

CAPÍTULO 7



Muestreo y distribuciones muestrales

CONTENIDO

LA ESTADÍSTICA EN LA PRÁCTICA: MEADWESTVACO CORPORATION

7.1 EL PROBLEMA DE MUESTREO DE ELECTRONICS ASSOCIATES

7.2 MUESTREO ALEATORIO SIMPLE

Muestreo de una población finita
Muestreo de una población infinita

7.3 ESTIMACIÓN PUNTUAL

7.4 INTRODUCCIÓN A LAS DISTRIBUCIONES MUESTRALES

7.5 DISTRIBUCIÓN MUESTRAL DE \bar{x}

Valor esperado de \bar{x}
Desviación estándar de \bar{x}
Forma de la distribución muestral de \bar{x}
Distribución muestral de \bar{x} en el problema EAI
Valor práctico de la distribución muestral de \bar{x}

Relación entre el tamaño de la muestra y la distribución muestral de \bar{x}

7.6 DISTRIBUCIÓN MUESTRAL DE \bar{p}

Valor esperado de \bar{p}
Desviación estándar de \bar{p}
Forma de la distribución muestral de \bar{p}
Valor práctico de la distribución muestral de \bar{p}

7.7 PROPIEDADES DE LOS ESTIMADORES PUNTUALES

Insesgadez
Eficiencia
Consistencia

7.8 OTROS MÉTODOS DE MUESTREO

Muestreo aleatorio estratificado
Muestreo por conglomerados
Muestreo sistemático
Muestreo de conveniencia
Muestreo subjetivo

LA ESTADÍSTICA *(en)* LA PRÁCTICA

MEADWESTVACO CORPORATION*

STAMFORD, CONNECTICUT

MeadWestvaco Corporation, líder mundial en la producción de embalajes y papeles especiales, productos de consumo y de oficina y de sustancias químicas especiales, emplea a más de 30 000 personas. Opera a nivel mundial en 29 países y atiende a clientes localizados en 100 países. MeadWestvaco tiene una posición líder en la producción de papel, con una capacidad de 1.8 millones de toneladas anuales. Entre los productos de la empresa se encuentran papel para libros de texto, papel para revistas, sistemas de embalaje para bebidas y productos de oficina. Los consultores internos de MeadWestvaco usan el muestreo para obtener diversas informaciones que permiten a la empresa ganar productividad y seguir siendo competitiva.

Por ejemplo, MeadWestvaco posee bosques que le proporcionan los árboles, o la materia prima, para muchos de los productos de la empresa. Los directivos necesitan información confiable y precisa acerca de los bosques maderables para evaluar las posibilidades de satisfacción de las futuras necesidades de materia prima. ¿Cuál es el volumen actual de los bosques? ¿Cuál ha sido el crecimiento de los bosques? ¿Cuál es el crecimiento proyectado de los bosques? Las respuestas a estas preguntas permiten a los directivos de la empresa elaborar los planes para el futuro, tales como planes a largo plazo y calendarios para la poda de árboles.

¿Cómo recolecta MeadWestvaco la información que necesita acerca de los amplios bosques que requiere? Los datos que obtiene de puntos muestrales en los bosques son la base para contar con información acerca de la población de árboles propiedad de la empresa. Para localizar estos puntos muestrales, primero se dividen los bosques en tres secciones de acuerdo con la localización y tipo de árboles. Mediante mapas y números aleatorios los analistas de MeadWestvaco identifican puntos muestrales aleatorios de 1/5 a 1/7 acres en cada sección del bosque. Los ingenie-



El muestreo aleatorio de sus bosques permite a MeadWestvaco satisfacer necesidades futuras de materia prima. © Walter Hodges/Corbis.

ros forestales de MeadWestvaco recogen los datos de estos puntos muestrales para obtener información acerca de la población forestal.

También participan en el proceso de campo de la recolección de datos. Con periodicidad, en equipos de dos personas, recolectan la información de cada árbol en todos los puntos muestrales. Los datos muestrales se ingresan en el sistema computacional de inventario forestal continuo (IFC) de la empresa. Los informes obtenidos del sistema IFC contienen información de distribuciones de frecuencia con estadísticos sobre los tipos de árboles, volumen de los bosques, tasas de crecimiento anteriores y crecimiento y volumen proyectados para el futuro. El muestreo y las correspondientes informaciones estadísticas de los datos muestrales proporcionan la información esencial para la adecuada administración de los bosques de MeadWestvaco.

En este capítulo se estudiará el muestreo aleatorio simple y el proceso de selección de muestras. Se verá también cómo se usan estadísticos como la media muestral y la proporción muestral para estimar la media de la población y la proporción de la población.

*Los autores agradecen al doctor Edgard P. Winkofsky por proporcionar la información para *La estadística en la práctica*.

En el capítulo 1 se definieron los términos población y muestra. Estas definiciones se retoman aquí.

1. Una *población* es el conjunto de todos los elementos que interesan en un estudio.
2. Una *muestra* es un subconjunto de la población.

A las características numéricas de una población, como la media y la desviación estándar, se les llama **parámetros**. El principal propósito de la inferencia estadística es hacer estimaciones y pruebas de hipótesis acerca de los parámetros poblacionales usando la información que propor-

ciona una muestra. Para empezar, se presentan dos situaciones en las que a partir de muestras se obtienen estimaciones de parámetros poblacionales.

1. Un fabricante de neumáticos elabora un nuevo modelo que tendrá mayor duración que los actuales neumáticos de la empresa. Para estimar la duración media, en millas, el fabricante selecciona una muestra de 120 neumáticos nuevos para probarlos. De los resultados de esta prueba se obtiene una duración media de 36 500 millas. Por tanto, una estimación de la duración media, en millas, de la población de nuevos neumáticos es 36 500 millas.
2. Los miembros de un partido político deseaban apoyar a un determinado candidato para senador, y los dirigentes del partido deseaban tener una estimación de la proporción de votantes registrados que podían estar a favor del candidato. El tiempo y el costo de preguntar a cada uno de los individuos de la población de votantes registrados eran prohibitivos. Por tanto, se seleccionó una muestra de 400 votantes registrados; 160 de los 400 votantes indicaron estar a favor del candidato. Una estimación de la proporción de la población de votantes registrados a favor del candidato es $160/400 = 0.40$.

Estos dos ejemplos ilustran algunas de las razones por las que se usan muestras. Observe que en el ejemplo de los neumáticos, obtener datos sobre su tiempo de duración implica usarlos hasta que se acaben. Es claro que no es posible probar toda la población de neumáticos; una muestra es la única manera factible de obtener los datos de duración deseados. En el ejemplo del candidato, preguntar a cada uno de los votantes registrados es, en teoría, posible, pero el tiempo y el costo para hacerlo son prohibitivos; de manera que se prefiere una muestra de los votantes registrados.

Es importante darse cuenta de que los resultados muestrales sólo proporcionan una *estimación* de los valores de las características de la población. No se espera que la media muestral de 36 500 millas sea exactamente igual al millaje medio de todos los neumáticos de la población, tampoco que 0.40, o 40% de la población de los votantes registrados esté a favor del candidato. La razón es simple, la muestra sólo contiene una parte de la población. Con métodos de muestreo adecuados, los resultados muestrales proporcionarán estimaciones “buenas” de los parámetros poblacionales. Pero ¿cuán buenos puede esperarse que sean los resultados muestrales? Por fortuna, existen procedimientos estadísticos para responder esta pregunta.

En este capítulo se enseña cómo emplear el muestreo aleatorio simple para seleccionar una muestra de una población. Después, cómo usar una muestra aleatoria simple para calcular estimaciones de una media poblacional, de una desviación estándar poblacional y de una proporción poblacional. Además, también se presenta el importante concepto de distribución muestral. Como verá, el conocimiento de la distribución muestral adecuada permite decir qué tan cerca se encuentran las estimaciones muestrales de los correspondientes parámetros poblacionales. En la última sección se estudian alternativas al muestreo aleatorio simple, usadas con frecuencia en la práctica.

7.1

El problema de muestreo de Electronics Associates

Al director de personal de Electronics Associates, Inc. (EAI), se le ha encargado la tarea de elaborar un perfil de los 2500 administradores de la empresa. Las características a determinar son el sueldo medio anual de los administradores y la proporción de administradores que ha terminado el programa de capacitación de la empresa.

Con los 2500 administradores de la empresa como la población para este estudio, es posible hallar el sueldo anual y la situación respecto al programa de capacitación de cada persona al consultar los archivos del personal. El archivo con los datos que contiene esta información para cada uno de los 2500 administradores que forman la población se encuentra en el disco compacto que se distribuye con el libro.

Con los datos de EAI y las fórmulas presentadas en el capítulo 3, se calcula la media poblacional y la desviación estándar poblacional de los salarios anuales.

$$\text{Media poblacional: } \mu = \$51\,800$$

$$\text{Desviación estándar poblacional: } \sigma = \$4000$$

Algunos de los costos de recopilar información de una muestra son sustancialmente menores que hacerlo de una población; especialmente cuando se deben realizar entrevistas personales para recopilar la información.

Los datos sobre la situación del programa de capacitación muestran que 1500 de los 2500 administradores han terminado el programa de capacitación. Si p denota la proporción de la población que ha terminado el programa de capacitación, se tiene que $p = 1500/2500 = 0.60$. La media poblacional de los sueldos anuales ($\mu = \$51\,800$), la desviación estándar poblacional de los sueldos anuales ($\sigma = \$4000$) y la proporción poblacional de quienes han terminado el programa de capacitación ($p = 0.60$) son parámetros de la población de administradores de EAI.

Ahora suponga que la información necesaria sobre todos los administradores de EAI no esté disponible en la base de datos de la empresa. La pregunta que se considera ahora es: ¿cómo puede obtener el director de personal de la empresa, estimaciones de los parámetros poblacionales usando una muestra de los administradores, en lugar de usar a los 2500 administradores de la población. Asuma que se va a emplear una muestra de 30 administradores. Es obvio que el tiempo y el costo de la elaboración de un perfil será mucho menor usando 30 administradores que la población entera. Si el director de personal tuviera la certeza de que una muestra de 30 administradores proporciona la información adecuada acerca de la población de los 2500 administradores, preferiría trabajar con una muestra que hacerlo con toda la población. Para explorar la posibilidad de usar una muestra para el estudio de EAI, primero se considerará cómo determinar una muestra de 30 administradores.

7.2

Muestreo aleatorio simple

Para seleccionar una muestra de una población hay diversos métodos; uno de los más comunes es el **muestreo aleatorio simple**. La definición de muestreo aleatorio simple y del proceso de seleccionar una muestra aleatoria simple dependen de si la población es *finita* o *infinita*. Como el problema de muestreo de EAI tiene una población finita de 2500 administradores, primero se considera el muestreo de una población finita.

Muestreo de una población finita

Una muestra aleatoria simple de tamaño n de una población finita de tamaño N se define como sigue.

MUESTREO ALEATORIO SIMPLE (POBLACIÓN FINITA)

Una muestra aleatoria simple de tamaño n de una población finita de tamaño N es una muestra seleccionada de manera que cada posible muestra de tamaño n tenga la misma probabilidad de ser seleccionada.

Un procedimiento para seleccionar una muestra aleatoria simple de una población finita es elegir los elementos para la muestra de uno en uno, de manera que, en cada paso, cada uno de los elementos que quedan en la población tenga la misma probabilidad de ser seleccionado. Al seleccionar n elementos de esta manera, será satisfecha la definición de muestra aleatoria simple seleccionada de una población finita.

Para seleccionar una muestra aleatoria simple de la población finita de administradores de EAI, primero se le asigna a cada administrador un número. Por ejemplo, se les asignan los números del 1 al 2500 en el orden en que aparecen sus nombres en el archivo de personal de EAI. A continuación se consulta la tabla de dígitos aleatorios que se muestran en la tabla 7.1. Al consultar el primer renglón de la tabla se da cuenta que cada dígito, 6, 3, 2, ... es un dígito aleatorio con la misma oportunidad de aparecer que cualquier otro. Como el número mayor en la lista de la población de administradores de EAI, 2500, tiene cuatro dígitos, se seleccionarán números aleatorios de la tabla en conjuntos o grupos de cuatro dígitos. Aun cuando para la selección de números aleatorios se puede empezar en cualquier lugar de la tabla y avanzar sistemáticamente en una de las cuatro direcciones, aquí se usará el primer renglón de la tabla 7.1 y se avanzará de izquierda a derecha. Los primeros 7 números aleatorios de cuatro dígitos son

6327 1599 8671 7445 1102 1514 1807

Los números aleatorios en la tabla aparecen en grupos de cinco para facilitar su lectura.

Los números aleatorios generados por computadora también sirven para realizar el proceso de selección de una muestra aleatoria. Excel proporciona una función para generar números aleatorios en sus hojas de cálculo.

Los números aleatorios en la tabla aparecen en grupos de cinco para facilitar su lectura.

TABLA 7.1 NÚMEROS ALEATORIOS

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15702
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782
55363	07449	34835	15290	76616	67191	12777	21861	68689	03263
69393	92785	49902	58447	42048	30378	87618	26933	40640	16281
13186	29431	88190	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337
84649	48968	75215	75498	49539	74240	03466	49292	36401	45525
63291	11618	12613	75055	43915	26488	41116	64531	56827	30825
70502	53225	03655	05915	37140	57051	48393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	67342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09226	64419	29457
10078	28073	85389	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	79607	52244	63303	10413	63839	74762	50289

Como los números de la tabla son aleatorios, estos números de cuatro dígitos son todos igualmente posibles. Ahora se pueden usar estos números aleatorios de cuatro dígitos para darle a cada uno de los administradores que constituyen la población la misma oportunidad de ser incluido en la muestra aleatoria. El primer número, 6327, es mayor que 2500. No corresponde a ninguno de los administradores numerados que forman la población y por tanto se descarta. El segundo número, 1599, está entre 1 y 2500. Por tanto, el primer administrador seleccionado para la muestra aleatoria es el administrador que tiene el número 1599 en la lista de los administradores de EAI. Siguiendo este proceso, se ignoran los números 8671 y 7445 antes de identificar a los administradores con los números 1102, 1514 y 1807 e incluirlos en la muestra aleatoria. Este proceso sigue hasta que se tiene la muestra aleatoria de 30 administrativos de EAI.

Al realizar este proceso para la selección de una muestra aleatoria simple, es posible que un número que ya haya sido usado se encuentre de nuevo en la tabla antes de completar la muestra de los 30 administradores de EAI. Como no se quiere seleccionar a un administrador más de una vez, cualquier número aleatorio que ya ha sido usado se ignora, porque el administrador correspondiente ya se ha incluido en la muestra. A este tipo de selección se le conoce como **muestreo sin reemplazo**. Cuando se selecciona una muestra en la que se acepten números aleatorios ya usados y los administradores correspondientes sean incluidos dos o más veces, se está **muestreando con reemplazo**. Muestrear con reemplazo es una forma válida de identificar una muestra aleatoria simple. Sin embargo, el muestreo sin reemplazo es el procedimiento de muestreo más usado. Cuando se habla de muestreo aleatorio simple, se asumirá que el muestreo es sin reemplazo.

Muestreo de una población infinita

En algunas situaciones la población o bien es infinita o tan grande que, para fines prácticos, se considera infinita. Por ejemplo, suponga que un restaurante de comida rápida desea obtener el

En la práctica, la población en estudio se considera infinita si se tiene un proceso continuo en el que sea imposible contar o enumerar cada uno de los elementos de la población.

perfil de su clientela seleccionando una muestra aleatoria de los mismos y pidiéndole a cada cliente que llene un breve cuestionario. En tales situaciones, el proceso continuo de clientes que visitan el restaurante puede verse como que los clientes provienen de una población infinita. La definición de muestra aleatoria simple tomada de una población infinita es la siguiente

MUESTRA ALEATORIA SIMPLE (POBLACIÓN INFINITA)

Una muestra aleatoria simple de una población infinita es una muestra seleccionada de manera que se satisfagan las condiciones siguientes.

1. Cada uno de los elementos seleccionados proviene de la población.
2. Cada elemento se selecciona independientemente.

En poblaciones infinitas un procedimiento para la selección de una muestra debe ser concebido especialmente para cada situación, de manera que permita seleccionar los elementos de manera independiente y evitar así un sesgo en la selección que dé mayores probabilidades de selección a ciertos tipos de elementos.

En poblaciones infinitas un procedimiento para la selección de una muestra debe ser concebido especialmente para cada situación, de manera que permita seleccionar los elementos de manera independiente y evitar así un sesgo en la selección que dé mayores probabilidades de selección a ciertos tipos de elementos. En el ejemplo de la selección de una muestra aleatoria simple entre los clientes de un restaurante de comida rápida, el primer requerimiento es satisfecho por cualquier cliente que entra en el restaurante. El segundo requerimiento es satisfecho seleccionando a los clientes de manera independiente. El objetivo del segundo requerimiento es evitar sesgos de selección. Habría un sesgo de selección, por ejemplo, si cinco clientes consecutivos que se seleccionaran fueran amigos. Es de esperar que estos clientes tengan perfiles semejantes. Dichos sesgos se evitan haciendo que la selección de un cliente no influya en la selección de cualquier otro cliente. En otras palabras, los clientes deben ser seleccionados de manera independiente.

McDonald's, el restaurante líder en comida rápida, realizó un muestreo aleatorio simple precisamente en una situación así. El procedimiento de muestreo se basó en el hecho de que algunos clientes presentaban cupones de descuento. Cada vez que un cliente presentaba un cupón descuento, al siguiente cliente que se atendía se le pedía que llenara un cuestionario sobre el perfil del cliente. Como los clientes que llegaban al restaurante presentaban cupones de descuento aleatoriamente, este plan de muestreo garantizaba que los clientes fueran seleccionados de manera independiente. Por tanto, los dos requerimientos para un muestreo aleatorio simple de una población infinita fueron satisfechos.

Las poblaciones infinitas suelen asociarse con un proceso que opera continuamente a lo largo del tiempo. Por ejemplo, partes fabricadas en una línea de producción, transacciones en un banco, llamadas que llegan a un centro de asesoría técnica y clientes que entran en las tiendas son considerados como provenientes de una población infinita. En tales casos un procedimiento de muestreo creativo garantiza que no haya sesgos de selección y que los elementos de la muestra sean seleccionados en forma independiente.

NOTAS Y COMENTARIOS

1. El número de muestras aleatorias simples distintas de tamaño n que pueden seleccionarse de una población finita de tamaño N es

$$\frac{N!}{n!(N-n)!}$$

En esta fórmula $N!$ y $n!$ son factoriales, vistos en el capítulo 4. Al usar esta expresión con los

datos del problema de EAI, en el que $N = 2500$ y $n = 30$, se ve que se pueden tomar 2.75×10^{69} muestras aleatorias simples distintas de 30 administradores de EAI.

2. Para tomar una muestra aleatoria pueden emplearse paquetes de software. En los apéndices del capítulo se muestra cómo usar Minitab y Excel para seleccionar una muestra aleatoria simple de una población finita

Ejercicios

Método

Autoexamen

- Dada una población finita que tiene cinco elementos A, B, C, D y E seleccione 10 muestras aleatorias simples de tamaño 2.
 - Enumere las 10 muestras empezando con AB, AC y así en lo sucesivo.
 - Usando el muestreo aleatorio simple, ¿cuál es la probabilidad que tiene cada muestra de tamaño 2 de ser seleccionada?
 - Si el número aleatorio 1 corresponde a A, el número 2 corresponde a B y así en lo sucesivo. Enliste la muestra aleatoria de tamaño 2 que será seleccionada al usar los números aleatorios 8 0 5 7 5 3 2.
- Suponga que una población finita tiene 350 elementos. A partir de los últimos tres dígitos de cada uno de los siguientes números aleatorios de cinco dígitos (por ejemplo: 601, 022, 448,...), determine los primeros cuatro elementos que se seleccionarán para una muestra aleatoria simple.

98601 73022 83448 02147 34229 27553 84147 93289 14209

Aplicaciones

Autoexamen

- Fortune* publicó datos sobre ventas, valor del activo, valor de mercado y ganancias por acción de las 500 corporaciones industriales más grandes de Estados Unidos (*Fortune 500*, 2003). Suponga que usted desea seleccionar una muestra aleatoria simple de 10 corporaciones de la lista *Fortune 500*. Use los tres últimos dígitos de la columna 9 de la tabla 7.1, empezando con 554. Leyendo hacia abajo por esa columna, identifique los números de las 10 corporaciones que se tomarán para la muestra.
- A continuación se presentan las 10 acciones más activas en la Bolsa de Nueva York del 6 de marzo del 2006 (*The Wall Street Journal*, 7 de marzo, 2006).

AT&T	Lucent	Nortel	Qwest
Pfizer	Texas Instruments	Gen. Elect.	iShrMSJpn
			LSI Logic

Las autoridades decidieron investigar las prácticas de negociación usando una muestra de tres de estas acciones.

- Empezando en el primer dígito aleatorio de la columna seis de la tabla 7.1,lea los números descendiendo por esa columna para seleccionar una muestra aleatoria simple de tres acciones para las autoridades.
- Con la información dada en la primera nota y comentario, determine cuántas muestras aleatorias simples diferentes de tamaño 3 pueden seleccionarse de una lista de 10 acciones.
- Una organización de estudiantes desean estimar la proporción de estudiantes que están a favor de una disposición de la escuela. Se cuenta con una lista con los nombres y direcciones de los 645 estudiantes inscritos el presente trimestre. Tomando números aleatorios de tres dígitos del renglón 10 de la tabla 7.1 y avanzando por ese renglón de izquierda a derecha, determine los 10 primeros estudiantes que serán seleccionados usando un muestreo aleatorio simple. Los números aleatorios de tres dígitos empiezan con 816, 283 y 610.
- El *County and City Data Book* del Census Bureau cuenta con información de los 3139 condados de Estados Unidos. Suponga que para un estudio nacional se recogerán datos de 30 condados seleccionados aleatoriamente. De la última columna de la tabla 7.1 extraiga números aleatorios de cuatro dígitos para determinar los primeros cinco condados seleccionados para la muestra. Ignore los primeros dígitos y empiece con los números aleatorios de cuatro dígitos 9945, 8364, 5702 y así sucesivamente.
- Suponga que se va a tomar una muestra aleatoria simple de 12 de los 372 médicos de una determinada ciudad. Una organización médica le proporciona los nombres de los médicos. De la tabla

- 7.1 use la columna ocho de números aleatorios de cinco dígitos para determinar cuáles serán los 12 médicos para la muestra. Ignore los primeros dos dígitos de cada grupo de cinco dígitos de números aleatorios. Este proceso empieza con el número aleatorio 108 y continúa descendiendo por la columna de números aleatorios.
8. La lista siguiente proporciona los 25 mejores equipos de fútbol de la NCAA en la temporada del 2002 (*NCAA News*, 4 de enero de 2003). De la tabla 7.1 use la novena columna que empieza con 13 554, para seleccionar una muestra aleatoria simple de seis equipos de fútbol. Empiece con el equipo 13 y use los primeros dos dígitos de cada renglón de la novena columna para el proceso de selección. ¿Cuáles son los seis equipos de fútbol seleccionados para la muestra aleatoria simple?
- | | |
|--------------------------|-------------------|
| 1. Ohio State | 14. Virginia Tech |
| 2. Miami | 15. Penn State |
| 3. Georgia | 16. Auburn |
| 4. Southern California | 17. Notre Dame |
| 5. Oklahoma | 18. Pittsburgh |
| 6. Kansas State | 19. Marshall |
| 7. Texas | 20. West Virginia |
| 8. Iowa | 21. Colorado |
| 9. Michigan | 22. TCU |
| 10. Washington State | 23. Florida State |
| 11. North Carolina State | 24. Florida |
| 12. Boise State | 25. Virginia |
| 13. Maryland | |
9. *The Wall Street Journal* proporciona el valor de activo neto, el rendimiento porcentual en lo que va del año y el rendimiento porcentual en tres años de 555 fondos mutualistas (*The Wall Street Journal*, 25 de abril de 2003). Suponga que se va a usar una muestra aleatoria simple de 12 de estos 555 fondos mutualistas para un estudio acerca de su tamaño y desempeño. Use la cuarta columna de números aleatorios en la tabla 7.1 empezando con el número 51102, para seleccionar la muestra aleatoria simple de 12 fondos mutualistas. Empiece con el fondo 102 y use los *últimos* tres dígitos de cada renglón de la cuarta columna para el proceso de selección. ¿Cuáles son los números de los 12 fondos mutualistas en esta muestra aleatoria simple?
10. Indique si las poblaciones siguientes se consideran finitas o infinitas.
- Todos los votantes registrados en el estado de California.
 - Todos los equipos de televisión que pueden ser producidos en una determinada fábrica.
 - Todas las órdenes que pueden ser procesadas por Allentown, Pensilvania, planta de TV-M Company.
 - Todas las llamadas de emergencia que pueden ser recibidas en una estación de policía.
 - Todas las piezas producidas por Fibercon, Inc., en el segundo turno el 17 de mayo.

7.3

Estimación puntual

Una vez descrito cómo seleccionar una muestra aleatoria simple, se vuelve al problema de EAI. En la tabla 7.2 se presenta una muestra aleatoria simple de 30 administradores con sus respectivos datos de sueldo anual y de participación en el programa de capacitación. La notación x_1 , x_2 , etc., se usa para denotar el sueldo anual del primer administrador de la muestra, del segundo, y así sucesivamente. La participación en el programa de capacitación se indica por un Sí en la columna programa de entrenamiento.

Para estimar el valor de un parámetro poblacional, la característica correspondiente se calcula con los datos de la muestra, a lo que se le conoce como **estadístico muestral**. Por ejemplo, para estimar la media poblacional μ y la desviación estándar poblacional σ de los salarios anuales de los administradores de EAI, se emplean los datos de la tabla 7.2 y se calculan los es-

TABLA 7.2 SALARIOS ANUALES Y SITUACIÓN RESPECTO AL PROGRAMA DE CAPACITACIÓN DE LOS ADMINISTRADORES PERTENECIENTES A UNA MUESTRA ALEATORIA SIMPLE DE 30 ADMINISTRADORES DE EAI

Salario anual	Programa de capacitación	Salario anual (\$)	Programa de capacitación
$x_1 = 49\ 094.30$	Sí	$x_{16} = 51\ 766.00$	Sí
$x_2 = 53\ 263.90$	Sí	$x_{17} = 52\ 541.30$	No
$x_3 = 49\ 643.50$	Sí	$x_{18} = 44\ 980.00$	Sí
$x_4 = 49\ 894.90$	Sí	$x_{19} = 51\ 932.60$	Sí
$x_5 = 47\ 621.60$	No	$x_{20} = 52\ 973.00$	Sí
$x_6 = 55\ 924.00$	Sí	$x_{21} = 45\ 120.90$	Sí
$x_7 = 49\ 092.30$	Sí	$x_{22} = 51\ 753.00$	Sí
$x_8 = 51\ 404.40$	Sí	$x_{23} = 54\ 391.80$	No
$x_9 = 50\ 957.70$	Sí	$x_{24} = 50\ 164.20$	No
$x_{10} = 55\ 109.70$	Sí	$x_{25} = 52\ 973.60$	No
$x_{11} = 45\ 922.60$	Sí	$x_{26} = 50\ 241.30$	No
$x_{12} = 57\ 268.40$	No	$x_{27} = 52\ 793.90$	No
$x_{13} = 55\ 688.80$	Sí	$x_{28} = 50\ 979.40$	Sí
$x_{14} = 51\ 564.70$	No	$x_{29} = 55\ 860.90$	Sí
$x_{15} = 56\ 188.20$	No	$x_{30} = 57\ 309.10$	No

tadísticos muestrales correspondientes; media muestral \bar{x} y desviación estándar muestral s . Con las fórmulas para la media muestral y la desviación estándar muestral presentadas en el capítulo 3 se obtiene que la media muestral es

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1\ 554\ 420}{30} = \$51\ 814$$

y la desviación estándar muestral es

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{325\ 009\ 260}{29}} = \$3348$$

Para estimar p , la proporción de administradores que han terminado el programa de capacitación, se usa la proporción muestral correspondiente \bar{p} . Sea x el número de administradores de la muestra que han terminado el programa de capacitación. De acuerdo con la tabla 7.2, $x = 19$. Por tanto, como el tamaño de la muestra es $n = 30$, la proporción muestral es

$$\bar{p} = \frac{x}{n} = \frac{19}{30} = 0.63$$

Al hacer los cálculos anteriores, se lleva a cabo el proceso estadístico conocido como *estimación puntual*. A la media muestral \bar{x} se le conoce como el **estimador puntual** de la media poblacional μ , a la desviación estándar muestral s como el estimador puntual de la desviación estándar poblacional σ y a la proporción muestral \bar{p} como el estimador puntual de la proporción poblacional p . Al valor numérico obtenido de \bar{x} , s , o \bar{p} se les conoce como **estimaciones puntuales**. Así, en la muestra aleatoria simple de 30 administradores de EAI que se presenta en la tabla 7.2, \$51 814 es la estimación puntual de μ , \$3 348 es la estimación puntual de σ y 0.63 es la estimación puntual de p . En la tabla 7.3 se resumen los resultados muestrales y se comparan las estimaciones puntuales con los valores de los parámetros poblacionales.

TABLA 7.3 INFORMACIÓN DE LAS ESTIMACIONES PUNTUALES OBTENIDAS DE UNA MUESTRA ALEATORIA SIMPLE DE 30 ADMINISTRADORES DE EAI

Parámetro poblacional	Valor del parámetro	Estimador puntual	Estimación puntual
μ = Media poblacional de los salarios anuales	\$51 800	\bar{x} = Media muestral de los salarios anuales	\$51 814
σ = Desviación estándar poblacional de los salarios anuales	\$4 000	s = Desviación estándar muestral de los salarios anuales	\$3 348
p = Proporción poblacional que ha terminado el programa de capacitación	0.60	\bar{p} = Proporción muestral que ha terminado el programa de capacitación	0.63

Como se observa en la tabla 7.3, las estimaciones puntuales difieren un poco de los correspondientes parámetros poblacionales. Estas diferencias son de esperarse ya que para elaborar las estimaciones muestrales se usa una muestra, y no un censo de toda la población. En el capítulo siguiente se verá cómo elaborar un intervalo de estimación para tener información acerca de qué tan cerca está la estimación muestral del parámetro poblacional.

Ejercicios

Métodos

Autoexamen

11. Los datos siguientes provienen de una muestra aleatoria simple.

5 8 10 7 10 14

- a. ¿Cuál es la estimación puntual de la media poblacional?
 - b. ¿Cuál es la estimación puntual de la desviación estándar poblacional?
12. Como respuestas a una pregunta de una encuesta a 150 individuos de una muestra se obtuvieron 75 Sí, 55 No y 20 individuos no dieron su opinión.
- a. ¿Cuál es la estimación puntual de la proporción de la población que responde Sí?
 - b. ¿Cuál es la estimación puntual de la proporción de la población que responde No?

Aplicaciones

Autoexamen

13. La siguiente información son datos obtenidos en una muestra aleatoria de las ventas de 5 meses:

Mes:	1	2	3	4	5
Unidades vendidas:	94	100	85	94	92

- a. Calcule una estimación puntual de la media poblacional del número medio de unidades vendidas por mes.
 - b. Calcule una estimación puntual de la desviación estándar del número de unidades vendidas por mes.
14. *BusinessWeek* publicó información sobre 283 fondos mutualistas (*BusinessWeek* 26 de enero de 2004). En el conjunto de datos MutualFunds se encuentra una muestra de 40 de estos fondos. Use este conjunto de datos para hacer lo que se pide en los incisos siguientes.
- a. Calcule una estimación puntual de la proporción de fondos de inversión de *BusinessWeek* que son fondos de cargo.
 - b. Calcule una estimación puntual de la proporción de fondos clasificados como de alto riesgo.
 - c. Calcule una estimación puntual de la proporción de fondos con una puntuación abajo del promedio para el riesgo.
15. Muchos de los medicamentos empleados en el tratamiento del cáncer son costosos. *BusinessWeek* informó de los costos de los tratamientos con Herceptin, un medicamento para tratar el cáncer de

mama (*BusinessWeek*, 30 de enero de 2006). Los siguientes son los costos de tratamientos con Herceptin en una muestra aleatoria de 10 pacientes.

4376	5578	2717	4920	4495
4798	6446	4119	4237	3814

- a. Calcule una estimación puntual del costo medio de un tratamiento con Herceptin.
 - b. Calcule una estimación puntual de la desviación estándar en los costos de los tratamientos con Herceptin.
16. En una muestra de 50 empresas de *Fortune 500*, 5 se encontraban en Nueva York, 6 en California, 2 en Minnesota y 1 en Wisconsin.
- a. Dé una estimación de la proporción de empresas de *Fortune 500* que se encuentran en Nueva York.
 - b. Dé una estimación del número de empresas de *Fortune 500* que se encuentran en Minnesota.
 - c. Dé una estimación de la proporción de empresas de *Fortune 500* que no se encuentran en ninguno de estos estados.
17. La American Association of Individuals Investors (AAII) hace sondeos semanales entre sus suscriptores para determinar cuántos se muestran optimistas, pesimistas o indiferentes respecto al mercado de acciones a corto plazo. Sus hallazgos en la semana que terminó el 2 de marzo de 2006 son consistentes con los resultados muestrales siguientes (www.aaii.com).

Optimistas 409 Indiferentes 299 Pesimistas 291

Dé una estimación puntual de los parámetros poblacionales siguientes.

- a. Proporción de suscriptores de AAII optimistas respecto al mercado de acciones.
- b. Proporción de suscriptores de AAII indiferentes respecto al mercado de acciones.
- c. Proporción de suscriptores de AAII pesimistas respecto al mercado de acciones.

7.4

Introducción a las distribuciones muestrales

En la sección anterior se dijo que la media muestral \bar{x} es el estimador puntual de la media poblacional μ y que la proporción muestral \bar{p} es el estimador puntual de la proporción poblacional p . En la muestra aleatoria simple de los 30 administradores de EAI que se presenta en la tabla 7.2, la estimación puntual de μ es $\bar{x} = \$51\,814$ y la estimación puntual de p es $\bar{p} = 0.63$. Suponga que se selecciona otra muestra aleatoria simple de 30 administradores de EAI y se obtienen las estimaciones puntuales siguientes:

Media muestral: $\bar{x} = \$52\,670$

Proporción muestral: $\bar{p} = 0.70$

Observe que se obtuvieron valores diferentes de \bar{x} y de \bar{p} . En efecto, otra muestra aleatoria simple de 30 administradores de EAI no se puede esperar que dé las mismas estimaciones puntuales que la primera muestra.

Ahora suponga que el proceso de seleccionar una muestra aleatoria simple de 30 administradores se repite una y otra vez y que cada vez se calculan los valores de \bar{x} y de \bar{p} . En la tabla 7.4 se muestra una parte de los resultados obtenidos en 500 muestras aleatorias simples y en la tabla 7.5 las distribuciones de frecuencias y distribuciones de frecuencias relativas de los valores de las 500 \bar{x} . En la figura 7.1 se muestra el histograma de las frecuencias de los valores de \bar{x} .

En el capítulo 5 se definió una variable aleatoria como una descripción numérica del resultado de un experimento. Si el proceso de seleccionar una muestra aleatoria simple se considera como un experimento, la media muestral \bar{x} es el valor numérico del resultado de ese experimento. Por tanto, la media muestral es una variable aleatoria. Entonces, como ocurre con otras variables aleatorias, \bar{x} tiene una media o valor esperado, una desviación estándar y una distribución

TABLA 7.4 VALORES DE \bar{x} Y DE \bar{p} OBTENIDOS EN 500 MUESTRAS ALEATORIAS SIMPLES DE 30 ADMINISTRADORES DE EAI CADA UNA

Muestra número	Media muestral (\bar{x})	Proporción muestral (\bar{p})
1	51 814	0.63
2	52 670	0.70
3	51 780	0.67
4	51 588	0.53
.	.	.
.	.	.
500	51 752	0.50

de probabilidad. Como los distintos valores que toma \bar{x} son resultado de distintas muestras aleatorias simples, a la distribución de probabilidad de \bar{x} se le conoce como **distribución muestral de \bar{x}** . Conocer esta distribución muestral y sus propiedades permitirá hacer declaraciones de probabilidad de qué tan cerca está la media muestral \bar{x} de la media poblacional μ .

De regreso a la figura 7.1, se necesitaría enumerar todas las muestras posibles de 30 administradores y calcular cada una de las medias muestrales para determinar totalmente la distribución muestral de \bar{x} . Sin embargo, el histograma de 500 valores \bar{x} da una aproximación a esta distribución muestral. En esta aproximación se observa la apariencia de curva de campana de esta distribución. Además, la mayor concentración de valores de \bar{x} y la media de los 500 valores de \bar{x} se encuentran cerca de la media poblacional $\mu = \$51\,800$. En la sección siguiente se describirán más detalladamente las propiedades de la distribución muestral de \bar{x} .

Los 500 valores de las proporciones muestrales \bar{p} se resumen en el histograma de frecuencias relativas de la figura 7.2. Como ocurre con \bar{x} , \bar{p} es una variable aleatoria. Si se tomara cada muestra posible de tamaño 30 y para cada muestra se calculara el valor \bar{p} , la distribución de probabilidad que se obtuviera sería la distribución muestral de \bar{p} . En la figura 7.2, el histograma de frecuencias relativas de los 500 valores muestrales da una idea general de la apariencia de la distribución muestral de \bar{p} .

En la práctica sólo se selecciona una muestra aleatoria simple de la población. En esta sección el proceso de muestreo se repitió 500 veces para ilustrar que es posible tomar muchas mues-

TABLA 7.5 DISTRIBUCIÓN DE FRECUENCIAS DE \bar{x} EN 500 MUESTRAS ALEATORIAS SIMPLES DE 30 ADMINISTRADORES DE EAI CADA UNA

Salario anual medio (\$)	Frecuencia	Frecuencia relativa
49 500.00–49 999.99	2	0.004
50 000.00–50 499.99	16	0.032
50 500.00–50 999.99	52	0.104
51 000.00–51 499.99	101	0.202
51 500.00–51 999.99	133	0.266
52 000.00–52 499.99	110	0.220
52 500.00–52 999.99	54	0.108
53 000.00–53 499.99	26	0.052
53 500.00–53 999.99	6	0.012
Totales	500	1.000

FIGURA 7.1 HISTOGRAMA DE LAS FRECUENCIAS RELATIVAS DE LOS VALORES DE \bar{x} OBTENIDOS EN 500 MUESTRAS ALEATORIAS SIMPLES DE 30 ADMINISTRADORES CADA UNA

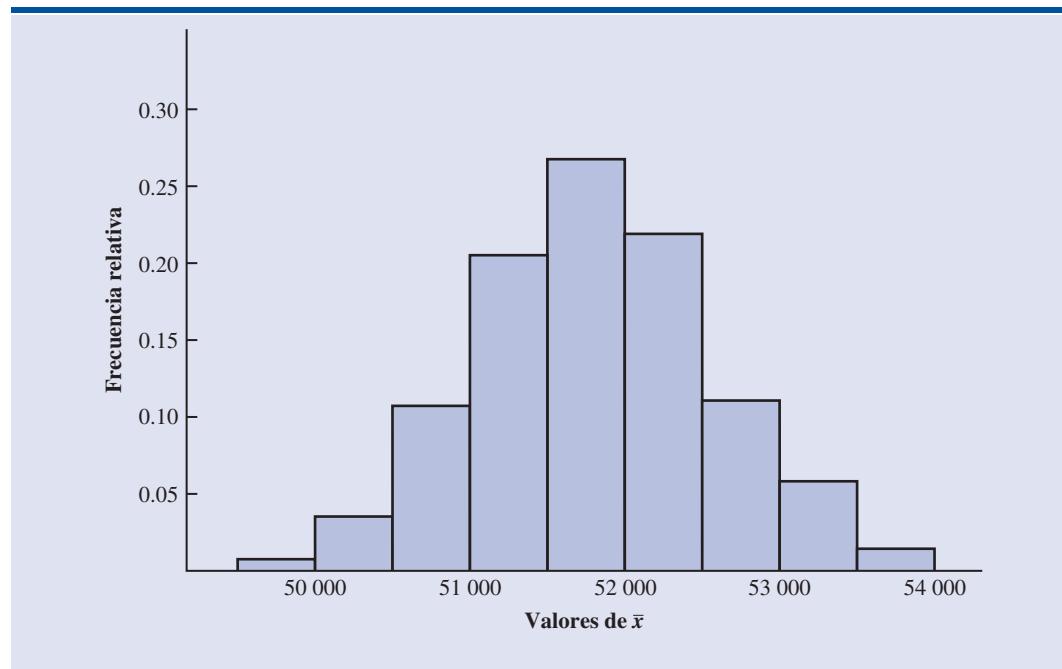
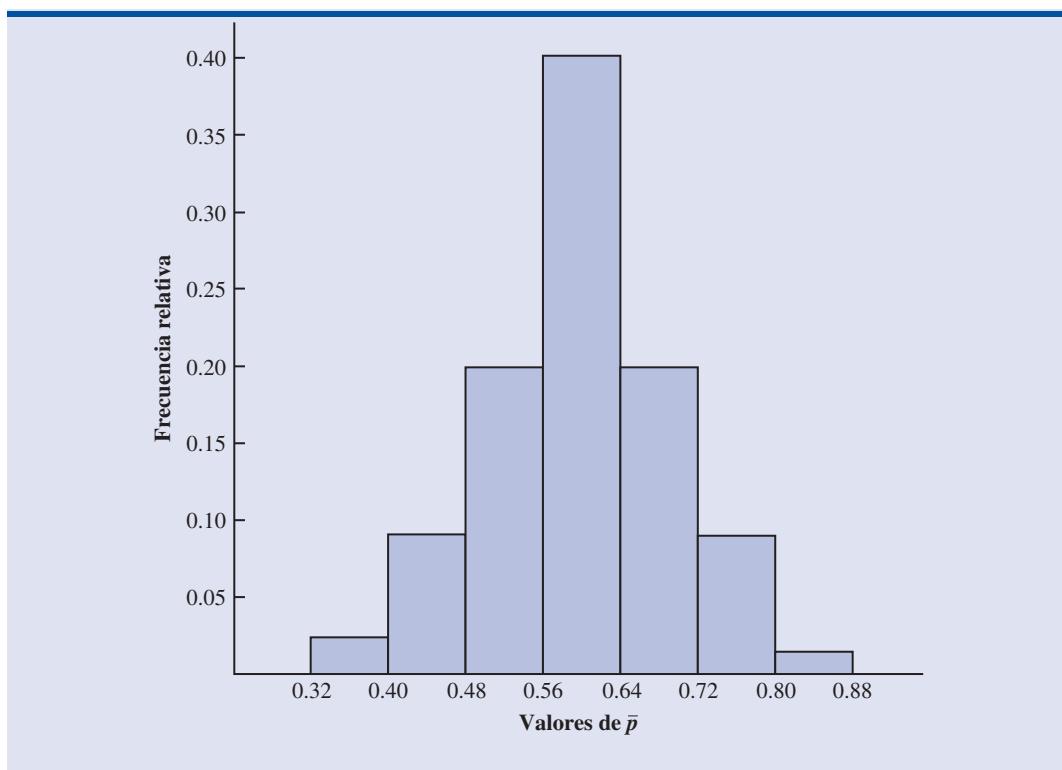


FIGURA 7.2 HISTOGRAMA DE LAS FRECUENCIAS RELATIVAS DE LOS VALORES DE \bar{p} OBTENIDOS EN 500 MUESTRAS ALEATORIAS SIMPLES DE 30 ADMINISTRADORES CADA UNA



tras diferentes y que diferentes muestras darán valores distintos de los estadísticos muestrales \bar{x} y \bar{p} . A la distribución muestral de cualquier estadístico determinado se le llama distribución muestral del estadístico. En la sección 7.5 se presentan las características de la distribución muestral de \bar{x} . En la sección 7.6 se muestran las características de la distribución muestral de \bar{p} .

7.5

Distribución muestral de \bar{x}

En la sección anterior se dijo que la media muestral \bar{x} es una variable aleatoria y que a su distribución de probabilidad se le llama distribución muestral de \bar{x} .

DISTRIBUCIÓN MUESTRAL DE \bar{x}

La distribución muestral de \bar{x} es la distribución de probabilidad de todos los valores de la media muestral \bar{x} .

En esta sección se describen las propiedades de la distribución muestral de \bar{x} . Como ocurre con otras distribuciones de probabilidad estudiadas, la distribución muestral de \bar{x} tiene un valor esperado, una desviación estándar y una forma característica. Para empezar se considerará la media de todos los valores de \bar{x} , a la que se conoce como valor esperado de \bar{x} .

Valor esperado de \bar{x}

En el problema de muestreo de EAI se vio que en distintas muestras aleatorias simples se obtienen valores diferentes para la media muestral \bar{x} . Como la variable aleatoria \bar{x} puede tener muchos valores diferentes, suele ser de interés conocer la media de todos los valores de \bar{x} que se obtienen con diferentes muestras aleatorias simples. La media de la variable aleatoria \bar{x} es el valor esperado de \bar{x} . Sea $E(\bar{x})$ el valor esperado de \bar{x} y μ la media de la población de la que se selecciona una muestra aleatoria simple. Se puede demostrar que cuando se emplea el muestreo aleatorio simple, $E(\bar{x})$ y μ son iguales.

VALOR ESPERADO DE \bar{x}

$$E(\bar{x}) = \mu \quad (7.1)$$

donde

$$\begin{aligned} E(\bar{x}) &= \text{valor esperado de } \bar{x} \\ \mu &= \text{media poblacional} \end{aligned}$$

El valor esperado de \bar{x} es igual a la media de la población de la que se tomó la muestra.

Esto enseña que usando el muestreo aleatorio simple, el valor esperado o media de la distribución muestral de \bar{x} es igual a la media de la población. En la sección 7.1 se vio que el sueldo anual medio de los administradores de EAI es $\mu = \$51\,800$. Por tanto, de acuerdo con la ecuación (7.1), la media de todas las medias muestrales en el estudio de EAI es también $\$51\,800$.

Cuando el valor esperado de un estimador puntual es igual al parámetro poblacional, se dice que el estimador puntual es **insesgado**. Por tanto, la ecuación (7.1) muestra que \bar{x} es un estimador insesgado de la media poblacional μ .

Desviación estándar de \bar{x}

Ahora se definirá la desviación estándar de la distribución muestral de \bar{x} . Se empleará la notación siguiente.

$$\begin{aligned}\sigma_{\bar{x}} &= \text{desviación estándar de } \bar{x} \\ \sigma &= \text{desviación estándar de la población} \\ n &= \text{tamaño de la muestra} \\ N &= \text{tamaño de la población}\end{aligned}$$

Es posible demostrar que usando el muestreo aleatorio simple, la desviación estándar de \bar{x} depende de si la población es finita o infinita. Las dos fórmulas para la desviación estándar son las siguientes.

DESVIACIÓN ESTÁNDAR DE \bar{x}

Población finita

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \left(\frac{\sigma}{\sqrt{n}} \right)$$

Población infinita

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

(7.2)

Al comparar las dos fórmulas se ve que el factor $\sqrt{(N-n)/(N-1)}$ se requiere cuando la población es finita, pero no cuando es infinita. A este factor se le conoce como **factor de corrección para una población finita**. En muchas de las situaciones prácticas de muestreo, se encuentra que aunque la población sea finita, es “grande”, mientras que el tamaño de la muestra es “pequeño”. En estos casos el factor de corrección para una población finita $\sqrt{(N-n)/(N-1)}$ es casi igual a 1. Por tanto, la diferencia entre el valor de la desviación estándar de \bar{x} en el caso de poblaciones finitas o infinitas se vuelve despreciable. Entonces, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ es una buena aproximación a la desviación estándar de \bar{x} , aun cuando la población sea finita. Esta observación lleva al siguiente lineamiento, o regla general, para calcular la desviación estándar de \bar{x} .

USO DE LA EXPRESIÓN SIGUIENTE PARA CALCULAR LA DESVIACIÓN ESTÁNDAR DE \bar{x}

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (7.3)$$

siempre que

1. La población sea infinita; o
2. La población sea finita y el tamaño de la muestra sea menor o igual a 5% del tamaño de la población; es decir, $n/N \leq 0.05$.

En el problema 21 se muestra que cuando $n/N \leq 0.05$, el factor de corrección para una población finita tiene poco efecto en el valor de $\sigma_{\bar{x}}$.

En los casos en que $n/N > 0.05$, para calcular $s_{\bar{x}}$ deberá usarse la versión para poblaciones finitas de la fórmula (7.2). En este libro, a menos que se indique otra cosa, se supondrá que el tamaño de la población es “grande”, $n/N \leq 0.05$, y se usará la expresión (7.3) para calcular $\sigma_{\bar{x}}$.

El término error estándar se usa en la inferencia estadística para referirse a la desviación estándar de un estimador puntual.

Para calcular $\sigma_{\bar{x}}$ se necesita conocer σ , la desviación estándar de la población. Para subrayar, aún más, la diferencia entre $\sigma_{\bar{x}}$ y σ , a la desviación estándar de \bar{x} , $\sigma_{\bar{x}}$ se le llama **error estándar** de la media. En general, el término *error estándar* se refiere a la desviación estándar de un estimador puntual. Más adelante se verá que el valor del error estándar de la media ayuda a determinar qué tan lejos puede estar la media muestral de la media poblacional. Ahora, de nuevo con el ejemplo de EAI se calcula el error estándar de la media correspondiente a las muestras aleatorias simples de 30 administradores de EAI.

En la sección 7.1 se halló que la desviación estándar de los sueldos anuales en la población de los 2500 administradores de EAI era $\sigma = 4000$. En este caso la población es finita, $N = 2500$. Sin embargo, como el tamaño de la muestra es 30, se tiene $n/N = 30/2500 = 0.012$. Como el tamaño de la muestra es menor que 5% del tamaño de la población, se puede ignorar el factor de corrección para una población finita y usar la ecuación (7.3) para calcular el error estándar.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.3$$

Forma de la distribución muestral de \bar{x}

Los resultados anteriores respecto al valor esperado y a la desviación estándar en la distribución muestral de \bar{x} son aplicables a cualquier población. El paso final en la identificación de las características de la distribución muestral de \bar{x} es determinar la forma de la distribución muestral. Se considerarán dos casos: 1. La población tiene distribución normal, y 2. La población no tiene distribución normal.

La población tiene distribución normal. En muchas situaciones es razonable suponer que la población de la que se seleccionó la muestra aleatoria simple tenga distribución normal o casi normal. Cuando la población tiene distribución normal, la distribución muestral de \bar{x} está distribuida normalmente sea cual sea el tamaño de la muestra.

La población no tiene distribución normal. Cuando la población de la que se tomó la muestra aleatoria simple no tiene distribución normal, el **teorema del límite central** ayuda a determinar la forma de la distribución muestral de \bar{x} . El enunciado del teorema del límite central aplicado a la distribución muestral de \bar{x} dice lo siguiente.

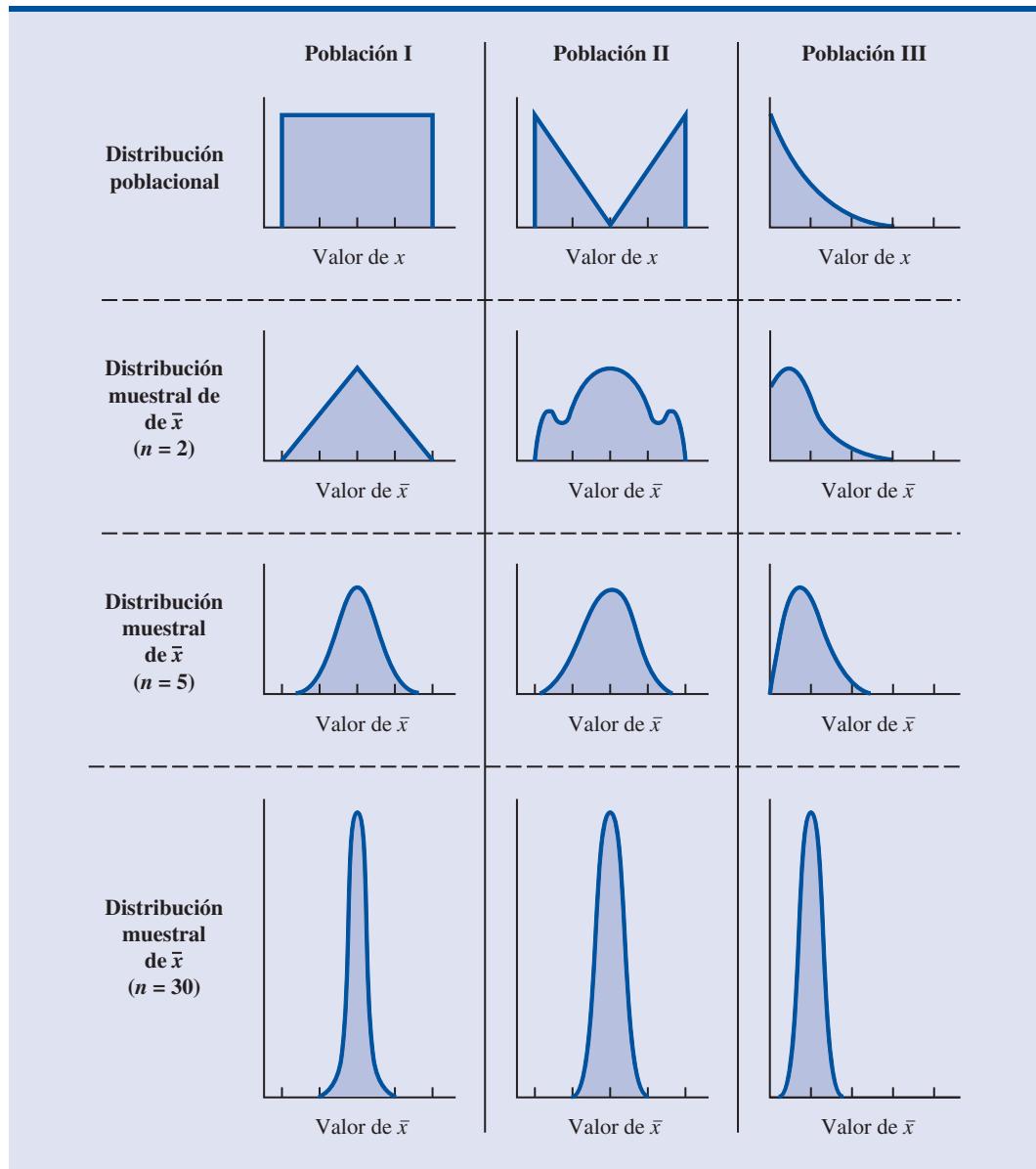
TEOREMA DEL LÍMITE CENTRAL

Cuando se seleccionan muestras aleatorias simples de tamaño n de una población, la distribución muestral de la media muestral \bar{x} puede aproximarse mediante una distribución normal a medida que el tamaño de la muestra se hace grande.

En la figura 7.3 se muestra cómo funciona el teorema del límite central en tres poblaciones diferentes; cada columna se refiere a una de las poblaciones. En el primer renglón de la figura se muestra que ninguna de las tres poblaciones está distribuida normalmente. La población I tiene una distribución uniforme. A la población II se le conoce como distribución en forma de orejas de conejo. Esta distribución es simétrica, pero los valores más probables se encuentran en las colas de la distribución. La forma de la población III se parece a una distribución exponencial; es sesgada a la derecha.

En los tres renglones siguientes de la figura 7.3 se muestran las formas de las distribuciones muestrales para tamaños de muestras $n = 2$, $n = 5$ y $n = 30$. Cuando el tamaño de la muestra es 2, se observa que cada distribución muestral tiene una forma diferente a la distribución poblacional correspondiente. Con muestras de tamaño 5, se observa que las formas de las distribuciones muestrales en los casos de las poblaciones I y II empiezan a parecerse a la forma de una distribución normal. En el caso de la población III, aun cuando la forma de la distribución muestral empieza a ser parecida a una distribución normal, todavía se observa cierto sesgo a la derecha.

FIGURA 7.3 ILUSTRACIÓN DEL TEOREMA DEL LÍMITE CENTRAL CON TRES POBLACIONES



Por último, para muestras de tamaño 30, las formas de cada una de las tres distribuciones muestrales es aproximadamente normal.

Desde el punto de vista de la práctica, será conveniente saber qué tan grande necesita ser el tamaño de la muestra para que aplique el teorema del límite central y pueda suponer que la forma de la distribución muestral es aproximadamente normal. En las investigaciones estadísticas se ha estudiado este problema en distribuciones muestrales de \bar{x} de muy diversas poblaciones y para muy diversos tamaños de muestras. Lo que se acostumbra hacer en la práctica es suponer que, en la mayor parte de las aplicaciones, la distribución muestral de \bar{x} se puede aproximar mediante una distribución normal siempre que la muestra sea de tamaño 30 o mayor. En los casos en que la población es muy sesgada o existen observaciones atípicas, pueden necesitarse muestras de tamaño 50. Por último, si la población es discreta, el tamaño de muestra necesario para la aproximación normal suele depender de la proporción poblacional. Más acerca de este tema se dirá cuando se estudie la distribución muestral de \bar{p} en la sección 7.6.

Distribución muestral de \bar{x} en el problema EAI

En el problema de EAI, para el que ya previamente se mostró que $E(\bar{x}) = \$51\,800$ y $\sigma_{\bar{x}} = 730.3$, no se cuenta con ninguna información acerca de la distribución de la población; puede estar o no distribuida normalmente. Si la población tiene una distribución normal, la distribución muestral de \bar{x} estará distribuida normalmente. Si la población no tiene una distribución normal, la muestra aleatoria simple de 30 administradores y el teorema del límite central permiten concluir que la distribución muestral de \bar{x} puede aproximarse mediante una distribución normal. En cualquiera de los casos, se concluye que la distribución muestral de \bar{x} se describe mediante una distribución normal como la que se muestra en la figura 7.4.

Valor práctico de la distribución muestral de \bar{x}

Siempre que se seleccione una muestra aleatoria simple y se use el valor de la media muestral para estimar el valor de la media poblacional μ , no se podrá esperar que la media muestral sea exactamente igual a la media poblacional. La razón práctica por la que interesa la distribución muestral de \bar{x} es que se puede usar para proporcionar información probabilística acerca de la diferencia entre la media muestral y la media poblacional. Para demostrar este uso, se retomará el problema de EAI.

Suponga que el director de personal cree que la media muestral será una estimación aceptable de la media poblacional si la primera está a más o menos de \$500 de la media poblacional. Sin embargo, no es posible garantizar que la media muestral esté a más o menos de \$500 de la media poblacional. En efecto, en la tabla 7.5 y en la figura 7.1 se observa que algunas de las 500 medias muestrales difieren en más de \$2000 de la media poblacional. Entonces hay que pensar en el requerimiento del director de personal en términos de probabilidad. Es decir, al director de personal le interesa la interrogante siguiente: “¿Cuál es la probabilidad de que la media muestral obtenida usando una muestra aleatoria simple de 30 administradores de EAI, se encuentre a más o menos de \$500 de la media poblacional?”

Como ya se han identificado las propiedades de la distribución muestral de \bar{x} (véase figura 7.4), se usará esta distribución para contestar esta interrogante probabilística. Observe la distribución muestral de \bar{x} que se muestra nuevamente en la figura 7.5. Como la media poblacional es \$51 800, el director de personal desea saber cuál es la probabilidad de que \bar{x} esté entre \$51 300 y \$52 300. Esta probabilidad corresponde al área sombreada de la distribución muestral que apa-

FIGURA 7.4 DISTRIBUCIÓN MUESTRAL DE \bar{x} PARA EL SUELDO ANUAL EN UNA MUESTRA ALEATORIA SIMPLE DE 30 ADMINISTRADORES

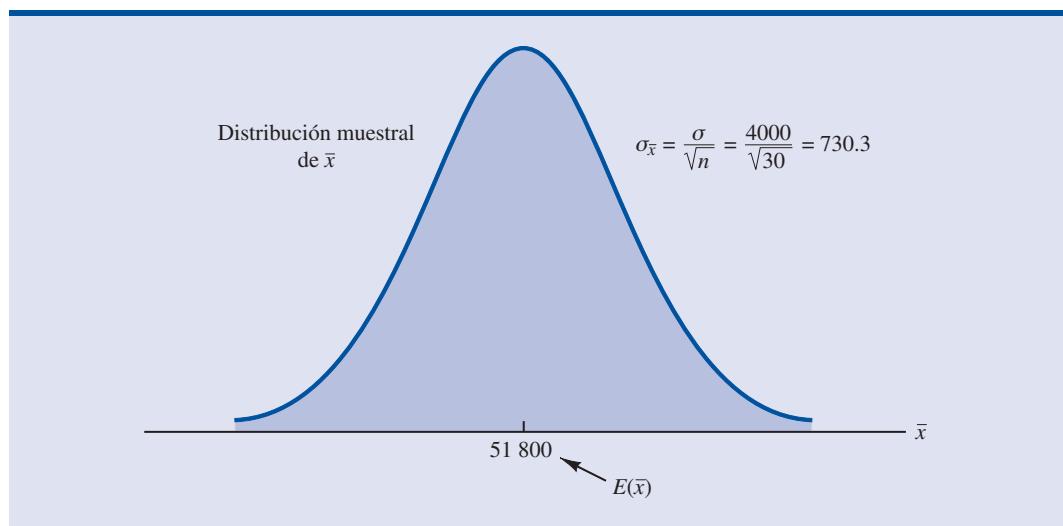
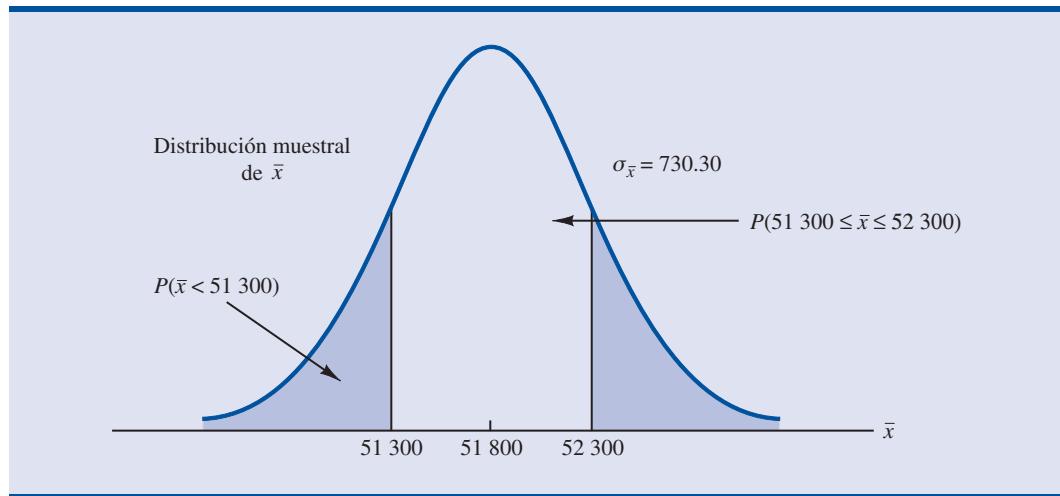


FIGURA 7.5 PROBABILIDAD DE QUE UNA MEDIA MUESTRAL DE UNA MUESTRA ALEATORIA SIMPLE DE 30 ADMINISTRADORES DE EAI SE ENCUENTRE ENTRE LOS \$500 DE LA MEDIA POBLACIONAL



rece en la figura 7.5. Como la distribución muestral está distribuida normalmente y su media es \$51 800 y el error estándar de la media es 730.3, se usa la tabla de probabilidad normal estándar para hallar el área o probabilidad.

Primero se calcula el valor de z en el extremo superior de este intervalo (52 300) y se usa la tabla para hallar el área bajo la curva a la izquierda de ese punto (área hacia la cola izquierda). Despues se calcula el valor de z en el extremo inferior de este intervalo (51 300) y se usa la tabla para hallar el área bajo la curva a la izquierda de este punto (otra área hacia la cola izquierda). Al restar la segunda área de la primera, se obtiene la probabilidad buscada.

En $\bar{x} = 52\,300$, se tiene

$$z = \frac{52\,300 - 51\,800}{730.30} = 0.68$$

En la tabla de probabilidad normal estándar la probabilidad acumulada (área a la izquierda de $z = 0.68$) es 0.7517.

En $\bar{x} = 51\,300$, se tiene

$$z = \frac{51\,300 - 51\,800}{730.30} = -0.68$$

El área bajo la curva a la izquierda de $z = -0.68$ es 0.2483. Por tanto, $P(51\,300 \leq \bar{x} \leq 52\,300) = P(z \leq 0.68) - P(z < -0.68) = 0.7517 - 0.2483 = 0.5034$.

Estos cálculos indican que hay una probabilidad de 0.5034 de que con una muestra aleatoria simple de 30 administradores de EAI se obtenga una media muestral \bar{x} que esté a más o menos de \$500 de la media poblacional. Por tanto, la probabilidad de que la diferencia entre \bar{x} y $\mu = \$51,800$ sea superior a \$500 es $1 - 0.5034 = 0.4966$. En otras palabras, una muestra aleatoria simple de 30 administradores de EAI tiene aproximadamente 50/50 oportunidades de tener una media muestral que no difiera de la media poblacional en más de los aceptables \$500. Quizá deba pensarse en una muestra de tamaño mayor. Se explorará esta posibilidad considerando la relación entre el tamaño de la muestra y la distribución muestral de \bar{x} .

La distribución muestral de \bar{x} se usa para obtener información probabilística acerca de qué tan cerca se encuentra la media muestral \bar{x} de la media poblacional μ .

Relación entre el tamaño de la muestra y la distribución muestral de \bar{x}

Suponga que en el problema de muestreo de EAI se toma una muestra aleatoria simple de 100 administradores en lugar de los 30 considerados. La intuición indica que teniendo más datos proporcionados por una muestra mayor, la media muestral basada en $n = 100$ proporcionará una mejor estimación de la media poblacional que una media muestral basada en $n = 30$. Para ver cuánto es mejor, se considerará la relación entre el tamaño de la muestra y la distribución muestral de \bar{x} .

Primero observe que $E(\bar{x}) = \mu$ independientemente del tamaño de la muestra. Entonces, la media de todos los valores posibles de \bar{x} es igual a la media poblacional μ independientemente del tamaño n de la muestra. Pero, el error estándar de la media, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, está relacionado con la raíz cuadrada del tamaño de la muestra. Siempre que el tamaño de la muestra aumente, el error estándar de la media $\sigma_{\bar{x}}$ disminuirá. Con $n = 30$, el error estándar de la media en el problema de EAI es 730.3. Sin embargo, aumentando el tamaño de la muestra $n = 100$, el error estándar de la media disminuye a

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{100}} = 400$$

En la figura 7.6 se muestran las distribuciones muestrales de \bar{x} correspondientes a $n = 30$ y a $n = 100$. Como la distribución muestral con $n = 100$ tiene un error estándar más pequeño, habrá menos variación entre los valores de \bar{x} y éstos tenderán a estar más cerca de la media poblacional que los valores de \bar{x} con $n = 30$.

La distribución muestral de \bar{x} , en el caso $n = 100$, puede emplearse para calcular la probabilidad de que una muestra aleatoria simple de 100 administradores de EAI dé una media muestral que no difiera de los \$500 de la media poblacional. Como la distribución muestral es normal y su media es \$51 800 y el error estándar de la media es 400, se emplea la tabla de probabilidad normal estándar para hallar el área o la probabilidad.

Para $\bar{x} = 52 300$ (véase figura 7.7) se tiene

$$z = \frac{52 300 - 51 800}{400} = 1.25$$

FIGURA 7.6 COMPARACIÓN ENTRE LAS DISTRIBUCIONES MUESTRALES DE \bar{x} CON MUESTRAS ALEATORIAS SIMPLES DE TAMAÑO $n = 30$ ADMINISTRADORES DE EAI Y CON MUESTRAS DE TAMAÑO $n = 100$ ADMINISTRADORES

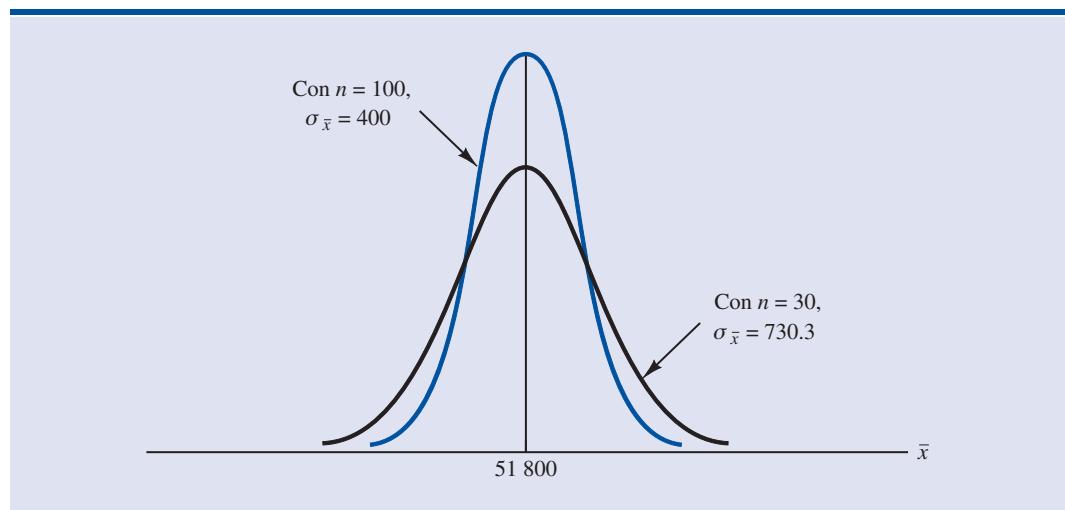
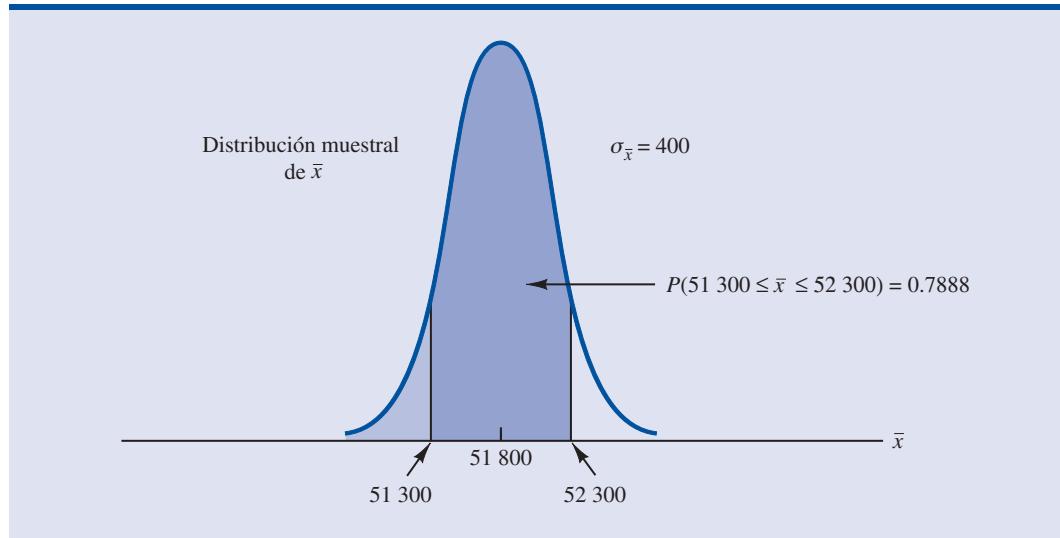


FIGURA 7.7 PROBABILIDAD DE QUE LA MEDIA MUESTRAL NO DIFIERA EN MÁS DE \$500 DE LA MEDIA POBLACIONAL USANDO UNA MUESTRA ALEATORIA SIMPLE DE 100 ADMINISTRADORES DE EAI



En la tabla de probabilidad normal estándar se encuentra que la probabilidad acumulada correspondiente a $z = 1.25$ es 0.8944.

Para $\bar{x} = 51 300$, se tiene

$$z = \frac{51 300 - 51 800}{400} = -1.25$$

La probabilidad acumulada correspondiente a $z = -1.25$ es 0.1056. Por tanto, $P(51 300 \leq \bar{x} \leq 52 300) = P(z \leq 1.25) - P(z \leq -1.25) = 0.8944 - 0.1056 = 0.7888$. Entonces, aumentando el tamaño de la muestra de 30 a 100 administradores de EAI, la probabilidad de obtener una muestra aleatoria simple que esté entre los \$500 de la media poblacional aumenta de 0.5034 a 0.7888.

Aquí, el punto importante es que cuando aumenta el tamaño de la muestra, el error estándar de la media disminuye. Así, una muestra de mayor tamaño proporciona mayor probabilidad de que la media muestral esté dentro de una distancia determinada de la media poblacional.

NOTAS Y COMENTARIOS

1. Al presentar la distribución muestral de \bar{x} para el problema de EAI, se aprovechó que se conocían la media poblacional $\mu = 51 800$ y la desviación estándar poblacional $\sigma = 4000$. Sin embargo, lo usual es que los valores de la media poblacional μ y de la desviación estándar poblacional σ , que se necesitan para determinar la distribución muestral de \bar{x} , no se conozcan. En el capítulo 8 se verá cómo se usan la media muestral \bar{x} y la desviación estándar muestral s cuando no se conocen μ y σ .
2. La demostración del teorema del límite central requiere observaciones independientes en la muestra. Esta condición se satisface cuando se trata de poblaciones infinitas y cuando se trata de poblaciones finitas, si el muestreo se hace con reemplazo. Aunque el teorema del límite central no se refiere directamente a muestreos sin reemplazo de poblaciones finitas, se aplican los hallazgos del teorema del límite central cuando la población es de tamaño grande.

Ejercicios

Métodos

18. La media de una población es 200 y su desviación estándar es 50. Se va a tomar una muestra aleatoria simple de tamaño 100 y se usará la media muestral para estimar la media poblacional.
 - a. ¿Cuál es el valor esperado de \bar{x} ?
 - b. ¿Cuál es la desviación estándar de \bar{x} ?
 - c. Muestre la distribución muestral de \bar{x}
 - d. ¿Qué muestra la distribución muestral de \bar{x} ?
19. La media de una población es 200 y su desviación estándar es 50. Suponga que se selecciona una variable aleatoria simple de tamaño 100 y se usa \bar{x} para estimar μ
 - a. ¿Cuál es la probabilidad de que la diferencia entre la media muestral y la media poblacional no sea mayor que ± 5 ?
 - b. ¿De qué la diferencia entre la media muestral y la media poblacional no sea mayor que ± 10 ?
20. Suponga que la desviación estándar poblacional es $\sigma = 25$. Calcule el error estándar de la media, $\sigma_{\bar{x}}$, con muestras de tamaño 50, 100, 150 y 200. ¿Qué puede decir acerca del error estándar de la media conforme el tamaño de la muestra aumenta?
21. Suponga que de una población en la que $\sigma = 10$ se toma una muestra aleatoria simple de tamaño 50. Halle el valor del error estándar de la media en cada uno de los casos siguientes (si es necesario use el factor de corrección para una población finita).
 - a. El tamaño de la población es infinito.
 - b. El tamaño de la población es $N = 50\,000$.
 - c. El tamaño de la población es $N = 5000$.
 - d. El tamaño de la población es $N = 500$.

Autoexamen

Aplicaciones

22. Regrese al problema de los administradores de EAI. Suponga que se usa una muestra aleatoria simple de 60 administradores.
 - a. Dibuje la distribución muestral de \bar{x} si se emplean muestras aleatorias simples de tamaño 60.
 - b. ¿Qué pasa con la distribución muestral de \bar{x} si se usan muestras aleatorias simples de tamaño 120?
 - c. ¿Qué puede decir acerca de lo que le pasa a la distribución muestral de \bar{x} conforme el tamaño de la muestra aumenta? ¿Parece ser lógica esta generalización? Explique.
23. En el problema de EAI (véase figura 7.5), se mostró que con $n = 30$, la probabilidad de que la media muestral no difiriera más de \$500 de la media poblacional era 0.5034.
 - a. ¿Cuál es la probabilidad de que la media muestral no difiera más de \$500 de la media poblacional si se usa una muestra de tamaño 60?
 - b. Responda el inciso a si el tamaño de la muestra es 120.
24. El costo medio de la colegiatura en una universidad estatal de Estados Unidos es \$4260 anuales. Considere este valor como media poblacional y asuma que la desviación estándar poblacional es $\sigma = \$900$. Suponga que selecciona una muestra aleatoria de 50 universidades.
 - a. Presente la distribución muestral de \bar{x} como media muestral de la colegiatura en las 50 universidades.
 - b. ¿Cuál es la probabilidad de que la muestra aleatoria simple proporcione una media muestral que no difiera de la media poblacional en más de \$250?
 - c. ¿Cuál es la probabilidad de que la muestra aleatoria simple proporcione una media muestral que no difiera de la media poblacional en más de \$100?
25. El College Board American College Testing Program informa que en el examen de admisión a las universidades, a nivel nacional, la media poblacional de las puntuaciones que se obtienen es $\mu = 1020$ (*The World Almanac 2003*). Suponga que la desviación estándar poblacional es $\sigma = 100$.

Autoexamen

- a. ¿Cuál es la probabilidad de que en una muestra aleatoria de 75 estudiantes la media muestral de las puntuaciones no difiera más de 10 puntos de la media poblacional?
- b. ¿Cuál es la probabilidad de que en una muestra aleatoria de 75 estudiantes la media muestral de las puntuaciones no difiera más de 20 puntos de la media poblacional?
26. El costo medio anual de un seguro para automóvil es de \$939 (*CNBC*, 23 de febrero de 2006). Suponga que la desviación estándar es $\sigma = \$245$.
 - a. ¿Cuál es la probabilidad de que una muestra aleatoria simple de pólizas de seguros de automóvil la media muestral no difiera más de \$25 de la media poblacional si el tamaño de la muestra es 30, 50, 100 y 400?
 - b. ¿Qué ventaja tiene una muestra grande cuando se quiere estimar la media poblacional?
27. *BusinessWeek* realizó una encuesta entre los estudiantes que terminaban sus estudios en los 30 programas de una maestría (*BusinessWeek*, 22 de septiembre de 2003). De acuerdo con esta encuesta el salario medio anual de una mujer y de un hombre 10 años después de terminar sus estudios es \$117 000 y \$168 000, respectivamente. Suponga que la desviación estándar entre los salarios de las mujeres es \$25 000 y entre los salarios de los hombres es \$40 000.
 - a. ¿Cuál es la probabilidad de que en una muestra aleatoria simple de 40 hombres la media muestral no difiera más de \$10 000 de la media poblacional de \$168 000?
 - b. ¿Cuál es la probabilidad de que en una muestra aleatoria simple de 40 mujeres la media muestral no difiera más de \$10 000 de la media poblacional de \$117 000?
 - c. ¿En cuál de los dos casos, inciso a o inciso b, hay más probabilidad de obtener una media muestral que no difiera en más de \$10 000 de la media poblacional? ¿Por qué?
 - d. ¿Cuál es la probabilidad de que en una muestra aleatoria simple de 100 hombres, la media muestral no difiera en más de \$4000 de la media poblacional?
28. Un hombre golfista tiene una puntuación promedio de 95 y una mujer de 106 (*Golf Digest*, abril de 2006). Considere estos valores como medias poblacionales de los hombres y de las mujeres y suponga que la desviación estándar poblacional es $\sigma = 14$ golpes en ambos casos. Se tomará una muestra aleatoria simple de 40 golfistas hombres y otra de 45 mujeres golfistas.
 - a. Dé la distribución muestral de \bar{x} correspondiente a los hombres golfistas.
 - b. ¿Cuál es la probabilidad de que, en el caso de los hombres golfistas, la media muestral no difiera en más de 3 golpes de la media poblacional?
 - c. ¿Cuál es la probabilidad de que, en el caso de las mujeres golfistas, la media muestral no difiera en más de 3 golpes de la media poblacional?
 - d. ¿En cuál de los casos, inciso a o inciso b, es mayor la probabilidad de que la media muestral no difiera en más de 3 golpes de la media poblacional? ¿Por qué?
29. En el norte de Kentucky (*The Cincinnati Enquirer*, 21 de enero de 2006) el precio promedio de la gasolina sin plomo era \$2.34. Use este precio como media poblacional y suponga que la desviación estándar poblacional es \$0.20.
 - a. ¿Cuál es la probabilidad de que el precio medio en una muestra de 30 gasolineras no difiera en más de \$0.30 de la media poblacional?
 - b. ¿Cuál es la probabilidad de que el precio medio en una muestra de 50 gasolineras no difiera en más de \$0.30 de la media poblacional?
 - c. ¿Cuál es la probabilidad de que el precio medio en una muestra de 100 gasolineras no difiera en más de \$0.30 de la media poblacional?
 - d. ¿Recomendaría usted alguno de los tamaños muestrales de los incisos a, b o c para que la probabilidad de que el precio muestral no difiriera en más de \$0.30 de la media muestral fuera \$0.95?
30. Para estimar la edad media de una población de 4000 empleados se toma una muestra de 40 empleados.
 - a. ¿Usted usaría el factor de corrección para una población finita en el cálculo del error estándar de la media? Explique.

- b. Si la desviación estándar poblacional es $\sigma = 8.2$ años, calcule el error estándar con y sin el factor de corrección para una población finita. ¿Cuál es la base para ignorar el factor de corrección para la población finita, si $n/N \leq 0.05$?
- c. ¿Cuál es la probabilidad de que la media muestral de las edades de los empleados no difiera en más de ± 2 años de la media poblacional de las edades?

7.6

Distribución muestral de \bar{p}

La proporción muestral \bar{p} es el estimador puntual de la proporción poblacional p . La fórmula para calcular la proporción muestral es

$$\bar{p} = \frac{x}{n}$$

donde

x = número de elementos de la muestra que poseen la característica de interés

n = tamaño de la muestra

Como se indicó en la sección 7.4, la proporción muestral \bar{p} es una variable aleatoria y su distribución de probabilidad se conoce como distribución muestral de \bar{p} .

DISTRIBUCIÓN MUESTRAL DE \bar{p}

La distribución muestral de \bar{p} es la distribución de probabilidad de todos los posibles valores de la proporción muestral \bar{p} .

Para determinar qué tan cerca está la proporción muestral \bar{p} de la proporción poblacional p , se necesita entender las propiedades de la distribución muestral de \bar{p} : el valor esperado de \bar{p} , la desviación estándar de \bar{p} y la forma de la distribución muestral de \bar{p} .

Valor esperado de \bar{p}

El valor esperado de \bar{p} , la media de todos los posibles valores de \bar{p} , es igual a la proporción poblacional p .

VALOR ESPERADO DE \bar{p}

$$E(\bar{p}) = p \quad (7.4)$$

donde

$E(\bar{p})$ = valor esperado de \bar{p}

p = proporción poblacional

Como $E(\bar{p}) = p$, \bar{p} es un estimador insesgado de p . Recuerde que en la sección 7.1 se encontró que en la población de EAI $p = 0.60$, siendo p la proporción de la población de administradores que han participado en el programa de capacitación de la empresa. Por tanto, el valor esperado de \bar{p} en el problema de muestreo de EAI es 0.60.

Desviación estándar de \bar{p}

Como en el caso de la desviación estándar de \bar{x} la desviación estándar de \bar{p} obedece a si la población es finita o infinita. Las dos fórmulas para calcular la desviación estándar de \bar{p} se presentan a continuación.

DESVIACIÓN ESTÁNDAR DE \bar{p}

Población finita	Población infinita	(7.5)
$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p(1-p)}{n}}$	$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$	

Al comparar las dos fórmulas (7.5) se aprecia que la única diferencia es el uso del factor de corrección para una población finita $\sqrt{(N-n)/(N-1)}$.

Como en el caso de la media poblacional \bar{x} , la diferencia entre las expresiones para una población finita y para una infinita es despreciable si el tamaño de la población finita es grande en comparación con el tamaño de la muestra. Se seguirá la misma regla recomendada para la media poblacional. Es decir, si la población es finita y $n/N \leq 0.05$ se usará $\sigma_{\bar{p}} = \sqrt{p(1-p)/n}$. Pero, si la población es finita y $n/N > 0.05$, entonces deberá usar el factor de corrección para una población finita. También, a menos que se especifique otra cosa, en este libro se supondrá que el tamaño de la población es grande en comparación al tamaño de la muestra y por tanto, el factor de corrección para una población finita no será necesario.

En la sección 7.5 se usó el error estándar de la media para referirse a la desviación estándar de \bar{x} . Se dijo que en general el término error estándar se refiere a la desviación estándar de un estimador puntual. Así, en el caso de proporciones se usa *el error estándar de la proporción* para referirse a la desviación estándar de \bar{p} . Ahora se vuelve al ejemplo de EAI para calcular el error estándar de la proporción en la muestra aleatoria simple de los 30 administradores de EAI.

En el estudio de EAI se sabe que la proporción poblacional de administradores que han participado en el programa de capacitación es $p = 0.60$. Como $n/N = 30/2\,500 = 0.012$ se puede ignorar el factor de corrección para una población finita al calcular el error estándar de la proporción. En la muestra aleatoria simple de 30 administradores, $\sigma_{\bar{p}}$ es

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.60(1-0.60)}{30}} = 0.0894$$

Forma de la distribución muestral de \bar{p}

Ahora que se conoce la media y la desviación estándar de la distribución muestral de \bar{p} , el último paso es determinar la forma de la distribución muestral. La proporción muestral es $\bar{p} = x/n$. En una muestra aleatoria simple de una población grande, el valor de x es una variable aleatoria binomial que indica el número de los elementos de la muestra que tienen la característica de interés. Como n es una constante, la probabilidad de x/n es la misma que la probabilidad de x , lo cual significa que la distribución muestral de \bar{p} también es una distribución de probabilidad discreta y que la probabilidad de cada x/n es la misma que la probabilidad de x .

En el capítulo 6 se mostró que una distribución binomial se aproxima mediante una distribución normal siempre que el tamaño de la muestra sea lo suficientemente grande para satisfacer las dos condiciones siguientes.

$$np \geq 5 \quad \text{y} \quad n(1 - p) \geq 5$$

Suponiendo que se satisfagan estas dos condiciones, la distribución de probabilidad de x en la proporción muestral, $\bar{p} = x/n$, puede aproximarse por medio de una distribución normal. Y como n es una constante, la distribución muestral de \bar{p} también se aproxima mediante una distribución normal. Esta aproximación se formula como sigue:

La distribución muestral de \bar{p} se aproxima mediante una distribución normal siempre que $np \geq 5$ y $n(1 - p) \geq 5$.

En las aplicaciones prácticas, cuando se requiere una estimación de la proporción poblacional, casi siempre se encuentra que el tamaño de la muestra es suficientemente grande para poder usar la aproximación normal para la distribución muestral de \bar{p} .

Recuerde que en el problema de muestreo de EAI la proporción poblacional de administradores que han participado en el programa de capacitación es $p = 0.60$. Con una muestra aleatoria simple de tamaño 30, se tiene $np = 30(0.60) = 18$ y $n(1 - p) = 30(0.40) = 12$. Por tanto, la distribución muestral de \bar{p} se calcula mediante la distribución normal que se muestra en la figura 7.8.

Valor práctico de la distribución muestral de \bar{p}

El valor práctico de la distribución muestral de \bar{p} es que permite obtener información probabilística acerca de la diferencia entre la proporción muestral y la proporción poblacional. Por ejemplo, en el problema de EAI, el director de personal desea saber cuál es la probabilidad de obtener un valor de \bar{p} que no difiera en más de 0.05 de la proporción poblacional de los administradores de EAI que han participado en el programa de capacitación. Es decir, ¿cuál es la probabilidad de tener una muestra en la que \bar{p} esté entre 0.55 y 0.65? El área sombreada de la figura 7.9 corres-

FIGURA 7.8 DISTRIBUCIÓN MUESTRAL DE \bar{p} , PROPORCIÓN DE ADMINISTRADORES QUE HAN PARTICIPADO EN EL PROGRAMA DE CAPACITACIÓN DE EAI

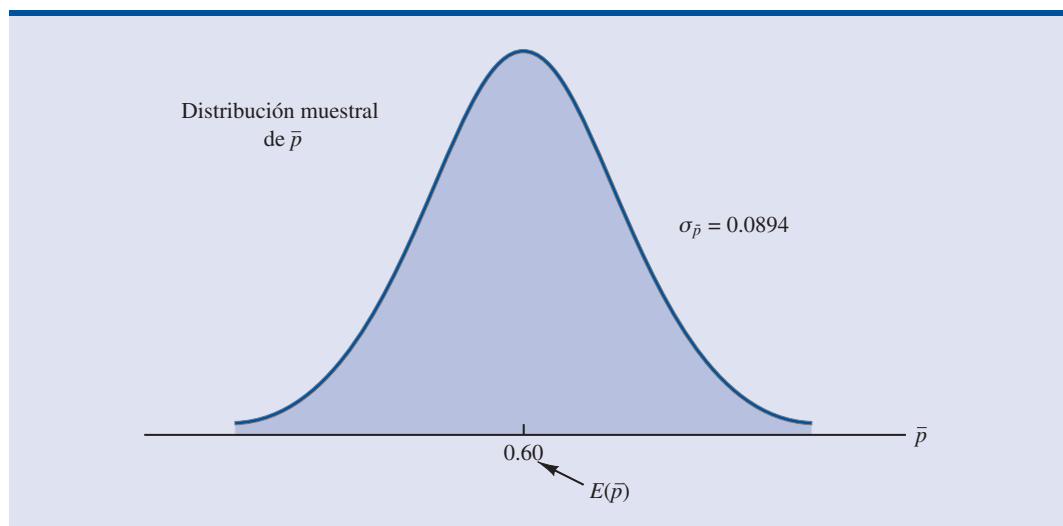
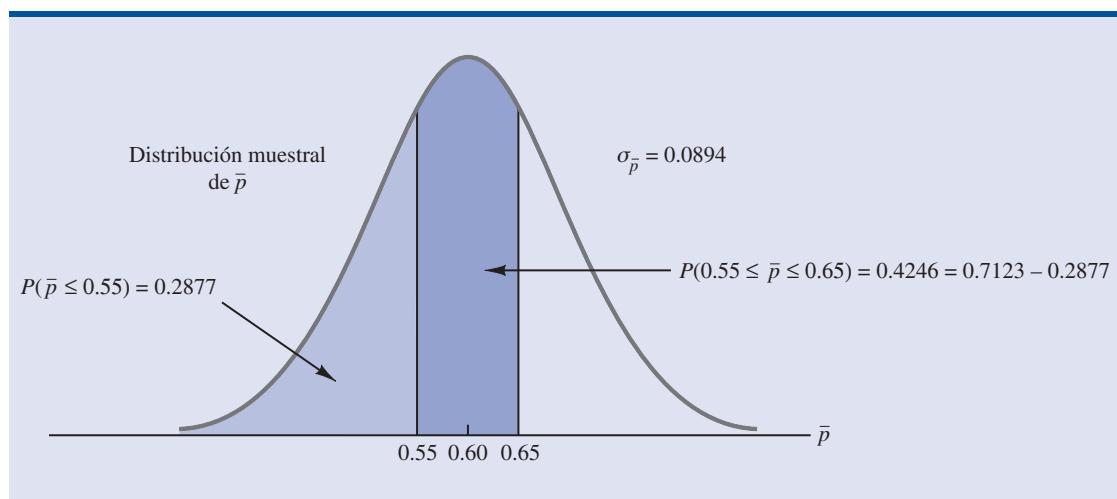


FIGURA 7.9 PROBABILIDAD DE QUE \bar{p} ESTÉ ENTRE 0.55 Y 0.65

ponde a esta probabilidad. A partir de que la distribución muestral de \bar{p} se aproxima mediante una distribución normal con media 0.60 y error estándar de la proporción $\sigma_{\bar{p}} = 0.0894$, se encuentra que la variable aleatoria normal estándar correspondiente a $\bar{p} = 0.65$ tiene el valor $z = (0.65 - 0.60)/0.0894 = 0.56$. En la tabla de probabilidad normal estándar aparece que la probabilidad acumulada que corresponde a $z = 0.56$ es 0.7123. De manera similar para $\bar{p} = 0.55$, se encuentra que $z = (0.55 - 0.60)/0.0894 = -0.56$. En la misma tabla y correspondiente a $z = -0.56$ es 0.2877. De esta manera, la probabilidad de seleccionar una muestra en la cual el valor de \bar{p} no difiera más de 0.05 de la proporción poblacional p está dada por $0.7123 - 0.2877 = 0.4246$.

Si se aumenta el tamaño de la muestra a $n = 100$, el error estándar de la proporción se convierte en

$$\sigma_{\bar{p}} = \sqrt{\frac{0.60(1 - 0.60)}{100}} = 0.049$$

Con una muestra de 100 administradores de EAI, se calcula ahora la probabilidad de que la proporción muestral tenga un valor que no difiera en más de 0.05 de la proporción poblacional. Como la distribución muestral es aproximadamente normal, con media 0.60 y desviación estándar 0.049, se puede usar la tabla de probabilidad normal estándar para hallar el área o probabilidad. Para $\bar{p} = 0.65$, se tiene $z = (0.65 - 0.60)/0.049 = 1.02$. La tabla de probabilidad normal estándar arroja que la probabilidad acumulada correspondiente a $z = 1.02$ es 0.8461. De manera similar, para $\bar{p} = 0.55$, se tiene que $z = (0.55 - 0.60)/0.049 = -1.02$. Se encuentra que la probabilidad acumulada correspondiente a $z = -1.02$ es 0.1539. Por tanto, si el tamaño de la muestra aumenta de 30 a 100, la probabilidad de que la proporción muestral \bar{p} no difiera en más de 0.05 de la proporción poblacional aumenta a $0.8461 - 0.1539 = 0.6922$.

Ejercicios

Métodos

31. De una muestra aleatoria de tamaño 100 de una población en la que $p = 0.40$.
 - a. ¿Cuál es el valor esperado de \bar{p} ?
 - b. ¿Cuál es el error estándar de \bar{p} ?

Autoexamen

- c. Exprese la distribución muestral de \bar{p} .
d. ¿Qué indica la distribución muestral de \bar{p} ?
32. Una proporción poblacional es 0.40. Se toma una muestra aleatoria de tamaño 200 y la proporción muestral \bar{p} se usa para estimar la proporción poblacional.
- ¿Cuál es la probabilidad de que la proporción muestral esté entre ± 0.03 de la proporción poblacional?
 - De qué la proporción muestral esté entre ± 0.05 de la proporción poblacional?
33. Suponga que la proporción poblacional es 0.55. Calcule el error estándar de la proporción, $\sigma_{\bar{p}}$, para los tamaños de muestra 100, 200, 500 y 1000. ¿Qué puede decir acerca del tamaño del error estándar a medida que el tamaño de la muestra aumenta?
34. La proporción poblacional es 0.30. ¿Cuál es la probabilidad de que las proporciones muestral y poblacional estén entre ± 0.04 con los tamaños de muestra siguientes?
- $n = 100$
 - $n = 200$
 - $n = 500$
 - $n = 1000$
 - ¿Qué ventaja tiene un tamaño grande de muestra?

Aplicaciones

Autoexamen

35. El director de una empresa piensa que 30% de los pedidos provienen de nuevos compradores. Para ver la proporción de nuevos compradores se usará una muestra aleatoria simple de 100 pedidos.
- Suponga que el director está en lo cierto y que $p = 0.30$. ¿Cuál es la distribución muestral de \bar{p} en este estudio?
 - ¿Cuál es la probabilidad de que la proporción muestral de \bar{p} esté entre 0.20 y 0.40?
 - ¿Cuál es la probabilidad que la proporción muestral de \bar{p} esté entre 0.25 y 0.35?
36. *The Cincinnati Enquirer* informa que en Estados Unidos 66% de los adultos y 87% de los jóvenes entre 12 y 17 años usan Internet (*The Cincinnati Enquirer*, 7 de febrero de 2007). Considere estos datos como proporciones poblacionales y suponga que se usará una muestra de 300 adultos y 300 jóvenes para obtener información respecto de su opinión acerca de la seguridad en Internet.
- Muestre la distribución muestral de \bar{p} , siendo \bar{p} la proporción muestral de adultos que usan Internet.
 - ¿Cuál es la probabilidad de que la diferencia entre la proporción muestral y la proporción poblacional de adultos que usan Internet no sea mayor que ± 0.04 ?
 - ¿Cuál es la probabilidad de que la diferencia entre la proporción muestral y la proporción poblacional de jóvenes que usan Internet no sea mayor que ± 0.04 ?
 - ¿Son diferentes las probabilidades del inciso b y del inciso c? Si es así, ¿por qué?
 - Responda al inciso b en el caso de que el tamaño de la muestra sea 600. ¿Es menor la probabilidad? ¿Por qué?
37. Los sondeos de *Time/CNN* entre los votantes siguieron la opinión del público respecto de los candidatos presidenciales en las votaciones del 2000. En uno de estos sondeos Yankelovich Partners empleó una muestra de 589 probables votantes (*Time*, 26 de junio de 2000). Suponga que la proporción poblacional a favor de un determinado candidato a la presidencia haya sido $p = 0.50$. Sea \bar{p} la proporción muestral en los posibles votantes que está a favor de ese candidato a la presidencia.
- Muestre la distribución muestral de \bar{p} .
 - ¿Cuál es la probabilidad de que los sondeos de *Time/CNN* indiquen que la diferencia entre las proporciones muestral y poblacional en uno de estos sondeos no sea mayor que ± 0.04 ?
 - ¿Cuál es la probabilidad de que los sondeos de *Time/CNN* indiquen que la diferencia entre las proporciones muestral y poblacional en uno de estos sondeos no sea mayor que ± 0.03 ?
 - ¿Cuál es la probabilidad de que los sondeos de *Time/CNN* indiquen que la diferencia entre las proporciones muestral y poblacional en uno de estos sondeos no sea mayor que ± 0.02 ?

38. Roper ASW realizó una encuesta para obtener información acerca de la opinión de los estadounidenses respecto al dinero y la felicidad (*Money*, octubre de 2003). Cincuenta y seis por ciento de los entrevistados dijeron revisar el estado de su bloc de cheques por lo menos una vez al mes.
- Suponga que se toma una muestra de 400 estadounidenses adultos. Indique la distribución muestral de la proporción de adultos que revisan el estado de su bloc de cheques por lo menos una vez al mes.
 - ¿Cuál es la probabilidad de que la diferencia entre la proporción muestral y la proporción poblacional no sea mayor que ± 0.02 ?
 - ¿Cuál es la probabilidad de que la diferencia entre las proporciones muestral y poblacional no sea mayor que ± 0.04 ?
39. El *Democrat and Chronicle* informa que 25% de los vuelos que llegaron al aeropuerto de San Diego en los primeros cinco meses de 2001, arribaron con retraso (*Democrat and Chronicle*, 23 de julio de 2001). Suponga que la proporción poblacional sea $p = 0.25$.
- Muestre la distribución muestral de \bar{p} , la proporción de vuelos retrasados en una muestra de 1 000 vuelos.
 - ¿Cuál es la probabilidad de que la diferencia entre las proporciones muestral y poblacional no sea mayor que ± 0.03 , si el tamaño de la muestra es 1000?
 - Responda el inciso b con una muestra de 500 vuelos.
40. The Grocery Manufacturers of America informa que 76% de los consumidores leen los ingredientes que se enumeran en la etiqueta de un producto. Suponga que la proporción poblacional es $p = 0.76$ y que de la población de consumidores se selecciona una muestra de 400 consumidores.
- Exprese la distribución muestral de la proporción muestral \bar{p} , si \bar{p} es la proporción de consumidores de la muestra que lee los ingredientes que se enumeran en la etiqueta.
 - ¿Cuál es la probabilidad de que la diferencia entre las proporciones muestral y poblacional no sea mayor que ± 0.03 ?
 - Conteste el inciso b si el tamaño de la muestra es 750 consumidores.
41. El Food Marketing Institute informa que 17% de los hogares gastan más de \$100 en productos de abarrotes. Suponga que la proporción poblacional es $p = 0.17$ y que de la población se toma una muestra aleatoria simple de 800 hogares.
- Exprese la distribución muestral de \bar{p} , la proporción muestral de hogares que gastan más de \$100 semanales en abarrotes.
 - ¿Cuál es la probabilidad de que la proporción poblacional no difiera en más de 0.02 de la proporción poblacional?
 - Conteste el inciso b en el caso de que el tamaño de la muestra sea 1600 hogares.

7.7

Propiedades de los estimadores puntuales

En este capítulo se ha mostrado que los estadísticos muestrales, como la media muestral \bar{x} , la desviación estándar muestral s y la proporción muestral \bar{p} sirven como estimadores puntuales de sus correspondientes parámetros poblacionales, μ , σ y p . Resulta interesante que cada uno de estos estadísticos muestrales sean los estimadores puntuales de sus correspondientes parámetros poblacionales. Sin embargo, antes de usar un estadístico muestral como estimador puntual, se verifica si el estimador puntual tiene ciertas propiedades que corresponden a un buen estimador puntual. En esta sección se estudian las propiedades que deben tener los buenos estimadores puntuales: insesgadez, eficiencia y consistencia.

Como hay distintos estadísticos muestrales que se usan como estimadores puntuales de sus correspondientes parámetros poblacionales, en esta sección se usará la notación general siguiente.

$$\theta = \text{el parámetro poblacional de interés}$$

$$\hat{\theta} = \text{el estadístico muestral o estimador puntual de } \theta$$

En esta notación θ es la letra griega theta y la notación $\hat{\theta}$ se lee “theta sombrero”. En general, θ representa cualquier parámetro poblacional como, por ejemplo, la media poblacional, la desvia-

ción estándar poblacional, la proporción poblacional, etc.; $\hat{\theta}$ representa el correspondiente estadístico muestral, por ejemplo, la media muestral, la desviación estándar muestral y la proporción muestral.

Insesgadez

Si el valor esperado del estadístico muestral es igual al parámetro poblacional que se estudia, se dice que el estadístico muestral es un *estimador insesgado* del parámetro poblacional.

INSESGADEZ

El estadístico muestral $\hat{\theta}$ es un estimado insesgado del parámetro poblacional θ si

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

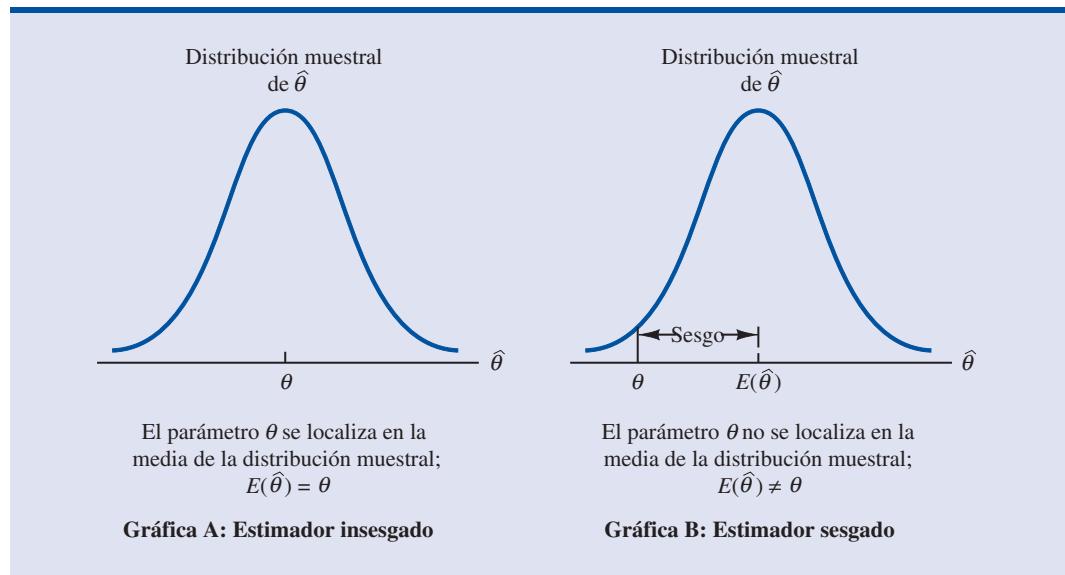
donde

$$E(\hat{\theta}) = \text{valor esperado del estadístico muestral } \hat{\theta}$$

Por tanto, el valor esperado, o media, de todos los posibles valores de un estadístico muestral insesgado es igual al parámetro poblacional que se estudia.

En la figura 7.10 se muestran los casos de los estimadores puntuales sesgado e insesgado. En la figura en que se muestra el estimador insesgado, la media de la distribución muestral es igual al valor del parámetro poblacional. En este caso los errores de estimación se equilibrان, ya que algunas veces el valor del estimador puntual $\hat{\theta}$ puede ser menor que θ y otras veces sea mayor que θ . En el caso del estimador sesgado, la media de la distribución muestral es menor o mayor que el valor del parámetro poblacional. En la gráfica B de la figura 7.10, $E(\hat{\theta})$ es mayor que θ ; así, la probabilidad de que los estadísticos muestrales sobreestimen el valor del parámetro poblacional es grande. En la figura se muestra la amplitud de este sesgo.

FIGURA 7.10 EJEMPLOS DE ESTIMADORES PUNTUALES SESGADO E INSESGADO



Al estudiar las distribuciones muestrales de la media muestral y de la proporción muestral, se vio que $E(\bar{x}) = \mu$ y que $E(\bar{p}) = p$. Por tanto, \bar{x} y \bar{p} son estimadores insesgados de sus correspondientes parámetros poblacionales μ y p .

En el caso de la desviación estándar muestral s y de la varianza muestral s^2 , se puede mostrar que $E(s^2) = \sigma^2$. Por tanto, se concluye que la varianza muestral s^2 es un estimador insesgado de la varianza poblacional σ^2 . En efecto, en el capítulo 3, cuando se presentaron las fórmulas para la varianza muestral y la desviación estándar muestral en el denominador se usó $n - 1$ en lugar de n para que la varianza muestral fuera un estimado insesgado de la varianza poblacional.

Eficiencia

Suponga que se usa una muestra aleatoria simple de n elementos para obtener dos estimadores puntuales insesgados de un mismo parámetro poblacional. En estas circunstancias preferirá usar el estimador puntual que tenga el menor error estándar, ya que dicho estimador tenderá a dar estimaciones más cercanas al parámetro poblacional. Se dice que el estimador puntual con menor error estándar tiene mayor **eficiencia relativa** que los otros.

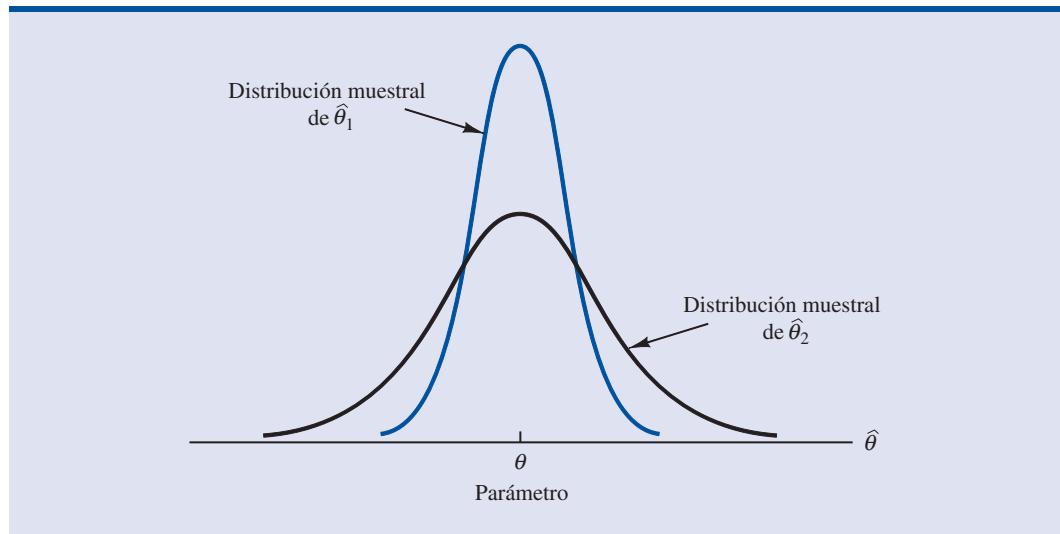
En la figura 7.11 se presentan las distribuciones muestrales de dos estimadores puntuales insesgados, $\hat{\theta}_1$ y $\hat{\theta}_2$. Observe que el error estándar de $\hat{\theta}_1$ es menor que el error estándar de $\hat{\theta}_2$; por tanto, los valores de $\hat{\theta}_1$ tienen más posibilidades de estar cerca del parámetro θ que los valores de $\hat{\theta}_2$. Como el error estándar del estimado puntual $\hat{\theta}_1$ es menor que el error estándar del estimado puntual $\hat{\theta}_2$, $\hat{\theta}_1$ es relativamente más eficiente que $\hat{\theta}_2$ y se prefiere como estimador puntual.

Consistencia

La tercera propiedad relacionada con un buen estimador puntual es la **consistencia**. Dicho de manera sencilla, un estimador puntual es consistente si el valor del estimador puntual tiende a estar más cerca del parámetro poblacional a medida que el tamaño de la muestra aumenta. En otras palabras, una muestra grande tiende a proporcionar mejor estimación puntual que una pequeña. Observe que en el caso de la media muestral \bar{x} , el error estándar de \bar{x} está dado por $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Puesto que $\sigma_{\bar{x}}$ está vinculado con el tamaño de la muestra, de manera que muestras mayores dan

Cuando se muestrean poblaciones normales, el error estándar de la media muestral es menor que el error estándar de la mediana muestral. Por tanto, la media muestral es más eficiente que la mediana muestral.

FIGURA 7.11 DISTRIBUCIONES MUESTRALES DE DOS ESTIMADORES PUNTUALES INSESGADOS



valores menores de $\sigma_{\bar{x}}$, entonces muestras de tamaño grande tienden a proporcionar estimadores puntuales más cercanos a la media poblacional μ . Mediante un razonamiento similar, concluya que la proporción muestral \bar{p} es un estimador consistente de la proporción poblacional p .

NOTAS Y COMENTARIOS

En el capítulo 3 se dijo que la media y la mediana son dos medidas de localización central. En este capítulo sólo se estudió la media. La razón es que cuando se muestrea de una población normal, en la cual la media y la mediana poblacionales son idénticas, el error estándar de la mediana es cerca de 25% mayor que el error estándar de la media. Re-

cuerde que en el problema de EAI con $n = 30$, el error estándar de la media fue $\sigma_{\bar{x}} = 730.3$. El error estándar de la mediana en este problema será $1.25 \times (730.7) = 913$. Por tanto, la media muestral es más eficiente y tendrá más probabilidad de estar dentro de una determinada distancia de la media poblacional.

7.8

Otros métodos de muestreo

Esta sección proporciona una breve introducción a otros métodos de muestreo distintos al muestreo aleatorio simple.

El muestreo aleatorio estratificado funciona mejor cuando la varianza entre los elementos de cada estrato es relativamente pequeña.

Se describió el procedimiento de muestreo aleatorio simple y se estudiaron las propiedades de las distribuciones muestrales de \bar{x} y de \bar{p} cuando se usa el muestreo aleatorio simple. Sin embargo, el muestreo aleatorio simple no es el único método de muestreo que existe. Hay otros métodos como el muestreo aleatorio estratificado, el muestreo por conglomerados y el muestreo sistemático que, en ciertas situaciones, tienen ventajas sobre el muestreo aleatorio simple. En esta sección se introducen brevemente estos métodos de muestreo. En el capítulo 22 que se encuentra en el CD que se distribuye con el texto se estudian estos métodos de muestreo con más detallamiento.

Muestreo aleatorio estratificado

En el **muestreo aleatorio estratificado** los elementos de la población primero se dividen en grupos, a los que se les llama *estratos*, de manera que cada elemento pertenezca a uno y sólo un estrato. La base para la formación de los estratos, que puede ser departamento, edad, tipo de industria, etc., está a discreción de la persona que diseña la muestra. Sin embargo, se obtienen mejores resultados cuando los elementos que forman un estrato son lo más parecido posible. La figura 7.12 es un diagrama de una población dividida en H estratos.

Una vez formados los estratos, se toma una muestra aleatoria simple de cada estrato. Existen fórmulas para combinar los resultados de las muestras de los varios estratos en una estimación

FIGURA 7.12 DIAGRAMA DE UN MUESTREO ALEATORIO ESTRATIFICADO

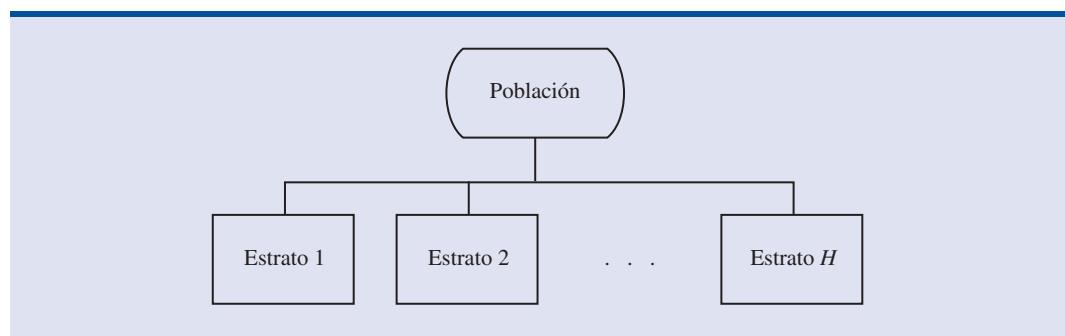
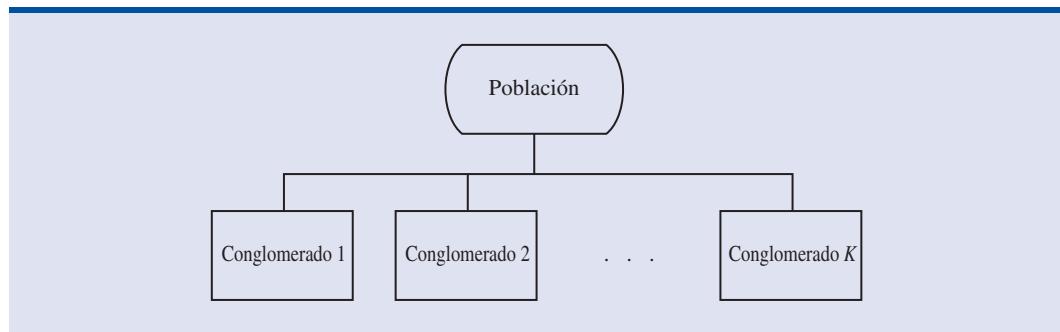


FIGURA 7.13 DIAGRAMA DEL MUESTREO POR CONGLOMERADOS

del parámetro poblacional de interés. El valor del muestreo aleatorio estratificado depende de qué tan homogéneos sean los elementos dentro de cada estrato. Si los elementos de un estrato son homogéneos, el estrato tendrá una varianza pequeña. Por tanto, con muestras relativamente pequeñas de los estratos se obtienen buenas estimaciones de las características de los estratos. Si los estratos son homogéneos, el muestreo aleatorio estratificado, proporciona resultados tan precisos como los de un muestreo aleatorio simple, pero con una muestra de tamaño total menor.

Muestreo por conglomerados

El muestreo por conglomerados funciona mejor cuando cada conglomerado proporciona una representación a menor escala de la población.

En el **muestreo por conglomerados** los elementos de la muestra primero se dividen en grupos separados, llamados *conglomerados*. Cada elemento de la población pertenece a uno y sólo un conglomerado (véase figura 7.13). Se toma una muestra aleatoria simple de los conglomerados. La muestra está formada por todos los elementos dentro de cada uno de los conglomerados que forman la muestra. El muestreo por conglomerados tiende a proporcionar mejores resultados cuando los elementos dentro de los conglomerados no son semejantes. Lo ideal es que cada conglomerado sea una representación, a pequeña escala, de la población. Si todos los conglomerados son semejantes en este aspecto, tomando en la muestra un número pequeño de conglomerados se obtendrá una buena estimación de los parámetros poblacionales.

Una de las principales aplicaciones del muestreo por conglomerados es el muestreo de áreas, en el que los conglomerados son las manzanas de una ciudad u otras áreas bien definidas. El muestreo por conglomerados requiere, por lo general, tamaños de muestra mayores que los requeridos en el muestreo aleatorio simple o en el muestreo aleatorio estratificado. Sin embargo, es posible reducir costos debido a que cuando se envía a un entrevistador a uno de los conglomerados de la muestra (por ejemplo, a una manzana de una ciudad), es posible obtener muchas observaciones en poco tiempo. Por tanto, se obtiene una muestra de tamaño grande a un costo significativamente menor.

Muestreo sistemático

Para ciertos muestreos, en especial en aquellos con poblaciones grandes, se necesita mucho tiempo para tomar una muestra aleatoria simple (hallando primero los números aleatorios y después contando y recorriendo toda una lista de la población hasta encontrar los elementos correspondientes). Una alternativa al muestreo aleatorio simple es el **muestreo sistemático**. Por ejemplo, si se quiere una muestra de tamaño 50 de una población que tiene 5000 elementos, se muestrea uno de cada $5\ 000/50 = 100$ elementos de la población. En este caso, un muestreo sistemático consiste en seleccionar en forma aleatoria uno de los primeros elementos de la lista de la población. Los otros elementos se identifican contando a partir del primer elemento 100 elementos para tomar el elemento que tenga la posición 100 en la lista de la población, a partir de este elemento se cuentan otros 100 y así se continúa. Por lo general, de esta manera es más fácil de identificar la muestra de 50 que si se usara el muestreo aleatorio simple. Como el primer elemento que se selecciona es elegido en forma aleatoria, se supone que una muestra sistemática tiene las

propiedades de una muestra aleatoria simple. Esta suposición es aplicable, en especial, cuando la lista de los elementos de la población es un orden aleatorio de los elementos.

Muestreo de conveniencia

Los métodos de muestreo hasta ahora vistos se conocen como técnicas *probabilísticas de muestreo*. Los elementos seleccionados de una población tienen una probabilidad conocida de ser incluidos en la muestra. La ventaja del muestreo probabilístico es que, por lo general, se identifica la distribución muestral del estadístico muestral correspondiente. Para determinar las propiedades de la distribución muestral se usan las fórmulas presentadas en este capítulo para el muestreo aleatorio simple. La distribución muestral permite hacer afirmaciones probabilísticas acerca del error al usar los resultados muestrales para hacer inferencias acerca de la población.

El **muestreo de conveniencia** es una técnica de *muestreo no probabilística*. Como el nombre lo indica, la muestra se determina por conveniencia. Los elementos se incluyen en la muestra sin que haya una probabilidad previamente especificada o conocida de que sean incluidos en la muestra. Por ejemplo, un profesor que realiza una investigación en una universidad puede usar estudiantes voluntarios para que constituyan una muestra; ¿la razón para elegirlos? simple, los tiene al alcance y participarán como sujetos a un costo bajo o sin costo. De manera similar, un inspector puede muestrear un cargamento de naranjas seleccionando al azar naranjas de varias de las cajas. Marcar cada naranja y usar un método probabilístico de muestreo puede no resultar práctico. Muestras como los paneles de voluntarios en investigaciones sobre los consumidores son también muestras de conveniencia.

Dichas muestras tienen la ventaja de que es relativamente fácil seleccionar la muestra y recoger los datos; sin embargo, es imposible evaluar la “bondad” de la muestra en términos de su representatividad de la población. Una muestra de conveniencia puede o no dar buenos resultados. Algunas veces los investigadores aplican los métodos estadísticos propios de muestras probabilísticas a las muestras de conveniencia, con el argumento de que la muestra de conveniencia se trata como si fuera una muestra probabilística. Sin embargo, estos argumentos no tienen fundamento y se debe tener cuidado al interpretar los resultados de muestreos de conveniencia que han sido usados para hacer inferencias acerca de la población.

Muestreo subjetivo

Otra técnica de muestreo no probabilística es el muestreo subjetivo. En este método la persona que más sabe sobre un asunto selecciona elementos de la población que considera los más representativos de la población. Este método suele ser una manera relativamente fácil de seleccionar una muestra. Por ejemplo, un reportero puede seleccionar dos o tres senadores considerando que estos senadores reflejan la opinión general de todos los senadores. Sin embargo, la calidad de los resultados muestrales depende de la persona que selecciona la muestra. Aquí también hay que tener mucho cuidado al hacer inferencias acerca de las poblaciones a partir de muestreos subjetivos.

NOTAS Y COMENTARIOS

Se recomienda usar métodos de muestreo probabilístico: muestreo aleatorio simple, muestreo aleatorio estratificado, muestreo por conglomerados o muestreo sistemático. Si se usan estos métodos existen fórmulas para evaluar la “bondad” de los resultados muestrales en términos de la cercanía de

los resultados a los parámetros poblacionales que se estiman. Con los muestreos de conveniencia o con los muestreos subjetivos no se puede estimar la bondad de los resultados. Por tanto, debe tenerse mucho cuidado al interpretar resultados basados en métodos de muestreo no probabilístico.

Resumen

En este capítulo se presentaron los conceptos de muestreo aleatorio simple y de distribución muestral. Se mostró cómo seleccionar una muestra aleatoria simple y la forma de usar los datos recolectados de la muestra para obtener estimadores puntuales de los parámetros poblacionales. Ya que distintas muestras aleatorias simples dan valores diferentes de los estimadores puntuales, los estimadores puntuales como \bar{x} y \bar{p} son variables aleatorias. A la distribución de probabilidad de una variable aleatoria de este tipo se le conoce como distribución muestral. En particular, se describieron la distribución muestral de la media muestral \bar{x} y la distribución muestral de la proporción muestral \bar{p} .

Al estudiar las características de las distribuciones muestrales de \bar{x} y de \bar{p} , se vio que $E(\bar{x}) = \mu$ y que $E(\bar{p}) = p$. Después de dar las fórmulas para la desviación estándar o error estándar de dichos estimadores, se describieron las condiciones necesarias para que las distribuciones muestrales de \bar{x} y de \bar{p} sigan una distribución normal. Otros métodos de muestreo fueron el muestreo aleatorio estratificado, el muestreo por conglomerados, el muestreo sistemático, el muestreo por conveniencia y el muestreo subjetivo.

Glosario

Parámetro Característica numérica de una población, por ejemplo, la media poblacional μ , la desviación estándar poblacional σ , la proporción poblacional p , etcétera.

Muestreo aleatorio simple Poblaciones finitas: muestra seleccionada de manera que cada una de las muestras de tamaño n tenga la misma probabilidad de ser seleccionada. Poblaciones infinitas: muestra seleccionada de manera que todos los elementos provengan de la misma población y los elementos se seleccionen de manera independiente.

Muestreo sin reemplazo Una vez que un elemento ha sido incluido en la muestra, se retira de la población y ya no se selecciona una vez más.

Muestreo con reemplazo Una vez que un elemento se ha incluido en la muestra, se regresa a la población. Un elemento ya seleccionado para la muestra puede ser seleccionado nuevamente y puede aparecer más de una vez en la muestra.

Estadístico muestral Característica muestral, por ejemplo, la media muestral \bar{x} , la desviación estándar muestral s , la proporción muestral \bar{p} , etc. El valor del estadístico muestral se usa para estimar el valor del correspondiente parámetro poblacional.

Estimador puntual Un estadístico muestral como \bar{x} , s , o \bar{p} que proporciona una estimación puntual del parámetro poblacional correspondiente.

Estimación puntual Valor de un estimador que se usa en una situación particular como estimación del parámetro poblacional.

Distribución muestral Distribución de probabilidad que consta de todos los posibles valores de un estadístico muestral.

Insesgado Propiedad de un estimador que consiste en que el valor esperado del estimador puntual es igual al parámetro poblacional que estima.

Factor de corrección para una población finita Es el factor $\sqrt{(N - n)/(N - 1)}$ que se usa en las fórmulas de $\sigma_{\bar{x}}$ y $\sigma_{\bar{p}}$ siempre que se muestre de una población finita y no de una población infinita. Sin embargo, hay una regla generalmente aceptada, ignorar el factor de corrección en una población finita siempre que $n/N \leq 0.05$.

Error estándar La desviación estándar de un estimador puntual.

Teorema del límite central Permite usar la distribución de probabilidad normal para aproximar la distribución muestral de \bar{x} siempre que la muestra sea grande.

Eficiencia relativa Dados dos estimadores puntuales insesgados de un mismo parámetro poblacional, el estimador puntual que tenga menor error estándar será más eficiente.

Consistencia Propiedad de un estimador puntual que está presente siempre que muestras más grandes dan estimaciones puntuales más cercanas al parámetro poblacional.

Muestreo aleatorio estratificado Método probabilístico en el que primero se divide la población en estratos y después se toma una muestra aleatoria simple de cada estrato.

Muestreo por conglomerados Método probabilístico en el que primero se divide la población en conglomerados y después se toma una muestra aleatoria de los conglomerados.

Muestreo sistemático Método probabilístico en el que primero se selecciona uno de los primeros k elementos de una población y después se selecciona cada k -ésimo elemento de la población.

Muestreo de conveniencia Método no-probabilístico en el que la selección de los elementos para la muestra es de acuerdo con la conveniencia.

Muestreo subjetivo Método no-probabilístico en el que la selección de los elementos para la muestra es de acuerdo con la opinión de la persona que hace el estudio.

Fórmulas clave

Valor esperado de \bar{x}

$$E(\bar{x}) = \mu \quad (7.1)$$

Desviación estándar de \bar{x} (error estándar)

$Población finita$	$Población infinita$
$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}} \left(\frac{\sigma}{\sqrt{n}} \right)$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

(7.2)

Valor esperado de \bar{p}

$$E(\bar{p}) = p \quad (7.4)$$

Desviación estándar de \bar{p} (error estándar)

$Población finita$	$Población infinita$
$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}} \sqrt{\frac{p(1-p)}{n}}$	$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$

(7.5)

Ejercicios complementarios

42. *BusinessWeek's Corporate Scoreboard* proporciona datos trimestrales sobre distintos aspectos de las acciones de 899 empresas (*BusinessWeek*, 14 de agosto de 2000). Las empresas son numeradas del 1 al 899 en el orden en que aparecen en la lista del Corporate Scoreboard. Remítase a la parte inferior de la segunda columna de dígitos aleatorios de la tabla 7.1, ignore los dos primeros dígitos de cada conjunto de números, use números de tres dígitos, empiece con el 112, lea hacia arriba de la columna para determinar las ocho primeras empresas a incluir en una muestra aleatoria simple.
43. Los estadounidenses están cada vez más preocupados por el aumento en los costos de Medicare. En 1990 el promedio de gastos anuales de un derechohabiente de Medicare era \$3267; en el 2003 el promedio de gastos anuales de un derechohabiente de Medicare era \$6883 (*Money*, otoño de

- 2003). Suponga que usted contrata a una empresa consultora para tomar una muestra de 50 de los derechohabientes de Medicare en 2003 con objeto de investigar los gastos. Asuma que la desviación estándar en los gastos de Medicare en 2003 haya sido de \$2000.
- a. Muestre la distribución muestral de la media, en muestras de tamaño cincuenta, de los gastos de derechohabientes de Medicare en 2003.
 - b. ¿Cuál es la probabilidad de que la media muestral no se aleje más de $\pm \$300$ de la media poblacional?
 - c. ¿Cuál es la probabilidad de que la media muestral sea mayor que \$7500? Si la empresa que contrató le dice que la media muestral en los derechohabientes que entrevistó es \$7500, ¿dudaría que la empresa contratada hubiera hecho un muestreo adecuado? ¿Por qué sí o por qué no?
44. *BusinessWeek* encuesta a ex alumnos de administración 10 años después de terminados sus estudios (*BusinessWeek*, 22 de septiembre de 2003). Uno de los hallazgos fue que gastan en promedio \$115.50 semanales en comidas sociales. A usted se le pide que realice un estudio con una muestra de 40 de estos ex alumnos.
- a. Muestre la distribución muestral de \bar{x} , la media muestral de los gastos de 40 ex alumnos.
 - b. ¿Cuál es la probabilidad de que la media muestral no se aleje en más o menos de \$10 de la media poblacional?
 - c. Suponga que encuentra una media muestral de \$100. ¿Cuál es la probabilidad de hallar una media muestral de \$100 o menos? ¿Consideraría que los ex alumnos de esta muestra son un grupo inusual respecto a estos gastos? ¿Por qué sí o por qué no?
45. El tiempo promedio que un estadounidense ve televisión es 15 horas por semana (*Money*, noviembre de 2003). Suponga que se toma una muestra de 60 estadounidenses para investigar con más detalle sus hábitos a este respecto. Asuma que la desviación estándar poblacional en las horas de televisión semanales es $\sigma = 4$ horas.
- a. ¿Cuál es la probabilidad de que la media muestral no se aleje más o menos de 1 hora de la media poblacional?
 - b. ¿Cuál es la probabilidad de que la media muestral no se aleje más o menos de 45 minutos de la media poblacional?
46. En Indiana el salario anual promedio de un empleado del gobierno federal es \$41 979 (*The World Almanac*, 2001). Use esta cifra como media poblacional y suponga que la desviación estándar poblacional es $\sigma = \$5000$. Suponga que se selecciona una muestra de 50 de estos empleados del gobierno federal.
- a. ¿Cuál es el valor del error estándar de la media?
 - b. ¿Cuál es la probabilidad de que la media muestral sea mayor que \$41 979?
 - c. ¿Cuál es la probabilidad de que la media muestral no se aleje más o menos de \$1000 de la media poblacional?
 - d. ¿Qué tanto variaría la probabilidad del inciso c si el tamaño de la muestra se aumentara a 100?
47. Tres empresas llevan inventarios de distintos tamaños. El inventario de la empresa A contiene 2000 artículos, el inventario de la empresa B tiene 5000 artículos y el inventario de la empresa C 10 000. La desviación estándar poblacional de los costos de los artículos en los inventarios de estas empresas es $\sigma = 144$. Un consultor de estadística recomienda que cada empresa tome una muestra de 50 artículos de su inventario para obtener una estimación estadística válida del costo promedio por artículo. Los administradores de la empresa más pequeña opinan que como su población es menor se podrá hacer la estimación con una muestra mucho más pequeña de la que se requiere para la empresa más grande. Sin embargo, el consultor opina que para tener el mismo error estándar y, por tanto, la misma precisión en los resultados muestrales, todas las empresas deberán emplear el mismo tamaño de muestra, sin importar el tamaño de la población.
- a. Con el factor de corrección para una población finita, calcule el error estándar de cada una de las tres empresas para un tamaño de muestra de 50.
 - b. ¿Cuál es la probabilidad en cada empresa de que la media muestral \bar{x} esté a no más de ± 25 de la media poblacional μ ?

48. Un investigador informa sobre sus resultados diciendo que el error estándar de la media es 20. La desviación estándar poblacional es 500.
- ¿De qué tamaño fue la muestra usada en esta investigación?
 - ¿Cuál es la probabilidad de que la estimación puntual esté a no más de ± 25 de la media poblacional?
49. Un inspector de control de calidad vigila periódicamente un proceso de producción. El inspector selecciona muestras aleatorias simples de artículos ya terminados y calcula la media muestral del peso del producto \bar{x} . Si en un periodo largo se encuentra que 5% de los valores de \bar{x} son mayores que 2.1 libras y 5% son menores que 1.9 libras. ¿Cuáles son la media y la desviación estándar de la población de los productos elaborados en este proceso?
50. Cerca de 28% de las empresas tienen como propietario a una mujer (*The Cincinnati Enquirer*, 26 de enero de 2006). Responda estas preguntas con base en una muestra de 240 empresas.
- Muestre la distribución muestral de \bar{p} , la proporción muestral de las empresas propiedad de una mujer.
 - ¿Cuál es la probabilidad de que la proporción muestral esté a no más de ± 0.04 de la proporción poblacional?
 - ¿Cuál es la probabilidad de que la proporción muestral esté a no más de ± 0.02 de la proporción poblacional?
51. Una empresa de investigación de mercado realiza encuestas telefónicas con una tasa de respuesta de 40%, de acuerdo con la experiencia. ¿Cuál es la probabilidad de que en una muestra de 400 números telefónicos 150 personas cooperen y respondan las preguntas? En otras palabras, ¿cuál es la probabilidad de que la proporción muestral sea al menos $150/400 = 0.375$?
52. Los publicistas contratan proveedores de servicios de Internet y motores de búsqueda para poner su publicidad en los sitios Web. Pagan una cuota de acuerdo con el número de clientes potenciales que hacen clic en su publicidad. Por desgracia, el fraude por clic —la práctica de hacer clic en una publicidad con el solo objeto de aumentar las ganancias— se ha convertido en un problema. Cuarenta por ciento de los publicistas se quejan de haber sido víctima de fraude por clic (*BusinessWeek*, 13 de marzo de 2006). Suponga que se toma una muestra aleatoria de 380 publicistas con objeto de tener más información acerca de cómo son afectados por este fraude por clic.
- ¿Cuál es la probabilidad de que la proporción muestral esté a no más de ± 0.04 de la proporción poblacional?
 - ¿Cuál es la probabilidad de que la proporción muestral sea mayor que 0.45?
53. La proporción de personas aseguradas con una compañía de seguros para automóviles que tienen una multa de tráfico en el periodo de un año es 0.15
- Indique la distribución muestral de \bar{p} si se emplea una muestra aleatoria de 150 asegurados para determinar la proporción de quienes han tenido por lo menos una multa en un año.
 - ¿Cuál es la probabilidad de que la proporción muestral esté a no más de ± 0.03 de la proporción poblacional?
54. Lori Jeffrey es un exitoso representante de ventas de libros universitarios, tiene éxito en sus recomendaciones de libros en 25% de sus llamadas. Considere sus llamadas de ventas de un mes como muestra de todas sus posibles llamadas, suponga que en el análisis estadístico de los datos se encuentra que el error estándar de la proporción es 0.0625.
- ¿De qué tamaño fue la muestra que se usó en el análisis? Es decir, ¿cuántas llamadas hizo Lori Jeffrey en ese mes?
 - Sea \bar{p} la proporción muestral de éxitos en sus recomendaciones de libros en ese mes. Muestre la distribución muestral de \bar{p} .
 - Mediante la distribución muestral de \bar{p} , calcule la probabilidad de que el vendedor tenga éxito en 30% o más de las llamadas de ventas en el lapso de un mes.

Apéndice 7.1 Valor esperado y desviación estándar de \bar{x}

En este apéndice se presentan las bases matemáticas de las expresiones $E(\bar{x})$, valor esperado de \bar{x} , ecuación (7.1), y $\sigma_{\bar{x}}$, desviación estándar de \bar{x} , ecuación (7.2).

Valor esperado de \bar{x}

Se tiene una población que tiene media μ y varianza σ^2 . Se selecciona una muestra aleatoria de tamaño n cuyas observaciones se denotan x_1, x_2, \dots, x_n . La media muestral \bar{x} se calcula como sigue.

$$\bar{x} = \frac{\sum x_i}{n}$$

Si se repiten los muestreos aleatorios de tamaño n , \bar{x} será una variable aleatoria que tomará diferentes valores dependiendo de los n elementos que formen la muestra. El valor esperado de la variable aleatoria \bar{x} es la media de todos los posibles valores \bar{x} .

$$\begin{aligned}\text{Media de } \bar{x} &= E(\bar{x}) = E\left(\frac{\sum x_i}{n}\right) \\ &= \frac{1}{n}[E(x_1 + x_2 + \dots + x_n)] \\ &= \frac{1}{n}[E(x_1) + E(x_2) + \dots + E(x_n)]\end{aligned}$$

Para cada x_i se tiene $E(x_i) = \mu$; por tanto,

$$\begin{aligned}E(\bar{x}) &= \frac{1}{n}(\mu + \mu + \dots + \mu) \\ &= \frac{1}{n}(n\mu) = \mu\end{aligned}$$

Este resultado indica que la media de todos los posibles valores de \bar{x} es igual a la media poblacional μ . Es decir $E(\bar{x}) = \mu$.

Desviación estándar de \bar{x}

Se tiene, de nuevo, una población con media μ y varianza σ^2 y una media muestral dada por

$$\bar{x} = \frac{\sum x_i}{n}$$

Se sabe que \bar{x} es una variable aleatoria que toma distintos valores en distintas muestras aleatorias de tamaño n , dependiendo de los elementos que constituyen la muestra. Lo que sigue es una deducción de la fórmula para la desviación estándar de los valores de \bar{x} , $\sigma_{\bar{x}}$, en el caso en el que la población sea infinita. La deducción de la fórmula para $\sigma_{\bar{x}}$ cuando la población es finita y el muestreo se hace sin reemplazo es más complicada y queda fuera de los alcances de este texto.

De regreso al caso de una población infinita, recuerde que una muestra aleatoria simple de una población infinita, consta de observaciones x_1, x_2, \dots, x_n que son independientes. Las dos expresiones siguientes son fórmulas generales para la varianza de una variable aleatoria.

$$\text{Var}(ax) = a^2 \text{Var}(x)$$

donde a es una constante y x es una variable aleatoria, y

$$\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y)$$

donde x y y son variables aleatorias *independientes*. Usando las ecuaciones anteriores, se puede deducir la fórmula para la varianza de la variable \bar{x} como sigue.

$$\text{Var}(\bar{x}) = \text{Var}\left(\frac{\sum x_i}{n}\right) = \text{Var}\left(\frac{1}{n}\sum x_i\right)$$

Entonces, como $1/n$ es una constante, se tiene

$$\begin{aligned}\text{Var}(\bar{x}) &= \left(\frac{1}{n}\right)^2 \text{Var}(\sum x_i) \\ &= \left(\frac{1}{n}\right)^2 \text{Var}(x_1 + x_2 + \dots + x_n)\end{aligned}$$

En el caso de una población infinita, las variables aleatorias x_1, x_2, \dots, x_n son independientes, lo que nos permite escribir

$$\text{Var}(\bar{x}) = \left(\frac{1}{n}\right)^2 [\text{Var}(x_1) + \text{Var}(x_2) + \dots + \text{Var}(x_n)]$$

Para toda x_i , se tiene $\text{Var}(x_i) = \sigma^2$; por tanto se tiene

$$\text{Var}(\bar{x}) = \left(\frac{1}{n}\right)^2 (\sigma^2 + \sigma^2 + \dots + \sigma^2)$$

Como en esta expresión hay n valores σ^2 , se tiene

$$\text{Var}(\bar{x}) = \left(\frac{1}{n}\right)^2 (n\sigma^2) = \frac{\sigma^2}{n}$$

Sacando ahora la raíz cuadrada, se obtiene la fórmula de la desviación estándar de \bar{x} .

$$\sigma_{\bar{x}} = \sqrt{\text{Var}(\bar{x})} = \frac{\sigma}{\sqrt{n}}$$

Apéndice 7.2 Muestreo aleatorio con Minitab

Si en un archivo se encuentra una lista con los elementos de una población, se puede usar Minitab para seleccionar una muestra aleatoria simple. Por ejemplo, en la columna 1 del conjunto de datos MetAreas se proporciona una lista de las 100 principales áreas metropolitanas de Estados Unidos y Canadá (*Places Rated Almanac-The Millennium Edition 2000*). La columna 2 contiene

TABLA 7.6 PUNTUACIÓN GENERAL PARA LAS PRIMERAS 10 ÁREAS METROPOLITANAS EN EL CONJUNTO DE DATOS METAREAS



Área metropolitana	Puntuación
Albany, NY	64.18
Albuquerque, NM	66.16
Appleton, WI	60.56
Atlanta, GA	69.97
Austin, TX	71.48
Baltimore, MD	69.75
Birmingham, AL	69.59
Boise City, ID	68.36
Boston, MA	68.99
Buffalo, NY	66.10

la puntuación general dada a cada área. En la tabla 7.6 se presentan las primeras 10 áreas metropolitanas con sus puntuaciones correspondientes.

Suponga que pretende seleccionar una muestra aleatoria simple de 30 áreas metropolitanas con objeto de hacer un estudio sobre el costo de la vida en Estados Unidos y Canadá. Para seleccionar la muestra aleatoria se siguen los pasos que se indica a continuación.

- Paso 1.** Seleccionar el menú desplegable **Calc**
- Paso 2.** Elegir **Random Data**
- Paso 3.** Elegir **Sample From Columns**
- Paso 4.** Cuando aparezca el cuadro de diálogo **Sample From Columns:**
 - Ingresar 30 en el cuadro **Sample**
 - Ingresar C1 C2 en el cuadro que se encuentra debajo
 - Ingresar C3 C4 en el cuadro **Store samples in**
- Paso 5.** Hacer clic en **OK**

La muestra aleatoria con las 30 áreas metropolitanas aparece en las columnas C3 y C4.

Apéndice 7.3 Muestreo aleatorio con Excel

Si en un archivo se encuentra una lista con los elementos de una población, Excel se podrá usar para seleccionar una muestra aleatoria simple. Por ejemplo, en la columna A del conjunto de datos MetAreas se proporciona una lista de las 100 principales áreas metropolitanas de Estados Unidos y Canadá (*Places Rated Almanac-The Millennium Edition 2000*).

La columna B contiene la puntuación general dada a cada área. En la tabla 7.6 se presentan las primeras 10 áreas metropolitanas con sus puntuaciones correspondientes.

Suponga que quiere seleccionar una muestra aleatoria simple de 30 áreas metropolitanas con objeto de hacer un estudio sobre el costo de la vida en Estados Unidos y Canadá.

Los renglones de cualquier conjunto de datos en Excel se pueden colocar en orden aleatorio agregando una columna al conjunto de datos y llenando la columna con números aleatorios mediante la función =ALEATORIO(); después con la herramienta de Excel Orden ascendente aplicada a la columna de números aleatorios, los renglones del conjunto de datos quedarán reordenados aleatoriamente. La muestra aleatoria de tamaño n aparecerá en los n primeros renglones del conjunto de datos reordenado.

En el conjunto de datos MetAreas, los encabezados aparecen en el renglón 1 y las 100 áreas metropolitanas se encuentran en los renglones 2 a 101. Para seleccionar una muestra aleatoria de 30 áreas metropolitanas siga los pasos siguientes.

- Paso 1.** Ingresar =ALEATORIO() en la celda C2.
- Paso 2.** Copiar la celda C2 a las celdas C3:C101
- Paso 3.** Seleccionar cualquier celda de la columna C
- Paso 4.** Clic en el botón **Orden ascendente** de la barra de herramientas.

La muestra aleatoria con 30 áreas metropolitanas aparecerá en los renglones 2 a 31 del conjunto de datos reordenado. Los números aleatorios de la columna C ya no se necesitan y pueden borrarse si se desea.