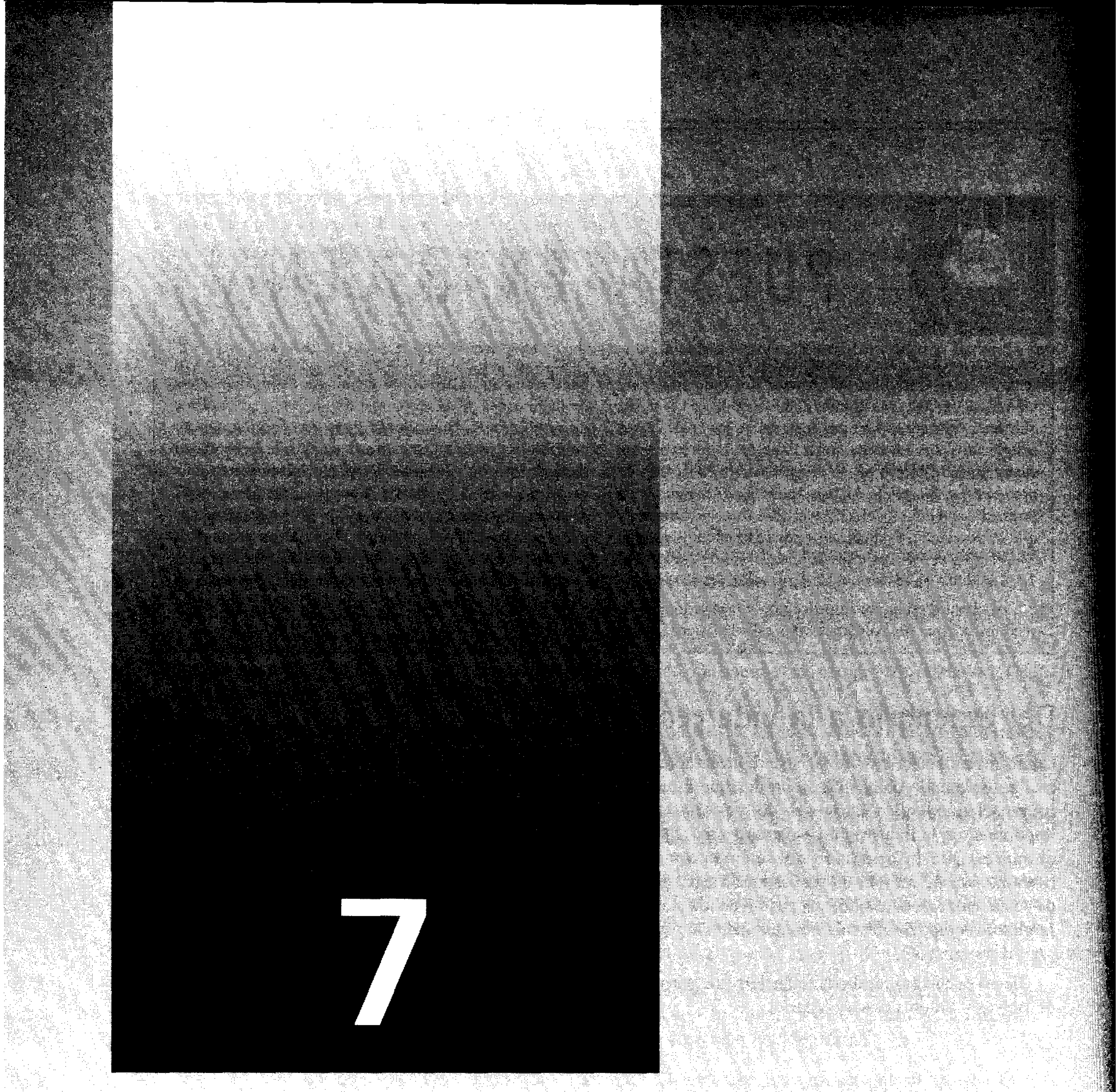
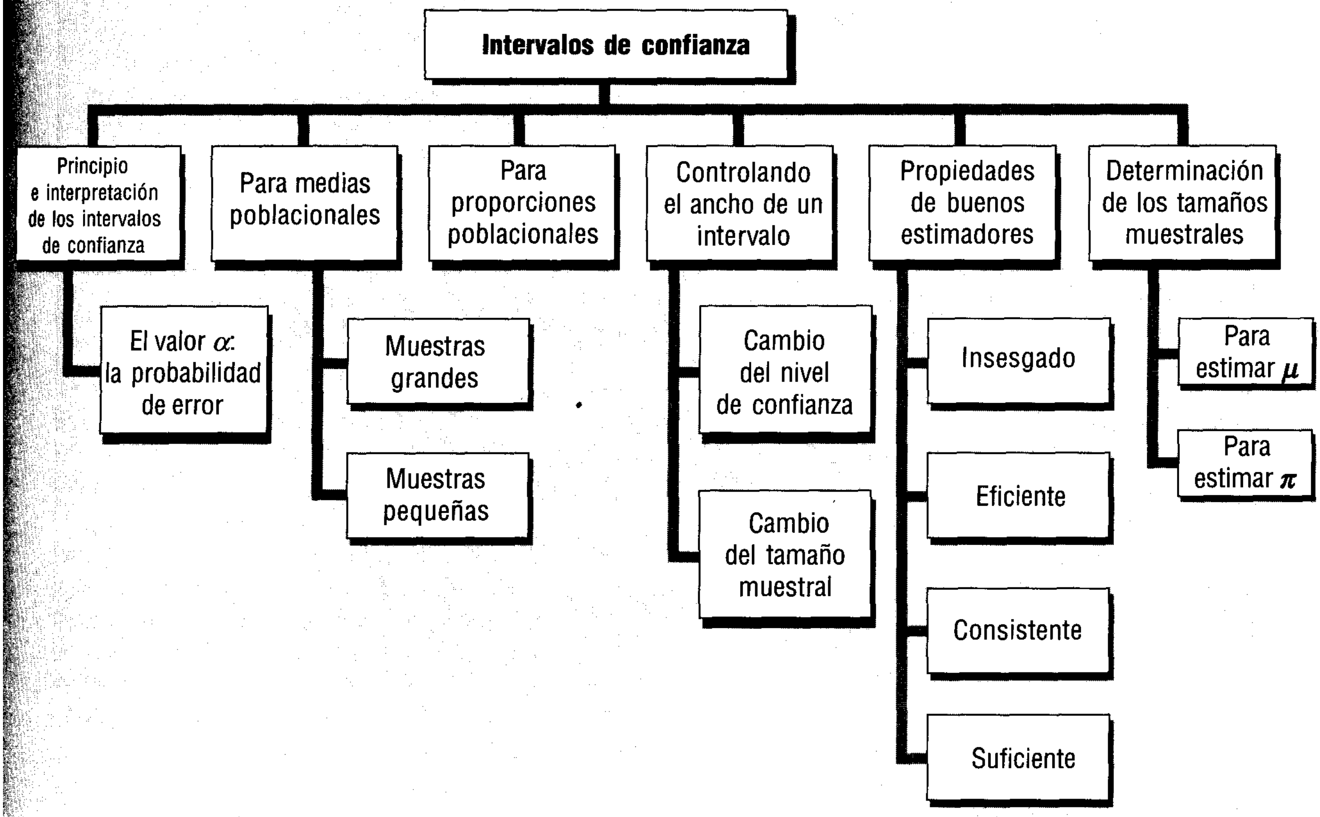


7

Estimación con intervalos de confianza

Plan del capítulo

El capítulo 6 mostró cómo las distribuciones muestrales de la media y de la proporción muestral pueden utilizarse para generar estimaciones puntuales de μ y π. Este capítulo muestra cómo pueden establecerse estimaciones por intervalo para estos dos parámetros y cómo pueden asignarse los niveles de confianza a estos intervalos.



ESCENARIO

En 1997 la Agencia Federal de Investigación (*Federal Bureau of Investigation*-FBI) implantó procedimientos revolucionarios para facilitar la captura de personas solicitadas por crímenes graves. Un vocero de la División de Estadísticas Criminales del FBI apareció en el programa en vivo *Larry King Live*, en la Red de Noticias por Cable (Cable News Network, CNN) para discutir los *procedimientos que harían más seguras las calles de la* ciudad. Ella mencionó varias estadísticas que el FBI había recolectado describiendo motivos, técnicas y frecuencia de los crímenes que ella consideraba que eran útiles para hacer el perfil de criminales que andan sueltos y que la agencia desea capturar.

Su discusión se centró en los esfuerzos de la agencia por mantener una gran base de datos sobre estadísticas criminales que pudieran utilizarse para predecir la actividad criminal y así poder anticipar dónde y cuándo puede ocurrir un acto ilegal. Ella mencionó varios casos que se habían resuelto, en gran parte gracias al trabajo realizado por estadísticos profesionales que proporcionaron estimaciones *sobre tasas de reincidencia de los infractores, así como* otras actividades que proporcionan alguna pista que ayude a su arresto. Esta información mostró ser de extrema utilidad para los agentes que trabajaban en el área y cuya función es ubicar a quienes están en la lista de los más buscados por el FBI.

7.1 Introducción

Actualmente se debe estar bien consciente de que las poblaciones son generalmente muy grandes como para ser estudiadas en su totalidad. Su tamaño requiere que se seleccionen muestras, las cuales se pueden utilizar más tarde para hacer inferencias sobre las poblaciones. Si un gerente de una tienda minorista desea saber sobre el gasto promedio de sus clientes durante el año anterior, podría encontrar difícil calcular el promedio de los cientos o quizá miles de clientes que pasaron por su tienda. Sería mucho más fácil estimar la media poblacional con la media de una muestra representativa.

Hay por lo menos dos tipos de estimadores que se utilizan más comúnmente para este propósito: un estimador puntual y un estimador por intervalo. Un **estimador puntual** utiliza un estadístico para estimar el parámetro en un solo valor o punto. El gerente de la tienda puede seleccionar una muestra de n = 500 clientes y hallar el gasto promedio de $\bar{X}$= US$37.10. Este valor sirve como una estimación puntual para la media poblacional.

Una **estimación por intervalo** especifica el rango dentro del cual está el parámetro desconocido. El gerente puede decidir que la media poblacional está en algún sitio entre US$35 y US$38. Tal intervalo con frecuencia va acompañado de una afirmación sobre el nivel de confianza que se da en su exactitud. Por tanto se llama **intervalo de confianza** (I.C.).

Estimador Un estimador puntual utiliza un número único o valor para localizar una estimación del parámetro. Un intervalo de confianza denota un **rango** dentro del cual puede encontrarse el parámetro, y el nivel de confianza que el intervalo contiene del parámetro.

En realidad hay tres niveles de confianza relacionados comúnmente con los intervalos de confianza: 99, 95, y 90%. No hay nada mágico sobre estos tres valores. Se podría calcular un intervalo de confianza del 82% si se deseara. Estos tres niveles de confianza, denominados **coeficientes de confianza**, son simplemente convencionales. El gerente mencionado anteriormente puede tener un 95% de confianza en que la media poblacional está entre US$35 y US$38.

Las estimaciones por intervalo gozan de ciertas ventajas sobre las estimaciones puntuales. Debido al error de muestreo, probablemente $\bar{X}$ no será igual a μ. Sin embargo, no hay manera de saber qué tan grande es el error de muestreo. Por tanto, los intervalos se utilizan para explicar esta discrepancia desconocida.

Se iniciará con una discusión sobre lo que es un intervalo de confianza y cómo interpretarlo.

A. El fundamento de un intervalo de confianza

Un intervalo de confianza tiene un **límite inferior de confianza** (LIC) y un **límite superior de confianza** (LSC). Estos límites se hallan calculando primero la media muestral, $\bar{X}$. Luego se suma una cierta cantidad a $\bar{X}$ para obtener el LSC, y la misma cantidad se resta de $\bar{X}$ para obtener el LIC. La determinación de dicha cantidad es el tema de este capítulo.

¿Cómo se puede construir un intervalo y luego argumentar que se puede tener un 95% de confianza en que contiene μ, si incluso no se sabe cuál es la media poblacional? Vale la pena recordar de la discusión anterior sobre la Regla Empírica que el 95.5% de todas las medias muestrales caen dentro de dos errores estándar de la media poblacional. Entonces la media poblacional está máximo a dos errores estándar del 95.5% de todas las medias muestrales. Por tanto, al comenzar con *cualquier* media muestral, si se pasa de dos errores estándar por encima de dicha media y dos errores estándar por debajo de ella, se puede tener un 95.5% de confianza en que el intervalo resultante contenga la media poblacional desconocida.

La discusión sobre distribuciones de muestreo mostró que de toda población se pueden obtener muchas muestras diferentes de un tamaño dado, cada una con su propia media. La figura 7.1 muestra seis de estas medias muestrales posibles. Si la muestra da $\bar{X}_1$, un intervalo que se extiende dos errores estándar por encima y dos errores estándar por debajo de $\bar{X}_1$ todavía incluye el valor desconocido de la media poblacional. De igual forma, si la muestra hubiese dado una media de $\bar{X}_2$, el intervalo resultante también incluiría la media poblacional. Vale la pena destacar que sólo $\bar{X}_3$ y $\bar{X}_5$ quedan tan lejos de la media poblacional que un intervalo de ± 2 errores estándar no incluye la media poblacional. Todas las otras muestras consideradas producirán un intervalo que contiene la media poblacional. Entonces, la clave para recordar es esta: como la media poblacional está a lo más a dos errores estándar para el 95.5% de todas las medias muestrales, entonces dada una media muestral cualquiera, se puede estar 95.5% seguro de que el intervalo de dos errores estándar alrededor de dicha media muestral contiene la media poblacional desconocida.

Si se desea construir un intervalo más convencional del 95% (en lugar del 95.5% discutido anteriormente), ¿cuántos errores estándar se debe mover por encima y por debajo de la media muestral? Como lo demuestra la figura 7.2, debido a que la tabla Z contiene valores sólo para el área que está por encima o por debajo de la media, se debe dividir el 95% por 2, produciendo 0.4750. Luego, se halla el valor de Z, correspondiente a un área de 0.4750, el cual es Z = 1.96. Así, para construir un intervalo de confianza del 95%, simplemente se especifica un intervalo de 1.96 errores estándar por encima y por debajo de la media muestral. Este valor del 95% es llamado **coeficiente de confianza**.

Figura 7.1
Posible intervalo de confianza del 95.5% para estimar μ

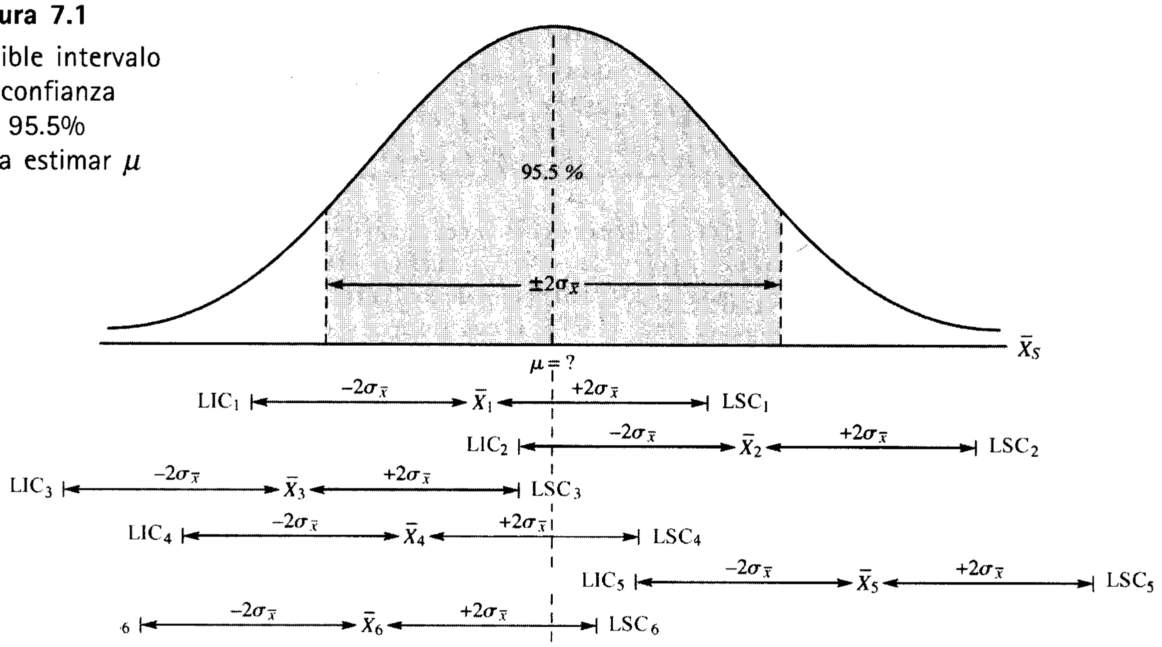
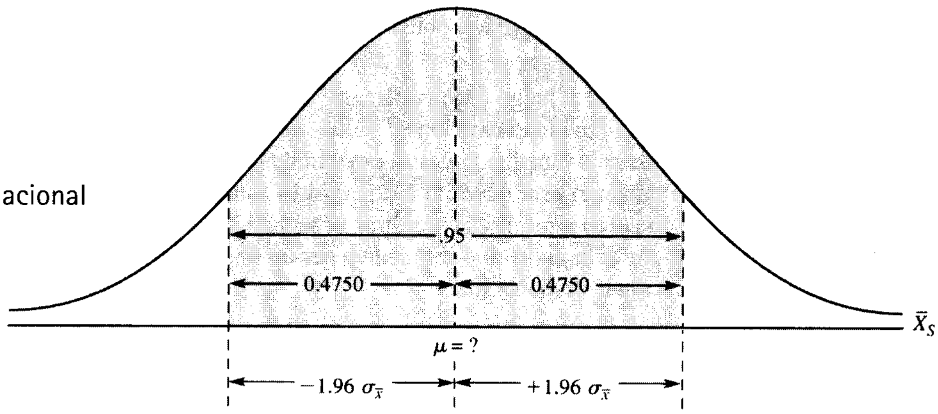


Figura 7.2
Intervalo de confianza del 95% para estimar la media poblacional



Coeficiente de confianza El coeficiente de confianza es el nivel de confianza que se tiene en el que el intervalo contenga el valor desconocido del parámetro.

Quizá se puede ilustrar mejor utilizando un ejemplo. Se comienza desarrollando una estimación por intervalo para la media poblacional con una muestra grande ($n \geq 30$).

7.2 Intervalo de confianza para la media poblacional – Muestras grandes

Uno de los usos más comunes de los intervalos de confianza es estimar la media poblacional. Un fabricante puede querer estimar la producción mensual promedio de su planta; un representante de mercadeo puede interesarse en la reducción en las ventas semanales promedio; el jefe financiero de una firma, que aparece entre las 500 mejores firmas en la revista *Fortune*, puede querer estimar los rendimientos trimestrales promedio que se tuvieron en operaciones corporativas. El número de circunstancias que se encuentran comúnmente en el mundo de los negocios y que requiere de una estimación de la media poblacional es casi ilimitado.

Se debe recordar que el intervalo se forma utilizando la media muestral como una estimación puntual para el cual se adiciona y se resta un cierto valor para obtener los límites superior e inferior del intervalo de confianza, respectivamente. Por tanto el intervalo es

Intervalo de confianza para estimar μ cuando σ es conocido	I.C. para estimar $\mu = \bar{X} \pm Z\sigma_{\bar{x}}$	[7.1]

Cuánto debe sumarse y restarse, depende en parte del nivel de confianza deseado, estipulado por el valor de Z en la fórmula (7.1). Un nivel de confianza del 95% requiere un valor de Z de 1.96 (0.95/2 = 0.4750). El área de 0.4750 corresponde a un valor de Z de 1.96.

Consideremos el caso de un promotor inmobiliario quien intenta construir un gran centro comercial. Puede estimar en el área el ingreso promedio por familia como indicador de las ventas esperadas. Una muestra de $n = 100$ familias da una media de $\bar{X}$ = US\$35,500. Se asume que la desviación estándar poblacional es σ = US\$7,200. Dado que $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, se estima un intervalo del 95% como

$$\begin{aligned} \text{I.C. para estimar } \mu &= 35{,}500 \pm (1.96)\frac{7{,}200}{\sqrt{100}} \\ &= 34{,}088.80 \leq \mu \leq 36{,}911.20 \end{aligned}$$

A. Interpretación de un intervalo de confianza

El promotor puede interpretar los resultados de su intervalo de confianza de dos formas. La primera, y la más común, establece que el promotor tiene un "95% de confianza en que la media poblacional real desconocida esté entre US\$34,088.80 y US\$36,911.20". Aunque el valor real para la media poblacional sigue siendo desconocido, el promotor tiene un 95% de confianza en que esté entre estos dos valores.

La segunda interpretación reconoce que se pueden desarrollar muchos intervalos de confianza diferentes. Otra muestra probablemente produciría una media muestral diferente debido al error de muestreo. Con una $\bar{X}$ diferente, el intervalo tendría límite superior e inferior distintos. Por tanto, la segunda interpretación establece que si se construyen todos los ${}_{N}C_{n}$ intervalos de confianza, el 95% de ellos contendrá la media poblacional desconocida.

Si una segunda muestra da una media de US\$35,600 en lugar de US\$35,500, el intervalo es

$$\begin{aligned} \text{I.C. para estimar } \mu &= \$35{,}600 \pm (1.96)\frac{\text{US}\$7{,}200}{\sqrt{100}} \\ &= \text{US}\$34{,}188.80 \leq \mu \leq \text{US}\$37{,}011.20 \end{aligned}$$

El promotor puede estar un 95% seguro de que la media poblacional está comprendida entre US\$34,188.80 y US\$37,011.20. Si todos los intervalos posibles se construyeran con base en todas las medias muestrales diferentes, el 95% de ellas contendría la media poblacional desconocida.

Esto por supuesto significa que el 5% de todos los intervalos estaría errado – no contendrían la media poblacional–. Este 5%, hallado como (1 – coeficiente de confianza), es denominado el **valor alfa** y representa la probabilidad de error. El valor alfa es la probabilidad de que cualquier intervalo dado no contenga la media poblacional.

Valor alfa Es la probabilidad de error o la probabilidad de que un intervalo dado no contenga la media poblacional desconocida.

B. Intervalo de confianza cuando σ es desconocida

La fórmula (7.1) requiere la suposición improbable que la desviación estándar poblacional σ es conocida. En el evento probable que σ sea desconocida, la desviación estándar de la muestra debe substituirse:

Intervalo de confianza para estimar μ cuando σ es desconocida

$$\text{I.C. para estimar } \mu = \bar{X} \pm Zs_{\bar{x}} \qquad [7.2]$$

en donde $s_{\bar{x}} = s/\sqrt{n}$

Gerry Gerber, CPA, acaba de registrar las declaraciones de impuestos de sus clientes. Desea estimar la cantidad promedio que deben al Servicio de Renta Interna. De los 50 clientes que seleccionó en su muestra, la cantidad promedio que se adeudaba era de US\$652.68. Ya que la desviación estándar de todos sus clientes σ es desconocida, Gerber debe *estimar* σ con la desviación estándar de la muestra de s = US\$217.43. Si se desea un nivel del 99% de confianza, el valor de Z apropiado es 2.58 (0.99/2 = 0.4950). De la tabla Z, un área de 0.4950 revela que Z = 2.58. Utilizando la fórmula (7.2)

$$\begin{aligned}\text{I.C. para estimar } \mu &= \bar{X} \pm Zs_{\bar{x}}\\ &= \text{US\$}652.68 \pm 2.58\,\frac{\text{US\$}217.43}{\sqrt{50}}\\ &= \text{US\$}573.35 \leq \mu \leq 732.01\end{aligned}$$

El señor Gerber puede tener un 99% de confianza en que la cantidad promedio que deben todos sus clientes al Servicio de Renta Interna (SRI) está entre US\$573.35 y US\$732.01.

¿Qué pasaría a este intervalo si el señor Gerber estuviera dispuesto a aceptar un nivel de confianza del 95%? Con un valor de Z de 1.96, el intervalo sería

$$\begin{aligned}&\text{US\$}652.68 \pm 1.96\,\frac{\text{US\$}217.43}{\sqrt{50}}\\ &\text{US\$}592.41 \leq \mu \leq \text{US\$}712.96\end{aligned}$$

Los resultados son tanto buenos como malos. Las buenas noticias son que el intervalo del 95% es más estrecho y ofrece mayor **precisión**. Un intervalo amplio no es especialmente útil. Revelaría muy poco si el profesor *le pidiera que la media del siguiente examen estuviera entre el 0 y el 100%. Entre más estrecho sea el intervalo, más* significativo es.

Las malas noticias son que el señor Gerber ahora está el 95% seguro de que el intervalo contiene en realidad μ. Aunque el intervalo es más preciso (más estrecho), la probabilidad de que contenga μ se ha reducido del 99 al 95%. El señor Gerber tuvo que abandonar algo de confianza y ganar más precisión.

La pantalla 7.1 muestra la impresión en Minitab de la declaración de impuestos del señor Gerber (TXRET). Haciendo clic en **Stat > Basic Statistics > 1-sample Z** e ingresando 99% para el **Confidence Interval Level** y 217.43 para **Sigma**, el señor Gerber obtiene su estimado por intervalo, el cual difiere sólo levemente por encima de sus cálculos manuales, debido a la aproximación.

Pantalla 7.1

```
MTB > ZIntervalo 99.0  217.434338 C1

Confidence intervals (Intervalos de confianza)

The assumed sigma (El sigma asumido) = 217
                               Mean        StDev                 SE Mean        C.I.
Variable                  N  (Media)(Desviación estándar)(Media SE) 99.0 % (I.C.)
TXRET (Declaración
de impuestos)             50  652.7        217.4                    30.7  ( 573.4,  731.9)
```

Ejemplo 7.1

Checkered Cabs planea comprar una flota de nuevos taxis para sus operaciones en Miami. La decisión depende de si el rendimiento del auto en consideración es por lo menos 27.5 millas por galón de gasolina. Los 36 carros que prueba la compañía reportan una media de 25.6 millas por galón (MPG), con una desviación estándar de 3.5 MPG. A un nivel de confianza del 99%, ¿qué aconsejaría a Checkered que hiciera?

Solución

Se tiene que el intervalo de confianza I.C. para estimar $\mu = 25.6 \pm (2.58)\dfrac{3.5}{\sqrt{36}}$

$$= 24.10 \leq \mu \leq 27.11$$

Interpretación

Puede estar un 99% seguro de que las MPG promedio de este carro es menor que el mínimo de 27.5 requerido. Usted debería aconsejar a Checkered que busque un modelo alternativo.

Ejercicios de la sección

1. ¿En qué se diferencian las estimaciones puntuales de las estimaciones por intervalo?
2. Si la media poblacional es desconocida, ¿cómo es posible dar un nivel de confianza a su estimado por intervalo? Incluya una gráfica en su respuesta.
3. Un intervalo del 90% para estimar la ganancia promedio de peso de los ratones de laboratorio oscila entre 0.93 onzas y 1.73 onzas. ¿Cómo interpretaría estos resultados? ¿Qué valor de Z se utilizó en el estimado?
4. Cien latas de 16 onzas de la salsa de tomate Jake's Mom's tienen un promedio de 15.2 onzas. La desviación estándar poblacional en peso es de 0.96 onzas. ¿A un nivel de confianza del 95% las latas parecen estar llenas con un promedio de 16 onzas?
5. Para estimar el gasto promedio de los clientes en el McDonald's local, los estudiantes de una clase de estadística toman una muestra de 200 clientes y encuentran un gasto promedio de US$5.67, con una desviación estándar de US$1.10. ¿Cuál es el intervalo de confianza del 95% para los gastos promedio de todos los clientes? Interprete sus resultados.

6. Un estudio realizado por los profesores en una universidad de Kansas está diseñado para ofrecer inferencias sobre las tasas de desempleo por condado en Estados Unidos. Una muestra de 200 condados reporta una tasa promedio del 6.2%, con una desviación estándar del 1.7%. A un nivel de confianza del 90%, ¿cuál es el estimado de la tasa de desempleo promedio por condado en la nación? Interprete sus resultados.
7. Después de observar 50 programas de televisión seleccionados aleatoriamente, la Asociación Nacional de Educación (*National Education Association*, NEA) reportó un promedio de 32.7 actos de violencia en 1997. Asuma una desviación estándar muestral de 10.1. ¿Cuál sería su estimación al 95% del número promedio de actos violentos por programa que los niños ven en la televisión?
8. Un teatro de cine local desea desarrollar un intervalo para estimar las cajas promedio de palomitas de maíz que se venden por sala de cine. Si los registros llevados para 70 salas revelan un promedio de 54.98 cajas y una desviación estándar de 12.7, calcule e interprete un intervalo de confianza del 92% para la media poblacional.
9. Una muestra de 121 llamadas al número 900 que usted maneja tiene una duración promedio de 16.6 minutos y una desviación estándar de 3.63 minutos. Usted pretende discontinuar el servicio a menos que la duración promedio sea superior a 18 minutos. En el nivel de confianza del 90% ¿cuál es su decisión?
10. ¿Cuál sería su decisión en el problema anterior a un nivel de confianza del 95%? ¿Por qué son diferentes los intervalos?
11. ¿Cuál sería su decisión si el ejercicio 9 utilizara una muestra de 200 llamadas? ¿Por qué los intervalos son diferentes?

7.3 Intervalo de confianza para la media en el caso de muestras pequeñas – la distribución *t*

En todos los ejemplos anteriores, el tamaño de la muestra era mayor ($n \geq 30$). Sin embargo, no siempre puede ser posible obtener por lo menos 30 observaciones. Para una compañía de seguros que prueba la resistencia al impacto de los autos, destruir a propósito 30 vehículos de lujo puede volverse un poco costoso. Un investigador médico que prueba una nueva medicina puede no encontrar 30 personas dispuestas a actuar como conejillo de *indias. En muchos casos una muestra grande no es posible.*

Cuando debe tomarse una muestra pequeña, la distribución normal puede no aplicarse. El teorema del límite central asegura normalidad en el proceso de muestreo sólo si la muestra es grande. Cuando se utiliza una muestra pequeña, puede ser necesaria una distribución alternativa, la **distribución *t* Student** (o simplemente la distribución *t*). Específicamente, la distribución *t* se utiliza cuando se cumplen las tres condiciones: (1) la muestra es pequeña, (2) σ es desconocida, y (3) la población es normal o casi normal. Si σ es conocida, la distribución Z se usa inclusive si la muestra es pequeña. Además, si no puede asumirse una población normal, se aumenta el tamaño de la muestra para utilizar la distribución Z y de no ser posible se debe confiar en las pruebas *no paramétricas*.

La distribución *t* Student fue desarrollada en 1908 por William S. Gosset (1876 – 1937), quien trabajó como experto cervecero para Guinnes Breweries en Dublin, Irlanda. Guiness no permitía que sus empleados publicaran su investigación, de manera que Gosset (a quien le gustaba "jugar con los números para relajarse") informó por primera vez sobre su distribución *t*, aunque publicó bajo el seudónimo de "Student" para proteger su trabajo.

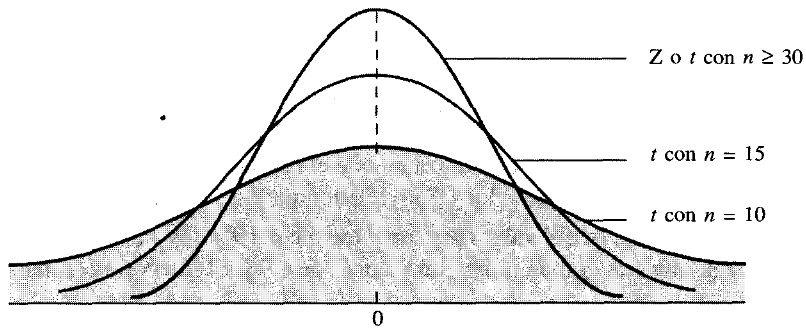
Al igual que la distribución Z, la distribución *t* tiene una media de cero, es simétrica con respecto a la media y oscila entre $-\infty$ y $+\infty$. Sin embargo, mientras que la distribución Z tiene una varianza de $\sigma^2 = 1$, la varianza de la distribución *t* es mayor que 1. Por tanto, es más plana y más dispersa que la distribución Z. La varianza para la distribución *t* es

Varianza de la distribución *t*

$$\sigma^2 = \frac{n-1}{n-3} \qquad [7.3]$$

En realidad la distribución t es una familia de distribuciones cada una con su propia varianza. La varianza depende de los *grados de libertad* (g.l.), definidos como el número de observaciones que se pueden escoger libremente. Es el número de observaciones menos el número de restricciones impuestas sobre tales observaciones, en donde una restricción es algún valor que tales observaciones deben poseer. Se asume que se tienen $n = 4$ observaciones que deben producir una media de 10. La media de 10 sirve como una restricción y hay $n - 1 = 3$ grados de libertad. Por tanto, se pueden escoger tres observaciones cualquiera; por ejemplo se puede escoger 8, 9 y 11. Después de que se seleccionan estos tres valores, ya no hay libertad para escoger la última observación. El cuarto valor *debe* ser 12 si se quiere tener un promedio de 10. Vale la pena destacar en la figura 7.3 que a medida que n aumenta, la distribución t se aproxima a la distribución Z. Es por esto que se puede utilizar la distribución Z cuando $n \geq 30$.

Figura 7.3
La familia de distribuciones t



Grados de libertad El número de observaciones menos el número de restricciones impuestas sobre tales observaciones.

Como se verá en breve, para todo conjunto de condiciones dadas la distribución t producirá un intervalo más amplio que la distribución Z, si ésta se utilizara. Este ancho adicional es necesario debido a que se pierde algo de precisión porque σ es desconocida y debe estimarse.

El estadístico t se calcula en gran parte como el estadístico Z.

$$t = \frac{\bar{X} - \mu}{s_{\bar{x}}} \qquad [7.4]$$

Reescribiendo (7.4) algebraicamente para expresarlo como un intervalo de confianza para estimar μ, se tiene que

Intervalo de confianza para estimar la media poblacional - muestras pequeñas

$$\text{I.C. para estimar } \mu = \bar{X} \pm (t)(s_{\bar{x}}) = \bar{X} \pm t\frac{s}{\sqrt{n}} \qquad [7.5]$$

El valor apropiado de t puede hallarse de la tabla F en el apéndice III. Para ilustrar, se asume que se desea un intervalo de confianza del 95% y se tiene una muestra de 20 observaciones. Debido a que $n = 20$, los grados de libertad son g.l. $= n - 1 = 19$. Bajando por la primera columna en la tabla F bajo "g.l." hasta 19. Se mueve a través de dicha fila hacia la columna encabezada por un nivel de confianza de 0.95 para las pruebas de dos colas. (Se ignoran las dos filas referentes a las pruebas de una cola. Estas se tratarán en el capítulo 8). La entrada resultante de 2.093 es el valor t apropiado para un intervalo de confianza del 95% con un tamaño muestral de 20 (g.l. = 19).

Consideremos el siguiente problema tomado de *The Wall Street Journal*. Una empresa de construcción fue culpada de inflar los comprobantes que registra para los contratos de construcción con el gobierno federal. El contrato estableció que un cierto tipo de trabajo debería promediar US$1,150. Por motivos de tiempo, los directivos de sólo 12 agencias del gobierno fueron llamados a dar testimonio ante la corte respecto a los comprobantes de la empresa. Si se descubrió a partir del testimonio una media de US$1,275 y una desviación estándar de US$235, ¿un intervalo de confianza del 95% apoyaría el caso legal de la empresa? Se asume que los montos de los comprobantes son normales.

Un nivel de confianza del 95% con g.l. = 12 -1 = 11 resulta de la tabla F un valor t de 2.201. Entonces

$$\begin{aligned}\text{I.C. para estimar } \mu &= \bar{X} \pm t\frac{s}{\sqrt{n}} \\ &= 1275 \pm (2.201)\frac{235}{\sqrt{12}} \\ &= 1275 \pm 149.31 \\ \text{US\$1,125.69} \leq \mu &\leq \text{US\$1,424.31}\end{aligned}$$

La corte puede tener un 95% de confianza en que el promedio de todos los comprobantes está entre US$1,125 y US$1,424. *Este intervalo contiene los US$1,150 acordados*, fortaleciendo la defensa de la empresa.

Vale la pena destacar que el valor t para un intervalo del 95% es 2.201 (dado g.l. = 11), mientras que un intervalo del 95% de una muestra grande utiliza un valor Z de 1.96. El intervalo con base en un valor t es, por tanto, más amplio.

Ejemplo 7.2

El contrato laboral realizado entre United Auto Workers (UAW) y Ford Motor Company (FMC) requirió que la producción promedio para una sección de producción se realizara a 112 unidades por mes, por empleado. Surgieron desacuerdos entre UAW y FMC respecto a si se mantenía este estándar o no. El contrato laboral especificó que si los niveles de producción promedio se reducían por debajo de la cantidad estipulada de $\mu = 112$, a FMC se le permitiría tomar "acciones correctivas". Debido al costo implicado, sólo se evaluaron 20 trabajadores, resultando una media de 102 unidades. Se asume que se encontró una desviación estándar de 8.5 unidades y que los niveles de producción están distribuidos normalmente. ¿Un intervalo de confianza del 90% tiende a sugerir una contravención del contrato laboral, permitiendo así una acción correctiva?

Solución

Con un nivel de confianza del 90% y n - 1 19 g.l., la tabla F da un valor t de 1.729.

$$\begin{aligned}\text{I.C. para estimar } \mu &= \bar{X} \pm t\frac{s}{\sqrt{n}} \\ &= 102 \pm (1.729)\frac{8.5}{\sqrt{20}} \\ &= 102 \pm 3.29 \\ 98.71 \leq \mu &\leq 105.29\end{aligned}$$

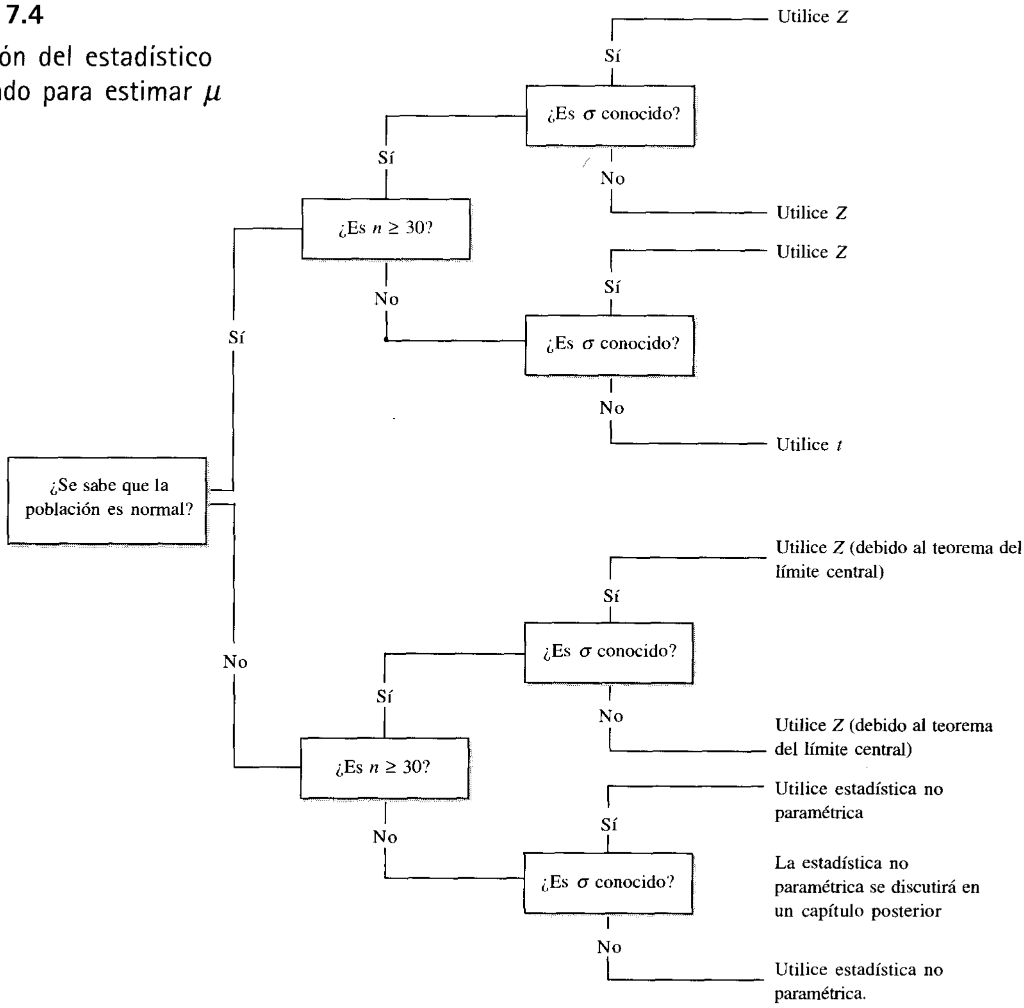
El nivel de producción promedio de 112 unidades, especificado en el contrato laboral, no está dentro del intervalo de confianza.

Interpretación

Hay un nivel de confianza del 90% en el que el contrato está siendo violado. FMC está en su derecho para interponer un recurso por retraso en la productividad.

Obviamente, decidir si utilizar un estadístico t o Z es crucial. La figura 7.4 ayudará en la selección del estadístico apropiado. Vale la pena recordar que la distribución t debería utilizarse cuando todas estas tres condiciones están presentes: (1) la población es normal, (2) se toma una muestra pequeña, y (3) σ es desconocido.

Figura 7.4
Selección del estadístico apropiado para estimar μ



Ejercicios de la sección

12. ¿Cuáles tres condiciones se deben cumplir antes de que se pueda utilizar la distribución t?

13. ¿En qué se diferencia la varianza de la distribución t de la de la distribución Z? Si un conjunto de datos tiene 17 observaciones, ¿cuál es la varianza de la distribución t?

14. The Lucky Lady, una tertulia estudiantil popular, vende vasos de cerveza de 16 onzas. Diez estudiantes compran un total de 22 vasos, y utilizando su propia taza de medida, estiman los contenidos promedio. La media muestral es de 15.2 onzas, con $s = 0.86$. ¿Con un nivel de confianza del 95% los estudiantes creen que su dinero lo vale? Interprete el intervalo.

15. Dell Publishings muestrea 23 paquetes para estimar el costo postal promedio. La media muestral es de US\$23.56, con s = US\$4.65.

a. El editor senior de Dell espera mantener el costo promedio por debajo de US$23.00. Calcule e interprete el intervalo de confianza del 99%. ¿El editor estará satisfecho?

b. Compare los resultados de la parte *a* con el intervalo de confianza del 99%, si s = US$2.05. Explique por qué existe diferencia.

c. Manteniendo s = US$4.65, compare los resultados de la parte *a* con el intervalo del 95%. Explique la diferencia.

16. Las bonificaciones para 10 nuevos jugadores de la Liga Nacional de Fútbol se utilizan para estimar la bonificación promedio para todos los nuevos jugadores. La media muestral es de US$65,890 con s = US$12,300. ¿Cuál es su estimación con un intervalo del 90% para la media poblacional?

17. Una muestra de 25 llamadas a *Psychic Friends Network* (Red de Amigos Síquicos) revela un costo promedio de US$23.87. Si la desviación estándar es US$9.56, ¿cuál es la estimación con un intervalo del 98% para el costo promedio de todos los que llaman para conocer su futuro?

18. Greenleaf Lawn Care descubre que el costo promedio de adornar los jardines de 20 casas del área es de US$2,365, con s = US$983. Al nivel de confianza del 99%, ¿qué costo promedio estimaría usted para adornar los jardines de todas las casas del área?

7.4 Intervalo de confianza para la proporción poblacional

Las decisiones dependen con frecuencia de parámetros que son binarios, parámetros con sólo dos posibles categorías dentro de las cuales pueden clasificarse las respuestas. En este evento, el parámetro de interés es la *proporción* poblacional. Una empresa puede desear saber qué proporción de sus clientes paga a crédito en oposición a quienes utilizan efectivo. Las corporaciones con frecuencia están interesadas en qué porcentaje de sus productos son defectuosos en oposición al porcentaje que no es defectuoso, o qué proporción de sus empleados renuncian después de un año en contraste con la proporción que no renuncia después de un año. En cada uno de estos casos, existen sólo dos posibles resultados. Por tanto, la preocupación se concentra en la proporción de respuestas que queda dentro de uno de estos dos resultados.

En el capítulo anterior se encontró que si $n\pi$ y $n(1 - \pi)$ son ambos mayores que 5, la distribución de las proporciones muestrales será normal y la distribución muestral de la proporción muestral tendrá una media igual a la proporción poblacional π y un error estándar de

El error estándar de la distribución muestral de las proporciones muestrales	$\sigma_p = \sqrt{\dfrac{\pi(1-\pi)}{n}}$	[7.6]

Sin embargo, la fórmula (7.6) contiene π, el parámetro que se desea estimar. Por tanto, la proporción muestral p se utiliza como estimador de π.

La fórmula (7.6) puede replantearse como

Estimación del error estándar de la distribución de las proporciones muestrales	$s_p = \sqrt{\dfrac{p(1-p)}{n}}$	[7.7]

El intervalo de confianza es entonces

Intervalo de confianza para estimar la proporción poblacional	I.C. para estimar $\pi = p \pm Zs_p$	[7.8]

El gerente de una estación de televisión debe determinar en la ciudad qué porcentaje de casas tiene más de un televisor. Una muestra aleatoria de 500 casas revela que 275 tienen dos o más televisores. ¿Cuál es el intervalo de confianza del 90% para estimar la proporción de todas las casas que tienen dos o más televisores? Dados estos datos, $p = 275 / 500 = 0.55$, y

$$s_p = \sqrt{\frac{(0.55)(0.45)}{500}}$$
$$= 0.022$$

La tabla E da un valor de Z de 1.65 para un intervalo de confianza del 90%.

$$\text{I.C. para estimar } \pi = 0.55 \pm (1.65)(0.022)$$
$$= 0.55 \pm 0.036$$
$$0.514 \leq \pi \leq 0.586$$

El gerente puede tener un 90% de confianza que entre el 51.4% y el 58.6% de las casas de la ciudad tienen más de un televisor.

Ejemplo 7.3

Las empresas de búsqueda de ejecutivos se especializan en ayudar a las empresas a ubicar y asegurar talento para la alta gerencia. Tales firmas denominadas "cazadoras de cabezas" son responsables de la ubicación de muchos de los mejores directores ejecutivos de la nación. *Business Week* reportó recientemente que "uno de cada cuatro directores ejecutivos es una persona de fuera –un ejecutivo con menos de 5 años en la compañía que maneja–". Si en una muestra de 350 compañías de los Estados Unidos, 77 tienen directores ejecutivos de fuera, ¿un intervalo del 99% de confianza apoyaría la afirmación?

Solución

$$p = \frac{77}{350} = 0.22$$
$$s_p = \sqrt{\frac{(0.22)(0.78)}{350}} = 0.022$$
$$\text{I.C. para estimar } \pi = p \pm Zs_p$$
$$= 0.22 \pm (2.58)(0.022)$$
$$0.163 \leq \pi \leq 0.277$$

Interpretación

Se confía en que al nivel del 99% entre el 16.3% y el 27.7% de las empresas de Estados Unidos tienen directores ejecutivos de fuera. La afirmación está apoyada por tales descubrimientos, ya que el 25% está contenido dentro del intervalo.

Ejercicios de la sección

19. Qué es s_p y qué mide?

20. CNN informó que el 68% de todos los estudiantes de secundaria tenía computadores en sus casas. Si una muestra de 1,020 estudiantes revela que 673 tienen computadores caseros, ¿un intervalo del 99% apoya a CNN?

21. Como respuesta al nuevo furor de fumar cigarrillo que arrasa la nación, el Instituto Nacional del Corazón (National Heart Institute) practicó encuestas a mujeres para estimar la proporción de quienes fumaban un cigarrillo ocasionalmente. De las 750 mujeres que respondieron, 287 respondieron que sí lo hacían. Con base en estos datos, ¿cuál es su estimación al 90% para la proporción de todas las mujeres que participan de este hábito?

22. La Asociación Nacional de Viajes (National Travel Association) tomó muestras de las personas que tomaban vacaciones en Irlanda para estimar la frecuencia con la cual los norteamericanos visitaban Emerald Isle. ¿Cuál es el intervalo de confianza del 96% para la proporción de turistas que son norteamericanos, si 1,098 de los 3,769 encuestados portaban pasaportes de Estados Unidos?

23. De los 1,098 turistas norteamericanos 684 habían registrado su viaje a través de un agente de viajes. Calcule e interprete el intervalo del 95% para la proporción de todos los norteamericanos que utilizan los servicios de agencias de viajes profesionales en Irlanda.

24. Si 896 de los turistas norteamericanos recomendaran al viaje a sus amigos, ¿qué porcentaje de todos los turistas norteamericanos harían lo mismo con un nivel del 99% de confianza?

25. Si 796 de los 1,098 turistas norteamericanos planean viajes para regresar a Irlanda, con un nivel de confianza del 92%, ¿qué proporción de todos los turistas norteamericanos repetirían sus vacaciones?

7.5 Control del ancho de un intervalo

Como se expresó anteriormente, es preferible un intervalo más estrecho debido a la precisión adicional que proporciona. Hay dos métodos principales para lograr un intervalo más preciso: (1) reducir el nivel de confianza y (2) incrementar el tamaño de la muestra.

A. Reducción del nivel de confianza

Ya se ha visto, en el intento del señor Gerber por estimar la declaración de impuestos promedio de sus clientes, que un incremento en la precisión puede obtenerse aceptando un nivel inferior de confianza. Su intervalo de confianza del 99% oscilaba entre US\$573 y US\$732, mientras que el intervalo del 95% era más estrecho de US\$594 a US\$712. Esto resultó del hecho que el intervalo de confianza del 99% requirió un valor de Z de 2.58 en lugar de 1.96 que utiliza el intervalo del 95%.

Sin embargo, había un costo involucrado en lograr esta precisión mayor: el nivel de confianza bajó a 95%, produciendo un 5% de probabilidad de error en lugar del 1% relacionado con el intervalo de confianza del 99%. ¿Existe alguna manera en la que se pueda reducir el intervalo sin sufrir una pérdida de confianza? Sí, incrementando el tamaño muestral.

B. Incremento del tamaño muestral

Incrementando el tamaño muestral se puede reducir el error estándar $\sigma/\sqrt{n}$. Si el tamaño muestral del señor Gerber se incrementa a 80, el intervalo del 99% presenta un grado de precisión similar al intervalo más estrecho del 95%, sin ninguna pérdida de confianza. Con $n = 80$, el intervalo del 99% es

$$\text{I.C. para estimar } \mu = \text{US\$}652.68 \pm 2.58 \frac{\text{US\$}217.43}{\sqrt{80}}$$

$$\text{US\$}589.96 \leq \mu \leq \text{US\$}715.39$$

Esto está muy cercano al intervalo más preciso del 95% de US$592,41 a US$712,96, pero mantiene un nivel de confianza del 99%.

Infortunadamente, esta ventaja no se gana sin un precio. El tamaño más grande de la muestra significa más tiempo y más dinero que deben gastarse al recolectar y manejar los datos. De nuevo, debe tomarse una decisión. Se vuelve una decisión gerencial respecto a qué método tomar.

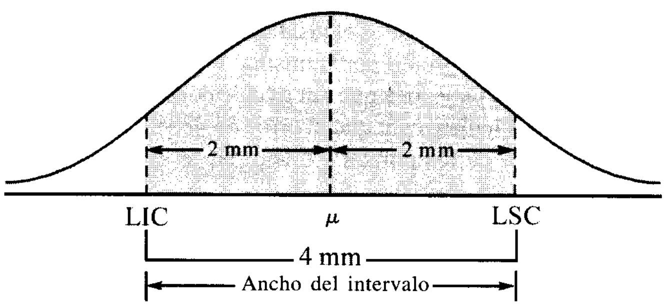
7.6 Determinación del tamaño apropiado de la muestra

El tamaño de la muestra juega un papel importante al determinar la probabilidad de error así como en la precisión de la estimación. Una vez que se ha seleccionado el nivel de confianza, dos factores importantes influyen en el tamaño muestral: (1) la varianza de la población σ^2 y (2) el tamaño del *error tolerable* que el investigador está dispuesto a aceptar. Mientras que el primer factor está más allá del control del investigador (no hay nada que se pueda hacer sobre la varianza de la población), sí es posible limitar el tamaño del error.

El tamaño del error que un investigador puede tolerar depende de qué tan crítico es el trabajo. Algunas tareas extremadamente delicadas requieren de resultados exactos: los procedimientos médicos vitales de los cuales dependen vidas humanas, o la producción de piezas de una máquina que deba cumplir medidas precisas, pueden tolerar sólo un pequeño error. En otros casos, los errores más grandes pueden tener consecuencias menos graves.

Se supone que en la fabricación de una pieza para los reproductores de discos compactos (CD), un error de 2 milímetros (mm) en el diámetro no causaría problema alguno; sin embargo, todo error superior a 2 mm resultará en un reproductor defectuoso. Si una pieza puede variar por encima y por debajo de algún diámetro deseado en 2 mm, se permite un intervalo de 4 mm. Todo intervalo dado es dos veces el error tolerable. Ver la figura 7.5 para observar una ilustración.

Figura 7.5
El error tolerable es la mitad del intervalo



En esta sección se considera la determinación del tamaño muestral apropiado bajo varias condiciones.

A. Tamaño de la muestra para estimar μ

Vale la pena recordar que la desviación normal Z puede expresarse como

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Esto puede reescribirse algebraicamente como

Tamaño muestral para intervalos de la media poblacional	$n = \dfrac{Z^2\sigma^2}{(\bar{X} - \mu)^2}$	[7.9]

en donde la diferencia entre la media muestral y la media poblacional ($\bar{X}$ - μ) es el error. En el ejemplo anterior de los reproductores de CD's, con un error tolerable de 2 mm, la fórmula (7.9) se escribiría como

$$n = \frac{Z^2\sigma^2}{(2)^2}$$

El valor de Z depende del nivel de confianza requerido. Esto deja por determinar sólo σ^2 para calcular el tamaño muestral apropiado. En el evento probable que σ^2 sea desconocido, puede estimarse mediante la desviación estándar muestral s, utilizando una **muestra piloto** de cualquier tamaño razonable ($n \geq 30$). La varianza calculada de esta muestra preliminar puede entonces utilizarse en la fórmula (7.9).

Por ejemplo, se asume que el fabricante de los reproductores de discos compactos desea construir un intervalo del 95% para el tamaño promedio de la pieza. Una muestra piloto ha revelado una desviación estándar de 6 mm. ¿Qué tan grande debería ser la muestra? Un intervalo del 95% da un valor de Z de 1.96. Por tanto,

$$n = \frac{(1.96)^2(6)^2}{(2)^2} = 34.5 \text{ o } 35$$

El fabricante debería seleccionar una muestra de 35 piezas. De esta muestra, un intervalo de 95% podría construirse para el tamaño promedio. El intervalo tendría un error no superior a 2 mm.

Ejemplo 7.4

El propietario de un centro de esquí en el sur de Winsconsin está considerando comprar una máquina para hacer nieve y ayudarle a la Madre Naturaleza a proporcionar una base apropiada para los entusiastas esquiadores. Si el promedio de nevadas parece insuficiente, piensa que la máquina debería pagarse muy pronto por sí misma. Planea estimar las pulgadas promedio de nieve que cae en el área, pero no tiene idea qué tan grande debería ser la muestra. Sólo sabe que desea un 99% de confianza en sus hallazgos y que el error no debe exceder de 1 pulgada. El propietario le promete tiquetes gratuitos de temporada. ¿Usted puede ayudarlo?

Solución

Usted comienza con una muestra piloto grande ($n \geq 30$) que produce una desviación estándar de 3.5 pulgadas. Por tanto,

$$\begin{aligned} n &= \frac{Z^2(s)^2}{(\text{error})^2} \\ &= \frac{(2.58)^2(3.5)^2}{(1)^2} \\ &= 81.5 \end{aligned}$$

u 82 nevadas durante los últimos años.

Interpretación

Ahora puede recolectar los datos sobre las últimas 82 nevadas que se utilizarán para estimar las nevadas promedio. Con esta información el propietario puede determinar si la Madre Naturaleza necesita ayuda. Lo más importante, usted puede pasar el resto del invierno esquiando gratuitamente.

B. Tamaño de la muestra para estimar π

En el capítulo 6 se encontró que

$$Z = \frac{p - \pi}{\sigma_p}$$

en donde

$$\sigma_p = \sqrt{\frac{\pi(1 - \pi)}{n}}$$

Se puede reescribir esto para producir una expresión para el tamaño muestral

Tamaño muestral para intervalos para la proporción poblacional	$n = \dfrac{Z^2(\pi)(1 - \pi)}{(p - \pi)^2}$	[7.10]

en donde $(p - \pi)$ es la diferencia entre la proporción muestral y la proporción poblacional, y por tanto es el error.

La fórmula (7.10) requiere el valor de π. Sin embargo, π es el parámetro que se desea estimar y es desconocido. Este problema puede tratarse en una de las dos maneras. Se podría tomar una muestra piloto para obtener un valor preliminar para π, tal y como se hizo al determinar el tamaño muestral apropiado para la media. O se puede determinar que $\pi = 0.5$, para efectos de determinar el tamaño muestral. Frecuentemente se prefiere este método porque es muy "seguro" o conservador – garantizará el tamaño muestral más grande posible, dado cualquier nivel de confianza y error deseados–. Esta muestra más grande resulta del hecho que el numerador de la fórmula (7.10), el cual contiene $\pi(1 - \pi)$ es máximo (por tanto, n se maximizará) cuando $\pi = 1 - \pi = 0.5$. No existe valor distinto a 0.5 que pueda asignarse a π que haga más grande $\pi(1 - \pi)$. Si $\pi = 0.5$, entonces $\pi(1 - \pi) = 0.25$. Todo valor distinto a 0.5 resultaría en $\pi(1 - \pi) < 0.25$. Por tanto, n sería más pequeño.

Wally Simpleton está postulado para gobernador. Él desea estimar dentro de 1 punto porcentual la proporción de personas que votarán por él. También desea tener el 95% de confianza en sus hallazgos. ¿Qué tan grande debería ser el tamaño muestral?

$$\begin{aligned} n &= \frac{(1.96)^2(0.5)(0.5)}{(0.01)^2} \\ &= 9{,}604 \text{ votantes} \end{aligned}$$

Una muestra de 9,604 votantes permitirá a Wally estimar π con un error de un 1% y un nivel de confianza del 95%.

Ejemplo 7.5

El consejo de la ciudad está planeando una ley que prohíbe fumar en edificios públicos incluyendo restaurantes, tabernas, y teatros. Sólo estará exenta la vivienda privada. Sin embargo, antes que dicha ley se lleve ante el consejo, este organismo desea estimar la proporción de residentes quienes apoyan dicho plan. La carencia de toda habilidad estadística obliga al consejo a contratarlo como consultor. Su primer paso será determinar el tamaño muestral necesario. Se le dice que su error no debe exceder del 2% y usted debe estar 95% seguro de sus resultados.

Solución

Debido a que no se tomó previamente una encuesta piloto, usted debe determinar temporalmente π en 0.5 para efectos de resolver el tamaño muestral.

$$
\begin{aligned}
n &= \frac{Z^2\pi(1-\pi)}{(\text{error})^2} \\
&= \frac{(1.96)^2(0.5)(0.5)}{(.02)^2} \\
&= 2{,}401 \text{ ciudadanos}
\end{aligned}
$$

Interpretación

Con los datos suministrados por las 2,401 personas usted puede proceder con su estimación de la proporción de todos los residentes quienes están a favor de la ley. El consejo puede tomar su determinación respecto a la política que se aplicará en toda la ciudad sobre el hecho de fumar.

Ejercicios de la sección

26. Days Inn desea desarrollar un intervalo de confianza del 99% para estimar el número promedio de habitaciones ocupadas cada noche en sus localidades de toda la nación. ¿Cuántas noches deben incluirse en la muestra si se puede tolerar un error de 50 habitaciones y una muestra piloto revela que $s = 165$ habitaciones?

27. ¿Qué pasaría a su respuesta si $s = 265$? ¿Por qué?

28. ¿Qué tan grande se requiere que sea una muestra para que proporcione una estimación del 90% del número promedio de graduados de las universidades de la nación con un error de 2,000 estudiantes si una muestra piloto reporta que $s = 8{,}659$?

29. Un estudio que usted está realizando requiere un intervalo del 95% para la tasa de rendimiento promedio que su empresa gana sobre los proyectos para presupuestar capital. ¿Cuántos proyectos debe tener su muestra si su supervisor especifica un error máximo de sólo el 5% y $s = 2.3\%$?

30. Como empleado recién contratado en la división de mercadeo para un importante asunto sobre ventas minoristas, a usted se le ha asignado la tarea de estimar la proporción de consumidores que prefieren su producto al de la competencia. ¿Cuántos consumidores se deben tomar en la muestra si usted desea restringir el error al 10%, pero sin embargo desea proporcionar un nivel de confianza del 99%?

31. ¿Qué tan grande debe ser la muestra del problema anterior si el error se restringe al 5%? Explique la diferencia.

32. La división de créditos de un banco comercial grande desea estimar con un nivel de confianza del 99% la proporción de sus créditos que están en mora. Si el ancho del intervalo es del 7%, ¿cuántos créditos deben revisarse? ¿Cuál es el error tolerable?

7.7 Propiedades de un buen estimador

Debe hacerse una distinción entre un estimador y una estimación. Un **estimador** es la regla o procedimiento, generalmente expresado como una fórmula, que se utiliza para derivar la **estimación**. Por ejemplo,

$$\bar{X} = \frac{\Sigma X_i}{n}$$

es el estimador para la media poblacional. Si el valor del estimador $\bar{X}$ es digamos 10, entonces 10 es la estimación de la media poblacional.

Estimadores y estimaciones Un estimador es el proceso mediante el cual se obtiene la estimación. Una estimación es el resultado numérico del estimador.

Para desempeñarse de manera confiable, los estimadores deben ser (1) insesgados, (2) eficientes, (3) consistentes, y (4) suficientes. Cada propiedad se discute a su tiempo en esta sección.

A. Estimador insesgado

Como se observó en el capítulo 6, es posible construir una distribución muestral seleccionado todas las muestras posibles de un tamaño dado de una población. Un estimador es insesgado si la media del estadístico calculado en todas estas muestras es igual al parámetro correspondiente.

Sea θ (la letra griega teta) el parámetro que se intenta estimar mediante $\hat{\theta}$ (que se lee "teta sombrero"). Entonces $\hat{\theta}$ es un estimador insesgado si su media, o valor esperado, E ($\hat{\theta}$), es igual a θ. Es decir,

$$E(\hat{\theta}) = \theta$$

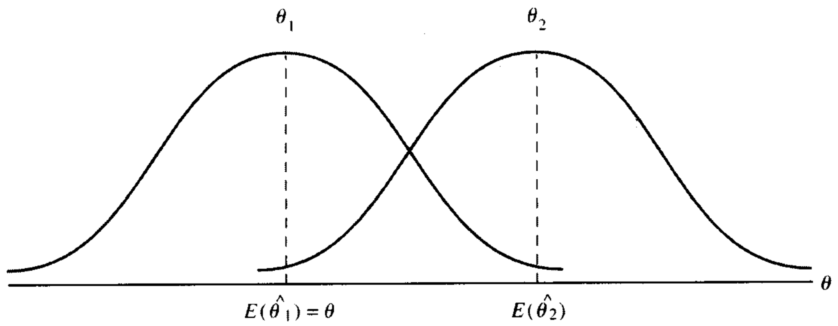
Para citar un ejemplo específico, $\bar{X}$ es un estimador insesgado de μ porque la media de la distribución muestral de las medias muestrales, $\bar{\bar{X}}$ es igual a μ. Por tanto,

$$E(\bar{X}) = \bar{\bar{X}} = \mu$$

Estimador insesgado Un estimador es insesgado si la media de su distribución muestral es igual al parámetro correspondiente.

La figura 7.6 ilustra cómo la media de una distribución muestral debe ser igual al parámetro correspondiente para garantizar un estimador insesgado.

Figura 7.6
Distribuciones para estimadores sesgados e insesgados



Aquí $\hat{\theta}_1$ es un estimador insesgado de θ debido a que su distribución está centrada en θ. Por tanto, E($\hat{\theta}_1$) = θ. Si se tomaran muchas muestras diferentes, produciendo muchos valores diferentes para $\hat{\theta}_1$, su media sería igual a θ. Y viceversa, si se toman muchas muestras y se calculara $\hat{\theta}_2$ cada vez, su media excedería θ. Por tanto, $\hat{\theta}_2$ es un estimador sesgado (hacia arriba) de θ. La medida de sesgo es la diferencia entre la media de $\hat{\theta}_2$ y θ. Vale la pena notar que

$$E(\hat{\theta}_1) - \theta = 0$$

mientras que

$$E(\hat{\theta}_2) - \theta \neq 0$$

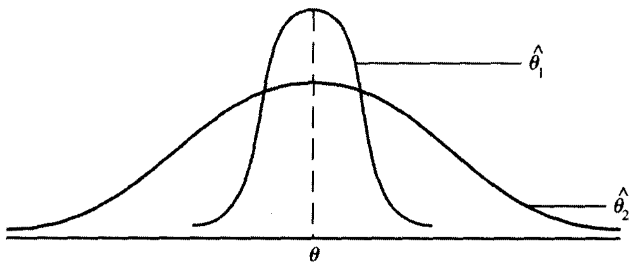
B. Estimador eficiente

La eficiencia de un estimador depende de su varianza. Sea $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de θ. Entonces $\hat{\theta}_1$ es un estimador más eficiente si, en muestreos repetidos con un tamaño muestral dado, su varianza es menor que el de $\hat{\theta}_2$. Es lógico que un estimador con una varianza más pequeña estimará de forma más próxima el parámetro. Considerando la figura 7.7, la cual muestra las distribuciones muestrales con un tamaño muestral dado de dos estadísticos $\hat{\theta}_1$ y $\hat{\theta}_2$. Tanto $\hat{\theta}_1$ como $\hat{\theta}_2$ son estimadores insesgados de θ debido a que sus distribuciones muestrales están centradas en θ, y

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$$

Sin embargo, la varianza de la distribución muestral de $\hat{\theta}_1$ es menor que la de $\hat{\theta}_2$. Los valores posibles para $\hat{\theta}_2$ están más dispersos. Cualquier estimado de θ utilizando $\hat{\theta}_2$ es probable que produzca un error de muestreo más grande que un estimado de θ que utiliza $\hat{\theta}_1$.

Figura 7.7
La varianza de los estimadores



> **Un estimador eficiente** Dado todo estimador insesgado, el estimador más eficiente es aquel que tenga la varianza más pequeña.

C. Estimador consistente

Un estimador es consistente cuando, a medida que n se incrementa, el estimador se aproxima al valor del parámetro.

> **Estimador consistente** Un estimador es consistente si, a medida que n aumenta, el valor del estadístico se aproxima al parámetro.

Para que un estimado sea consistente, debe ser insesgado y su varianza debe aproximarse a cero a medida que n aumenta. La varianza de la distribución muestral de las medias muestrales es $\sigma^2_{\bar{x}}$, es σ^2/n. A medida que n aumenta, $\sigma^2_{\bar{x}}$ se aproximará a cero. Por tanto, se puede decir que $\bar{X}$ es un estimador consistente de μ.

Si un estadístico no es un estimador consistente, tomar una muestra más grande para mejorar su estimado será infructuosa.

D. Estimador suficiente

Un estimador es suficiente si utiliza toda la información relevante sobre el parámetro contenida en la muestra. Si un estimador es suficiente, nada puede ganarse utilizando cualquier otro estimador.

Estimador suficiente Un estimador es suficiente si ningún otro estimador puede proporcionar más información sobre el parámetro.

Esta discusión de las propiedades de un estimador, por ningún motivo es una cuenta completa. Sin embargo, proporciona una base suficiente para una evaluación sobre la estimación de parámetros mediante la construcción de intervalos de confianza.

Problemas resueltos

Artesian Spring Water proporciona agua embotellada, en contenedores de 15 galones, a las casas de un sector de tres condados. El gerente desea estimar el número promedio de contenedores que una casa típica utiliza cada mes. Se toma una muestra de 75 casas y se registra el número de contenedores. La media es $\bar{X} = 3.2$, con $s = 0.78$.

a. ¿Qué revelaría un intervalo de confianza del 92%?

Solución

El valor de Z para un intervalo de confianza del 92% se halla dividiendo primero 0.92 por 2: así 0.92/2 = 0.46. El área de 0.46 requiere un valor de Z de 1.75. entonces,

$$\begin{aligned}\text{I.C. para estimar } \mu &= \bar{X} \pm (Z)s_{\bar{x}} \\ &= 3.2 \pm (1.75)\frac{0.78}{\sqrt{75}} \\ 3.04 \leq \mu &\leq 3.36\end{aligned}$$

El gerente puede estar 92% seguro de que el número de contenedores promedio por mes está entre 3.04 y 3.36.

b. Sin embargo, el gerente siente que este intervalo es demasiado amplio. ¿Cuántas casas deben tomar como muestra para estar 99% seguro de que el intervalo no está errado en más de 0.10 contenedores?

Solución

$$n = \frac{Z^2 s^2}{(\text{error})^2} = \frac{(2.58)^2(0.78)^2}{(0.1)^2} = 405$$

Una muestra de 405 daría un intervalo de confianza del 99%, con un error no superior a 0.10 contenedores.

c. Se selecciona una muestra más pequeña de 10 casas para estimar el número promedio de miembros de la familia por casa. Los resultados son 1, 3, 4, 7, 2, 2, 3, 5, 6 y 6 personas en cada casa. ¿Cuáles son los resultados de un intervalo de 99% para el número promedio de miembros de la familia?

Solución

La desviación estándar de la muestra es $s = 2.02$ con una media de 3.9. Dada la muestra pequeña, debe utilizarse un valor t de $t_{0.01,9} = 3.250$. Por tanto,

$$\begin{aligned} \text{I.C. para estimar } \mu &= \bar{X} \pm ts_{\bar{x}} \\ &= 3.9 \pm (3.250)\frac{2.02}{\sqrt{10}} \\ 1.82 \leq \mu &\leq 5.98 \end{aligned}$$

d. De las 75 casas de la muestra, 22 tienen ablandadores de agua en casa. ¿Cuál es el estimado del intervalo del 95% de la proporción de todas las casas en el sector de tres condados que tiene ablandadores?

Solución

Debido a que la proporción muestral es $p = 22/75 = 0.29$, el error estándar es

$$s_p = \sqrt{\frac{(p)(1-p)}{n}} = \sqrt{\frac{(0.29)(0.71)}{75}} = 0.052$$

Entonces,

$$\begin{aligned} \text{I.C. para estimar } \pi &= p \pm Zs_p \\ &= 0.29 \pm (1.96)(0.052) \\ 0.188 \leq \pi &\leq 0.392 \end{aligned}$$

e. Si un intervalo oscila entre el 18.8% y el 39.2% de todas las casas que tienen ablandadores y carece de precisión, ¿qué tan grande debe tomarse una muestra para producir un intervalo de sólo el 10%?

Solución

Si el intervalo de confianza ha de ser del 10%, el error puede ser sólo del 5%. Entonces,

$$n = \frac{Z^2(0.5)(0.5)}{(\text{error})^2} = \frac{(1.96)^2(0.5)(0.5)}{(0.05)^2} = 385 \text{ casas}$$

Lista de fórmulas

[7.1]	$\text{I.C. para estimar } \mu = \bar{X} \pm Z\sigma_{\bar{x}}$	Intervalo de confianza para estimar μ cuando σ es conocida
[7.2]	$\text{I.C. para estimar } \mu = \bar{X} \pm Zs_{\bar{x}}$	Intervalo de confianza para estimar μ cuando σ es desconocida
[7.3]	$\sigma^2 = \dfrac{n-1}{n-3}$	Varianza para la distribución t
[7.5]	$\text{I.C. para estimar } \mu = \bar{X} \pm ts_{\bar{x}}$	Intervalo de confianza para muestras pequeñas
[7.7]	$s_p = \sqrt{\dfrac{p(1-p)}{n}}$	Estimado del error estándar de la distribución de las proporciones muestrales
[7.8]	$\text{I.C. para estimar } \pi = p \pm Zs_p$	Intervalo de confianza para la proporción poblacional

[7.9] $$n = \frac{Z^2\sigma^2}{(\bar{X} - \mu)^2}$$ Tamaño muestral para estimar μ

[7.10] $$n = \frac{Z^2(\pi)(1 - \pi)}{(p - \pi)^2}$$ Tamaño muestral para estimar π

Ejercicios del capítulo

33. Se construye un intervalo del 95% de confianza, que da un límite inferior de confianza de 62 y un límite superior de confianza de 69. ¿Usted puede concluir de esto que existe un 95% de probabilidad que el parámetro esté entre 62 y 69? Explique.

34. En una encuesta de 500 gerentes, Posner Products encontró que 200 gerentes carecían de suficiente capacitación estadística.

a. ¿Cuál es el estimado puntual de la proporción de todos los gerentes de Posner que requieren trabajo adicional en análisis estadístico?

b. ¿Cuál es el estimado del error estándar de la proporción?

c. ¿Cuál es el intervalo de confianza del 90% para la proporción poblacional? Interprete su respuesta.

35. José tiene un negocio próspero en Acapulco que vende restos incas de plástico a los turistas norteamericanos. Él selecciona $n = 60$ días para estimar su utilidad diaria. Sin embargo, José no sabe si la población de utilidades diarias está distribuida normalmente, y no está seguro sobre cómo proceder. ¿Que debe hacer?

36. Como experto en control de calidad usted desea estimar el grosor promedio de los lentes ópticos producidos por su empresa. Una muestra de 120 lentes revela una media de 0.52 milímetros (mm). La desviación estándar poblacional es de 0.17 mm. Usted considera que puede arriesgarse a una probabilidad de error de sólo el 1%. Construya el intervalo de confianza apropiado.

37. ¿Cómo cambiaría el problema anterior si σ fuera desconocida y la desviación estándar de la muestra fuera de 0.17 mm? Calcule el intervalo.

38. Georgia Pacific (GP), una empresa papelera de Estados Unidos, decide sembrar un bosque maderero si puede obtener un promedio de por lo menos 700 pies cuadrados de tabla (b.f). Una muestra de 1,000 árboles da una media de 695 b.f, con una desviación estándar de 22.1 b.f.

a. Calcule el intervalo de confianza del 90%.

b. Interprete su respuesta.

c. ¿Debería GP sembrar el bosque?

39. En una encuesta a 6,000 personas realizada por *U.S. News & World Report* se encontró que, en su tiempo de vida, el norteamericano promedio gasta seis meses sentado en los semáforos. Tomando esto como la media muestral, y asumiendo que la desviación estándar es de 2.2 meses, ¿cuál es el intervalo de confianza del 90% para la media poblacional? Interprete su respuesta.

40. *The Journal Retail Management* reportó que una muestra de 600 compradores pasaban un promedio de 1.79 horas en un centro comercial de visita. La desviación estándar era de 0.83 hora. ¿Cuál es la estimación por intervalo del número promedio de horas que todos los compradores pasan en el centro comercial? Sea $\alpha = 0.10$.

41. Su producto requiere que un cierto componente utilizado en su fabricación promedie 15.2 gramos. Si usted compra 100 componentes y encuentra que $\bar{X} = 14.8$ gramos, con $s = 3.2$ gramos, ¿qué le diría un intervalo de confianza sobre lo aconsejable de comprarle más a este proveedor? Su producto es muy delicado y usted siente que puede tolerar sólo el 1% de probabilidad de error.

42. Si dadas las condiciones del problema anterior, la muestra hubiera dado una media de 14.1 gramos, ¿qué concluiría?

43. Wally desea comprarle a su esposa para su cumpleaños un nuevo tanque séptico. Siendo un comprador cuidadoso, él examina 40 modelos diferentes y encuentra un precio promedio de US$712, con una desviación estándar de US$215. ¿Cuál es el intervalo de confianza del 95% para estimar el precio promedio de todos los tanques sépticos?

44. Un fabricante de esquís de nieve desea estimar el número promedio de viajes para esquiar que realizan los esquiadores ávidos. Una muestra de 1,100 esquiadores da $\bar{X} = 15.3$ viajes por temporada, con $s = 5.1$ viajes. ¿Cuál es el intervalo de confianza del 99% para estimar la media poblacional?

45 Considere los datos del ejercicio anterior:

a. Sin resolver el problema, explique qué pasaría al intervalo si el nivel de confianza se redujera al 90%.

b. Solucione el problema con $\alpha = 0.10$ y demuestre cómo la respuesta sustenta su respuesta a la parte *a.*

46. Un investigador descubrió que una muestra de 100, con $\bar{X} = 50.3$ y $s = 10.1$, generó un intervalo de confianza de 48,3204 a 52.2796. ¿Qué nivel de confianza puede atribuirse a este intervalo?

47. Los pesos de 25 paquetes enviados a través de United Parcel Service (UPS) tuvieron una media de 3.7 libras y una desviación estándar de 1.2 libras. Halle el intervalo de confianza del 95% para estimar el peso promedio de todos los paquetes. Los pesos de los paquetes están distribuidos normalmente.

48. Se registró una muestra de 12 donaciones por parte de comités de acción política (PAC's) para los fondos de las campañas del congreso, en miles de dólares, de 12.1, 8.3, 15.7, 9.35, 14.3, 12.9, 13.2, 9.73, 16.9, 15.5, 14.3 y 12.8. Calcule el intervalo de confianza del 90% para estimar la donación promedio realizada por PAC's. Se asume que las donaciones son normales.

49. Las ganancias por acción para 10 acciones industriales cotizadas en el Dow Jones fueron US$1.90, US$2.15, US$2.01, US$0.89, US$1.53, US$1.89, US$2.12, US$2.05, US$1.75 y US$2.22. Calcule un intervalo de confianza del 99% de los EPS de todas las acciones industriales cotizadas en el Indice Dow Jones. ¿Qué suposición debe hacer usted?

50. El Dr. Bottoms, el proctólogo local, descubrió que la edad promedio de 75 de sus pacientes era de 47.3 con una desviación estándar de 10.9 años. Calcule el intervalo de confianza del 99% para la edad promedio de todos sus pacientes bajo la suposición de que las edades no están distribuidas normalmente.

51. Durante el último Superbowl Sunday (Domingo de Fútbol) Sammy Salami y sus compañeros ordenaron 27 pizzas de Pizzas On Wheels. El tiempo promedio de entrega era de 23.7 minutos, con una desviación estándar de 10.7 minutos. Al considerar que ésta era una demora muy grande en su propósito culinario, Sammy y sus amigos decidieron comprar la pizza número 28 en otra parte, si el tiempo de entrega de POW (Pizzas On Wheels) era superior a 30 minutos. Sea α 1%. ¿Ordenarán en otra parte?

52. Una gran empresa de contabilidad contrató un psicólogo industrial para medir la satisfacción laboral de sus socios más antiguos. A 17 socios se les practicó una prueba para medir la satisfacción; los puntajes de las pruebas son normales y la varianza para todos sus socios es de 120. ¿Cuál es el intervalo de confianza del 90% para el puntaje promedio?

53. Para estimar cuántos de los 350,000 residentes de Peoria desean un nuevo centro cívico, el alcalde descubre que 1,570 de los 2,100 residentes a quienes se tomó como muestra expresaron su apoyo. Ayude al alcalde a construir e interpretar un intervalo de confianza del 90%.

54. De 209 clientes, 183 expresaron su satisfacción con los servicios bancarios ofrecidos por First of America en Peoria, Illinois. ¿Cómo se compara esto con los resultados de un estudio anterior por parte de River Valley Savings, el cual estimó al 99% del nivel de confianza que entre el 74.1 y el 83.7% de sus clientes estaban satisfechos?

55. Una empresa textil descubrió que soporta un factor de desperdicio total del 19%. Un nuevo proceso se implementó y una muestra de 1.376 unidades reveló un factor promedio de desperdicio de 11.1%, con una desviación estándar del 1.12%. ¿Qué conclusión puede sacarse con un intervalo de confianza del 95% acerca del beneficio del nuevo proceso?

56. A los golfistas profesionales se les pidió calificar un nuevo palo de grafito en una escala de 0 a 10. Veinticinco pros generaron un rango promedio de 7,3, con $s = 1.01$. Construya e interprete el intervalo de confianza del 90%.

57. El Premio Nacional de Calidad Baldrige, llamado así en honor a Malcolm Baldrige, quien sirvió como secretario de comercio a finales de los años 80's, es un reconocimiento ampliamente codiciado del compromiso de una corporación con un sistema de gerencia de calidad total (TQM). Se especifican siete criterios mediante los cuales se juzgan las empresas. Uno de tales criterios, el liderazgo, permite a las organizaciones sumar hasta 100 puntos para lograr este honor. Chrysler Corporation tomó muestras en 19 plantas y descubrió un promedio de 73.2 puntos, con $s = 10.1$. Construya e interprete el intervalo de confianza del 99% para todas las plantas de Chrysler.

58. Pizza Pub está considerando incrementar el precio de su pizza grande de encurtidos si el precio promedio de su competencia excede el precio de Pizza Pub de US$12.95. Otras 37 pizzerías reportan un precio promedio de US$12.50, con una desviación estándar de US$1.01. Pizza Pub desea estar un 90% segura de sus hallazgos.

59. En una encuesta de 673 tiendas minoristas, 521 reportaron un problema de robos por parte de los empleados. Puede usted concluir con un 99% de confianza, que estos datos sugieren que el 78% de todas las tiendas tiene una dificultad similar, como presentó un informe recientemente CNN?

60. Work and Roll, un restaurante de comida china para llevar, deseaba determinar qué porcentaje de sus clientes seleccionan la sopa de nido de aves como parte de su cena placentera.

   a. En una muestra de 320 clientes, 220 se llevaron a casa esta delicia. Calcule e interprete un intervalo del 99%.

   b. Utilizando los datos anteriores, construya el intervalo del 90%.

   c. ¿Por qué obtuvo un intervalo más pequeño? ¿Sería siempre deseable reducir el ancho de intervalo de esta manera?

61. The Jesse James First National Bank lo contrató como consultor estadístico para analizar las operaciones de sus cajeros automáticos. Una muestra de 15 mostró las transacciones promedio de US$4,810, con una desviación estándar de US$1,202 por día. Su supervisor insiste en que usted debe tener un 99% de confianza en su estimado del volumen diario promedio. ¿Qué puede decirle?

62. El propietario de un negocio pequeño desea estimar el tiempo promedio requerido para finalizar cierto trabajo. Él debe asegurar que tiene un 90% de confianza en que el error sea menor que 0.5 minutos. La desviación estándar es de 3.2 minutos. ¿Cuántas observaciones de tiempos de finalización debe hacer?

63. El decano de una universidad privada desea estimar el número de estudiantes registrados que vienen de otro estado. Debe estar un 95% seguro de que el error es menos del 3%. ¿Qué tan grande debe tomar la muestra? Si la muestra revela una proporción del 31% de estudiantes de otro estado, y hay 12,414 estudiantes, ¿cuántos estudiantes estima usted que provienen de otros estados?

64. Al director de una sucursal de banco se le pide estimar el tiempo promedio que un cliente gasta en las instalaciones del auto-banco. Debe estar un 99% seguro de que el estimado no tenga un error superior a 15 segundos. ¿Cuántas observaciones debe hacer si la desviación estándar conocida es de 2.7 minutos?

65. En un esfuerzo por reducir el comercio de iniciado, la Comisión de Bolsas y Valores (*Securities and Exchange Commission*- SEC) solicitó información respecto a la proporción de bancos de empresas cuyos funcionarios poseen más del 50% del capital en circulación de sus bancos. De las 200 compañías seleccionadas aleatoriamente, 79 reportaron que sus funcionarios tenían una mayoría de sus acciones. ¿Cuál es el intervalo de confianza del 90% para la proporción de todos los bancos de empresas cuyos funcionarios tienen por lo menos un 50% de las acciones?

66. Un investigador de la Administración Federal de Aviación (*Federal Aviation Administration*, FAA) fue mencionado en una emisión de febrero de *The Washington Post*, en la que dijo que de los 112 accidentes de aerolíneas, "73 involucraban algún tipo de problema estructural con la aeronave". Si tales cifras son representativas, ¿cuál es el intervalo de confianza para la proporción de accidentes que involucran tal defecto estructural? Sea $\alpha = 0.01$.

67. United Airlines hizo encuestas a 93 pasajeros en un vuelo de Cincinnati a Atlanta. Sesenta y cuatro dijeron que les hubiera gustado ir en un vuelo posterior si hubiese habido espacio disponible. United había decidido que si más del 50% de los pasajeros expresaba su interés en salidas en horas siguientes durante el día, consideraría la posibilidad de poner tales vuelos a disposición. Dados los resultados de la encuesta, ¿un intervalo de confianza del 90% sugiere que deben hacerlo?

68. *The Wall Street Journal* informó los esfuerzos realizados por Nestlé, la compañía de alimentos más grande del mundo, para introducir un nuevo producto. La gerencia decidió utilizar el sector de Chicago como mercado de prueba. Si más del 30% de las personas expresaban un deseo por el producto, ellos considerarían el comercializarlo en un sector más amplio. De las 820 personas a quienes se les practicó la prueba, 215 expresaron una reacción positiva. ¿Un intervalo de confianza del 90% para la proporción de todos los clientes quienes prefieren el producto hace que la gerencia continúe con sus planes de comercialización?

69. *Business Week* presentó uña historia sobre los esfuerzos realizados por los 12 países miembros del Mercado Común para abreviar la creciente ola de fusiones que se pensaba era "indeseable económicamente para los intereses internacionales". Se va a seleccionar una muestra para estimar el tamaño promedio de las empresas (tal y como se mide en valor neto corporativo) involucradas en las fusiones. Si el intervalo debe ser de US$5.2 millones y tener un nivel de confianza del 95%, ¿qué tan grande debería ser la muestra si la desviación estándar del valor neto corporativo se considera que es US$21.7 millones?

70. El director de su división le pide, como analista de mercadeo recientemente contratado, que estime las ventas semanales promedio. Le advierte que debe mantener el error dentro de US$100 y mantener un nivel de confianza del 90%. ¿Cuántas semanas de datos debe recolectar si la desviación estándar es US$750?

71. Una encuesta sobre la violencia en los colegios está diseñada para estimar el porcentaje de estudiantes de sexo masculino que fueron amenazados con violencia en los campos del colegio durante el año pasado. El error tolerable se fija en 1%, y el nivel de confianza es del 99%. ¿Cuál es el tamaño apropiado de la muestra?

72. La Asociación de Finanzas Estudiantiles en Faber College está planeando una "Feria Primaveral" en la cual intentan vender camisetas impresas con su logo. El tesorero desea un estimado de la proporción de estudiantes que comprarán una camiseta. El estimado debe proporcionar un nivel de confianza del 90% y el error no debe exceder del 3%. ¿Qué tan grande debe tomarse la muestra?

73. Si un fabricante desea desarrollar un intervalo del 99% para la proporción de defectos con un error de menos del 1%, ¿qué tan grande se necesita la muestra?

Ejercicios por computador

Telmark Technologies planteó recientemente nuevas metas para que la compañía incrementara el tamaño promedio de ventas por cliente y mejorara el servicio ofrecido a los mismos. En el último trimestre las ventas presentaron un promedio de US$373.10 por pedido del cliente y tomó un promedio de 4.5 días para hacer el envío a los clientes. Además, los ejecutivos de alto nivel en Telmark desean reducir el peso promedio de los envíos por debajo de la media de 55.8 libras para reducir los costos de envío. Hal Ketchum, vicepresidente para Relaciones del Producto, ha sido encargado de la responsabilidad de lograr tales objetivos. Él dirige su *staff* de estadísticos a recolectar todos los datos necesarios y a preparar un informe detallando las condiciones actuales.

Ingrese al archivo TELMARK de su disco de datos. Contiene los datos de los pedidos de ventas de 75 clientes. El tamaño de las ventas en cientos de dólares está registrado como Size (tamaño). El número de días que tomó cada pedido en ser enviado al cliente se registra como Days (días) y la variable Weight (peso) proporciona datos de los pesos de los envíos en libras. Finalmente, la variable SAT indica si el cliente quedó satisfecho con el servicio de Telmark. Está codificado como "1" si el cliente quedó satisfecho y "0" si no quedó satisfecho.

Utilizando los datos del archivo TELMARK, proporcione los estimados de intervalo para Size, Days y Weight de los envíos. ¿A qué conclusión llega? Calcule el estimado del intervalo de la proporción de clientes satisfechos. Presente su informe estadístico terminado tal y como se describe en el apéndice. Incluya todos los hallazgos relevantes, las conclusiones y las recomendaciones.

PUESTA EN ESCENA

Durante muchos años el FBI (Federal Bureau of Investigation) ha servido como modelo para las agencias que hacen cumplir la ley. Los esfuerzos revolucionarios por parte del FBI por integrar de manera más completa el uso del análisis estadístico en los esfuerzos por combatir el crimen se mencionaron en la sección Escenario a comienzos de este capítulo.

El método para combatir el crimen basado en la estadística incluye datos sobre una gran variedad de crímenes, así como las características y los hábitos personales de los infractores. Aunque la vocera de Larry King Live no ofreció detalles específicos, estableció que se guardaban los datos sobre el número de crímenes que cada transgresor de la ley comete, el número de días que transcurren entre crímenes, y el número de personas que son asesinadas por el organismo ejecutor de la ley en el intento de arresto.

Se asume que se recolectan datos para una muestra de 1,000 personas que han cometido un delito grave. Los datos revelan que, en promedio, los criminales comenten 12.4 actos ilegales con una desviación estándar de 4.7 antes de que sean arrestados o asesinados. Además, los datos reportan que un promedio de 19.6 días transcurren entre actos criminales, con una desviación estándar de 11.5 días. Además, 212 de los 1,000 criminales que aparecen en la base perdieron sus vidas intentando eludir a la policía.

Se asume además que se toma una muestra pequeña de un tipo específico de criminales. Quince observaciones de robo a mano armada reportan las siguientes cantidades, en miles de dólares, que se tomaron en los robos: 45.6, 23.8, 45.3, 27.9, 54.3, 27.5, 63.4, 12.6, 75.9, 64.6, 54.7, 17.5, 21.4, 56.4, y 34.9.

Con base en estos datos y las herramientas aprendidas en este capítulo, proporcione un perfil de la actividad criminal a lo largo de la nación, tal como la que usted considera que un estadístico del FBI puede preparar. Seleccione y justifique sus propios niveles de confianza y calcule e interprete todos los intervalos de confianza. Haga comentarios sobre los tamaños muestrales. ¿Son suficientes para lograr su objetivo? Especifique claramente qué tamaños muestrales son necesarios para lograr el nivel de confianza y el error tolerable que usted requiere. Presente el informe final en forma de un reporte de negocios tal y como se describe en el apéndice I.

Del escenario a la vida real

El sitio web del FBI (*www.fbi.gov*) proporciona estadísticas sobre crímenes seleccionados, similares a las utilizadas en el análisis de la sección *Puesta en Escena*. De la página de presentación (home page) del FBI, haga clic en "Publications". Bajo el encabezado de Estadísticas de Crímenes Uniformes, haga clic en "Final 19xx Crime Statistics" (datos completos de los años más recientes). Haga una lista de las categorías de estadísticas sobre crímenes que estén disponibles en este sitio.

Mire la información sobre robo con allanamiento de morada y sobre asaltos agravados. Note tanto el conteo de crímenes como la tasa por cada 100,000 para estas dos categorías. ¿Qué información se da sobre el incremento o reducción en el tipo de crimen sobre los datos del año anterior? ¿Estos datos son una muestra o un censo? ¿Los datos disponibles proporcionan suficientes detalles para evaluar si el cambio proveniente del año anterior es estadísticamente significativo? Si no, ¿qué otra información adicional necesita para construir dicha prueba?