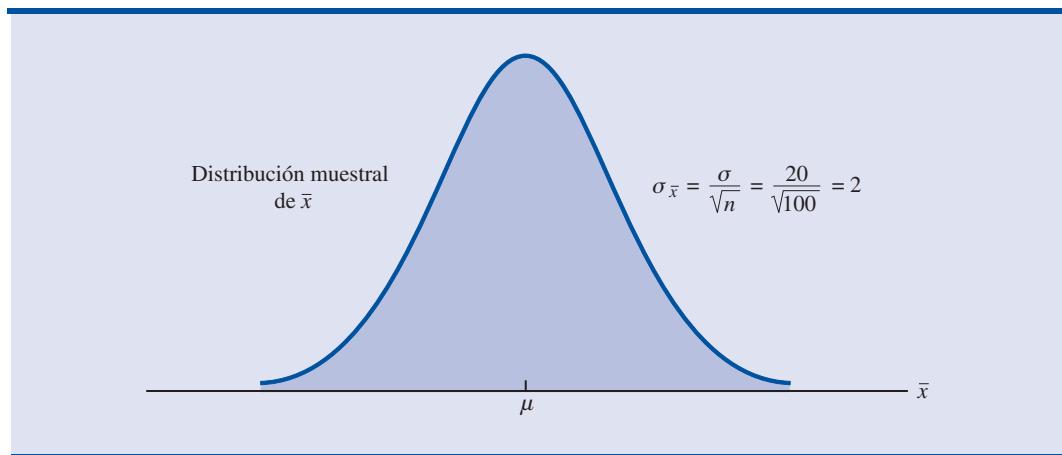
En la última semana, en su estudio de 100 clientes ( $n = 100$ ), Lloyd's obtuvo como media muestral  $\bar{x} = \$82$ . La media muestral de la cantidad gastada permite una estimación puntual de la media poblacional de la cantidad gastada en cada visita,  $\mu$ . A continuación se muestra cómo calcular un margen de error para esta estimación y cómo obtener una estimación por intervalo para la media poblacional.

### Margen de error y estimación por intervalo

En el capítulo 7 se mostró que la distribución muestral de  $\bar{x}$  sirve para calcular la probabilidad de que  $\bar{x}$  esté dentro de una distancia dada de  $\mu$ . En el ejemplo de Lloyd's, los datos históricos muestran que la población constituida por las cantidades gastadas está distribuida normalmente y que su desviación estándar es  $\sigma = 20$ . De manera que, usando lo aprendido en el capítulo 7, se puede concluir que la distribución muestral de  $\bar{x}$  es una distribución normal con error estándar de  $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 20/\sqrt{100} = 2$ . En la figura 8.1 se muestra esta distribución muestral.\* Puesto que la

\* Se aprovecha que las cantidades gastadas tienen una distribución normal para concluir que la distribución muestral de  $\bar{x}$  tiene una distribución normal. Si la población no tuviera una distribución normal, de acuerdo con el teorema del límite central y dado que el tamaño de la muestra  $n = 100$  se puede concluir que la distribución muestral de  $\bar{x}$  es aproximadamente normal. De cualquier manera, la distribución muestral de  $\bar{x}$  es como se muestra en la figura 8.1.

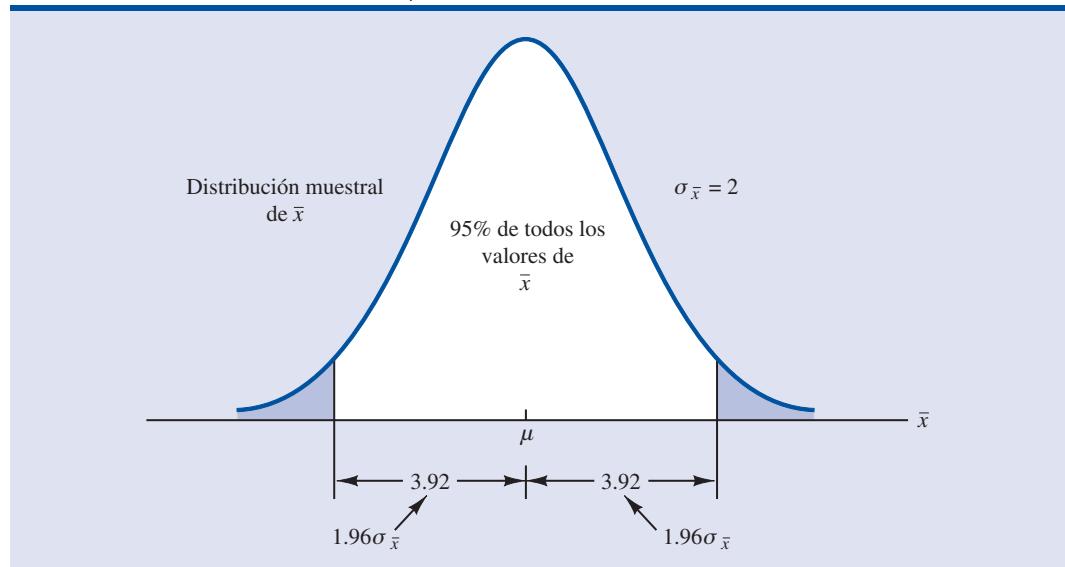
**FIGURA 8.1** DISTRIBUCIÓN MUESTRAL DE LA MEDIA MUESTRAL DE LAS CANTIDADES GASTADAS EN MUESTRAS ALEATORIAS SIMPLES DE 100 CLIENTES



distribución muestral indica cómo están distribuidos los valores de  $\bar{x}$  en torno a la media poblacional  $\mu$ , la distribución muestral de  $\bar{x}$  proporciona información acerca de la posible diferencia entre  $\bar{x}$  y  $\mu$ .

En la tabla de probabilidad normal estándar se encuentra que 95% de los valores de cualquier variable aleatoria distribuida normalmente aparecen dentro de  $\pm 1.96$  desviaciones estándar de la media. Por tanto, si la distribución muestral se encuentra distribuida normalmente, 95% de los valores de  $\bar{x}$  deben estar dentro de  $\pm 1.96\sigma_{\bar{x}}$  de la media  $\mu$ . En el ejemplo de Lloyd's, se sabe que la distribución muestral de  $\bar{x}$  está distribuida normalmente y que el error estándar es  $\sigma_{\bar{x}} = 2$ . Como  $\pm 1.96\sigma_{\bar{x}} = 1.96(2) = 3.92$ , se puede concluir que 95% de los valores de  $\bar{x}$  obtenidos usando muestras de  $n = 100$  estarán dentro de  $\pm 3.92$  de la media poblacional. Véase figura 8.2.

**FIGURA 8.2** DISTRIBUCIÓN MUESTRAL DE  $\bar{x}$  EN LA QUE SE MUESTRA LA LOCALIZACIÓN DE LAS MEDIAS MUESTRALES QUE SE ENCUENTRAN DENTRO DE 3.92 DE  $\mu$ .

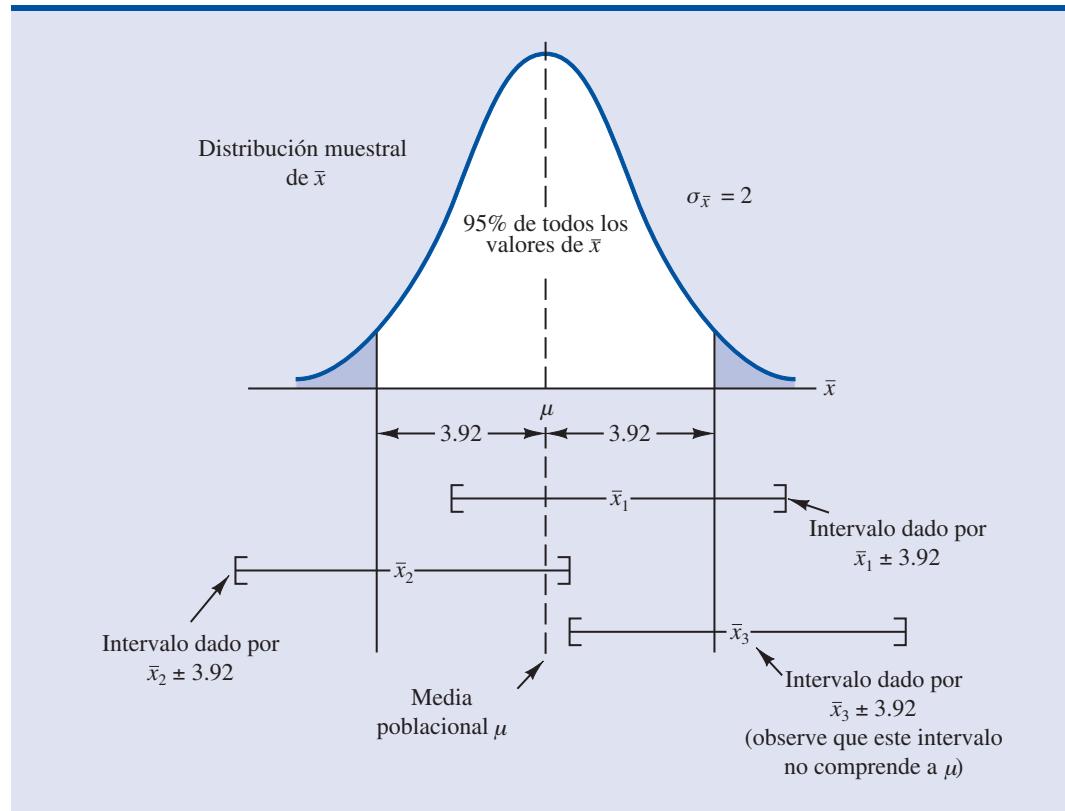


En la introducción a este capítulo se dijo que la fórmula general para estimar un intervalo de la media poblacional  $\mu$  es  $\bar{x} \pm$  margen de error. En el ejemplo de Lloyd's, suponga que se establece 3.92 como margen de error y se calcula una estimación por intervalo para  $\mu$  usando  $\bar{x} \pm 3.92$ . Para ver cómo se interpreta dicha estimación por intervalo, considere los valores de  $\bar{x}$  que podrían obtenerse si se tomaran tres muestras aleatorias simples *diferentes*, cada una de 100 clientes de Lloyd's. La primera media muestral puede que dé el valor  $\bar{x}_1$  de la figura 8.3. En este caso, como se ve en la figura, el intervalo que se obtiene restando 3.92 de  $\bar{x}_1$  y sumando 3.92 a  $\bar{x}_1$  abarca a la media poblacional  $\mu$ . Ahora razoné lo que pasa si la segunda media muestral resulta tener el valor  $\bar{x}_2$  que se muestra en la figura 8.3. Aunque esta media muestral difiere de la primera media muestral, el intervalo obtenido restando 3.92 de  $\bar{x}_2$  y sumando 3.92 a  $\bar{x}_2$  también abarca a la media poblacional  $\mu$ . Pero, considere lo que pasa si la tercera media muestral resulta tener el valor  $\bar{x}_3$  que se muestra en la figura 8.3. En este caso el intervalo obtenido restando 3.92 de  $\bar{x}_3$  y sumando 3.92 a  $\bar{x}_3$  no abarca a la media poblacional  $\mu$ . Como  $\bar{x}_3$  cae en la cola superior de la distribución muestral y dista más de 3.92 de  $\mu$ , restando y sumando 3.92 de  $\bar{x}_3$  se obtiene un intervalo que no abarca a  $\mu$ .

Con cualquier media muestral  $\bar{x}$  que se encuentre dentro de la región sombreada en la figura 8.3, se obtendrá un intervalo que contenga a la media poblacional  $\mu$ . Como 95% de todas las posibles medias muestrales se encuentran en la región sombreada, 95% de todos los intervalos que se obtengan restando 3.92 de  $\bar{x}$  y sumando 3.92 a  $\bar{x}$  abarcarán a la media poblacional  $\mu$ .

Recuerde que en la última semana el equipo encargado del aseguramiento de la calidad de Lloyd's encuestó a 100 clientes y obtuvo una media muestral de  $\bar{x} = 82$ . Usando  $\bar{x} \pm 3.92$  para

**FIGURA 8.3** INTERVALOS OBTENIDOS A PARTIR DE ALGUNAS MEDIAS MUESTRALES LOCALIZADAS EN  $\bar{x}_1$ ,  $\bar{x}_2$ , Y EN  $\bar{x}_3$



Esto aclara por qué se le llama intervalo de confianza de 95%.

obtener la estimación por intervalo, se obtiene  $82 \pm 3.92$ . Por tanto, la estimación por intervalo que se basa en los datos de la última semana es el intervalo que va de  $82 - 3.92 = 78.08$  a  $82 + 3.92 = 85.92$ . Como 95% de todos los intervalos construidos usando  $\bar{x} \pm 3.92$  contendrán a la media poblacional se tiene 95% de confianza de que el intervalo 78.08 a 85.92 contenga a la media poblacional  $\mu$ . Entonces dicho intervalo tiene un **nivel de confianza** de 95%. Al valor 0.95 se le conoce como **coeficiente de confianza** y al intervalo 78.08 a 85.92 como el **intervalo de confianza de 95%**.

Como el margen de error está dado por  $z_{\alpha/2}(\sigma/\sqrt{n})$ , la fórmula general de una estimación por intervalo de la media poblacional con  $\sigma$  conocida, es la siguiente.

#### ESTIMACIÓN POR INTERVALO DE UNA MEDIA POBLACIONAL: $\sigma$ CONOCIDA

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (8.1)$$

donde  $(1 - \alpha)$  es el coeficiente de confianza y  $z_{\alpha/2}$  es el valor de  $z$  que proporciona un área  $\alpha/2$  en la cola superior de la distribución de probabilidad normal estándar.

En el ejemplo de Lloyd's, mediante la expresión (8.1), se construye un intervalo de confianza de 95% con un coeficiente de confianza  $(1 - \alpha) = 0.95$  y, por tanto,  $\alpha = 0.05$ . En la tabla de la distribución normal estándar aparece que un área de  $\alpha/2 = 0.05/2 = 0.025$  en la cola superior corresponde a  $z_{0.025} = 1.96$ . Como en el ejemplo de Lloyd's, la media muestral es  $\bar{x} = 82$ ,  $\sigma = 20$  y el tamaño de la muestra es  $n = 100$ , se obtiene

$$82 \pm 1.96 \frac{20}{\sqrt{100}} \\ 82 \pm 3.92$$

Por tanto, empleando la expresión (8.1), el margen de error es 3.92 y el intervalo de confianza de 95% va de  $82 - 3.92 = 78.08$  a  $82 + 3.92 = 85.92$ .

Aunque a menudo se usa un nivel de confianza de 95%, también suelen usarse otros niveles de confianza como 90% y 99%. En la tabla 8.1 se muestran los valores de  $z_{\alpha/2}$  correspondientes a los niveles de confianza más utilizados. A partir de estos valores y de la expresión (8.1), el intervalo de confianza de 90% en el ejemplo de Lloyd's es

$$82 \pm 1.645 \frac{20}{\sqrt{100}} \\ 82 \pm 3.29$$

**TABLA 8.1** VALORES DE  $z_{\alpha/2}$  PARA LOS NIVELES DE CONFIANZA MÁS USADOS

Nivel de confianza	$\alpha$	$\alpha/2$	$z_{\alpha/2}$
90%	.10	.05	1.645
95%	.05	.025	1.960
99%	.01	.005	2.576

Por tanto, para 90% de confianza, el margen de error es 3.29 y el intervalo de confianza es  $82 - 3.29 = 78.71$  a  $82 + 3.29 = 85.29$ . De manera similar, el intervalo de confianza de 99% es

$$82 \pm 2.576 \frac{20}{\sqrt{100}} \\ 82 \pm 5.15$$

Entonces, para 99% de confianza el margen de error es 5.15 y el intervalo de confianza es  $82 - 5.15 = 76.85$  a  $82 + 5.15 = 87.15$ .

Al comparar los resultados para los niveles de 90%, 95% y 99% de confianza, es claro que para tener mayor grado de confianza, el margen de error y con esto la amplitud del intervalo de confianza debe ser mayor.

## Recomendación práctica

Si la población tiene una distribución normal, el intervalo de confianza que se obtiene con la expresión (8.1) es exacto. En otras palabras, si la expresión (8.1) se usa repetidas veces para generar intervalos de 95% de confianza, exactamente 95% de los intervalos generados contendrán la media poblacional. Si la población no tiene una distribución normal, el intervalo de confianza obtenido con la expresión (8.1) será aproximado. En este caso la calidad de la aproximación depende tanto de la distribución de la población como del tamaño de la muestra.

En la mayoría de las aplicaciones, cuando se usa la expresión (8.1), un tamaño de muestra  $n \geq 30$  es adecuado para obtener una estimación por intervalo de la media poblacional. Si la población no está distribuida normalmente, pero es más o menos simétrica, tamaños de muestra hasta de 15 puede esperarse que proporcionen una buena aproximación del intervalo de confianza. Con tamaños de muestra menores, la expresión (8.1) sólo se debe usar si el analista cree, o está dispuesto a suponer, que la distribución de la población es cuando menos aproximadamente normal.

## NOTAS Y COMENTARIOS

1. El procedimiento de estimación por intervalo visto en esta sección se basa en la suposición de que la desviación estándar poblacional  $\sigma$  sea conocida. Decir que  $\sigma$  es conocida significa que se cuenta con datos históricos o con otra información que permita obtener una buena estimación de la desviación estándar antes de tomar la muestra que se usará para obtener la estimación de la media poblacional. De manera que, técnicamente esto no significa que  $\sigma$  se conozca con seguridad. Lo que significa es que sólo se obtuvo una buena estimación de la desviación estándar antes de tomar la muestra y que de esta manera no se usará la misma muestra para estimar tanto la media poblacional como la desviación estándar poblacional.
2. El tamaño de la muestra  $n$  aparece en el denominador de la expresión (8.1) para la estimación por intervalo. En consecuencia, si un determinado tamaño de muestra da un intervalo demasiado amplio, para que tenga utilidad práctica, se aumenta el tamaño de la muestra. Si  $n$  está en el denominador, con un tamaño de muestra mayor se obtendrá un margen de error menor, un intervalo más estrecho y mayor precisión. El procedimiento para determinar el tamaño de una muestra aleatoria simple que se necesita para obtener una determinada precisión se verá en la sección 8.3.

## Ejercicios

### Métodos

1. En una muestra aleatoria simple de 40 artículos la media muestral obtenida es 25. La desviación estándar poblacional es  $\sigma = 5$ .
  - a. ¿Cuál es el error estándar de la media  $\sigma_{\bar{x}}$ ?
  - b. ¿Cuál es el margen de error para tener 95% de confianza?

**Autoexamen**

2. En una muestra aleatoria simple de 50 artículos de una población en la que  $\sigma = 6$  la media muestral fue 32.
  - a. Proporcione un intervalo de confianza de 90% para la media poblacional.
  - b. Establezca un intervalo de confianza de 95% para la media poblacional.
  - c. Proporcione un intervalo de confianza de 99% para la media poblacional.
3. En una muestra aleatoria simple de 60 artículos la media muestral fue 80. La desviación estándar poblacional es  $\sigma = 15$ .
  - a. Calcule el intervalo de confianza de 95% para la media poblacional.
  - b. Suponga que la misma media muestral se obtuvo de una muestra de 120 artículos. Dé el intervalo de confianza de 95% para la media poblacional.
4. Para la media poblacional se dio el siguiente intervalo de confianza de 95%, de 152 a 160. Si  $\sigma = 15$ , ¿cuál es el tamaño de la muestra que se usó en este estudio?

**Aplicaciones****Autoexamen**

5. Con objeto de estimar la cantidad media que gasta un cliente en una comida en un importante restaurante, se recogieron los datos de una muestra de 49 clientes. Suponga que la desviación estándar de la población es \$5.
  - a. ¿Cuál es el margen de error para 95% de confianza?
  - b. Si la media poblacional es \$24.80, ¿cuál es el intervalo de confianza de 95% para la media poblacional?
6. Nielsen Media Research llevó a cabo un estudio para saber cuánto tiempo se veía televisión en los hogares, en el horario de 8:00 a 11:00 de la noche. Los datos que se encuentran en el archivo Nielsen del disco compacto son consistentes con los hallazgos reportados (*The World Almanac, 2003*). Con base en estudios anteriores, la desviación estándar poblacional se considera conocida y es  $\sigma = 3.5$  horas. Dé una estimación mediante un intervalo de confianza de 95% para la media del tiempo que se ve televisión por semana en el horario de 8:00 a 11:00 de la noche.
7. En una investigación sobre los negocios pequeños que tienen un sitio en la Web se encontró que la cantidad promedio que se gasta en un sitio es \$11 500 por año. Dada una muestra de 60 negocios y una desviación estándar  $\sigma = \$4000$ , ¿cuál es el margen de error? Use 95% de confianza. ¿Qué recomendaría si el estudio requiere un margen de error de \$500?
8. The National Quality Research Center de la University of Michigan proporciona medidas trimestrales de las opiniones de los consumidores acerca de productos y servicios (*The Wall Street Journal*, 18 de febrero de 2003). En una encuesta sobre 10 restaurantes de comida rápida y pizza la media del índice de satisfacción de los consumidores fue 71. Datos anteriores indican que la desviación estándar ha sido relativamente estable y es  $\sigma = 5$ .
  - a. ¿Qué debe estar dispuesto a suponer el investigador si desea un margen de error?
  - b. Con 95% de confianza, ¿cuál es el margen de error?
  - c. ¿Cuál es el margen de error si se desea 99% de confianza?
9. El puntuaje promedio en el examen de admisión de los estudiantes que ingresaron a la escuela de negocios fue 3.37 (*Best Graduate Schools, U.S. News and World Report*, 2001). Suponga que dicha estimación se basó en una muestra de 120 estudiantes. De acuerdo con datos anteriores se admite que se conoce la desviación estándar y que es  $\sigma = 0.28$ . ¿Cuál es la estimación mediante un intervalo de confianza de 95% para la media del puntuaje promedio de los alumnos que ingresaron a la escuela de negocios?
10. La revista *Playbill* reportó que el ingreso familiar anual medio de sus suscriptores es \$119 155 (*Playbill*, enero de 2006). Suponga que la estimación del ingreso familiar anual medio está basada en una muestra de 80 familias y que por datos de estudios anteriores la desviación estándar poblacional es conocida y es  $\sigma = \$30\ 000$ .

- Desarrolle un intervalo de estimación de 90% de confianza para la media poblacional.
- Dé un intervalo de estimación de 95% de confianza para la media poblacional.
- Dé un intervalo de estimación de 99% de confianza para la media poblacional.
- ¿Qué le pasa a la amplitud del intervalo de confianza a medida que el nivel de confianza aumenta? ¿Parece esto razonable? Explique.

## 8.2

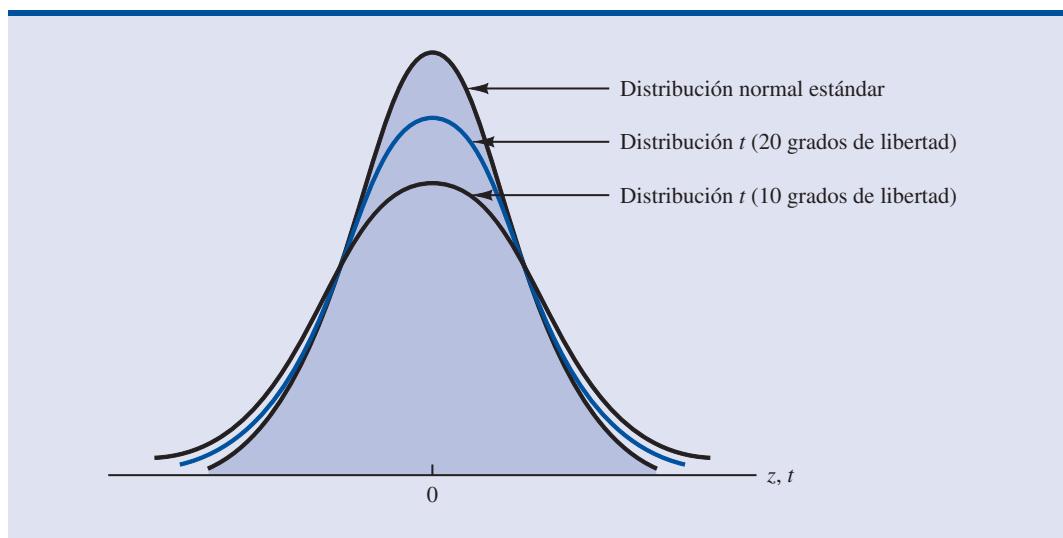
## Media poblacional: $\sigma$ desconocida

*William Sealy Gosset, quien publicaba bajo el seudónimo "Student" estableció la distribución t. Gosset, que había estudiado matemáticas en Oxford, trabajaba para Guinness Brewery en Dublín, Irlanda. Elaboró la distribución t cuando trabajaba con materiales a pequeña escala y hacía experimentos de temperatura.*

Cuando se calcula un intervalo de confianza para la media poblacional, suele no contarse con una buena estimación de la desviación estándar poblacional. En tales casos se usa la misma muestra para estimar  $\mu$  y  $\sigma$ . Esta situación es el caso que se conoce como  **$\sigma$  desconocida**. Cuando se usa  $s$  para estimar  $\sigma$ , el margen de error y la estimación por intervalo de la media poblacional se basan en una distribución de probabilidad conocida como **distribución t**. Aunque la elaboración matemática de la distribución t parte de la suposición de que la población de la que se muestrea tiene una distribución normal, las investigaciones han demostrado que la distribución t se aplica en muchas situaciones en que la población se desvía significativamente de una población normal. Más adelante, en esta misma sección, se proporcionan lineamientos para usar la distribución t cuando la población no está distribuida normalmente.

La distribución t es una familia de distribuciones de probabilidad similares; cada distribución t depende de un parámetro conocido como **grados de libertad**. La distribución t para un grado de libertad es única, como lo es la distribución t para dos grados de libertad, para tres grados de libertad, etc. A medida que el número de grados de libertad aumenta, la diferencia entre la distribución t y la distribución normal estándar se va reduciendo. En la figura 8.4 se muestran las distribuciones t para 10 y 20 grados de libertad y su relación con la distribución de probabilidad normal estándar. Observe que una distribución t para más grados de libertad exhibe menos va-

**FIGURA 8.4** COMPARACIÓN DE LA DISTRIBUCIÓN NORMAL ESTÁNDAR CON LAS DISTRIBUCIONES t PARA 10 Y 20 GRADOS DE LIBERTAD



riabilidad y un mayor parecido con la distribución normal estándar, también que la media de toda distribución  $t$  es cero.

Para indicar el área en la cola superior de la distribución  $t$ , a la  $t$  se le pone un subíndice. Por ejemplo, así como se usó  $z_{0.025}$  para indicar el valor de  $z$  que deja en la cola superior de la distribución normal estándar un área de 0.025, se usará  $t_{0.025}$  para indicar el valor de  $t$  que deja en la cola superior de la distribución  $t$  un área de 0.025. En general se usará la notación  $t_{\alpha/2}$  para representar el valor de  $t$  que deja un área de  $\alpha/2$  en la cola superior de la distribución  $t$ . Véase figura 8.5.

La tabla 2 del apéndice B contiene una tabla con la distribución  $t$ . En la tabla 8.2 se muestra parte de esa tabla. Cada renglón de la tabla corresponde a una distribución  $t$  distinta con los grados de libertad que se indican. Por ejemplo, en la distribución  $t$  con 9 grados de libertad,  $t_{0.025} = 2.262$ . De manera similar en la distribución  $t$  con 60 grados de libertad,  $t_{0.025} = 2.000$ . A medida que los grados de libertad aumentan,  $t_{0.025}$  se aproxima a  $z_{0.025} = 1.96$ . En efecto, el valor  $z$  de la distribución normal estándar se encuentra en el renglón correspondiente a infinitos grados de libertad (cuyo encabezado es  $\infty$ ) de la tabla de las distribuciones  $t$ . Si los grados de libertad son más de 100, se puede usar el renglón correspondiente a infinitos grados de libertad para aproximar el verdadero valor de  $t$ ; en otras palabras, para más de 100 grados de libertad, el valor  $z$  normal estándar proporciona una buena aproximación al valor  $t$ .

*A medida que los grados de libertad aumentan, la distribución  $t$  se approxima más a la distribución normal estándar.*

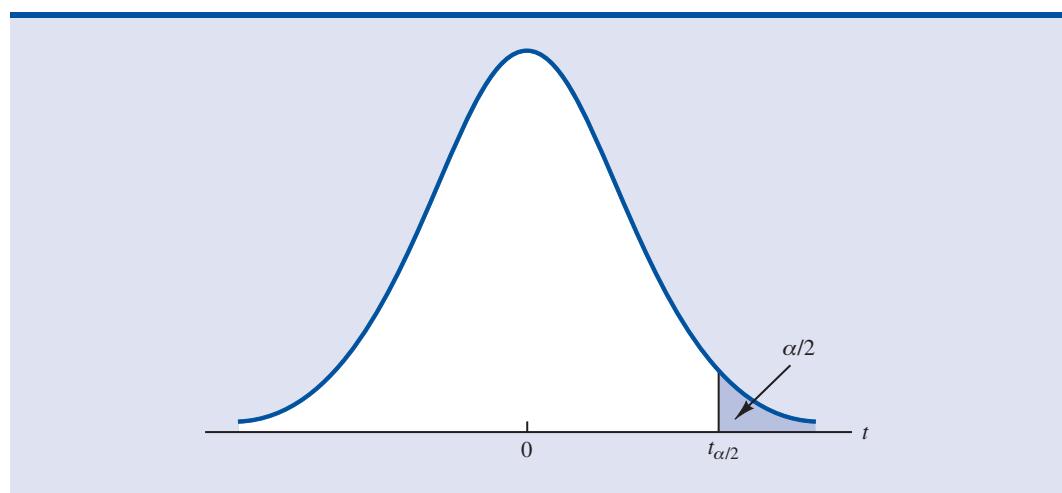
## Margen de error en estimación por intervalo

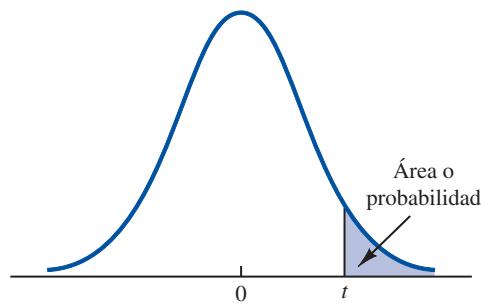
En la sección 8.1 se mostró que la estimación por intervalo de la media poblacional cuando  $\sigma$  es conocida es

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Para calcular una estimación por intervalo para  $\mu$  en el caso en que no se conoce  $\sigma$  se usa la desviación estándar muestral  $s$  para estimar  $\sigma$ , y  $z_{\alpha/2}$  se sustituye por el valor  $t_{\alpha/2}$  de la distribución  $t$ .

**FIGURA 8.5** ÁREA DE DISTRIBUCIÓN  $t$  CON UN ÁREA O PROBABILIDAD  $\alpha/2$  EN LA COLA SUPERIOR



**TABLA 8.2** ALGUNOS VALORES DE LA TABLA DE LA DISTRIBUCIÓN  $t^*$ 

Grados de libertad	Área en la cola superior					
	0.20	0.10	0.05	0.025	0.01	0.005
1	1.376	3.078	6.314	12.706	31.821	63.656
2	1.061	1.886	2.920	4.303	6.965	9.925
3	0.978	1.638	2.353	3.182	4.541	5.841
4	0.941	1.533	2.132	2.776	3.747	4.604
5	0.920	1.476	2.015	2.571	3.365	4.032
6	0.906	1.440	1.943	2.447	3.143	3.707
7	0.896	1.415	1.895	2.365	2.998	3.499
8	0.889	1.397	1.860	2.306	2.896	3.355
9	0.883	1.383	1.833	2.262	2.821	3.250
60	0.848	1.296	1.671	2.000	2.390	2.660
61	0.848	1.296	1.670	2.000	2.389	2.659
62	0.847	1.295	1.670	1.999	2.388	2.657
63	0.847	1.295	1.669	1.998	2.387	2.656
64	0.847	1.295	1.669	1.998	2.386	2.655
65	0.847	1.295	1.669	1.997	2.385	2.654
66	0.847	1.295	1.668	1.997	2.384	2.652
67	0.847	1.294	1.668	1.996	2.383	2.651
68	0.847	1.294	1.668	1.995	2.382	2.650
69	0.847	1.294	1.667	1.995	2.382	2.649
90	0.846	1.291	1.662	1.987	2.368	2.632
91	0.846	1.291	1.662	1.986	2.368	2.631
92	0.846	1.291	1.662	1.986	2.368	2.630
93	0.846	1.291	1.661	1.986	2.367	2.630
94	0.845	1.291	1.661	1.986	2.367	2.629
95	0.845	1.291	1.661	1.985	2.366	2.629
96	0.845	1.290	1.661	1.985	2.366	2.628
97	0.845	1.290	1.661	1.985	2.365	2.627
98	0.845	1.290	1.661	1.984	2.365	2.627
99	0.845	1.290	1.660	1.984	2.364	2.626
100	0.845	1.290	1.660	1.984	2.364	2.626
$\infty$	0.842	1.282	1.645	1.960	2.326	2.576

\*Nota: Una tabla más extensa se encuentra en la tabla 2 del apéndice B.

El margen de error está dado, entonces, por  $t_{\alpha/2}s/\sqrt{n}$ . Con este margen de error, la expresión general para una estimación por intervalo de la media poblacional cuando  $\sigma$  no se conoce es la siguiente.

#### ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL: $\sigma$ DESCONOCIDA

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad (8.2)$$

donde  $s$  es la desviación estándar muestral,  $(1 - \alpha)$  es el coeficiente de confianza y  $t_{\alpha/2}$  es el valor de  $t$  que proporciona un área de  $\alpha/2$  en la cola superior de la distribución  $t$  para  $n - 1$  grados de libertad.

La razón de que el número de grados de libertad para el valor de  $t$  en la expresión (8.2) sea  $n - 1$  se debe al uso de  $s$  como estimación de la desviación estándar poblacional  $\sigma$ . La expresión para calcular la desviación estándar muestral es

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

Los grados de libertad se refieren al número de valores independientes en el cálculo de  $\sum(x_i - \bar{x})^2$ . Los  $n$  valores en el cálculo de  $\sum(x_i - \bar{x})^2$  son los siguientes:  $x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}$ . En la sección 3.2 se indicó que en cualquier conjunto de datos  $\sum(x_i - \bar{x}) = 0$ . Por tanto, únicamente  $n - 1$  de los valores  $x_i - \bar{x}$  son independientes; es decir, si se conocen  $n - 1$  de estos valores, el valor restante puede determinarse exactamente usando el hecho de que los valores  $x_i - \bar{x}$  deben sumar 0. Entonces,  $n - 1$  es el número de grados de libertad en la suma  $\sum(x_i - \bar{x})^2$  y de ahí el número de grados de libertad para la distribución  $t$  en la expresión (8.2).

Para ilustrar la estimación por intervalo en el caso  $\sigma$  desconocida, se verá un estudio realizado para estimar la media del adeudo en las tarjetas de crédito en la población de familias de Estados Unidos. En la tabla 8.3 se presentan los saldos en las tarjetas de crédito de una muestra de  $n = 70$  familias. En esta ocasión no se cuenta con una estimación previa de la desviación estándar poblacional  $\sigma$ . De manera que los datos muestrales deberán usarse para estimar tanto la media poblacional como la desviación estándar poblacional. Con los datos de la tabla 8.3 se calcula

**TABLA 8.3** SALDOS EN LAS TARJETAS DE CRÉDITO DE UNA MUESTRA DE 70 FAMILIAS

9 430	14 661	7 159	9 071	9 691	11 032
7 535	12 195	8 137	3 603	11 448	6 525
4 078	10 544	9 467	16 804	8 279	5 239
5 604	13 659	12 595	13 479	5 649	6 195
5 179	7 061	7 917	14 044	11 298	12 584
4 416	6 245	11 346	6 817	4 353	15 415
10 676	13 021	12 806	6 845	3 467	15 917
1 627	9 719	4 972	10 493	6 191	12 591
10 112	2 200	11 356	615	12 851	9 743
6 567	10 746	7 117	13 627	5 337	10 324
13 627	12 744	9 465	12 557	8 372	
18 719	5 742	19 263	6 232	7 445	

la media muestral  $\bar{x} = \$9312$  y la desviación estándar muestral  $s = \$4007$ . Ahora se usa la tabla 8.2 para obtener el valor de  $t_{0.025}$  correspondiente a 95% de confianza y  $n - 1 = 69$  grados de libertad. El valor de  $t$  que se necesita está en el renglón correspondiente a 69 grados de libertad y en la columna correspondiente a 0.025 en la cola superior. El valor que se encuentra es  $t_{0.025} = 1.995$ .

Con la expresión (8.2), para calcular la estimación por intervalo de la media poblacional de los saldos en las tarjetas de crédito, se tiene:

$$9312 \pm 1.995 \frac{4007}{\sqrt{70}} \\ 9312 \pm 955$$

La estimación puntual de la media poblacional es \$9312, el margen de error es \$955 y el intervalo de confianza de 95% va de  $9312 - 955 = \$8357$  a  $9312 + 955 = \$10 267$ . En consecuencia, 95% de confianza de la media de los saldos en las tarjetas de crédito de la población de todas las familias está entre \$8357 y \$10 267.

En los apéndices 8.1 y 8.2 se describen los procedimientos para obtener un intervalo de confianza para la media poblacional usando Minitab y Excel. En la figura 8.6 se muestran los resultados para el estudio de los saldos en las tarjetas de crédito que da el procedimiento de Minitab para la estimación por intervalo. Con la muestra de 70 familias se obtiene una media muestral de \$9312 para los saldos en las tarjetas de crédito, una desviación estándar muestral de \$4007, una estimación del error estándar de la media de \$479 (valor redondeado) y un intervalo de confianza de 95% que va de \$8357 a \$10 267.

## Recomendación práctica

Si la población tiene una distribución normal, el intervalo de confianza suministrado en la expresión (8.2) es exacto y se puede usar con cualquier tamaño de muestra. Si la población no sigue una distribución normal, el intervalo de confianza en la expresión (8.2) será aproximado. En este caso la calidad de la aproximación depende tanto de la distribución de la población como del tamaño de la muestra.

En la mayoría de las aplicaciones un tamaño de muestra  $n \geq 30$  es suficiente al usar la expresión (8.2), para obtener una estimación por intervalo de la media poblacional. Sin embargo, si la distribución de la población es muy sesgada o si hay observaciones atípicas, la mayoría de los especialistas en estadística recomienda un tamaño de muestra de 50 o más. Si la población no tiene una distribución normal pero es más o menos simétrica, ya con un tamaño de muestra de 15 puede esperarse una buena aproximación al intervalo de confianza. Con muestras más pequeñas la expresión (8.2) sólo debe usarse si el analista cree, o está dispuesto a suponer, que la distribución de la población es por lo menos aproximadamente normal.

*Cuando la distribución de la población es muy sesgada o cuando hay observaciones atípicas se necesitan muestras grandes.*

## Uso de una muestra pequeña

En el ejemplo siguiente se obtiene una estimación por intervalo para una media poblacional teniendo una muestra pequeña. Como ya se indicó, conocer la distribución de la población es importante para decidir si mediante una estimación por intervalo se obtendrán resultados aceptables.

Scheer Industries está considerando un nuevo programa asistido por computadora con el fin de capacitar a los empleados de mantenimiento para realizar la reparación de las máquinas. Con

**FIGURA 8.6** INTERVALO DE CONFIANZA DE MINITAB PARA EL ESTUDIO DE LOS SALDOS EN LAS TARJETAS DE CRÉDITO

Variable	N	Mean	StDev	SE Mean	95% CI
NewBalance	70	9312.00	4007.00	478.93	(8356.56, 10 267.44)

**TABLA 8.4** DURACIÓN DE LA CAPACITACIÓN, EN DÍAS, EN LA MUESTRA DE 20 EMPLEADOS DE SCHEER INDUSTRIES

archivo en CD  
Scheer

52	59	54	42
44	50	42	48
55	54	60	55
44	62	62	57
45	46	43	56

objeto de evaluar este programa, el director de manufactura solicita una estimación de la media poblacional del tiempo requerido para que los empleados de mantenimiento completen la capacitación asistida por computadora.

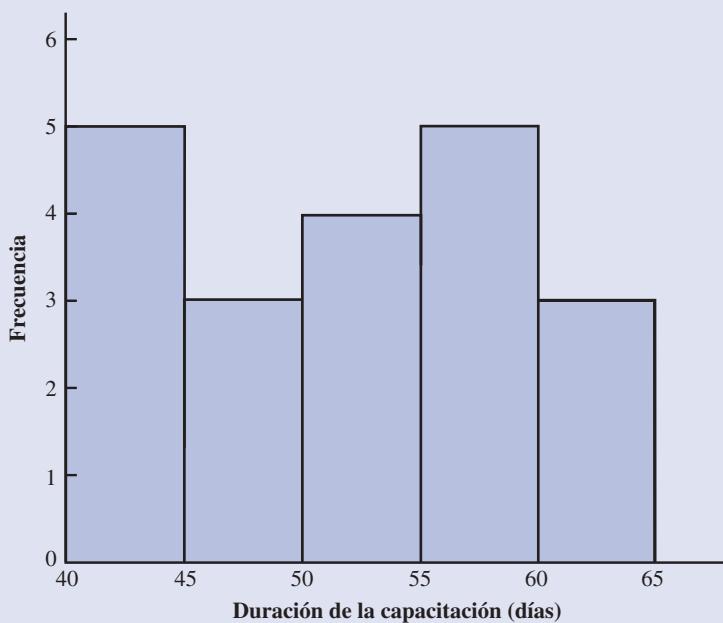
Considere una muestra de 20 empleados que siguen el programa de capacitación. En la tabla 8.4 se muestran los datos del tiempo, en días, que necesitó cada uno de los empleados para el programa de capacitación. En la figura 8.7 aparece un histograma de los datos. De acuerdo al histograma, ¿qué se puede decir de la distribución de estos datos? Primero, de acuerdo con los datos muestrales, no es posible concluir que la población sea normal, si bien no se observan evidencias de sesgo o de observaciones atípicas. Por tanto, mediante los lineamientos de la subsección anterior, se concluye que una estimación por intervalo basada en la distribución *t* parece ser aceptable para esta muestra de 20 empleados.

A continuación se calcula la media muestral y la desviación estándar muestral.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1030}{20} = 51.5 \text{ días}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{889}{20-1}} = 6.84 \text{ días}$$

**FIGURA 8.7** HISTOGRAMA SOBRE LA DURACIÓN DE LA CAPACITACIÓN EN LA MUESTRA DE SCHEER INDUSTRIES



Para dar un intervalo de confianza de 95%, se usa la tabla 2 del apéndice B y  $n - 1 = 19$  grados de libertad y se obtiene  $t_{0.025} = 2.093$ . La expresión (8.2) suministra la estimación por intervalo de la media poblacional.

$$51.5 \pm 2.093 \left( \frac{6.84}{\sqrt{20}} \right)$$

$$51.5 \pm 3.2$$

La estimación puntual de la media poblacional es 51.5 días. El margen de error es 3.2 días y el intervalo de confianza de 95% va de  $51.5 - 3.2 = 48.3$  días a  $51.5 + 3.2 = 54.7$  días.

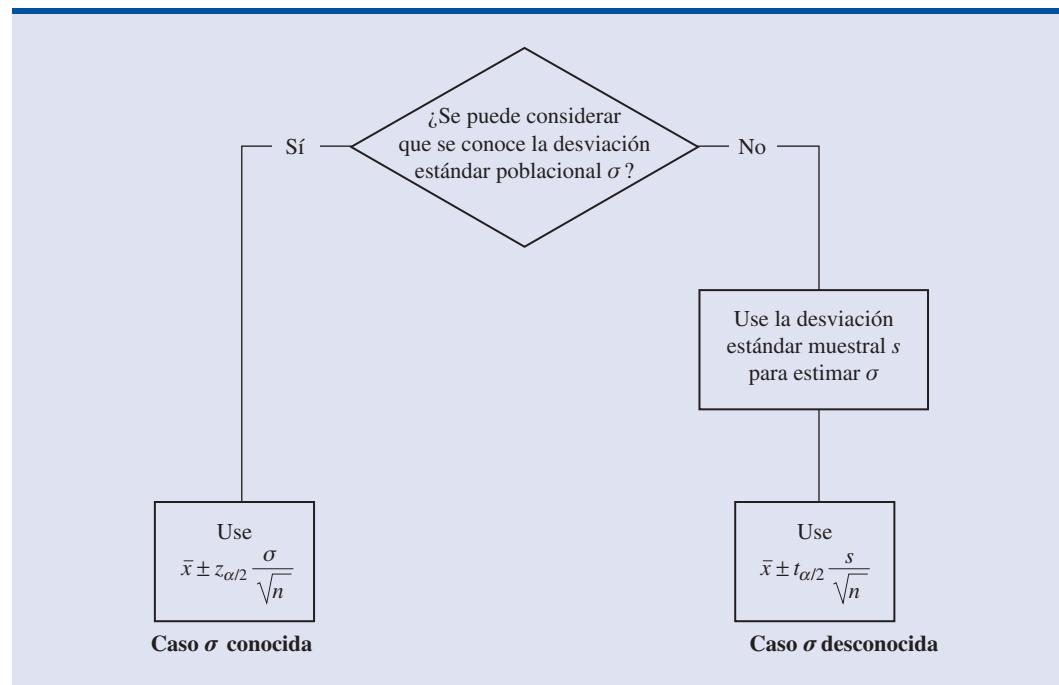
Usar un histograma de los datos muestrales para tener información acerca de la distribución de la población no es siempre concluyente, pero en muchos casos es la única información disponible. El histograma, junto con la opinión del analista, suele usarse para decidir si es adecuado usar la expresión (8.2) para obtener una estimación por intervalo.

## Resumen de los procedimientos de estimación por intervalo

Se presentaron dos métodos para calcular una estimación por intervalo para la media poblacional. En el caso en que  $\sigma$  es conocida, en la expresión (8.1) se usan  $\sigma$  y la distribución normal estándar para calcular el margen de error y dar la estimación por intervalo. En el caso en que  $\sigma$  no es conocida, en la expresión (8.2) se usan la desviación estándar muestral  $s$  y la distribución  $t$  para calcular el margen de error y dar una estimación por intervalo.

En la figura 8.8 se presenta un resumen de los procedimientos para la estimación por intervalo de los dos casos. En la mayoría de las aplicaciones un tamaño de muestra  $n \geq 30$  es adecuado. Sin embargo, si la población tiene distribución normal o aproximadamente normal, se pueden usar tamaños de muestra menores. En caso de que no se conozca  $\sigma$  y si la distribución de la población es muy sesgada o existen observaciones atípicas, se recomienda que el tamaño de la muestra sea  $n \geq 50$ .

**FIGURA 8.8** RESUMEN DE LOS PROCEDIMIENTOS PARA LA ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL



## NOTAS Y COMENTARIOS

1. Cuando se conoce  $\sigma$ , el margen de error,  $z_{\alpha/2}(\sigma/\sqrt{n})$ , es fijo y es el mismo para todas las muestras de tamaño  $n$ . Cuando  $\sigma$  no se conoce, el margen de error,  $t_{\alpha/2}(s/\sqrt{n})$ , varía de una muestra a otra. Esta variación se debe a que la desviación estándar muestral  $s$  varía con la muestra que se seleccione. Si  $s$  es grande, se obtiene un margen de error grande, mientras que si  $s$  es pequeña, se obtiene un margen de error pequeño.
2. ¿Qué pasa con las estimaciones por intervalo cuando la población es sesgada? Considere una población sesgada a la derecha, en la cual los datos con valores grandes jalan la distribución hacia la derecha. Cuando existe un sesgo así, hay una correlación positiva entre la media muestral  $\bar{x}$  y la desviación estándar muestral  $s$ . Valores mayores de  $s$  tienden a corresponder a

valores mayores de  $\bar{x}$ . De esta manera, cuando  $\bar{x}$  es mayor que la media poblacional,  $s$  tiende a ser mayor que  $\sigma$ . Este sesgo hace que el margen de error,  $t_{\alpha/2}(s/\sqrt{n})$ , sea mayor de lo que sería si se conociera  $\sigma$ . Un intervalo de confianza con un margen de error mayor tenderá a incluir con más frecuencia a la media poblacional  $\mu$  que si se usara el verdadero valor  $\sigma$ . Pero cuando  $\bar{x}$  es menor que la media poblacional, la correlación entre  $\bar{x}$  y  $s$  hace que el margen de error sea menor. En este caso, dichos intervalos de confianza con menor margen de error incluirán a la media poblacional menos veces que si se conociera  $\sigma$  y se usara  $\sigma$ . Por esta razón se recomienda usar tamaños de muestra más grandes cuando la distribución de la población es muy sesgada.

## Ejercicios

### Métodos

11. En la distribución  $t$  con 16 grados de libertad, encuentre el área, o la probabilidad, de cada una de las regiones siguientes:
  - a. A la derecha de 2.120
  - b. A la izquierda de 1.337
  - c. A la izquierda de -1.746
  - d. A la derecha de 2.583
  - e. Entre -2.120 y 2.120
  - f. Entre -1.746 y 1.746
12. Encuentre los valores de  $t$  para las situaciones siguientes.
  - a. Un área de 0.025 en la cola superior, con 12 grados de libertad
  - b. Un área de 0.05 en la cola inferior, con 50 grados de libertad
  - c. Un área de 0.01 en la cola superior, con 30 grados de libertad
  - d. Entre los que queda 90% del área, con 25 grados de libertad
  - e. Entre los que queda 95% del área, con 45 grados de libertad
13. Los datos muestrales siguientes provienen de una población normal: 10, 8, 12, 15, 13, 11, 6, 5.
  - a. ¿Cuál es la estimación puntual de la media poblacional?
  - b. ¿Cuál es la estimación puntual de la desviación estándar poblacional?
  - c. Con 95% de confianza, ¿cuál es el margen de error para la estimación de la media poblacional?
  - d. ¿Cuál es el intervalo de confianza de 95% para la media poblacional?
14. En una muestra aleatoria simple con  $n = 54$  la media muestral fue 22.5 y la desviación estándar muestral 4.4.
  - a. Encuentre un intervalo de confianza de 90% para la media poblacional.
  - b. Dé un intervalo de confianza de 95% para la media poblacional.
  - c. Dé un intervalo de confianza de 99% para la media poblacional.
  - d. ¿Qué pasa con el margen de error y con el intervalo de confianza a medida que aumenta el nivel de confianza?

## Aplicaciones



15. Los agentes de ventas de una empresa presentan un informe semanal que enumera los clientes contactados durante la semana. En una muestra de 65 informes semanales la media muestral es 19.5 clientes por semana. La desviación estándar es 5.2. Dé intervalos de confianza de 90% y 95% para la media poblacional del número de clientes contactados semanalmente por el personal de ventas.
16. El número medio de horas de vuelo de los pilotos de Continental Airlines es 49 horas por mes (*The Wall Street Journal*, 25 de febrero de 2003). Suponga que esta media se basó en las horas de vuelo de una muestra de 100 pilotos de esa empresa y que la desviación estándar muestral haya sido 8.5 horas.
- A 95% de confianza, ¿cuál es el margen de error?
  - Dé el intervalo de estimación de 95% para la media poblacional de las horas de vuelo de los pilotos.
  - La media en las horas de vuelo de los pilotos de United Airlines es 36 horas por mes. Use los resultados del inciso b para analizar la diferencia entre la cantidad de horas de vuelo de los pilotos en las dos líneas aéreas. *The Wall Street Journal* informa que United Airlines tiene el costo laboral más elevado de todas las aerolíneas. La información dada en estos ejercicios, ¿sirve para entender por qué se puede esperar que United Airlines tenga los costos más elevados?
17. La International Air Transport Association realiza encuestas entre los viajeros de negocios en las que se califica la calidad de los aeropuertos de salida internacional. La calificación máxima es 10. Se seleccionó una muestra aleatoria simple de 50 viajeros de negocios y a cada uno se le pidió su calificación para el aeropuerto internacional de Miami. Las calificaciones que dieron estos 50 viajeros fueron las que se muestran a continuación.

archivo  
en CD  
Miami

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

- Calcule el intervalo de confianza de 95% para la media poblacional de las calificaciones al aeropuerto de Miami.

archivo  
en CD  
FastFood

18. Durante el verano de 2000 fueron visitados 30 restaurantes de comida rápida entre los que se encontraban Wendy's, McDonald's y Burger King (*The Cincinnati Enquirer*, 9 de julio de 2000). Se registró el tiempo que transcurría entre que el cliente hiciera su pedido y la recepción del mismo. Los tiempos en los 30 restaurantes visitados fueron los siguientes:

0.9	1.0	1.2	2.2	1.9	3.6	2.8	5.2	1.8	2.1
6.8	1.3	3.0	4.5	2.8	2.3	2.7	5.7	4.8	3.5
2.6	3.3	5.0	4.0	7.2	9.1	2.8	3.6	7.3	9.0

- Dé una estimación puntual de la media poblacional.
  - ¿Cuál es el margen de error con 95% de confianza?
  - ¿Cuál es la estimación por intervalo de confianza de 95% para la media poblacional?
  - Analice el sesgo que puede encontrarse en esta población. ¿Qué sugeriría para la repetición de este estudio?
19. En un estudio de National Retail Foundation se encontró que las familias estaban dispuestas a gastar en promedio \$649 durante las vacaciones decembrinas (*The Wall Street Journal*, 2 de diciembre de 2002). Suponga que en el estudio participaron 600 familias y que la desviación estándar muestral fue \$175.
- Con 95% de confianza cuál es el margen de error?
  - ¿Cuál es el intervalo de confianza de 95% para estimar la media poblacional?
  - El año anterior, la media poblacional de gastos por familia fue \$632. Analice la variación en el gasto en las vacaciones decembrinas en este periodo de un año.



20. ¿Los comerciales interrumpen constantemente su programa de televisión favorito? CNBC presentó datos estadísticos sobre la cantidad promedio de minutos de programa en media hora de transmisión (CNBC, 23 de febrero de 2006). Los datos siguientes (en minutos) son representativos de sus hallazgos.

21.06	22.24	20.62
21.66	21.23	23.86
23.82	20.30	21.52
21.52	21.91	23.14
20.02	22.20	21.20
22.37	22.19	22.34
23.36	23.44	

Suponga que la población es aproximadamente normal. Dé una estimación puntual y un intervalo de confianza de 95% para la cantidad media de minutos de programa en media hora de transmisión.

21. El consumo de las mujeres en edad de tomar bebidas alcohólicas ha aumentado en el Reino Unido, Estados Unidos y Europa (*The Wall Street Journal*, 15 de febrero de 2006). Datos (de consumo anual en litros) reportados por *The Wall Street Journal* hallados en una muestra de 20 mujeres jóvenes europeas son:

266	82	199	174	97
170	222	115	130	169
164	102	113	171	0
93	0	93	110	130

Si la población es más o menos simétrica, dé un intervalo de confianza de 95% para el consumo medio anual de bebidas alcohólicas entre las mujeres europeas jóvenes.

22. Las primeras semanas del 2004 fueron buenas para el mercado de acciones. En una muestra de 25 fondos abiertos se encontraron las siguientes ganancias obtenidas desde principio del año al 24 de enero del 2004 (*Barron's*, 19 de enero de 2004).

7.0	3.2	1.4	5.4	8.5
2.5	2.5	1.9	5.4	1.6
1.0	2.1	8.5	4.3	6.2
1.5	1.2	2.7	3.8	2.0
1.2	2.6	4.0	2.6	0.6

- a. ¿Cuál es la estimación puntual de la media poblacional de las ganancias en fondos abiertos desde principio del año hasta esa fecha?
- b. Puesto que la población tiene una distribución normal, calcule un intervalo de confianza de 95% para la media poblacional de las ganancias en fondos abiertos desde principio del año hasta esa fecha.



OpenEndFunds

### 8.3 Determinación del tamaño de la muestra

En las recomendaciones prácticas de las dos secciones anteriores, se habló del papel del tamaño de la muestra para obtener una buena aproximación a los intervalos de confianza en los casos en que la población no tiene una distribución normal, ahora se enfoca la atención a otro aspecto relacionado con el tamaño de la muestra. Se describe cómo elegir un tamaño de muestra suficientemente grande para obtener un margen de error deseado. Para explicar esto, se vuelve al caso de la sección 8.1 en el que se tenía una  $\sigma$  conocida. Con la expresión (8.1), el intervalo de estimación está dado por

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

La cantidad  $z_{\alpha/2}(\sigma/\sqrt{n})$  es el margen de error. De manera que, como se ve,  $z_{\alpha/2}$ , la desviación estándar poblacional  $\sigma$ , y el tamaño de la muestra  $n$  se combinan para determinar el margen de

*El procedimiento que se presenta en esta sección se emplea para determinar el tamaño de muestra que se necesita para tener un determinado margen de error que se ha establecido antes de tomar la muestra.*

error. Una vez que se selecciona el coeficiente de confianza  $1 - \alpha$ , se determina  $z_{\alpha/2}$ . Por tanto, si se tiene el valor de  $\sigma$ , es posible encontrar el tamaño de muestra  $n$  necesario para proporcionar cualquier margen de error deseado. A continuación se presenta la deducción de la fórmula que se usa para calcular el tamaño  $n$  de muestra deseado.

Sea  $E =$  el margen de error deseado

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Despejando  $\sqrt{n}$ , se tiene

$$\sqrt{n} = \frac{z_{\alpha/2}\sigma}{E}$$

Al elevar al cuadrado ambos lados de esta ecuación, se obtiene la expresión siguiente para el tamaño de la muestra.

*La ecuación (8.3) se usa para determinar el tamaño de muestra adecuado. Sin embargo, la opinión del analista deberá usarse para determinar si el tamaño de muestra final debe ser mayor.*

#### TAMAÑO DE MUESTRA PARA UNA ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} \quad (8.3)$$

Este tamaño de muestra proporciona el margen de error deseado al nivel de confianza elegido.

En la ecuación (8.3)  $E$  es el margen de error que el usuario está dispuesto a aceptar, y el valor  $z_{\alpha/2}$  es consecuencia directa del nivel de confianza que se va usar para calcular la estimación por intervalo. A reserva de la decisión del usuario, 95% de confianza es el valor más usado ( $z_{0.025} = 1.96$ ).

Por último, para usar la ecuación (8.3) es necesario contar con el valor de la desviación estándar poblacional  $\sigma$ . Sin embargo, aun cuando este valor no se conozca, puede usarse la ecuación (8.3) siempre que se tenga un valor preliminar o un *valor planeado* de  $\sigma$ . En la práctica, se suele usar alguno de los procedimientos siguientes para obtener este valor planeado de  $\sigma$ .

*El valor planeado de la desviación estándar poblacional  $\sigma$  debe especificarse antes de determinar el tamaño de la muestra. Aquí se muestran tres métodos para obtener este valor planeado de  $\sigma$ .*

1. Usar como valor planeado de  $\sigma$  una estimación de la desviación estándar poblacional calculada a partir de datos de estudios anteriores.
2. Emplear un estudio piloto seleccionando una muestra preliminar. La desviación estándar muestral obtenida de la muestra preliminar puede usarse como valor planeado de  $\sigma$ .
3. Use su juicio para el valor de  $\sigma$ . Por ejemplo, se puede empezar por estimar el mayor y el menor valor en los datos de la población. Esta diferencia entre el mayor y el menor valor proporciona una estimación del rango de los datos. Por último, este valor dividido entre 4 suele considerarse como una aproximación burda a la desviación estándar y tomarse como un valor planeado aceptable de  $\sigma$ .

Para demostrar cómo se usa la ecuación (8.3) en la determinación del tamaño de la muestra, se considera el ejemplo siguiente. En un estudio previo para investigar el costo de la renta de automóviles en Estados Unidos se encontró que el costo medio de la renta de un automóvil mediano era aproximadamente \$55 por día. Suponga que la organización que realizó dicho estudio quiere realizar un nuevo estudio para estimar la media poblacional de las rentas por día de automóviles medianos en Estados Unidos. Antes de iniciar, especificó que la media poblacional de las rentas por día debe estimarse con un margen de error de \$2 y que se desea un nivel de 95% de confianza.

El margen de error especificado es  $E = 2$ , el nivel 95% de confianza indica que  $z_{0.025} = 1.96$ . Por tanto, sólo falta un valor planeado de la desviación estándar poblacional  $\sigma$  para calcular el tamaño de muestra deseado. El analista revisó los datos muestrales del estudio anterior y encon-

*La ecuación (8.3) proporciona el tamaño de muestra mínimo necesario para obtener el margen de error deseado. Si el tamaño de muestra calculado no es un número entero, se redondea al siguiente número entero, con lo que se tendrá un margen de error ligeramente menor al requerido.*



tró que la desviación estándar poblacional del costo de la renta diaria era \$9.65. Usando \$9.65 como valor planeado de  $\sigma$ , se tiene

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} = \frac{(1.96)^2 (9.65)^2}{2^2} = 89.43$$

De esta manera, el tamaño de la muestra necesario para obtener un margen de error de \$2 debe ser de por lo menos 89.43 rentas de automóviles medianos. En casos como éste, en los que el valor de  $n$  no es un número entero, se redondea al siguiente valor entero; así que el tamaño de muestra que se aconseja es 90 rentas de automóviles medianos.

## Ejercicios

### Métodos

23. ¿Qué tan grande debe seleccionarse una muestra para tener un intervalo de confianza de 95% con un margen de error de 10? Suponga que la desviación estándar poblacional es 40.
24. En un conjunto de datos se estima que el rango es 36.
  - a. ¿Cuál es el valor planeado para la desviación estándar poblacional?
  - b. ¿De qué tamaño deberá ser la muestra para que el margen de error en un intervalo de confianza de 95% sea 3?
  - c. ¿De qué tamaño deberá ser la muestra para que el margen de error en un intervalo de confianza de 95% sea 2?

### Aplicaciones

25. Remítase al ejemplo de Scheer Industries de la sección 8.2. Como valor planeado para la desviación estándar poblacional use 6.84 día.
  - a. Asuma 95% de confianza, ¿de qué tamaño deberá ser la muestra para tener un margen de error de 1.5 días?
  - b. Si se desea un intervalo de 90% de confianza, ¿de qué tamaño deberá ser la muestra para tener un margen de error de 2 días?
26. El costo promedio de la gasolina sin plomo en Grater Cincinnati es \$2.41 (*The Cincinnati Enquirer*, 3 de febrero de 2006). En una época de cambios en los precios, un periódico muestrea las gasolineras y presenta un informe sobre los precios de la gasolina. Suponga que en los precios del galón de la gasolina sin plomo la desviación estándar es \$0.15; dé el tamaño de muestra  $n$  que debe usar este periódico para tener 95% de confianza con cada uno de los márgenes de error siguientes.
  - a. Un margen de error de \$0.07
  - b. Un margen de error de \$0.05
  - c. Un margen de error de \$0.03
27. Los salarios anuales iniciales de estudiantes que acaban de terminar una carrera en administración se espera que estén entre \$30 000 y \$45 000. Suponga que quiere dar un intervalo de confianza de 95% para estimar la media poblacional de los salarios iniciales. ¿Cuál es el valor planeado de la desviación estándar poblacional? ¿Cuán grande deberá ser la muestra si quiere que el margen de error sea
  - a. \$500?
  - b. \$200?
  - c. \$100?
  - d. ¿Recomendaría usted tratar de tener \$100 como margen de error?
28. Smith Travel Research proporciona información sobre los precios por noche de las habitaciones de hotel en Estados Unidos (*USA Today*, 8 de julio de 2002). Use \$2 como el margen de error deseado y \$22.50 como valor planeado para la desviación estándar poblacional, y encuentre los tamaños de muestra que se solicitan en los incisos a, b y c.
  - a. Para un intervalo de confianza de 90% estime el precio medio de las habitaciones de hotel.
  - b. Para un intervalo de confianza de 95% estime el precio medio de las habitaciones de hotel.

- c. Para un intervalo de confianza de 99% estime el precio medio de las habitaciones de hotel.
  - d. Cuando se tiene un margen de error fijo, ¿qué pasa con el tamaño de la muestra a medida que el nivel de confianza aumenta? ¿Le recomendaría a Smith Travel Research que use 99% como nivel de confianza? Discuta.
29. El 2003 *Information Please Almanac* presenta los tiempos que para transportarse al trabajo son requeridos en las 15 ciudades más grandes de Estados Unidos. Suponga que usa una muestra aleatoria simple preliminar de los habitantes de San Francisco y como valor planeado para la desviación estándar poblacional obtiene 6.25 minutos.
- a. Si desea estimar la media poblacional del tiempo que necesitan en San Francisco para transportarse al trabajo, con un margen de error de 2 minutos, ¿cuál debe ser el tamaño de la muestra? Suponga que el nivel de confianza es de 95%.
  - b. Si desea estimar la media poblacional del tiempo que se necesita en San Francisco para transportarse al trabajo, con un margen de error de 1 minuto, ¿cuál debe ser el tamaño de la muestra? Suponga que el nivel de confianza es de 95%.
30. El primer trimestre del 2003, la proporción precio/ganancia P/G en las acciones de la Bolsa de Nueva York iba de 5 a 60 (*The Wall Street Journal*, 7 de marzo de 2003). Si se desea estimar la media poblacional de esta relación P/G en todas las acciones de la Bolsa de Nueva York, ¿cuántas acciones habrá que tomar en la muestra, si se quiere que el margen de error sea 3? Use 95% de confianza.

## 8.4

## Proporción poblacional

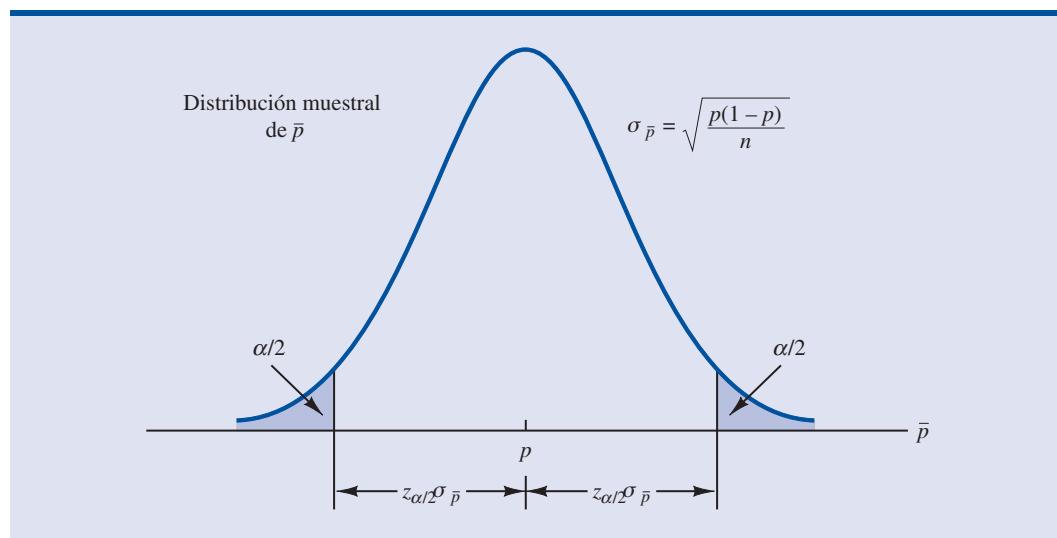
En la introducción a este capítulo se dijo que para obtener una estimación por intervalo de la proporción poblacional  $p$ , la fórmula general era

$$\bar{p} \pm \text{margen de error}$$

En el cálculo de esta estimación por intervalo la distribución muestral es importante.

En el capítulo 7 se dijo que la distribución muestral de  $\bar{p}$  se aproxima mediante una distribución normal siempre que  $np \geq 5$  y  $n(1-p) \geq 5$ . En la figura 8.9 se muestra una aproximación

**FIGURA 8.9** APROXIMACIÓN NORMAL A LA DISTRIBUCIÓN MUESTRAL DE  $\bar{p}$



normal a la distribución muestral de  $\bar{p}$ . La media de la distribución muestral de  $\bar{p}$  es la proporción poblacional  $p$  y el error estándar de  $\bar{p}$  es

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \quad (8.4)$$

Como la distribución muestral de  $\bar{p}$  es una distribución normal, si en la estimación por intervalo de la proporción poblacional se elige como margen de error  $z_{\alpha/2}\sigma_{\bar{p}}$ , entonces  $100(1 - \alpha)\%$  de los intervalos que se obtengan contendrán la verdadera proporción poblacional. Pero para calcular el margen de error no se puede usar  $\sigma_{\bar{p}}$ , ya que no se conoce  $p$ ;  $p$  es lo que se está tratando de estimar. Lo que se hace, es que  $p$  se sustituye por  $\bar{p}$  y de esta manera se calcula el margen de error y la estimación por intervalo de la proporción poblacional queda dada por

$$\text{Margen de error} = z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (8.5)$$

Con este margen de error la expresión general para la estimación por intervalo de la proporción poblacional es la siguiente.

#### ESTIMACIÓN POR INTERVALO DE UNA PROPORCIÓN POBLACIONAL

*El margen de error para un intervalo de confianza para la proporción poblacional está dado por la cantidad  $z_{\alpha/2}\sqrt{\bar{p}(1-\bar{p})}/n$*

$$\bar{p} \pm z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \quad (8.6)$$

donde  $1 - \alpha$  es el coeficiente de confianza y  $z_{\alpha/2}$  es el valor de  $z$  que deja un área  $\alpha/2$  en la cola superior de la distribución normal estándar.



En el siguiente ejemplo se ilustra el cálculo del margen de error y de la estimación por intervalo para una proporción muestral. Un estudio en Estados Unidos encuestó a 900 golfistas para conocer su opinión acerca de cómo se les trataba en los cursos de golf. En el estudio se encontró que 396 golfistas estaban satisfechas con la disponibilidad de horarios de salida. Por tanto, la estimación puntual de la proporción poblacional de golfistas satisfechas con la disponibilidad de horarios de salida es  $396/900 = 0.44$ . Usando la expresión (8.6) y el nivel de 95% de confianza,

$$\begin{aligned} \bar{p} &\pm z_{\alpha/2} \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\ 0.44 &\pm 1.96 \sqrt{\frac{0.44(1-0.44)}{900}} \\ 0.44 &\pm 0.0324 \end{aligned}$$

En consecuencia, el margen de error es 0.0324 y la estimación por intervalo de confianza de 95% de la proporción poblacional es 0.4076 a 0.4724. Empleando porcentajes, los resultados de la investigación permiten decir con 95% de confianza que entre 40.76 y 47.24% de las golfistas están satisfechas con la disponibilidad de horarios de salida.

## Determinación del tamaño de la muestra

Ahora se considera cuál debe ser el tamaño de la muestra para obtener una estimación de la proporción poblacional con una precisión determinada. La función que tiene el tamaño de la muestra en la determinación de la estimación por intervalo de  $p$  es semejante a la que tiene en la estimación de la media poblacional estudiada en la sección 8.3. Ya en esta sección se dijo que el margen de error en la estimación por intervalo de la proporción poblacional es  $z_{\alpha/2}\sqrt{\bar{p}(1 - \bar{p})/n}$ . El margen de error se basa en el valor de  $z_{\alpha/2}$ , en la proporción muestral  $\bar{p}$ , y en el tamaño de la muestra  $n$ . Muestras mayores proporcionan márgenes de error menores y mejor precisión.

Sea  $E$  el margen de error deseado.

$$E = z_{\alpha/2} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

Despejando  $n$  de esta fórmula se obtiene la fórmula para calcular el tamaño de la muestra con el que se tendrá el margen de error deseado,  $E$ .

$$n = \frac{(z_{\alpha/2})^2 \bar{p}(1 - \bar{p})}{E^2}$$

Sin embargo, debido a que no se conocerá  $\bar{p}$  sino hasta que se tome la muestra, no es posible usar esta fórmula para calcular el tamaño de la muestra con el que se obtendrá el margen de error deseado. Lo que se necesita, entonces, es un valor planeado de  $\bar{p}$  útil para hacer este cálculo. Con  $p^*$  como valor planeado de  $\bar{p}$  la fórmula para calcular el tamaño de la muestra con el que se obtendrá el error  $E$  queda como se presenta a continuación.

### TAMAÑO DE LA MUESTRA PARA UNA ESTIMACIÓN POR INTERVALO DE LA PROPORCIÓN POBLACIONAL

$$n = \frac{(z_{\alpha/2})^2 p^*(1 - p^*)}{E^2} \quad (8.7)$$

En la práctica, el valor planeado  $p^*$  se determina mediante alguno de los métodos siguientes.

1. Utilizar la proporción poblacional de una muestra previa de las mismas unidades o de unidades similares.
2. Utilizar un estudio piloto y elegir una muestra preliminar. La proporción muestral de esta muestra se usa como valor planeado,  $p^*$ .
3. Proponer una “mejor aproximación” para el valor de  $p^*$ .
4. Si no aplica ninguna de las alternativas anteriores, emplear como valor planeado  $p^* = 0.50$ .

De regreso al estudio con mujeres golfistas, suponga que la empresa desea llevar a cabo otro estudio para determinar la proporción en la población de golfistas que está satisfecha con la disponibilidad de horarios de salida. ¿De qué tamaño deberá ser la muestra si se desea que en la estimación de la proporción poblacional el margen de error sea 0.025 a 95% de confianza? Como  $E = .025$  y  $z_{\alpha/2} = 1.96$ , se necesita un valor planeado  $p^*$  para responder esta pregunta sobre el tamaño de la muestra. Utilizando como valor planeado  $p^*$ , el resultado del estudio anterior,  $\bar{p} = 0.44$ , con la ecuación (8.7) se obtiene

$$n = \frac{(z_{\alpha/2})^2 p^*(1 - p^*)}{E^2} = \frac{(1.96)^2(0.44)(1 - 0.44)}{(0.025)^2} = 1514.5$$

**TABLA 8.5** ALGUNOS VALORES DE  $p^*(1 - p^*)$ 

$p^*$	$p^*(1 - p^*)$	
0.10	(0.10)(0.90) = 0.09	
0.30	(0.30)(0.70) = 0.21	
0.40	(0.40)(0.60) = 0.24	
0.50	(0.50)(0.50) = 0.25	← Mayor valor de $p^*(1 - p^*)$
0.60	(0.60)(0.40) = 0.24	
0.70	(0.70)(0.30) = 0.21	
0.90	(0.90)(0.10) = 0.09	

Así, el tamaño de la muestra debe ser por lo menos 1514.5 golfistas para tener el margen de error requerido. Redondeando al valor entero siguiente, se tiene que se necesitan 1515 golfistas para obtener el margen de error deseado.

La cuarta alternativa sugerida para seleccionar un valor planeado  $p^*$  es elegir  $p^* = 0.50$ . Cuando no se cuenta con ninguna otra información suele usarse este valor. Para entender la razón, observe que el numerador de la ecuación (8.7) indica que el tamaño de la muestra es proporcional a la cantidad  $p^*(1 - p^*)$ . Si el valor de  $p^*(1 - p^*)$  es grande, el tamaño de la muestra será grande. En la tabla 8.5 se muestran algunos de los valores que puede tener  $p^*(1 - p^*)$ . El mayor valor de  $p^*(1 - p^*)$  se presenta cuando  $p^* = 0.50$ . De esta manera, en caso de duda acerca del valor planeado apropiado, se sabe que  $p^* = 0.50$  dará el mayor tamaño de muestra que se puede recomendar. En efecto, recomendando el mayor tamaño de muestra posible se va a lo seguro. Si resulta que la proporción muestral es diferente del valor planeado, el margen de error será menor que el deseado. De manera que al usar  $p^* = 0.50$ , se garantiza que el tamaño de la muestra será suficiente para obtener el margen de error deseado.

En el ejemplo del estudio de las golfistas, si se usa como valor planeado  $p^* = 0.50$ , el tamaño de muestra que se obtiene es

$$n = \frac{(z_{\alpha/2})^2 p^*(1 - p^*)}{E^2} = \frac{(1.96)^2 (0.50)(1 - 0.50)}{(0.025)^2} = 1536.6$$

Es decir, una muestra ligeramente mayor, 1537 golfistas.

### NOTAS Y COMENTARIOS

El margen de error deseado para calcular una proporción poblacional casi siempre es 0.10 o menos. En las encuestas de opinión pública a nivel nacional en Estados Unidos, conducidas por Gallup y Harris, un margen de error de 0.03 o 0.04 es común. Con dichos márgenes de error, la ecuación

(8.7) suministra un tamaño de la muestra que es suficiente para satisfacer los requerimientos de  $np \geq 5$  y  $n(1 - p) \geq 5$  para usar una distribución normal como una aproximación de la distribución muestral de  $\bar{x}$ .

### Ejercicios

### Métodos

31. En una muestra aleatoria simple de 400 individuos, 100 de las respuestas fueron Sí.
- Dé la estimación puntual de la proporción poblacional de individuos cuya respuesta será Sí.
  - Dé la estimación del error estándar de la proporción  $\sigma_{\hat{p}}$ .
  - Calcule el intervalo de confianza de 95% para la proporción poblacional.

32. En una muestra aleatoria de 800 elementos se obtiene una proporción muestral,  $\bar{p} = 0.70$ .
  - a. Dé un intervalo de 90% de confianza para la proporción poblacional.
  - b. Encuentre un intervalo de confianza de 95% para la proporción poblacional.
33. En un estudio el valor planeado para la proporción poblacional es  $p^* = 0.35$ . ¿De qué tamaño se debe tomar la muestra para dar un intervalo de confianza de 95% con un margen de error de 0.05?
34. Para 95% de confianza, ¿de qué tamaño deberá tomar la muestra para obtener un margen de error de 0.03 en la estimación de una proporción poblacional? Suponga que no se cuenta con datos anteriores para obtener un valor planeado de  $p^*$ .

## Aplicaciones



35. Se hizo un estudio con 611 oficinistas para investigar su atención al teléfono, el estudio registraba la frecuencia con que contestaban el teléfono y la frecuencia con que dejaban que la llamada pase al buzón de voz (*USA Today*, 21 de abril de 2002). De estos oficinistas, 281 indicaron constatar siempre las llamadas y no utilizar el buzón de voz.
  - a. Dé la estimación puntual de la proporción poblacional de oficinistas que siempre responden el teléfono.
  - b. A 90% de confianza, ¿cuál es el margen de error?
  - c. Dé el intervalo de 90% de confianza para la proporción de la población de oficinistas que siempre contestan el teléfono.
36. De acuerdo con estadísticas publicadas por la CNBC, la cantidad de vehículos que no están asegurados es sorprendente (CNBC, 23 de febrero de 2006). Los resultados muestrales de la CNBC indican que 46 de 200 vehículos no estaban asegurados.
  - a. ¿Cuál es la estimación puntual de la proporción de vehículos no asegurados?
  - b. Dé un intervalo de confianza de 95% para la proporción poblacional.
37. Towers Perrin, una empresa de recursos humanos, realizó un estudio con 1100 empleados de empresas medianas y grandes para determinar qué tan insatisfechos estaban con sus trabajos (*The Wall Street Journal*, 29 de enero de 2003). En el archivo JobSatisfaction se muestran datos representativos. Un Sí como respuesta indica que al empleado le desagrada mucho su empleo actual.
  - a. Dé la estimación puntual de la proporción poblacional de empleados a quienes les disgusta mucho su empleo actual.
  - b. A 95% de confianza, ¿cuál es el margen de error?
  - c. ¿Cuál es el intervalo de confianza de 95% para la proporción de la población de empleados a quienes les desagrada mucho su empleo actual?
  - d. Towers Perrin estima que a los empleadores les cuesta un tercio de un sueldo anual por hora hallar un sucesor y hasta 1.5 veces el sueldo anual encontrar un sucesor para un empleado que recibe una compensación elevada. ¿Cuál es el mensaje de esta investigación para los empleadores?
38. Según Thomson Financial, hasta el 25 de enero de 2006, la mayor parte de las empresas que informaban tener ganancias habían superado las estimaciones (*BusinessWeek*, 6 de febrero de 2006). En una muestra de 162 empresas, 104 superaron las estimaciones, 29 coincidieron y 29 se quedaron cortas.
  - a. ¿Cuál es la estimación puntual de la proporción de empresas que se quedaron cortas?
  - b. Determine el margen de error y dé un intervalo de confianza de 95% para la proporción que superó las estimaciones.
  - c. ¿De qué tamaño debe de ser la muestra si el margen de error es 0.05?
39. En 2003 el porcentaje de personas que no tenía un seguro médico (en Estados Unidos) era 15.6% (*Statistical Abstract of the United States*, 2006). Se le pide a un comité del Congreso realizar un estudio para obtener información actualizada.
  - a. ¿Qué tamaño de muestra le recomienda usted al comité, si el objetivo es que en la estimación de la proporción actual de individuos que no tienen seguro médico el margen de error sea 0.03? Use 95% de confianza.
  - b. Repita el inciso a usando 99% de confianza.



JobSatisfaction



40. En el béisbol profesional el récord de cuadrangulares era de 61 conectados en una temporada y perteneció durante 37 años a Roger Maris de los Yankees de Nueva York. Sin embargo, de 1998 a 2001, tres jugadores —Mark McGwire, Sammy Sosa y Barry Bonds— rompieron este récord, siendo Bonds quien tiene el récord actual de 73 cuadrangulares en una sola temporada. Debido al rompimiento de este récord después de tanto tiempo, y también al rompimiento de otros récords, se sospecha que los jugadores de béisbol emplean medicamentos ilegales conocidos como esteroides para la formación de músculo. En una encuesta realizada por *USA Today/CNN/Gallup* se encontró que 86% de los aficionados al béisbol opinaba que a los jugadores profesionales se les debería someter a un examen para detectar esteroides (*USA Today*, 18 de julio de 2002). Si en un estudio se seleccionan 650 aficionados al béisbol, calcule el margen de error y el intervalo de confianza de 95% para la proporción poblacional de aficionados que opinan que los jugadores de béisbol profesional deberían ser sometidos a un examen para detectar esteroides.
41. La juventud de Estados Unidos usa Internet intensamente; el 87% de los jóvenes entre 12 y 17 años son usuarios de Internet (*The Cincinnati Enquirer*, 7 de febrero de 2006). En una muestra de usuarios de Internet de esta edad, 9% votó por MySpace como el sitio de Internet más popular. Suponga que en este estudio participaron 1 400 jóvenes. ¿Cuáles son los márgenes de error y la estimación por intervalo de la proporción poblacional de quienes consideran que este sitio es el más popular? Use 95% de confianza.
42. Una encuesta realizada por *USA Today/CNN/Gallup* durante la campaña presidencial tomó en junio una muestra de 491 votantes potenciales (*USA Today*, 9 de junio de 2000). El objetivo de esta encuesta era estimar la proporción de votantes potenciales a favor de cada candidato. Suponga que el valor planeado es  $p^* = 0.50$  con un nivel de confianza de 95%.
- Si  $p^* = 0.50$ . ¿Cuál fue el margen de error planeado en la encuesta de junio?
  - Al acercarse la elección de noviembre se busca una mejor precisión y un menor margen de error. Suponga que los márgenes de error que se piden son los que se muestran en la tabla siguiente. Calcule el tamaño de muestra que se recomienda para cada estudio.

Estudio	Margen de error
Septiembre	0.04
Octubre	0.03
Comienzo de noviembre	0.02
Un día antes de la elección	0.01

43. Phoenix Wealth Management/Harris realizó un estudio con 1500 individuos cuyo patrimonio era de un millón o más de dólares, obtuvo diversos estadísticos sobre la gente rica (*BusinessWeek*, 22 de septiembre de 2003). Los tres años anteriores habían sido malos para el mercado de acciones, lo que motivó algunas de las preguntas realizadas.
- En este estudio se encontró que 53% de los encuestados perdió 25% o más del valor de su portafolio en los últimos tres años. Dé un intervalo de confianza de 95% para la proporción de gente rica que perdió 25% o más del valor de su portafolio en los últimos tres años.
  - El estudio indicó que 31% de los encuestados siente que deberá ahorrar más para su retiro para compensar lo perdido. Dé un intervalo de confianza de 95% para la proporción poblacional.
  - Cinco por ciento de los encuestados hicieron una donación de \$25 000 o más para obras de caridad el año anterior. Dé un intervalo de confianza de 95% para la proporción de quienes hicieron una donación de \$25 000 o más para obras de caridad.
  - Compare los márgenes de error de las estimaciones por intervalo de los incisos a, b y c. ¿Cuál es la relación entre margen de error y  $\bar{p}$ ? Si usa la misma muestra para obtener varias proporciones, ¿cuál de las proporciones debe usarse para estimar el valor planeado  $p^*$ ? ¿Por qué considera que en estos casos suela usarse  $p^* = 0.50$ ?

## Resumen

En este capítulo se presentaron los métodos para obtener estimaciones por intervalo de la media poblacional y de la proporción poblacional. Un estimador puntual puede o no proporcionar una buena estimación de un parámetro poblacional. Un intervalo de estimación suministra una medida de

la precisión de una estimación. Tanto la estimación por intervalo de una media poblacional como la de una proporción poblacional tienen la forma: estimación puntual  $\pm$  margen de error.

Para la media poblacional se presentaron estimaciones por intervalo en dos casos. En el caso  $\sigma$  conocida, se usan datos históricos o alguna otra información para obtener una estimación de  $\sigma$  antes de tomar la muestra. Entonces, el análisis de nuevos datos muestrales se realiza bajo la suposición de que se conoce  $\sigma$ . En el caso  $\sigma$  desconocida, los datos muestrales se usan para estimar tanto la media poblacional como la desviación estándar poblacional. La decisión final de qué procedimiento de estimación por intervalo usar depende de que el analista decida qué método proporciona una mejor estimación de  $\sigma$ .

En el caso  $\sigma$  conocida, el procedimiento de estimación por intervalo se basa en el valor supuesto para  $\sigma$  y en el uso de la distribución normal estándar. En el caso  $\sigma$  desconocida, para el procedimiento de estimación por intervalo se usa la desviación estándar muestral  $s$  y la distribución  $t$ . En ambos casos, la calidad de la estimación por intervalo depende de la distribución de la población y del tamaño de la muestra. Si la población tiene una distribución normal, la estimación por intervalo será exacta en ambos casos, aun cuando los tamaños de las muestras sean pequeños. Si la población no tiene una distribución normal, la estimación por intervalo resultante será aproximada. Tamaños de muestras mayores proporcionarán mejores aproximaciones, pero entre más sesgada sea la población, mayor será el tamaño de la muestra necesario para obtener una buena aproximación. En las secciones 8.1 y 8.2 se dieron consejos prácticos respecto al tamaño de muestra necesario para obtener buenas aproximaciones. En la mayoría de los casos, una muestra de tamaño 30 o mayor proporcionará una buena aproximación para el intervalo de confianza.

La forma general de una estimación por intervalo para la proporción poblacional es  $\bar{p} \pm$  margen de error. En la práctica, los tamaños de muestra empleados en estimaciones por intervalo de una proporción poblacional suelen ser grandes. Entonces, el procedimiento de estimación por intervalo se basa en la distribución normal estándar.

Algunas veces se suele especificar un determinado margen de error antes de llevar a cabo el plan de muestreo. También se explicó cómo elegir el tamaño de muestra adecuado para obtener la precisión deseada.

## Glosario

**Estimación por intervalo** Estimación de un parámetro poblacional que suministra un intervalo que se cree contiene el valor del parámetro. Para las estimaciones por intervalo vistas en este capítulo tiene la forma: estimación puntual  $\pm$  margen de error.

**Margen de error** Valor que se resta y se suma a la estimación puntual con objeto de obtener un intervalo de estimación para el parámetro poblacional.

**$\sigma$  conocida** Caso en el que datos históricos u alguna otra información proporciona un buen valor para ser considerado como desviación estándar poblacional antes de tomar la muestra. Este valor conocido de  $\sigma$  se usa en la estimación por intervalo para calcular el margen de error.

**Nivel de confianza** Confianza correspondiente a la estimación por intervalo. Por ejemplo, si un procedimiento para obtener una estimación por intervalo proporciona intervalos tales que, 95% de ellos contendrán al parámetro poblacional, se dice que esa estimación por intervalo tiene un nivel de confianza de 95%.

**Coeficiente de confianza** El nivel de confianza expresado como valor decimal. Por ejemplo 0.95 es el coeficiente de confianza correspondiente al nivel de confianza de 95%.

**Intervalo de confianza** Otro nombre para una estimación por intervalo.

**$\sigma$  desconocida** El caso más común cuando no existen bases sólidas para estimar la desviación estándar poblacional antes de tomar la muestra. En la estimación por intervalo se usa la desviación estándar muestral para calcular el margen de error.

**Distribución  $t$**  Una familia de distribuciones de probabilidad que se usa para obtener una estimación por intervalo de la media poblacional cuando la desviación estándar poblacional  $\sigma$  no se conoce y se estima mediante la desviación estándar muestral  $s$ .

**Grados de libertad** Parámetro de las distribuciones  $t$ . Cuando se usa una distribución  $t$  para calcular una estimación por intervalo de la media poblacional, la distribución  $t$  correspondiente tiene  $n - 1$  grados de libertad, donde  $n$  es el tamaño de la muestra aleatoria simple.

## Fórmulas clave

### Estimación por intervalo de una media poblacional: $\sigma$ conocida

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (8.1)$$

### Estimación por intervalo de una media poblacional: $\sigma$ desconocida

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \quad (8.2)$$

### Tamaño de la muestra para una estimación de la media poblacional

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} \quad (8.3)$$

### Estimación por intervalo de una proporción poblacional

$$\bar{p} \pm z_{\alpha/2} \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} \quad (8.6)$$

### Tamaño de la muestra para una estimación de la proporción poblacional

$$n = \frac{(z_{\alpha/2})^2 p^*(1 - p^*)}{E^2} \quad (8.7)$$

## Ejercicios complementarios

44. En un estudio realizado con 54 corredores de bolsa con descuento, se encontró que la media de los precios cobrados por una transacción de 100 acciones a \$50 la acción, fue \$33.77 (*AAII Journal*, febrero de 2006). Este estudio se realiza anualmente. De acuerdo con los datos históricos disponibles, considere que la desviación estándar poblacional conocida es \$15.
- Según los datos muestrales, ¿cuál es el margen de error en un intervalo de confianza de 95%?
  - Dé un intervalo de confianza de 95% para la media de los precios cobrados por una transacción de 100 acciones a \$50 la acción.
45. En un estudio realizado por la American Automobile Association se encontró que una familia de cuatro miembros en vacaciones gasta en promedio \$215.60 por día. Suponga que en una muestra de 64 familias de vacaciones en las Cataratas del Niágara la media muestral encontrada haya sido \$252.45 por día y la desviación estándar muestral \$74.50.
- Dé una estimación, mediante un intervalo de confianza de 95% para la media de la cantidad que gasta por día una familia de cuatro, que está de vacaciones en las Cataratas del Niágara.
  - De acuerdo con el intervalo de confianza del inciso a, ¿parece que la media poblacional de la cantidad gastada por día por las familias que visitan las Cataratas del Niágara es diferente de la media reportada por la American Automobile Association? Explique.
46. La película *Harry Potter y la piedra filosofal* echó por tierra el récord de taquilla en estreno de *El mundo perdido: Parque Jurásico* (*The WallStreet Journal*, 19 de noviembre de 2001). En una muestra de 100 cines se encontró que la media de la recaudación bruta en los tres días del primer fin de semana fue de \$25 467 por cine. La desviación estándar fue \$4980.
- ¿Cuál es el margen de error en este estudio? Use 95% de confianza.
  - ¿Cuál es la estimación del intervalo de confianza de 95% de la media poblacional de la recaudación por cine?
  - El mundo perdido* obtuvo \$72.1 millones en los tres días del primer fin de semana. *Harry Potter y la piedra filosofal* se presentó en 3672 cines. Dé una estimación del total obtenido por *Harry Potter y la piedra filosofal* en los tres días del primer fin de semana.
  - En un artículo de Associated Press se dijo que *Harry Potter* “echó por tierra” el récord de taquilla en el debut de *El mundo perdido*. ¿Coincidieron sus resultados con esto?

47. Muchos observadores de los mercados de acciones aseguran que cuando la proporción P/E en las acciones es superior a 20, el mercado está sobrevaluado. La proporción P/E es el precio de una acción dividido entre las ganancias de los últimos 12 meses. Suponga que usted desea saber si actualmente el mercado está sobrevaluado y qué proporción de las empresas pagan dividendos. A continuación aparece una lista de 30 empresas que cotizan en la Bolsa de Nueva York (NYSE) (*Barron's*, 19 de enero de 2004).

Empresa	Dividendos	Proporción P/E	Empresa	Dividendos	Proporción P/E
Albertsons	Sí	14	NY Times A	Sí	25
BRE Prop	Sí	18	Omnicare	Sí	25
CityNtl	Sí	16	PallCp	Sí	23
DelMonte	No	21	PubSvcEnt	Sí	11
EnrgzHldg	No	20	SensientTch	Sí	11
Ford Motor	Sí	22	SmtProp	Sí	12
Gildan A	No	12	TJX Cos	Sí	21
HudsnUtdBcp	Sí	13	Thomson	Sí	30
IBM	Sí	22	USB Hldg	Sí	12
JeffPilot	Sí	16	US Restr	Sí	26
KingswayFin	No	6	Varian Med	No	41
Libbey	Sí	13	Visx	No	72
MasoniteIntl	No	15	Waste Mgt	No	23
Motorola	Sí	68	Wiley A	Sí	21
Ntl City	Sí	10	Yum Brands	No	18

- a. Dé una estimación puntual para la proporción poblacional P/E en las acciones que cotizan en la Bolsa de Nueva York. Dé un intervalo de confianza de 95%.
- b. De acuerdo con su respuesta al inciso a, ¿considera usted que el mercado está sobrevaluado?
- c. Dé una estimación puntual de la proporción de empresas en la NYSE que pagan dividendos. ¿El tamaño de la muestra es suficientemente grande para justificar el empleo de la distribución normal en el cálculo de un intervalo de confianza para esta proporción? ¿Por qué sí o por qué no?
48. US Airways llevó a cabo diversos estudios que indican ahorros importantes si los viajeros frecuentes del programa Dividend Miles realizaran en línea el canje de millas y programaran los vuelos ganados (*US Airways Attaché*, febrero de 2003). En un estudio se recogieron datos sobre el tiempo que se requiere para realizar por teléfono el canje de millas y la programación de un vuelo ganado. En el conjunto de datos Flights se encuentra una muestra de tiempos en minutos requeridos para programar por teléfono cada uno de 150 vuelos ganados. Use Minitab o Excel para contestar las preguntas siguientes.
- a. ¿Cuál es la media muestral del número de minutos que se requiere para programar por teléfono los vuelos ganados?
- b. Dé el intervalo de confianza de 95% para la media poblacional del tiempo requerido para programar por teléfono los vuelos.
- c. Suponga que un agente de boletos por teléfono trabaja 7.5 horas. ¿Cuántos vuelos ganados se espera que atienda dicho agente en un día?
- d. Diga cómo apoya esta información al plan de US Airways de usar un sistema en línea para reducir costos.
49. En un estudio se les pidió a 200 ejecutivos de una muestra proporcionar datos sobre la cantidad de minutos por día que pierden los oficinistas tratando de localizar cosas mal guardadas, mal archivadas o mal clasificadas. Los datos de esta investigación se encuentran en el conjunto de datos ActTemps.
- a. Use ActTemps para dar una estimación puntual de los minutos por día perdidos por los oficinistas en localizar cosas mal guardadas, mal archivadas o mal clasificadas.
- b. ¿Cuál es la desviación estándar muestral?
- c. Dé un intervalo de confianza de 95% para la cantidad de minutos perdidos por día.
50. Se hacen pruebas de rendimiento de gasolina con un determinado modelo de automóvil. Si se quiere dar un intervalo de confianza de 98% con un margen de error de 1 milla por galón, ¿cuán-

archivo  
en  
**CD**

NYSEStocks

archivo  
en  
**CD**

Flights

archivo  
en  
**CD**

ActTemps

- tos automóviles deberán usarse? Suponga que por pruebas anteriores se sabe que la desviación estándar del rendimiento es 2.6 millas por galón.
51. Un centro médico quiere estimar la media del tiempo que se necesita para programar una cita de un paciente. ¿De qué tamaño deberá ser la muestra si se quiere que el margen de error sea de dos minutos y que el nivel de confianza sea 95%? ¿De qué tamaño deberá tomarse la muestra si se quiere que el nivel de confianza sea 99%? Para la desviación estándar poblacional use como valor planeado, 8 minutos.
  52. *Business Week* presenta datos sobre el salario anual más bonos de presidentes de consejos de administración. En una muestra preliminar la desviación estándar es \$675; los datos se dan en miles de dólares. ¿De cuántos presidentes de consejos de administración deberá constar la muestra si se quiere estimar el salario anual más bonos con un margen de error de \$100 000? (*Nota:* El margen de error deseado será  $E = 100$  si los datos están dados en miles de dólares.) Use 95% de confianza.
  53. El National Center for Education Statistics (informa que 47% de los estudiantes universitarios trabaja para pagar sus estudios y su sustento. Suponga que se empleó una muestra de 450 estudiantes universitarios en ese estudio.
    - a. Dé un intervalo de confianza de 95% para la proporción poblacional de estudiantes que trabajan para mantenerse y pagar sus estudios.
    - b. Dé un intervalo de confianza de 99% para la proporción poblacional de estudiantes que trabajan para mantenerse y pagar sus estudios.
    - c. ¿Qué ocurre con el margen de error cuando el nivel de confianza aumenta de 95% a 99%?
  54. En un estudio de *USA Today/CNN/Gallup*, realizado con 369 padres que trabajan, se encontró que 200 consideran que pasan muy poco tiempo con sus hijos debido al trabajo.
    - a. Dé una estimación puntual de la proporción poblacional de padres que trabajan y piensan que pasan muy poco tiempo con sus hijos debido al trabajo.
    - b. ¿Cuál es el margen de error para 95% de confianza?
    - c. ¿Cuál es el intervalo de confianza de 95% para la proporción poblacional de padres que trabajan y piensan que pasan muy poco tiempo con sus hijos debido al trabajo?
  55. ¿De qué le sería más difícil deshacerse: de su televisión o de su computadora? En un estudio reciente efectuado con 1677 usuarios de Internet en Estados Unidos, se encontró que a 74% de la élite juvenil (edad promedio de 22 años) le sería más difícil deshacerse de su computadora (*PC Magazine*, 3 de febrero de 2004). Sólo para el 48% sería más difícil deshacerse de su televisión.
    - a. Dé un intervalo de confianza de 95% para la proporción de jóvenes a quienes les sería más difícil deshacerse de su computadora.
    - b. Encuentre un intervalo de confianza de 99% para la proporción de jóvenes a quienes les sería más difícil deshacerse de su televisión.
    - c. ¿En cuál de los incisos a o b es mayor el margen de error? Explique por qué.
  56. El aeropuerto internacional Cincinnati/Northern Kentucky tuvo en 2005 el segundo lugar en puntualidad en la llegada de vuelos entre los aeropuertos con más actividad del país (*The Cincinnati Enquirer*, 3 de febrero de 2003). Suponga que esto se basa en una muestra de 550 vuelos de los cuales 455 llegaron a tiempo.
    - a. Elabore una estimación puntual de la tasa de llegadas a tiempo (proporción de vuelos que llegan a tiempo) al aeropuerto.
    - b. Construya un intervalo de confianza de 95% para la proporción de llegadas a tiempo en la población de todos los vuelos del aeropuerto en 2005.
  57. El *2003 Statistical Abstract of the United States* da el porcentaje de personas de 18 años o más que fuma. Asuma que en un nuevo estudio para recoger datos sobre los fumadores y no fumadores se usa como estimación preliminar de la proporción que fuma, 0.30.
    - a. ¿De qué tamaño deberá tomarse la muestra para estimar la proporción de fumadores con un margen de error de 0.02? Use 95% de confianza.
    - b. Suponga que el estudio usa su recomendación para el tamaño de la muestra del inciso a y encuentra 520 fumadores. ¿Cuál es la estimación puntual de la proporción de fumadores en la población?
    - c. ¿Cuál es el intervalo de confianza de 99% para la proporción de fumadores en la población?

58. Una firma de tarjetas de crédito de un banco conocido desea estimar la proporción de tarjetahabientes que al final del mes tienen un saldo distinto de cero que ocasiona cargos. Suponga que el margen de error deseado es 0.03 con 98% de confianza.
- ¿De qué tamaño deberá tomarse la muestra si se cree que 70% de los tarjetahabientes de la firma tienen un saldo distinto de cero al final del mes?
  - ¿De qué tamaño deberá tomarse la muestra si no se puede dar ningún valor planeado para la proporción?
59. En un estudio se le pidió a 200 personas que indicaran su mayor fuente de información de noticias; 110 indicaron que su principal fuente de noticias eran los noticieros de televisión.
- Dé un intervalo de confianza de 95% para la proporción poblacional de personas que tienen como principal fuente de noticias a la televisión.
  - ¿Cuál será el tamaño de muestra necesario para estimar la proporción poblacional con un margen de error de 0.05 y 95% de confianza?
60. Aunque para los viajeros de negocios, los horarios y los costos son factores importantes al elegir una línea aérea, en un estudio realizado por *USA Today* se encontró que para los viajeros de negocios el factor más importante es que la línea tenga un programa de viajero frecuente. En una muestra de  $n = 1993$  pasajeros que participaron en el estudio, 618 indicaron como factor más importante un programa de pasajero frecuente.
- ¿Cuál es la estimación puntual de la proporción poblacional de viajeros de negocios que consideran al programa de viajero frecuente como el factor más importante al elegir una línea aérea?
  - Dé un intervalo de confianza de 95% para estimar la proporción poblacional.
  - ¿De qué tamaño deberá ser la muestra para que el margen de error sea 0.01 con 95% de confianza? ¿Aconsejaría que *USA Today* tratara de tener esta precisión? ¿Por qué sí o por qué no?

## Caso problema 1 La revista *Young Professional*

La revista *Young Professional* fue creada para un público formado por personas que se encuentran en los 10 primeros años de su carrera profesional en negocios. En sus dos primeros años de publicación, la revista ha tenido bastante éxito. Ahora el editor está tratando de aumentar la base publicitaria de su revista. Los anunciantes potenciales preguntan continuamente sobre los datos demográficos e intereses de los suscriptores de *Young Professional*. Para recabar esta información, la revista realizó un estudio para elaborar el perfil de sus suscriptores. Los resultados de dicho estudio se usarán para ayudar a elegir artículos de interés y para proporcionar a los anunciantes un perfil de los suscriptores. Como nuevo empleado de la empresa se le pide a usted su ayuda para analizar los resultados del estudio.

A continuación se presentan algunas de las preguntas del estudio.



1. Edad \_\_\_\_
2. Sexo \_\_\_\_
3. ¿Piensa comprar algún bien inmueble en los próximos dos años? Sí \_\_\_\_ No \_\_\_\_
4. ¿Cuál es el valor aproximado de las inversiones financieras, excluyendo su casa, que son propiedad suya o de otro miembro de su hogar?
5. ¿Cuántas transacciones de acciones/bonos/fondos mutualistas realizó el año pasado?
6. ¿Tiene en casa acceso de banda ancha a Internet? Sí \_\_\_\_ No \_\_\_\_
7. Por favor indique cuál fue el ingreso de su hogar el año pasado.
8. ¿Tiene hijos? Sí \_\_\_\_ No \_\_\_\_

El archivo Professional contiene las respuestas a estas preguntas. En la tabla 8.6 se muestra la parte de este archivo correspondiente a las respuestas de los primeros cinco entrevistados. El archivo completo se encuentra en el disco compacto que se distribuye con el libro.

**TABLA 8.6** RESULTADOS PARCIALES DEL ESTUDIO DE LA REVISTA YOUNG PROFESSIONAL

Edad	Género	Compra de inmuebles	Valor de inversiones (\$)	Número de transacciones	Conexión de banda ancha	Ingreso del hogar (\$)	Hijos
38	Femenino	No	12 200	4	Sí	75 200	Sí
30	Masculino	No	12 400	4	Sí	70 300	Sí
41	Femenino	No	26 800	5	Sí	48 200	No
28	Femenino	Sí	19 600	6	No	95 300	No
31	Femenino	Sí	15 100	5	No	73 300	Sí
:	:	:	:	:	:	:	:

### Informe administrativo

Elabore un informe administrativo dando los resultados del estudio. Además de los resúmenes estadísticos, analice cómo la revista puede usar estos resultados para atraer más anunciantes. También presente una recomendación a los editores para que empleen los resultados del estudio en la elección de los temas de interés para sus suscriptores. Su informe debe contener los siguientes puntos, pero no limite su análisis a estas áreas.

1. Dé los estadísticos descriptivos adecuados para resumir los datos.
2. Muestre los intervalos de 95% de confianza para la edad promedio y para el ingreso promedio por hogar de los suscriptores.
3. Encuentre intervalos de confianza de 95% para la proporción de suscriptores que tienen acceso de banda ancha y para la proporción de éstos que tienen niños.
4. ¿Será *Young Professional* un buen sitio para la publicidad de agentes de bolsa? Justifique su conclusión con datos estadísticos.
5. ¿Será esta revista un buen lugar para la publicidad de empresas que venden software educativo y juegos de computadora para niños?
6. Haga un comentario sobre el tipo de artículos que crea usted son de interés para los lectores de la revista.

## Caso problema 2 Gulf Real Estate Properties

Gulf Real Estate Properties, Inc., es una empresa inmobiliaria ubicada en el suroeste de Florida. Esta empresa, que se anuncia como “experta en el mercado de bienes raíces”, monitorea las ventas de condominios recabando datos sobre ubicación, precio de lista, precio de venta y días necesarios para vender cada unidad. Los condominios están calificados como *con vista al golfo* o *sin vista al golfo* dependiendo de su ubicación hacia el Golfo de México. Multiple Listing Service Naples, Florida, proporciona datos muestrales sobre 40 condominios con vista al golfo y 18 condominios sin vista al golfo.\* Los precios están dados en miles de dólares. Estos datos se presentan en la tabla 8.7.

### Informe administrativo

1. Use los estadísticos descriptivos adecuados para resumir cada una de las tres variables de los 40 condominios con vista al golfo.
2. Aplique los estadísticos descriptivos adecuados para resumir cada una de las tres variables de los 18 condominios sin vista al golfo.
3. Compare los resultados. Analice cualquier estadístico específico que ayude al agente de ventas inmobiliarias a saber más sobre el mercado de los condominios.

\*Datos sustentados en las ventas de condominios reportadas en el Naples MLS (Coldwell Banker, junio de 2000).

**TABLA 8.7** DATOS DE VENTA DE PROPIEDADES VENDIDAS POR GULF REAL ESTATE PROPERTIES

Condominios con vista al golfo			Condominios sin vista al golfo		
Precio de lista	Precio de venta	Días hasta la venta	Precio de lista	Precio de venta	Días hasta la venta
495.0	475.0	130	217.0	217.0	182
379.0	350.0	71	148.0	135.5	338
529.0	519.0	85	186.5	179.0	122
552.5	534.5	95	239.0	230.0	150
334.9	334.9	119	279.0	267.5	169
550.0	505.0	92	215.0	214.0	58
169.9	165.0	197	279.0	259.0	110
210.0	210.0	56	179.9	176.5	130
975.0	945.0	73	149.9	144.9	149
314.0	314.0	126	235.0	230.0	114
315.0	305.0	88	199.8	192.0	120
885.0	800.0	282	210.0	195.0	61
975.0	975.0	100	226.0	212.0	146
469.0	445.0	56	149.9	146.5	137
329.0	305.0	49	160.0	160.0	281
365.0	330.0	48	322.0	292.5	63
332.0	312.0	88	187.5	179.0	48
520.0	495.0	161	247.0	227.0	52
425.0	405.0	149			
675.0	669.0	142			
409.0	400.0	28			
649.0	649.0	29			
319.0	305.0	140			
425.0	410.0	85			
359.0	340.0	107			
469.0	449.0	72			
895.0	875.0	129			
439.0	430.0	160			
435.0	400.0	206			
235.0	227.0	91			
638.0	618.0	100			
629.0	600.0	97			
329.0	309.0	114			
595.0	555.0	45			
339.0	315.0	150			
215.0	200.0	48			
395.0	375.0	135			
449.0	425.0	53			
499.0	465.0	86			
439.0	428.5	158			

4. Dé un intervalo de confianza de 95% para estimar las medias poblacionales del precio de venta y del número de días necesarios para vender los condominios con vista al golfo. Interprete los resultados.
5. Encuentre un intervalo de confianza de 95% para estimar las medias poblacionales del precio de venta y el número de días necesarios para vender los condominios sin vista al golfo. Interprete los resultados.
6. Suponga que se necesita estimar el precio medio de venta de los condominios con vista al golfo con un margen de error de \$40 000 y la media del precio de venta de los condo-

- minios sin vista al golfo con un margen de error de \$15 000. Si se usa 99% de confianza, ¿de qué tamaño deberán ser las muestras?
7. La inmobiliaria Golfo Real firmó contratos para dos nuevos catálogos: un condominio con vista al Golfo con un precio de lista de \$585 000 y un condominio sin vista al Golfo con un precio de \$285 000. ¿Cuál es su estimado del precio final de venta y el número de días requerido para vender cada una de estas unidades?

## Caso problema 3 Metropolitan Research, Inc.

Metropolitan Research, Inc., una organización para la investigación del consumo, realiza estudios que tienen por objeto evaluar una amplia variedad de productos y servicios para los consumidores. En uno de sus estudios, Metropolitan se enfocó en la satisfacción de los consumidores con el funcionamiento de los automóviles producidos por el principal fabricante de Detroit. En un cuestionario enviado a propietarios de automóviles de esta empresa se encontraron varias quejas relacionadas con problemas tempranos en la transmisión. Para tener más información acerca de los problemas en la transmisión, Metropolitan empleó una muestra de reparaciones de la transmisión proporcionada por empresas en Detroit, dedicadas a la reparación de transmisiones. Los datos siguientes son el número de milla recorridos por 50 automóviles hasta el momento en que se presentaron los problemas con la transmisión.

archivo  
en  
**CD**  
Auto

85 092	32 609	59 465	77 437	32 534	64 090	32 464	59 902
39 323	89 641	94 219	116 803	92 857	63 436	65 605	85 861
64 342	61 978	67 998	59 817	101 769	95 774	121 352	69 568
74 276	66 998	40 001	72 069	25 066	77 098	69 922	35 662
74 425	67 202	118 444	53 500	79 294	64 544	86 813	116 269
37 831	89 341	73 341	85 288	138 114	53 402	85 586	82 256
77 539	88 798						

### Informe administrativo

1. Use los estadísticos descriptivos adecuados para resumir los datos sobre los problemas en la transmisión.
2. Dé un intervalo de confianza de 95% para estimar el número de millas promedio, en la población de automóviles con fallas en la transmisión, recorridas hasta que se presenta el problema. Haga una interpretación administrativa del intervalo estimado.
3. Analice las consecuencias de sus hallazgos en términos de la creencia de que algunos propietarios de automóviles tuvieron problemas tempranos con la transmisión.
4. ¿Cuántos registros deben tomarse en la muestra si se desea estimar la media poblacional del número de millas recorridas hasta la aparición de problemas en la transmisión con un margen de error de 5000 millas? Use 95% de confianza.
5. ¿Qué otra información desearía recolectar para evaluar mejor los problemas con la transmisión?

## Apéndice 8.1 Estimación por intervalo con Minitab

A continuación se describe cómo usar Minitab para obtener intervalos de confianza para la media poblacional y para la proporción poblacional.

### Media poblacional: $\sigma$ conocida

La estimación por intervalo se ilustra mediante el ejemplo de Lloyd's de la sección 8.1. En una muestra de 100 clientes las cantidades gastadas en cada visita a la tienda están en la columna C1 de la hoja de cálculo de Minitab. Se supone que la desviación estándar poblacional se conoce y que es  $\sigma = 20$ . Los pasos siguientes se usan para calcular un intervalo de confianza de 95% para estimar la media poblacional. Los datos están en la columna C1 de la hoja de cálculo de Minitab.

archivo  
en  
**CD**  
Lloyd's

**Paso 1.** Seleccionar el menú **Stat****Paso 2.** Elegir **Basic Statistics****Paso 3.** Elegir **1-Sample Z****Paso 4.** Cuando aparezca el cuadro de diálogo 1-Sample Z:Ingresar C1 en el cuadro **Samples in column**Ingresar 20 en el cuadro **Standard deviation****Paso 5.** Clic en **OK**

Si no se especifica otra cosa, Minitab emplea 95% como nivel de confianza. Para especificar otro nivel de confianza, por ejemplo 90%, al paso 4 hay que agregar lo siguiente.

Seleccionar **Options**

Cuando aparezca el cuadro de diálogo 1-Sample Z:

Ingresar 90 en el cuadro **Confidence level**Clic en **OK****Media poblacional:  $\sigma$  desconocida**

La estimación por intervalo se ilustra empleando los datos de la tabla 8.2 que dan los saldos en las tarjetas de crédito en una muestra de 70 hogares. Los datos están en la columna C1 de la hoja de cálculo de Minitab. En este caso se estima la desviación estándar poblacional  $\sigma$  mediante la desviación estándar muestral  $s$ . Con los pasos siguientes se obtiene un intervalo de confianza de 95% para estimar la media poblacional.

**Paso 1.** Seleccionar el menú **Stat****Paso 2.** Elegir **Basic Statistics****Paso 3.** Elegir **1-Sample t****Paso 4.** Cuando aparezca el cuadro de diálogo 1-Sample tIngresar C1 en el cuadro **Samples in columns****Paso 5.** Clic en **OK**

Si no se especifica otra cosa, Minitab emplea 95% como nivel de confianza. Para especificar otro nivel de confianza, por ejemplo 90% al paso 4 hay que agregar lo siguiente.

Seleccionar **Options**

Cuando aparezca el cuadro de diálogo 1-Sample t-Options:

Ingresar 90 en el cuadro **Confidence level**Clic en **OK****Proporción poblacional**

La estimación por intervalo se ilustra empleando los datos de los golfistas, presentados en la sección 8.4. Los datos están en la columna C1 de la hoja de cálculo de Minitab. Las respuestas se registraron como Sí si la golfista está satisfecha con la disponibilidad de horarios de salida y No, si no es el caso. Usando los pasos siguientes se calcula un intervalo de confianza de 95% para estimar la proporción de golfistas satisfechas con la disponibilidad de los horarios de salida.

**Paso 1.** Seleccionar el menú **Stat****Paso 2.** Elegir **Basic Statistics****Paso 3.** Elegir **1 Proportion****Paso 4.** Cuando aparezca el cuadro de diálogo 1 Proportion:Ingresar C1 en el cuadro **Samples in columns****Paso 5.** Seleccionar **Options****Paso 6.** Cuando aparezca el cuadro de diálogo 1 Proportion-Options:Seleccionar **Use test and interval based on normal distribution**Clic en **OK****Paso 7.** Clic en **OK**

Si no se especifica otra cosa, Minitab emplea 95% como nivel de confianza. Para especificar otro nivel de confianza, como 90%, cuando aparezca el cuadro de diálogo 1 Proportions-Options en el paso 6, ingresar 90 en el cuadro **Confidence Level**.

*Nota:* La rutina 1 Proporción de Minitab usa de las respuestas puestas en orden alfabético y selecciona la *segunda respuesta* como la proporción poblacional de interés. En el ejemplo de las golfistas, Minitab usa el orden alfabético No-Sí y de esta manera da el intervalo de confianza para la proporción de respuestas Sí. Como Sí era la respuesta de interés, los resultados de Minitab fueron los adecuados. Sin embargo, si el orden alfabético de Minitab no da la respuesta de interés, se selecciona cualquier celda de la columna y se usa la secuencia: Editor > Column > Value Order. Minitab le proporcionará la opción de usar un orden especificado por el usuario, pero usted debe poner en segundo lugar de la lista la respuesta de interés en el cuadro define-an-order.

## Apéndice 8.2 Estimación por intervalo usando Excel



A continuación se describe el uso de Excel para calcular intervalos de confianza para la media poblacional y para la proporción poblacional.

### Media poblacional: $\sigma$ conocida

La estimación por intervalo se ilustra empleando el ejemplo de Lloyd's de la sección 8.1. Se supone que se conoce la desviación estándar poblacional y que es  $\sigma = 20$ . Las cantidades gastadas por los integrantes de la muestra de tamaño 100 se encuentran en la columna A de la hoja de cálculo de Excel. Para calcular el margen de error para la estimación de la media poblacional se emplean los pasos que se indican a continuación. Se empieza usando la herramienta para estadística descriptiva de Excel, descrita en el capítulo 3.

**Paso 1.** Seleccionar el menú **Herramientas**

**Paso 2.** Elegir **Análisis de datos**

**Paso 3.** Elegir **Estadística descriptiva**

**Paso 4.** Cuando aparezca el cuadro de diálogo de Estadística descriptiva:

Ingresar A1:A101 en el cuadro **Rango de entrada**

Seleccionar **Agrupado por columnas**

Seleccionar **Rótulos en el primer renglón**

Seleccionar **Rango de salida**

Ingresar C1 en el cuadro **Rango de salida**

Seleccionar **Resumen de estadísticas**

Clic en **OK**

El resumen de estadísticas aparecerá en las columnas C y D. Continúe con el cálculo del margen de error usando la función INTERVALO.CONFIANZA como sigue:

**Paso 5.** Seleccione la celda C16 e ingrese el título Margen de error

**Paso 6.** Seleccione la celda D16 e ingrese la fórmula de Excel = INTERVALO.CONFIANZA(0.50,20,100)

Los tres parámetros de esta función son

$$\text{Alfa} = 1 - \text{coeficiente de confianza} = 1 - 0.95 = 0.05$$

$$\text{La desviación estándar poblacional} = 20$$

El tamaño de la muestra = 100 (*Nota:* Este parámetro aparece como Cuenta en la celda D15.)

La estimación puntual de la media poblacional se encuentra en la celda D3 y el margen de error se encuentra en la celda D16. La estimación puntual de la media poblacional (82) y el margen de error (3.92) permiten calcular fácilmente el intervalo de confianza para la media poblacional.



## Media poblacional: $\sigma$ conocida

La estimación por intervalo se ilustra con los datos de la tabla 8.2 en la que se muestran los saldos en las tarjetas de crédito de 70 hogares. Los datos se encuentran en la columna A de la hoja de cálculo de Excel. Para calcular una estimación puntual y el margen de error de una estimación por intervalo de la media poblacional se siguen los pasos que se indican a continuación. Se emplea la herramienta para estadística descriptiva, vista en el capítulo 3.

**Paso 1.** Seleccionar el menú **Herramientas**

**Paso 2.** Elegir **Análisis de datos**

**Paso 3.** Elegir **Estadística descriptiva**

**Paso 4.** Cuando aparezca el cuadro de diálogo de Estadística descriptiva:

Ingresar A1:A71 en el cuadro **Rango de entrada**

Seleccionar **Agrupado por columnas**

Seleccionar **Rótulos en el primer renglón**

Seleccionar **Rango de salida**

Ingresar C1 en el cuadro Rango de salida

Seleccionar **Resumen de estadísticas**

Seleccionar **Nivel de confianza para la media**

Ingresar 95 en el cuadro de Nivel de confianza para la media

Clic en **OK**

El resumen estadístico aparecerá en las columnas C y D. La estimación puntual de la media poblacional aparecerá en la celda D3. El margen de error aparecerá como “Nivel de confianza (95.0%)” en la celda D16. La estimación puntual (\$9312) y el margen de error (\$955) permiten estimar fácilmente el intervalo de confianza para la media poblacional. En la figura 8.10 se muestra el resultado de este procedimiento de Excel.

**FIGURA 8.10 ESTIMACIÓN POR INTERVALO DE LA MEDIA POBLACIONAL DE SALDOS EN TARJETAS DE CRÉDITO USANDO EXCEL**

	A	B	C	D	E	F
<b>1</b>	<b>NewBalance</b>		<b>NewBalance</b>			
<b>2</b>	9430					
<b>3</b>	7535		Mean	9312		
<b>4</b>	4078		Standard Error	478.9281		
<b>5</b>	5604		Median	9466		
<b>6</b>	5179		Mode	13627		
<b>7</b>	4416		Standard Deviation	4007		
<b>8</b>	10676		Sample Variance	16056048		
<b>9</b>	1627		Kurtosis	-0.296		
<b>10</b>	10112		Skewness	0.18792		
<b>11</b>	6567		Range	18648		
<b>12</b>	13627		Minimum	615		
<b>13</b>	18719		Maximum	19263		
<b>14</b>	14661		Sum	651840		
<b>15</b>	12195		Count	70		
<b>16</b>	10544		Confidence Level(95.0%)	955.4354		
<b>17</b>	13659					
<b>70</b>	9743					
<b>71</b>	10324					
<b>71</b>						

Nota: Los renglones 18 a 69 están ocultos.

archivo  
en  
**CD**  
Interval p

## Proporción poblacional

Esta estimación por intervalo se ilustra usando los datos del estudio de las golfistas presentado en la sección 8.4. Los datos se encuentran en la columna A de la hoja de cálculo de Excel. En los datos recabados, una respuesta Sí implica que la golfista está satisfecha con los horarios de salida disponibles y No, que no es el caso. Excel no proporciona una rutina ya elaborada para la estimación de una proporción poblacional; sin embargo, es fácil elaborar una plantilla para usarla con tal propósito. La plantilla que se muestra en la figura 8.11 proporciona un intervalo de confianza de 95% para la estimación de la proporción de golfistas satisfechas con los horarios de salida disponibles. Observe que en la figura 8.11, en las celdas de la hoja de cálculo que aparece

**FIGURA 8.11** PLANTILLA DE EXCEL PARA LA ESTIMACIÓN POR INTERVALO DE UNA PROPORCIÓN POBLACIONAL

	A	B	C	D	E
1	<b>Response</b>		<b>Interval Estimate of a Population Proportion</b>		
2	Yes				
3	No		<b>Sample Size</b>	=COUNTA(A2:A901)	
4	Yes		<b>Response of Interest</b>	Yes	
5	Yes		<b>Count for Response</b>	=COUNTIF(A2:A901,D4)	
6	No		<b>Sample Proportion</b>	=D5/D3	
7	No				
8	No		<b>Confidence Coefficient</b>	0.95	
9	Yes		<b>z Value</b>	=NORMSINV(0.5+D8/2)	
10	Yes				
11	Yes		<b>Standard Error</b>	=SQRT(D6*(1-D6)/D3)	
12	No		<b>Margin of Error</b>	=D9*D11	
13	No				
14	Yes		<b>Point Estimate</b>	=D6	
15	No		<b>Lower Limit</b>	=D14-D12	
16	No		<b>Upper Limit</b>	=D14+D12	
17	Yes				
18	No				
901	Yes		<b>Interval Estimate of a Population Proportion</b>		
902					

	A	B	C	D	E	F	G
1	<b>Response</b>		<b>Interval Estimate of a Population Proportion</b>				
2	Yes						
3	No		<b>Sample Size</b>	900			
4	Yes		<b>Response of Interest</b>	Yes	Ingrese la respuesta de interés		
5	Yes		<b>Count for Response</b>	396			
6	No		<b>Sample Proportion</b>	0.4400			
7	No						
8	No		<b>Confidence Coefficient</b>	0.95	Ingrese el coeficiente de confianza		
9	Yes		<b>z Value</b>	1.960			
10	Yes						
11	Yes		<b>Standard Error</b>	0.0165			
12	No		<b>Margin of Error</b>	0.0324			
13	No						
14	Yes		<b>Point Estimate</b>	0.4400			
15	No		<b>Lower Limit</b>	0.4076			
16	No		<b>Upper Limit</b>	0.4724			
17	Yes						
18	No						
901	Yes						
902							

Nota Los renglones 19 a 900 están ocultos.

en segundo plano, se presentan las fórmulas que proporcionan los resultados de la hoja de cálculo que aparece en primer plano. Los pasos para usar la plantilla con este conjunto de datos son los que se dan a continuación.

**Paso 1.** Ingresar el rango de datos A2:A901 en la fórmula =CONTARA de la celda D3

**Paso 2.** Ingresar Sí como respuesta de interés en la celda D4

**Paso 3.** Ingresar el rango de datos A2:A901 en la fórmula =CONTAR.SI de la celda D5

**Paso 4.** Ingresar 0.95 como coeficiente de confianza en la celda D8.

Esta plantilla proporciona automáticamente el intervalo de confianza en las celdas D15 y D16.

Esta plantilla se usa para calcular un intervalo de confianza para la proporción poblacional en otras aplicaciones. Por ejemplo, para calcular la estimación por intervalo de un nuevo conjunto de datos, se ingresan los nuevos datos muestrales en la columna A de la hoja de cálculo y después se modifican las cuatro celdas que se muestran. Si el nuevo conjunto de datos ya ha sido resumido, no es necesario ingresar los datos muestrales en la hoja de cálculo. En este caso se ingresa el tamaño de la muestra en la celda D3 y la proporción muestral en la celda D6; la plantilla proporcionará el intervalo de confianza para la proporción poblacional. La hoja de cálculo de la figura 8.11 se encuentra en el archivo p del disco compacto que se distribuye con el libro.