7

MUESTREO Y DISTRIBUCIONES MUESTRALES

CONTENIDO

LA ESTADÍSTICA EN LA PRÁCTICA

Una máquina de papel de Mead, llamada "Espíritu de Escanaba", es una de las más grandes y modernas en el mundo.

MEAD CORPORATION*
Dayton, Ohio

Esta empresa localizada en Dayton, Ohio, se ha diversificado dentro del ramo de productos forestales y de papel; fabrica papel y celulosa, produce madera y transforma cartón en cajas de empaque y cartones para envasar bebidas. La capacidad de distribución de la compañía le permite llevar al mercado muchos de sus propios productos, que incluyen papel, artículos escolares y de papelería, la empresa cuenta con un departamento de mercadotecnia. Su grupo consultor interno recurre al muestreo para tomar decisiones y proporcionar una diversidad de información que le permite obtener grandes beneficios en la productividad y así permanecer competitivamente en su ramo.

Por ejemplo, Mead mantiene extensas propiedades forestales, que suministran los árboles, que constituyen la materia prima de muchos de sus productos. Los directores necesitan información confiable y exacta acerca de las áreas boscosas para evaluar las posibilidades que tiene la empresa de satisfacer sus necesidades futuras de materia prima. ¿Cuál es el volumen actual en los bosques? ¿Cuál fue el crecimiento de ellos en el pasado? ¿Cuál es su crecimiento proyectado a futuro? Al contar con respuestas a estas importantes preguntas, los gerentes de Mead pueden desarrollar planes para el futuro, incluyendo la planeación a largo plazo y los calendarios de tala de los árboles.

¿Cómo obtiene Mead la información que necesita sobre sus vastos bosques? Los datos reunidos a partir de muestras de porciones de los bosques son la base para conocer la población de árboles que posee la compañía. Para identificar las porciones de la muestra, se dividen primero los bosques en tres secciones, con base en la ubicación y tipos de árboles. Con mapas y tablas de números aleatorios, los analistas de Mead identifican muestras aleatorias de plantaciones de 1/5 a 1/7 de acre (entre 1500 y 1800 m^2) en cada sección. En esas porciones es donde los silvicultores de Mead reúnen datos y conocen la población del bosque.

Los silvicultores de la organización participan en el proceso de recolección de datos de campo. Periódicamente, equipos de dos personas reúnen información de cada árbol en cada porción de la muestra. Los datos obtenidos se capturan en el sistema computacional de inventario forestal continuo, cuyos informes incluyen varios resúmenes de distribución de frecuencias, que contienen medidas estadísticas sobre tipos de árboles, volumen actual del bosque, tasas de crecimiento en el pasado, y volumen y crecimiento proyectados en el futuro. El muestreo y los resúmenes estadísticos asociados de los datos muestrales se usan en los informes que son esenciales para la administración eficaz de los activos de Mead en bosques y demás zonas forestales.

En este capítulo aprenderá el muestreo aleatorio simple y el proceso de selección de la muestra. Además, aprenderá cómo se usan estadísticos como la media y la proporción de una muestral para estimar la media y la proporción de la población. También se presenta el concepto de distribución muestral, tan importante en esta área de estudio.

---

*El Dr. Edward P. Winkofsky, de Mead Corporation proporcionó esta "Estadística en la práctica".

En el capítulo 1 definimos a la *población* y la *muestra* como dos aspectos importantes de un estudio estadístico. Esas definiciones son:

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

El objetivo de la *inferencia estadística* es obtener información acerca de una población, partiendo de la información que contiene una muestra. Comenzaremos con la descripción de dos casos en los que se lleva a cabo un muestreo para proporcionar información, sobre una población, a un gerente o a quien tome decisiones.

1. Un fabricante de neumáticos ha desarrollado un nuevo producto que, según cree, tendrá una mayor duración en relación con las millas recorridas comparado con la línea actual de neumáticos. Para evaluar el nuevo neumático, los gerentes necesitan un estimado (o una estimación) de la media de las millas que dura el nuevo producto. Selecciona una muestra de 120 neumáticos nuevos para probarlos. El resultado de la prueba es una media de la muestra de 36,500 millas. En consecuencia, se usan 36,500 millas como estimado de la media para la población de neumáticos nuevos.
2. Los miembros de un partido político desean respaldar a determinado candidato en la elección senatorial. Para decidir si el candidato participará en la elección primaria, los dirigentes necesitan un estimado de la proporción de votantes empadronados que respaldan al candidato. Por el tiempo y el costo necesarios para recabar los datos de cada individuo de la población de votantes registrados resultan prohibitivos. En consecuencia, se selecciona una muestra de 400 votantes registrados. Si 160 de ellos indican su preferencia hacia el candidato, un estimado de la proporción de la población de votantes registrados que favorece al candidato es 160/400 = .40.

Los ejemplos anteriores muestran cómo se pueden emplear el muestreo y los resultados de una muestra para obtener estimados de las características de una población. Observe que en el ejemplo de la duración en millas, el reunir datos sobre la vida implica gastar cada neumático probado. Claramente no es posible probar todos los neumáticos de la población; una muestra es el único método realista para obtener los datos buscados de duración. En el ejemplo de la elección primaria, teóricamente es posible entrevistar a cada votante registrado en la población, pero el tiempo y costo de hacerlos son prohibitivos, así que se prefiere una muestra de los votantes registrados.

Una media de muestra suministra un estimado de una media poblacional, y una proporción de muestra suministra un estimado de una proporción poblacional. Con estimados como los anteriores cabe esperar cierto error de muestreo, o valor ±. En este capítulo, un argumento clave es que se pueden aplicar métodos estadísticos para hacer afirmaciones probabilísticas acerca del tamaño del error de muestreo.

Los ejemplos ilustran algunas de las razones por las que se recurre a las muestras. Sin embargo, es importante darse cuenta de que los resultados de la muestra sólo dan *estimados* (o estimaciones) de los valores de las características de la población. Esto es, no esperamos que la media de la muestra 36,500 millas sea exactamente igual que la media, en millas, de todos los neumáticos de la población; tampoco esperamos que sea exactamente el 40% de la población registrada de votantes el que favorezca al candidato. La razón simplemente es que la muestra contiene sólo una parte de la población. Con los métodos adecuados de muestreo, los resultados muestrales darán "buenos" estimados de las características de la población. Pero, ¿qué tan buenos esperamos que sean esos resultados? Por fortuna, disponemos de procedimientos estadísticos para contestar esta pregunta.

En este capítulo indicaremos cómo se puede usar el muestreo aleatorio simple para seleccionar una muestra de una población. A continuación mostraremos cómo se pueden emplear los datos obtenidos de una muestra aleatoria simple para calcular los estimados de la media de una población, una desviación estándar de ésta y una proporción de la misma. Además, describiremos el concepto de distribución muestral. Como veremos, el conocer la distribución muestral pertinente es lo que nos permite afirmar la bondad de los resultados de la muestra. La última parte describe algunas alternativas del muestreo aleatorio simple que se emplean con frecuencia en la práctica.

7.1 EL PROBLEMA DE MUESTREO EN ELECTRONICS ASSOCIATES

Al director de personal de Electronics Associates, Inc. (EAI) se le ha asignado la tarea de elaborar un perfil de los 2500 gerentes de la empresa. Las características por identificar son, entre otras, el sueldo anual promedio y la proporción de gerentes que terminaron el programa de adiestramiento administrativo de la empresa.

Si definimos a los 2500 gerentes como la población a estudiar, podemos determinar el salario anual y el estado de adiestramiento en el programa para cada individuo, consultando los registros del personal que tiene la empresa. Supongamos que ya se hizo lo anterior y que contamos con información de todos los 2500 gerentes.

Si empleamos las fórmulas de la media y la desviación estándar de la población, que presentamos en el capítulo 3, podremos calcular esas medidas para el salario anual. Supongamos que ya se hicieron esos cálculos y que los resultados fueron:

$$\text{Media de la población:} \quad \mu = \$51{,}800 \text{ dólares}$$

$$\text{Desviación estándar de la población:} \quad \sigma = \$4000 \text{ dólares}$$

Además, supongamos que 1500 de los 2500 gerentes han terminado con el programa de adiestramiento. Si p representa la proporción de la población que terminó el programa, vemos que $p = 1500/2500 = .60$.

Un *parámetro* es una característica numérica de una población. Por ejemplo, la media del salario anual de la población ($\mu = \$51{,}800$), su desviación estándar ($\sigma = \$4000$), y la proporción que terminó el programa de adiestramiento ($p = .60$) son parámetros de la población de los gerentes de EAI.

El asunto que deseamos considerar es cómo el director de personal puede obtener estimados de esos parámetros poblacionales con una muestra de gerentes, en lugar de hacerlo con los 2500 individuos de la población. Supongamos que se usará una muestra de 30 gerentes. Es claro que el tiempo y el costo de desarrollar un perfil para 30 gerentes serían mucho menores que para toda la población. Si el director de personal pudiera estar seguro de que la muestra de 30 gerentes suministra la información adecuada sobre la población de 2500 gerentes, preferirá trabajar con la muestra que con toda la población. Investiguemos la posibilidad de usar una muestra para el estudio de EAI, describiendo primero cómo identificar una muestra de 30 gerentes.

7.2 MUESTREO ALEATORIO SIMPLE

Se pueden usar varios métodos para seleccionar una muestra a partir de una población; uno de los más comunes es el *muestreo aleatorio simple*. La definición de este método y el proceso de seleccionar una muestra aleatoria simple dependen de si la población es *finita o infinita*. Como en el problema de muestreo de EAI interviene una población finita de 2500 gerentes, primero describiremos el muestreo para poblaciones finitas.

Muestreo para poblaciones finitas

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

Muestra aleatoria simple (población finita)

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

Un procedimiento para identificar una muestra aleatoria simple a partir de una población finita es seleccionar uno por uno los elementos que constituyen a la muestra, de tal modo que cada uno de los elementos que aún queden en la población tengan la misma probabilidad de ser seleccionados. Al muestrear n elementos en esa forma, se satisfará la definición de una muestra aleatoria simple de una población finita.

Para seleccionar una muestra aleatoria simple de la población finita en el problema de EAI, supondremos primero que se han numerado los 2500 gerentes, asignándoles un número progresivo, es decir, 1, 2, 3, . . . , 2499 y 2500, en el orden en que aparecen en el archivo del personal de la empresa. A continuación, consultaríamos la tabla 7.1 de números aleatorios.* En el primer renglón de la tabla cada dígito: 6, 3, 2, . . . , es aleatorio, porque tiene igual probabilidad de presentarse. Como el número máximo en la lista de la población de gerentes de EAI es 2500 y tiene 4 dígitos, seleccionaremos números aleatorios de la tabla en grupos o conjuntos de cuatro dígitos. Al usar el primer renglón de la tabla 7.1, los números aleatorios de 4 dígitos son:

También se pueden emplear números aleatorios generados en computadora, para implementar el proceso de selección de muestras aleatorias simples.

6327 1599 8671 7445 1102 1514 1807

En vista de que los números de la tabla son aleatorios, esos grupos de cuatro dígitos son igualmente probables.

Ahora podemos usar los números de cuatro dígitos para dar, a cada elemento de la población, la misma oportunidad de ser seleccionado para la muestra. El primer número, 6327, es mayor que 2500, no corresponde a un elemento de la población y, en consecuencia, lo descartamos. El segundo número, 1599, está entre 1 y 2500. Así, el primer individuo seleccionado para la muestra es el gerente número 1599 en la lista de los gerentes de EAI. Al continuar el proceso, pasamos por alto 8671 y 7445 e identificamos a los individuos 1102, 1514 y 1807 como los siguientes de la muestra. Este proceso de selección de gerentes continúa hasta haber obtenido la muestra aleatoria simple de tamaño 30.

Al llevar a cabo el proceso de selección de la muestra aleatoria simple, es posible que aparezca de nuevo, en la tabla, un número aleatorio que antes ya se haya usado, antes de completar la selección de la muestra de 30 gerentes de EAI. Como se trata de seleccionar tan sólo una vez a los gerentes, todos los números aleatorios que ya se hayan empleado no se vuelven a tomar en cuenta, porque los gerentes correspondientes ya son parte de la muestra. La selección de una muestra en esta forma se llama muestrear *sin remplazo*. Si hubiéramos seleccionado la muestra en tal forma que los números aleatorios usados antes se aceptaran, y algunos gerentes se pudieran incluir dos o más veces en la muestra, hubiéramos muestreado *con remplazo*. El muestreo con remplazo es una forma válida de identificar una muestra aleatoria simple. Sin embargo, lo que se usa con más frecuencia es el muestreo con remplazo. Cuando digamos muestreo aleatorio simple supondremos que ese muestreo es sin remplazo.

*Las tablas de números aleatorios están disponibles en una variedad de manuales como The Rand Corporation's. *A Millon Random Digits with 100,000 Normal Deviates*, The Free Press, 1983.

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

Se pueden seleccionar números aleatorios en cualquier lugar de la tabla. Aunque comenzamos por el primer renglón como ejemplo, podríamos haber comenzado en cualquier otro punto de la tabla y continuado en cualquier dirección. Una vez elegido el punto de partida, se recomienda usar un procedimiento sistemático predeterminado, por ejemplo, avanzar por columnas o por renglones, para seleccionar los números siguientes.

Muestreo para poblaciones infinitas

En la práctica, se suele considerar infinita la población que se estudia si interviene un proceso dinámico que haga imposible el listado o el conteo de cada elemento de la población.

La mayoría de los casos de muestreo en los negocios y la economía son de poblaciones finitas, pero en algunas situaciones, la población es infinita o (si es finita) es tan grande que, para fines prácticos, se puede considerar como infinita. Al muestrear una población infinita debemos usar una nueva definición de muestra aleatoria simple. Además, como no se pueden numerar los elementos de una población infinita, debemos emplear un proceso distinto para seleccionar los elementos de la muestra.

Supongamos que se desea estimar el tiempo promedio que transcurre entre colocar un pedido y recibirlo, para los clientes de un restaurante de comida rápida, entre las 11:30 A.M. y la 1:30 P.M., que es el horario del almuerzo. Si consideramos la población de todos los posibles clientes, vemos que no sería factible especificar un límite finito de la cantidad de posibles visitas. De hecho, si definimos que la población es todas las visitas de clientes que se pudieran recibir concebiblemente durante las horas del almuerzo, podemos considerar que la población es infinita. Nuestra tarea consiste en seleccionar una muestra aleatoria simple de n clientes de esa población. A continuación vemos la definición de muestra aleatoria simple de una población infinita.

Muestra aleatoria simple (población infinita)

Una muestra aleatoria simple de una población infinita es aquella que se selecciona en tal forma que se satisfacen las siguientes condiciones.

1. Cada elemento seleccionado proviene de la misma población.
2. Cada elemento se selecciona en forma independiente.

No se puede usar un procedimiento de selección con números aleatorios para una población infinita, porque es imposible hacer una lista de esa población. En este caso se debe determinar especialmente un procedimiento de selección de muestra, para determinar los elementos en forma independiente y evitar un prejuicio de selección que haga posible mayores probabilidades de selección para ciertos artículos.

En el caso del problema de seleccionar una variable aleatoria simple de asistencias de clientes a un restaurante de comida rápida, vemos que se satisface la primera condición antes mencionada, para cualquier asistencia de cliente durante el horario del almuerzo, estando trabajando el restaurante con su personal regular y bajo condiciones "normales" de funcionamiento. La segunda condición se satisface asegurando que la selección de determinado cliente no influye sobre la selección de cualquier otro. Esto es, los clientes son seleccionados en forma independiente.

Un restaurante muy conocido de comida rápida implementó un procedimiento de muestreo aleatorio simple para este caso. El procedimiento de muestreo se basa en el hecho de que algunos clientes presentan cupones de descuento en precios de emparedados, bebidas, papas fritas, etc. Siempre que un cliente presenta un cupón de descuento, se selecciona para la muestra al siguiente cliente que llega. Como los clientes que presentan cupones de descuento lo hacen al azar e independientemente, la empresa considera que el plan de muestreo satisface las dos condiciones del muestreo aleatorio simple de poblaciones infinitas.

N O T A S *y comentarios*

1. Con frecuencia se definen las poblaciones finitas mediante listas, como membresía a asociaciones, registros de inscripción de alumnos, listas de cuentahabientes, números de inventario de los productos, etc. Las poblaciones infinitas se definen en muchos casos mediante un proceso dinámico mediante el cual los elementos de la población consisten en cosas generadas, como si el proceso fuera a trabajar indefinidamente bajo las mismas condiciones. En tales casos es imposible obtener una lista de todos los elementos de la población. Por ejemplo, las poblaciones formadas por todas las partes posibles de fabricar, todas las visitas posibles a clientes, todas las transacciones bancarias posibles, etc., se pueden clasificar como poblaciones infinitas.
2. La cantidad de muestras aleatorias simples distintas, de tamaño n que se pueden seleccionar de una población finita de tamaño n es

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

En esta fórmula, $N!$ y $n!$ indican los factoriales que describimos en el capítulo 4. Para el problema de EAI, con $N = 2500$ y $n = 30$, esta expresión se puede usar para demostrar que, aproximadamente hay 2.75×10^{69} muestras aleatorias distintas de 30 empleados de EAI.

EJERCICIOS

MÉTODOS

AUTOEXAMEN

1. Se tiene una pequeña población con cinco artículos, identificados con A, B, C, D y E. Se pueden seleccionar diez muestras aleatorias simples de tamaño dos.
   a. Haga una lista de las 10 muestras, comenzando con AB, AC, y así sucesivamente.
   b. Si se usa muestreo aleatorio simple, ¿cuál es la probabilidad de seleccionar cada muestra aleatoria de tamaño dos? 1/10
   c. Suponga que el número aleatorio 1 corresponde a A, el número aleatorio 2 a B, y así sucesivamente. Haga una lista de la muestra aleatoria simple de dos artículos que se seleccionan con los dígitos aleatorios 8 0 5 7 5 3 2.

2. Suponga que una población tiene 350 elementos. Con los tres últimos dígitos de los siguientes números aleatorios de cinco dígitos, determine las cuatro primeras unidades que se seleccionarán para la muestra aleatoria simple.

98601 73022 83448 02147 34229 27553 84147 93289 14209

APLICACIONES

AUTOEXAMEN

3. La revista *Fortune* publica datos sobre ventas, utilidades, activos, propiedad de acciones, valor de mercado y ganancias por acción de las 500 empresas industriales más grandes en Estados Unidos (The *Fortune* 500, 1998). Suponga que desea seleccionar una muestra aleatoria simple de 10 compañías de la lista de 500 de *Fortune*. Use la columna 9 de la tabla 7.1, comenzando con 554. Avance hacia abajo por la columna e identifique los números de las 10 corporaciones que seleccionaría.

4. La Asociación de Transporte Aeroportuario de Estados Unidos proporcionó la siguiente lista de las 10 aerolíneas más grandes del mundo (*The Book of Mosts*, 1997).

   1. Aeroflot (Rusia)
   2. Air France (Francia)
   3. American Airlines (EUA)
   4. British Airways (Reino Unido)
   5. Delta Airlines (EUA)
   6. JAL (Japón)
   7. Lufthansa (Alemania)
   8. Northwest Airlines (EUA)
   9. United Airlines (EUA)
   10. USAir (EUA)

   a. Suponga que se debe seleccionar una muestra aleatoria de 5 de ellas, para un estudio de detalle de factores, como por ejemplo, la cantidad de aviones en servicio, el total anual de pasajeros-milla volados, etc. Comenzando con el primer dígito aleatorio de la tabla 7.1, que es 6, y avanzando hacia debajo de la columna, use los números aleatorios de un dígito para seleccionar una muestra aleatoria simple de 5 aerolíneas que entrarán en el estudio.

   b. Según lo que dicen las Notas y Comentarios (2), en la página 253, ¿cuántas muestras aleatorias simples de tamaño 5 se pueden tomar de una lista de 10 aerolíneas?

5. Una organización estudiantil desea estimar la proporción de alumnos que favorecen la política obligatoria "pasa-no pasa" para los cursos optativos. En la oficina de registro se dispone una lista de nombres y direcciones de los 645 alumnos inscritos durante el trimestre actual. Use números aleatorios de tres dígitos del renglón 10 en la tabla 7.1, y avance de izquierda a derecha, para identificar a los 10 primeros alumnos que se seleccionarían en un muestreo aleatorio simple. Los números aleatorios de tres dígitos comienzan con 816, 283 y 610.

6. El libro *County and City Data Book*, publicado por la Oficina del Censo en Estados Unidos, publica información acerca de 3139 condados en ese país. Suponga que un estudio nacional debe reunir datos de 30 condados seleccionados al azar. Use números aleatorios de la última columna de la tabla 7.1 para identificar los números que corresponden a los cinco primeros condados seleccionados para la muestra. No tome en cuenta los primeros dígitos, y comience con los números aleatorios de cuatro dígitos 9945, 8364, 5702, etcétera.

7. Suponga que se desea identificar una muestra aleatoria simple de 12 de los 372 doctores que ejercen en determinada ciudad. Los nombres de ellos se consiguen en una organización médica local. Use la octava columna de números aleatorios con cinco dígitos en la tabla 7.1 para identificar a los 12 doctores de la muestra. Ignore los dos primeros dígitos aleatorios en cada grupo de cinco. Este proceso comienza con el número aleatorio 108 y avanza hacia abajo de la columna de números aleatorios.

8. Los resultados de una encuesta sobre los 25 mejores entrenadores de baloncesto de NCAA, en enero de 1988, aparece a continuación (*USA Today*, 9 de enero de 1998). Use la novena columna de los números aleatorios en la tabla 7.1, comenzando con 13554, para seleccionar una muestra aleatoria simple de 6 equipos de baloncesto. Comience con el equipo 13 y use los dos primeros dígitos de cada renglón de la novena columna en el proceso de selección. ¿Cuáles equipos de baloncesto seleccionó para esa muestra aleatoria simple?

1. Duke	14. Mississippi
2. North Carolina	15. Syracuse
3. Kansas	16. Michigan
4. Utah	17. South Carolina
5. Arizona	18. Xavier
6. Stanford	19. Arkansas
7. Kentucky	20. Florida State
8. Connecticut	21. West Virginia
9. Purdue	22. Rhode Island
10. UCLA	23. Clemson
11. Princeton	24. Hawaii
12. Iowa	25. Cincinnati
13. New Mexico	

9. La revista *Business Week* presenta datos de funcionamiento y calificaciones anuales de 895 fondos de ahorro (*Business Week*, 3 de febrero de 1997). Suponga que se ha de seleccionar una muestra aleatoria simple de 12 de esos fondos, para un estudio de seguimiento. Use la cuarta columna de los números aleatorios en la tabla 7.1, comenzando con 51102, para seleccionar esa muestra. Comience con el fondo número 511 y use los tres primeros dígitos en cada renglón de la cuarta columna en el proceso de selección. ¿Cuáles son los números de los 12 fondos de ahorro que forman la muestra aleatoria simple?

10. Haskell Public Opinion Poll, Inc., efectúa encuestas telefónicas acerca de diversos asuntos de interés político y público en general. Las familias que se incluyen en las encuestas se identifican tomando una muestra aleatoria simple de los directorios telefónicos en algunas áreas metropolitanas. El directorio telefónico de una gran ciudad contiene 853 páginas con 400 renglones por página.

a. Describa un procedimiento de selección aleatoria en dos etapas que se pueda usar para identificar una muestra aleatoria simple de 200 familias. En el proceso se debe seleccionar primero una página al azar (etapa 1) y después, un renglón de la página muestreada (etapa 2). Use los números aleatorios de la tabla 7.1 para ilustrar este proceso. Seleccione su propio y arbitrario punto de partida en la tabla.

b. ¿Qué haría usted si el renglón seleccionado en el inciso *a* fuera claramente inadecuado para el estudio (esto es, si en ese lugar hubiera un número telefónico de una empresa, un restaurante, etcétera)?

11. El grupo de investigación sobre el tema del parque Paramont's King's Island, en Kings Mills, Ohio, recurre a encuestas para determinar qué les gusta a los visitantes del parque.

a. Suponga que el grupo de investigación considera que la población de visitantes es infinita. ¿Se puede aceptar? Explique por qué.

b. Suponga que después de terminar una entrevista con un visitante, el entrevistador regresa a la puerta de entrada del parque y comienza a contar a las personas que lo visitan. Seleccione al vigesimoquinto individuo contado, como la siguiente persona de la muestra en la encuesta. Después de terminar su entrevista, el entrevistador regresa a la entrada y de nuevo selecciona al vigesimoquinto individuo que entra al parque. Este procedimiento de selección, ¿parece dar como resultado una muestra aleatoria simple? Explique su respuesta.

12. Indique si las siguientes poblaciones se deben considerar como finitas o infinitas.

a. Todos los votantes registrados en un estado.

b. Todos los televisores que podría producir la planta de Córdoba de la empresa TV-M.

c. Todos los pedidos que pueda procesar una empresa de pedidos por correo.

d. Todas las llamadas telefónicas de emergencia que pudieran entrar a una estación local de policía.

e. Todos los componentes que produjo TV-M, en el segundo turno del 17 de mayo.

7.3 ESTIMACIÓN PUNTUAL

Ahora que ya describimos cómo seleccionar una muestra aleatoria simple, regresemos al problema de EAI. Supongamos que se ha seleccionado una muestra aleatoria simple de 30 gerentes, y que los datos correspondientes sobre salario anual y participación en el programa de adiestramiento gerencial son los que aparecen en la tabla 7.2. Se usa la notación x_1, x_2, etc., para indicar el salario anual del primer gerente de la muestra, el del segundo, y así sucesivamente. La participación en dicho programa de adiestramiento se indica con un Sí en la columna correspondiente.

Para estimar el valor de un parámetro de población se calcula una correspondiente característica de la muestra, que se denomina *estadístico de la muestra* o *medida muestral*. Por ejemplo, para estimar la media de la población μ y la desviación estándar de la población σ del salario anual de los gerentes, tan sólo se toman los datos de la tabla 7.2 para calcular los correspondientes estadísticos de la muestra: la media de la muestra $\bar{x}$ y la desviación estándar de la muestra s. De acuerdo con las fórmulas de la media y la desviación estándar de la muestra que vimos en el capítulo 3, la media de la muestra es

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{1{,}554{,}420}{30} = \$51{,}814.00$$

y la desviación estándar de la muestra es

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{325{,}009{,}260}{29}} = \$3347.72$$

Tabla 7.2 SALARIO ANUAL Y ESTADO EN EL PROGRAMA DE ADIESTRAMIENTO PARA UNA MUESTRA ALEATORIA SIMPLE DE 30 GERENTES DE EAI

Salario anual ($)	¿Programa de adiestramiento gerencial?	Salario anual ($)	¿Programa de adiestramiento gerencial?
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

Además, al calcular la proporción de gerentes en la muestra que contestaron Sí, podemos estimar la proporción de gerentes, en la población, que terminaron el programa de adiestramiento gerencial. La tabla 7.2 indica que 19 de los 30 gerentes de la muestra terminaron el adiestramiento. Entonces, la proporción de la muestra, representada por $\bar{p}$, es

$$\bar{p} = \frac{19}{30} = .63$$

Este valor se usa como estimado de la proporción p en la población.

Al hacer los cálculos anteriores hemos efectuado el procedimiento estadístico denominado *estimación puntual*. En él usamos los datos de la muestra para calcular un valor de un estadístico de la muestra que sirva como estimado de un parámetro de población. Continuando con la terminología de la estimación puntual, se dice que $\bar{x}$ es el *estimador puntual* de la media poblacional μ, s es el *estimador puntual* de la desviación estándar σ poblacional, y que $\bar{p}$ es el *estimador puntual* de la proporción p poblacional. A los valores numéricos obtenidos para $\bar{x}$, s o $\bar{p}$ en una determinada muestra se les llama *estimados puntuales* del parámetro. Así, para la muestra de 30 gerentes de EAI, el estimado puntual de μ es de \$51,814.00 dólares, el de σ es de \$3,347.72 dólares y el de p es .63. La tabla 7.3 contiene un resumen de los resultados de la muestra y compara los estimados puntuales con los valores reales de los parámetros de la población.

Tabla 7.3 RESUMEN DE ESTIMADOS PUNTUALES OBTENIDOS A PARTIR DE UNA MUESTRA ALEATORIA SIMPLE DE 30 GERENTES DE EAI

Parámetro de población	Valor del parámetro en dólares	Estimador puntual	Estimado puntual
μ = Salario promedio anual poblacional	\$51,800.00	$\bar{x}$ = Salario promedio anual de la muestra	\$51,814.00
σ = Desviación estándar poblacional del salario anual	\$4,000.00	s = Desviación estándar muestral del salario anual	\$3,347.72
p = Proporción poblacional que terminó el programa de adiestramiento gerencial	.60	$\bar{p}$ = Proporción muestral de quienes terminaron el programa de adiestramiento gerencial	.63

NOTAS y comentarios

En nuestra descripción de los estimados puntuales usamos $\bar{x}$ para indicar una media de la muestra, y a $\bar{p}$ para indicar una proporción de la muestra. Nuestro uso de $\bar{p}$ se basa en el hecho de que la proporción de la muestra también es una media de la muestra. Por ejemplo, supongamos que en una muestra de n elementos, con valores $x_1, x_2, \ldots, x_n$, definimos a $x_i = 1$ cuando en el i-ésimo elemento existe una característica de interés, y con $x_i = 0$ cuando ésta no existe. Entonces, la proporción de la muestra se calcula mediante $\sum x_i/n$, que es la fórmula de una media de la muestra. También nos gusta la consistencia de poner la raya sobre la letra para recordar al lector que la proporción de la muestra $\bar{p}$ estima la proporción de la población exactamente de la misma manera que la media de la muestra $\bar{x}$ estima a la media poblacional. Algunos textos de estadística usan $\hat{p}$ en lugar de $\bar{p}$ para indicar la proporción de la muestra.

EJERCICIOS

MÉTODOS

AUTOEXAMEN

13. Se han reunido los siguientes datos de una muestra aleatoria simple.

5 8 10 7 10 14

   a. ¿Cuál es el estimado puntual de la media de la población?
   b. ¿Cuál es el estimado puntual de la desviación estándar poblacional?

14. Una pregunta de una encuesta, en una muestra de 150 personas, obtuvo 75 respuestas Sí, 55 respuestas No y 20 Sin opinión.
   a. ¿Cuál es el estimado puntual de la proporción de la población que responde Sí? 0.5
   b. ¿Cuál es el estimado puntual de la proporción de la población que responde No? 0.36

APLICACIONES

AUTOEXAMEN

15. Una muestra aleatoria simple de datos de cinco meses de ventas da la siguiente información:

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

   a. ¿Cuál es el estimado puntual de la media de la población de la cantidad de unidades vendidas por mes?
   b. ¿Cuál es el estimado puntual de la desviación estándar de la población?

16. Se preguntó a 784 niños de una muestra, cuyas edades eran de 9 a 14 años, en qué forma conseguían dinero de sus padres (*Consumer Reports,* enero de 1997). Se obtuvieron las siguientes respuestas:

Fuente de ingresos	Frecuencia
Sólo domingos	149
Quehaceres, dádivas y domingo	219
Quehaceres y dádivas, sin domingo	251
Nada	165
Total	784

   a. ¿Qué proporción de niños recibe domingo como única fuente de dinero? 0.17
   b. ¿Qué proporción de niños recibe dinero por quehaceres y dádivas, pero no lo recibe como domingo?
   c. ¿Respecto a todas las fuentes de ingresos, ¿qué proporción de niños reciben al menos algo de dinero de sus padres?

17. La Patrulla de Carreteras de California cuenta con registros que muestran el tiempo que pasa entre la notificación de un accidente y la llegada de un oficial a la escena del mismo accidente. Una muestra aleatoria simple de 10 casos indica los siguientes tiempos en minutos

12.6 3.4 4.8 5.0 6.8 2.3 3.6 8.1 2.5 10.3

   a. ¿Cuál es un estimado puntual de la media de la población del tiempo entre una notificación de accidente y la llegada de un oficial?
   b. ¿Cuál es un estimado puntual de la desviación estándar de la población del tiempo transcurrido entre una notificación de accidente y la llegada de un oficial?

18. El Departamento de Transporte en Estados Unidos, publica estadísticas de llegadas, antes o después del horario programado, de los principales vuelos (*USA Today*, 26 de junio de 1997). Suponga que la proporción estimada de vuelos que llegan a tiempo, para todas las aerolíneas, se basa en una muestra de 1400 vuelos. Si 1117 llegan a tiempo, ¿cuál es el estimado puntual de la proporción de vuelos que llegan puntuales?

19. Una encuesta efectuada en 1996 por Louis Harris se basó en 1005 adultos, para averiguar el uso que la gente da a Internet (*Business Week*, 26 de agosto de 1996). Se obtuvieron las respuestas siguientes:

874 adultos sabían que Internet existe
503 adultos usaban computadora
191 adultos habían entrado a Internet durante los últimos 12 meses

Determine el estimado puntual de los siguientes parámetros de la población:
a. La proporción de adultos que saben que Internet existe.
b. La proporción de adultos que usan computadoras.
c. La proporción de adultos que han entrado a Internet durante los últimos 12 meses.

7.4 INTRODUCCIÓN A LAS DISTRIBUCIONES MUESTRALES

En la sección anterior usamos una muestra aleatoria simple de 30 gerentes de EAI para determinar estimados puntuales de la media y la desviación estándar del salario anual de la población de todos los gerentes de EAI, y también de la proporción de los gerentes, en la población, que terminaron el programa de adiestramiento gerencial de la empresa. Suponga que seleccionamos otra muestra aleatoria simple de 30 gerentes de EAI, y que al analizar los datos de ella obtenemos la siguiente información.

$$\text{Media de la muestra } \bar{x} = \$52{,}669.70$$
$$\text{Desviación estándar de la muestra } s = \$4{,}239.07$$
$$\text{Proporción de la muestra } \bar{p} = .70$$

Estos resultados indican que hemos obtenido distintos valores de $\bar{x}$, s y $\bar{p}$ con la segunda muestra. En general, debemos esperar que sea así, porque no es probable que la segunda muestra aleatoria simple contenga los mismos elementos que la primera. Imaginemos que llevamos a cabo el mismo proceso de selección de una nueva muestra aleatoria simple de 30 gerentes, una y otra vez, calculando en cada ocasión los valores de $\bar{x}$, s y $\bar{p}$. De este modo podríamos comenzar a identificar la variedad de valores que pueden tener esos estimadores puntuales. Por ejemplo, repetimos el proceso de muestreo aleatorio simple del problema de EAI hasta haber obtenido 500 muestras de 30 gerentes cada una, con sus valores correspondientes $\bar{x}$, s y $\bar{p}$. En la tabla 7.4 se muestra una parte de los resultados. La tabla 7.5 muestra las distribuciones de frecuencia y de frecuencia relativa de los 500 valores de $\bar{x}$. La figura 7.1 muestra el histograma de frecuencia relativa de los resultados de $\bar{x}$.

El concepto de una distribución muestral, o de muestreo, es uno de los temas más importantes en este capítulo. El poder comprender lo que se explica en los capítulos siguientes depende mucho de la capacidad de comprender y aplicar las distribuciones de muestreo que se explican en este capítulo.

En el capítulo 5 definimos a una variable aleatoria como una descripción numérica del resultado de un experimento. Si consideramos que un experimento es el proceso de elegir una muestra aleatoria simple, la media de la muestra $\bar{x}$ es la descripción numérica del resultado del experimento. En consecuencia, la media de la muestra $\bar{x}$ es una variable aleatoria. Por lo tanto, al igual que otras variables aleatorias, $\bar{x}$ tiene una media o valor esperado, una varianza y una distribución de probabilidades. Como los diversos valores posibles de $\bar{x}$ son el resultado de distintas muestras aleatorias simples, a la distribución de $\bar{x}$ se le llama *distribución muestral* de $\bar{x}$. El conocimiento de esta distribución muestral y de sus propiedades nos permitirá hacer afirmaciones probabilísticas acerca de lo cercano que se encuentre la media de la muestra $\bar{x}$ de la media de población μ.

Regresemos a la figura 7.1. Necesitaríamos enumerar cada muestra posible de 30 gerentes y calcular cada media de la muestra para determinar totalmente la distribución muestral de $\bar{x}$. Sin embargo, el histograma de 500 valores de $\bar{x}$ proporciona una aproximación a esa distribución muestral, en la cual observamos la apariencia acampanada de la misma. También vemos que la media de los 500 valores de $\bar{x}$ se encuentra cerca de la

Tabla 7.4 VALORES DE $\bar{x}$, s, Y $\bar{p}$ A PARTIR DE 500 MUESTRAS ALEATORIAS SIMPLES DE 30 GERENTES DE EAI

Muestra número	Promedio muestral ($\bar{x}$)	Desviación estándar muestral (s)	Proporción muestral ($\bar{p}$)
1	\$51,814.00	\$3,347.72	.63
2	\$52,669.70	\$4,239.07	.70
3	\$51,780.30	\$4,433.43	.67
4	\$51,587.90	\$3,985.32	.53
.	.	.	.
.	.	.	.
.	.	.	.
500	\$51,752.00	\$3,857.82	.50

Tabla 7.5 DISTRIBUCIÓN DE FRECUENCIAS DE $\bar{x}$ A PARTIR DE 500 MUESTRAS ALEATORIAS SIMPLES DE 30 GERENTES DE EAI

Promedio de salario anual ($)	Frecuencia	Frecuencia relativa
49,500.00–49,999.99	2	.004
50,000.00–50,499.99	16	.032
50,500.00–50,999.99	52	.104
51,000.00–51,499.99	101	.202
51,500.00–51,999.99	133	.266
52,000.00–52,499.99	110	.220
52,500.00–52,999.99	54	.108
53,000.00–53,499.99	26	.052
53,500.00–53,999.99	6	.012
Totales	500	1.000

media de la población $\mu = \$51,800$ dólares. En la siguiente sección estudiaremos las propiedades de la distribución muestral de $\bar{x}$.

Los 500 valores de la desviación estándar s y los 500 de la proporción de la muestra $\bar{p}$ se resumen en los histogramas de frecuencia relativa de las figuras 7.2 y 7.3. Como en el caso de $\bar{x}$, tanto s como $\bar{p}$ son variables aleatorias que proporcionan descripciones numéricas del resultado de una muestra aleatoria simple. Si se seleccionara toda muestra posible de tamaño 30 de la población, y si se calculara el valor de s y el de $\bar{p}$ de cada muestra, las distribuciones resultantes de probabilidad se llamarían distribución muestral de s y distribución muestral de $\bar{p}$, respectivamente. Los histogramas de los 500 valores muestrales de las figuras 7.2 y 7.3 dan una idea general de la apariencia de esas dos distribuciones muestrales.

En la práctica, tan sólo seleccionamos una sola muestra aleatoria simple de la población. En esta sección repetimos 500 veces el proceso de muestreo únicamente para ilustrar que es posible tener muchas muestras distintas, y que éstas generan una diversidad de valores para los estadísticos de la muestra $\bar{x}$, s y $\bar{p}$. La distribución de probabilidad de cualquier estadístico de la muestra se llama distribución muestral del estadístico. En la sección 7.5 mostraremos las características de la distribución muestral de $\bar{x}$, en la sección

Figura 7.1 HISTOGRAMA DE FRECUENCIAS RELATIVAS DE VALORES DE $\bar{x}$ A PARTIR DE 500 MUESTRAS ALEATORIAS SIMPLES, CADA UNA DE TAMAÑO 30

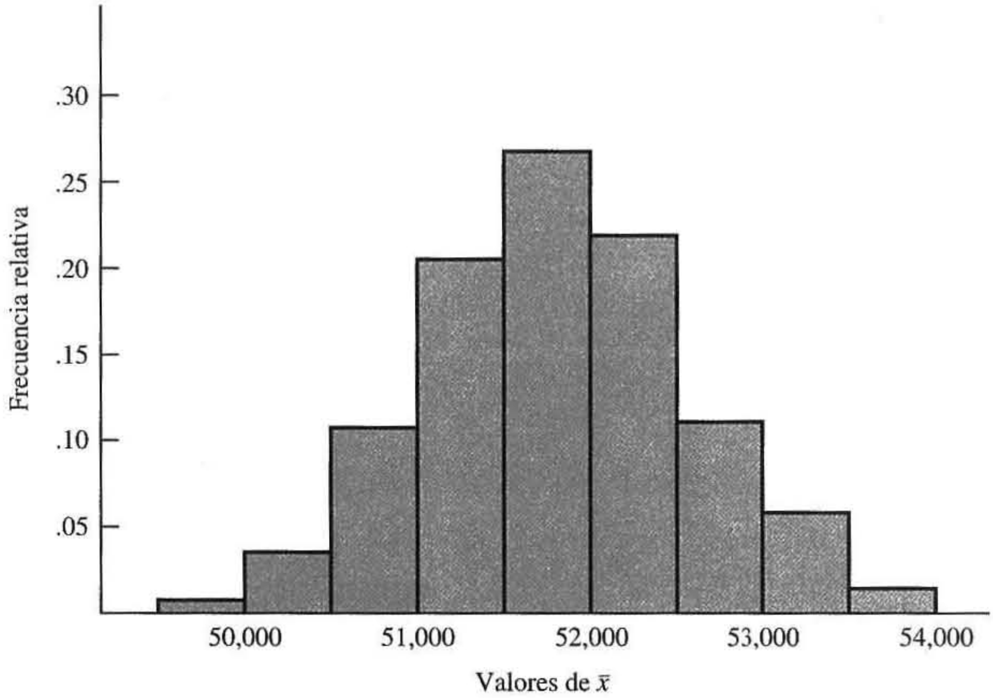


Figura 7.2 HISTOGRAMA DE FRECUENCIAS RELATIVAS DE VALORES DE s A PARTIR DE 500 MUESTRAS ALEATORIAS SIMPLES, CADA UNA DE TAMAÑO 30

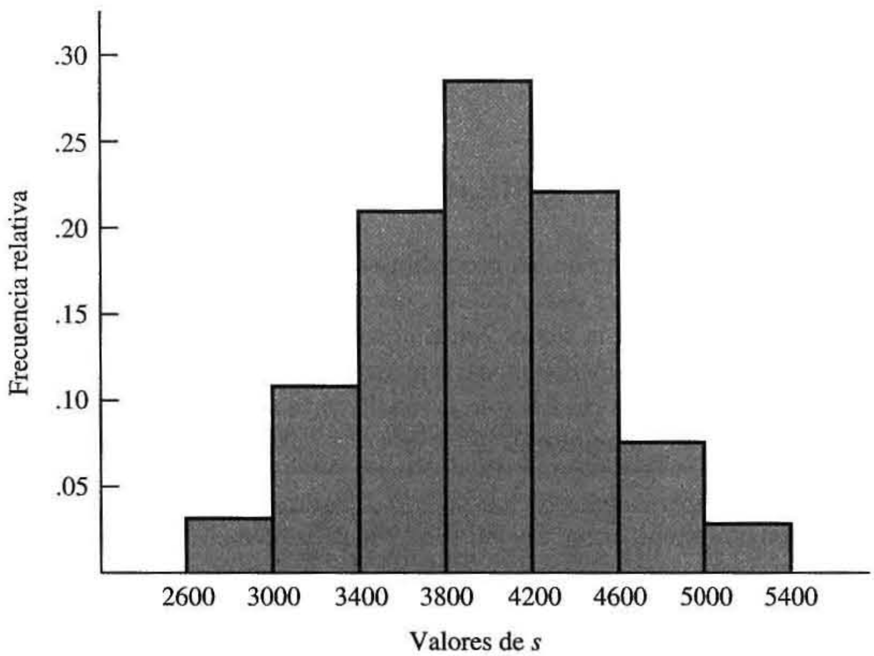
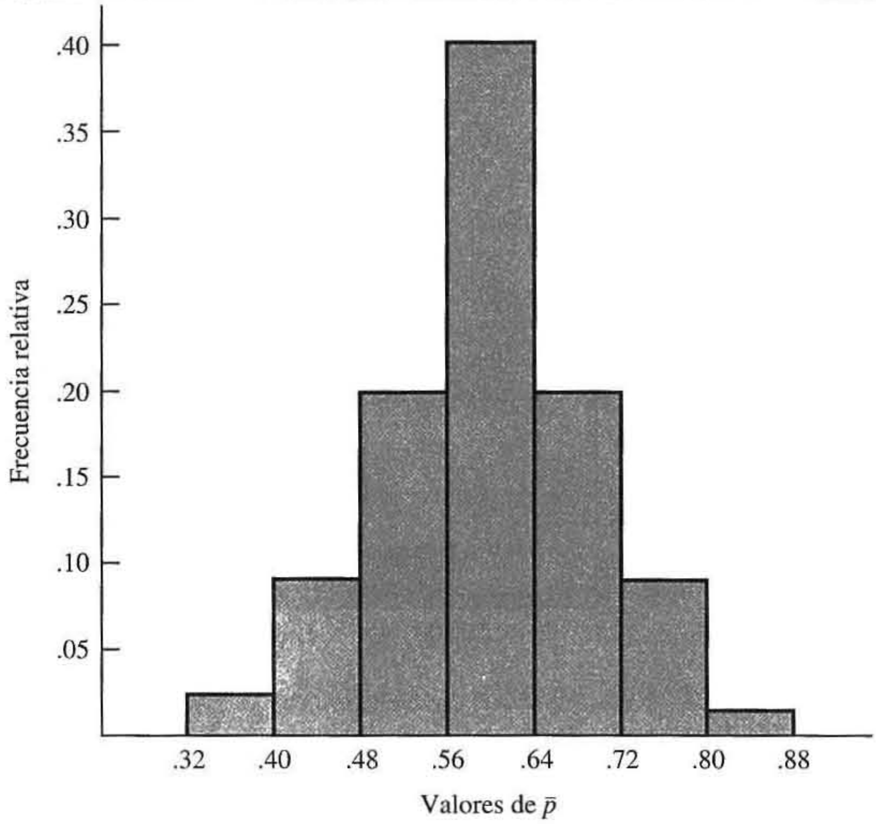


Figura 7.3 HISTOGRAMA DE FRECUENCIAS RELATIVAS DE VALORES DE $\bar{p}$ A PARTIR DE 500 MUESTRAS ALEATORIAS SIMPLES, CADA UNA DE TAMAÑO 30



7.6 las de $\bar{p}$, y diferiremos mayor descripción de la distribución muestral de s hasta que expliquemos las distribuciones muestrales que pertenecen a las varianzas de la muestral, que describiremos en el capítulo 11.

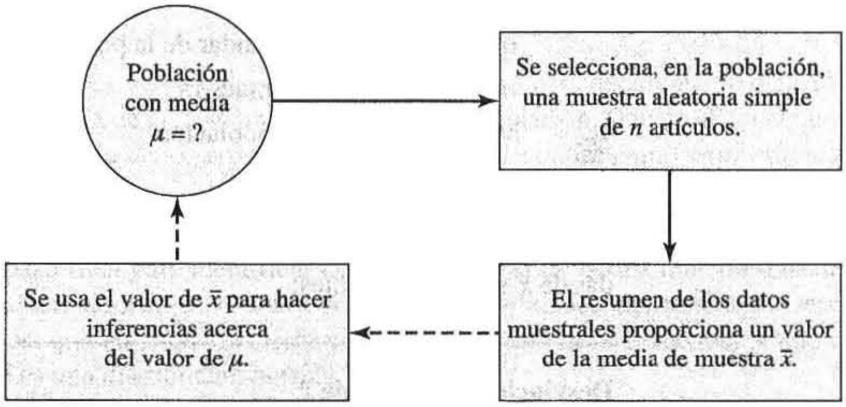
7.5 DISTRIBUCIÓN MUESTRAL DE $\bar{x}$

Uno de los procedimientos estadísticos más comunes es usar la media de la muestra $\bar{x}$ para hacer inferencias acerca de una media de la población μ. Este proceso se muestra en la figura 7.4. En cada repetición del proceso podemos anticipar la obtención de un valor distinto de la media de la muestra $\bar{x}$. La distribución de probabilidad de todos los valores posibles de la media de la muestra $\bar{x}$ se llama distribución muestral de la media de la muestra $\bar{x}$.

Distribución muestral de $\bar{x}$

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

Figura 7.4 PROCESO ESTADÍSTICO PARA EMPLEAR UNA MEDIA DE MUESTRA PARA HACER INFERENCIAS ACERCA DE UNA MEDIA POBLACIONAL



El objetivo de esta sección es describir las propiedades de la distribución muestral de $\bar{x}$, incluyendo el valor esperado o media de $\bar{x}$, su desviación estándar y la forma de la distribución misma. Como veremos, el conocimiento de la distribución muestral de $\bar{x}$ nos permitirá hacer afirmaciones probabilísticas acerca del error incurrido cuando se usa $\bar{x}$ para estimar μ. Comencemos considerando la media de todos los valores posibles de $\bar{x}$ o, simplemente, el valor esperado de $\bar{x}$.

Valor esperado de $\bar{x}$

En el problema de muestreo de EAI vimos que distintas muestras aleatorias simples dan como resultado varios valores de la media de la muestra $\bar{x}$. Como son posibles muchos valores distintos de la variable aleatoria $\bar{x}$, nos interesa la media de todos los valores posibles de $\bar{x}$ que se puedan generar mediante las diversas muestras aleatorias simples. Esa media es el valor esperado de $\bar{x}$. Sean $E(\bar{x})$ el valor esperado de $\bar{x}$, y μ la media de la población de donde se toma la muestra. Se puede demostrar que, para muestreo aleatorio simple, los dos valores son iguales.

Valor esperado de $\bar{x}$

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

donde

$E(\bar{x})$ = el valor esperado de la variable aleatoria $\bar{x}$

μ = la media de la población

Este resultado se deduce en el apéndice del capítulo. Indica que, con muestreo aleatorio simple, el valor esperado, o media de $\bar{x}$ es igual a la media de la población. En la sección 7.1 vimos que el sueldo anual promedio de la población de gerentes de EAI es μ = \$51,800 dólares. Entonces, según la ecuación (7.1), la media de todas las posibles medias de la muestra del estudio de EAI también es de \$51,800 dólares.

Desviación estándar de $\bar{x}$

Definiremos la desviación estándar de la distribución muestral de $\bar{x}$. Usaremos la siguiente notación:

$\sigma_{\bar{x}}$ = la desviación estándar de la distribución muestral de $\bar{x}$
σ = la desviación estándar de la población
n = el tamaño de la muestra
N = el tamaño de la población.

Se puede demostrar que, con muestreo aleatorio simple, la desviación estándar de $\bar{x}$ depende de si la población es finita o infinita. Las dos ecuaciones de la desviación estándar de $\bar{x}$ son las siguientes:

Desviación estándar de $\bar{x}$

Población finita | *Población infinita*

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right) \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{7.2}$$

En el apéndice 7.1 se describe la deducción de las fórmulas de $\sigma_{\bar{x}}$. Al comparar las dos ecuaciones vemos que se requiere un factor $\sqrt{(N-n)/(N-1)}$ para la población finita, pero ninguno para la infinita. Este factor se llama *factor de corrección para población finita*. En muchos casos prácticos de muestreo se ve que la población que se maneja, aunque finita, es "grande", mientras que el tamaño de la muestra es relativamente "pequeño". En esos casos el factor de corrección para población finita, $\sqrt{(N-n)/(N-1)}$ es cercano a 1. En consecuencia, la diferencia entre los valores de la desviación estándar de $\bar{x}$ para los casos de población finita e infinita se hace despreciable. Cuando esto sucede, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ se vuelve una aproximación muy buena a la desviación estándar de $\bar{x}$ aunque la población sea finita. Un lineamiento o regla general para calcular la desviación estándar de $\bar{x}$ es el siguiente:

Usar la siguiente ecuación para calcular la desviación estándar de $\bar{x}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{7.3}$$

siempre que

1. La población sea infinita, o bien
2. La población sea finita y también el tamaño de la muestra sea menor o igual que el 5% del tamaño de la población —esto es, $n/N \leq .05$.

En el problema 25 se ve que cuando $n/N \leq .05$, el factor de población finita tiene poco efecto sobre el valor de $\sigma_{\bar{x}}$.

En los casos en que $n/N > .05$, se debe usar la ecuación (7.2) para poblaciones finitas en el cálculo de $\sigma_{\bar{x}}$. A menos que digamos otra cosa, en este libro supondremos que el tamaño de la población es "grande", que es innecesario el factor de corrección por población finita, y que se puede usar la ecuación (7.3) para calcular $\sigma_{\bar{x}}$.

Regresemos ahora al estudio de EAI y determinemos la desviación estándar de todas las medias de la muestra posibles que se pueden generar con muestras de 30 gerentes de esa negociación. En la sección 7.1 vimos que la desviación estándar poblacional de los datos de salario anual es $\sigma = 4000$. En este caso, la población es finita y $N = 2500$. Sin embargo, con un tamaño de la muestra de 30 tenemos que $n/N = 30/2500 = .012$. Siguiendo la regla general de la ecuación (7.3), podemos pasar por alto el factor de corrección para población finita y usar esta ecuación para calcular la desviación estándar de $\bar{x}$.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.30$$

Después veremos que el valor de $\sigma_{\bar{x}}$ es útil para determinar lo alejado que pueda estar la media de la muestra de la media de la población. Debido al papel que desempeña $\sigma_{\bar{x}}$ en el cálculo de los errores posibles, a $\sigma_{\bar{x}}$ se le conoce como *error estándar de la media.*

Teorema del límite central

El paso final para identificar las características de la distribución muestral de $\bar{x}$ es determinar la forma de la distribución de probabilidades, también de $\bar{x}$. Describiremos dos casos: uno en el que la distribución de la población se desconoce y uno en que se sabe que es una distribución normal.

Cuando se desconoce la distribución de la población nos basamos en uno de los teoremas más importantes de la estadística: el teorema del límite central. Uno de sus enunciados, aplicado a la distribución muestral de $\bar{x}$ es el siguiente:

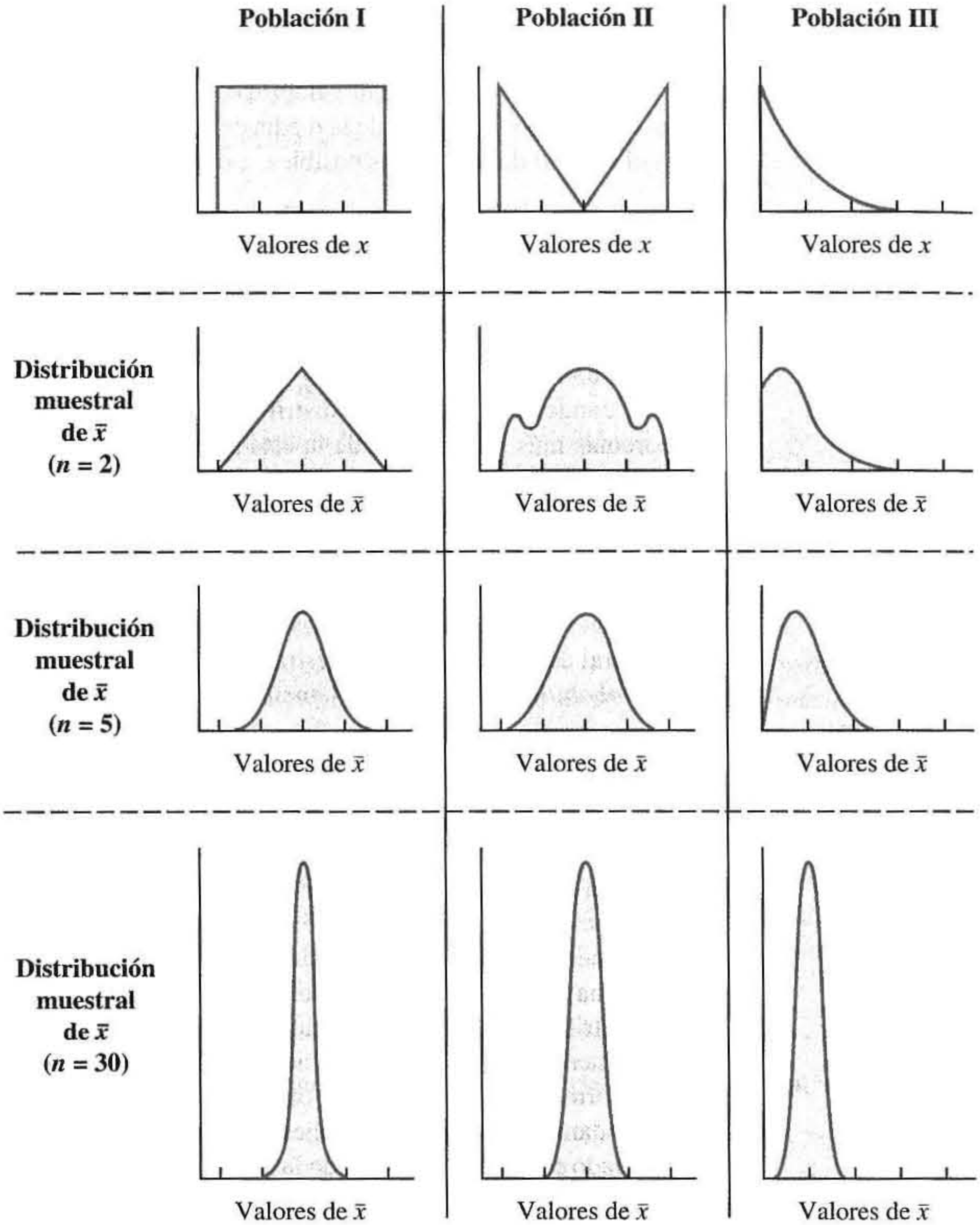
Teorema del límite central

Al seleccionar muestras aleatorias simples de tamaño n de una población, la distribución muestral de la media de la muestra $\bar{x}$ se puede aproximar con una *distribución normal de probabilidades*, cuando el tamaño de la muestra es grande.

La figura 7.5 muestra cómo se aplica el teorema del límite central para tres poblaciones distintas, y en cada caso se ve claramente que la población no es normal. Sin embargo, observamos lo que le comienza a suceder a la distribución muestral de $\bar{x}$ cuando aumenta el tamaño de la muestra. Cuando las muestras son de tamaño dos, vemos que la distribución muestral de $\bar{x}$ comienza a tener una apariencia distinta de la distribución de la población. Para muestras de tamaño 5 vemos que las tres distribuciones comienzan a tener una apariencia acampanada. Finalmente, las muestras de tamaño 30 hacen que las tres distribuciones muestrales sean aproximadamente normales. Así, para muestras suficientemente grandes, la distribución muestral de $\bar{x}$ se puede aproximar con una distribución normal de probabilidades. Sin embargo, ¿de qué tamaño debe ser la muestra para que podamos suponer que se aplica el teorema del límite central? En estadística se ha investigado este asunto estudiando la distribución muestral de $\bar{x}$ para distintas poblaciones y distintos tamaños de muestra. Siempre que la distribución de la población tiene forma de colina y es simétrica, las muestras de tamaño de 5 a 10 pueden bastar para que se aplique el teorema del límite central. Sin embargo, si la distribución poblacional es muy asimétrica y decididamente no normal, se necesitan tamaños mayores de muestra. La práctica general de la estadística es suponer que para la mayoría de las aplicaciones, la distribución muestral de $\bar{x}$ se puede aproximar mediante una distribución normal de probabilidades siempre que el tamaño de la muestra sea de 30 o más. En efecto, se supone que este tamaño, de 30 o más, satisface la condición de muestra grande del teorema del límite central. Esta observación es tan importante que conviene mencionarla.

La distribución muestral de $\bar{x}$ se puede aproximar mediante una distribución normal de probabilidades siempre que el tamaño de la muestra sea grande. Se puede suponer que la condición de muestra grande se cumple para muestras aleatorias simples de tamaño 30 o mayor.

Figura 7.5 ILUSTRACIÓN DEL TEOREMA DEL LÍMITE CENTRAL PARA TRES POBLACIONES



El teorema del límite central es la clave para identificar la forma de la distribución muestral de $\bar{x}$ cuando se desconoce la distribución de la población. Sin embargo, nos podemos encontrar con ciertos casos de muestreo en los que se supone que la población tiene distribución normal. Cuando suceden estos casos, el siguiente resultado identifica la forma de la distribución muestral de $\bar{x}$.

Siempre que la población tiene una distribución normal de probabilidades, la distribución muestral de $\bar{x}$ es distribución normal de probabilidades para cualquier tamaño de la muestra.

En resumen, si usamos una muestra aleatoria simple grande ($n \geq 30$) el teorema del límite central nos permite decir que la distribución muestral de $\bar{x}$ se puede aproximar con una distribución normal de probabilidades. Cuando la muestra aleatoria simple es pequeña ($n < 30$), sólo se puede considerar que la distribución muestral de $\bar{x}$ es normal si se supone que la población tiene una distribución normal de probabilidades.

Distribución muestral de $\bar{x}$ para el problema de EAI

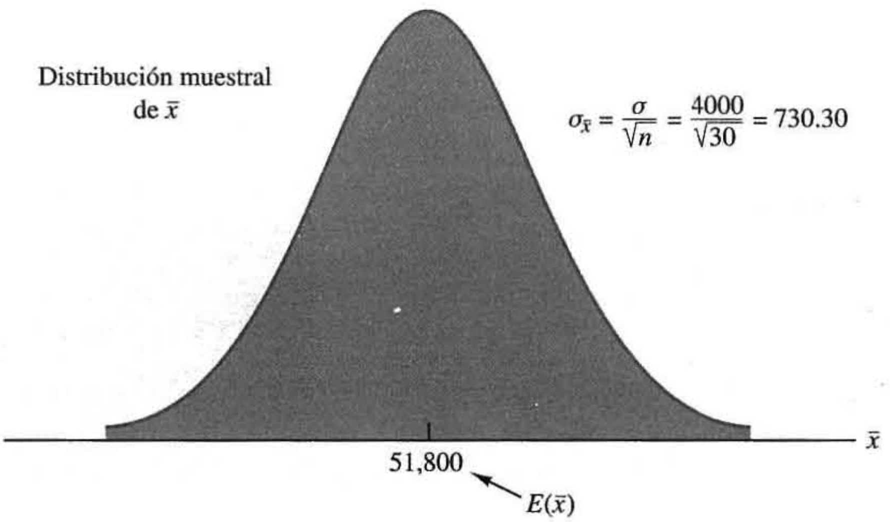
En el estudio de EAI hemos demostrado que $E(\bar{x}) = 51{,}800$ y $\sigma_{\bar{x}} = 730.30$. Como estamos usando una muestra aleatoria simple de 30 gerentes, el teorema del límite central nos permite decir que la distribución muestral de $\bar{x}$ es aproximadamente normal, como vemos en la figura 7.6.

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

Siempre que se selecciona una muestra aleatoria sencilla y se calcula el valor de la media $\bar{x}$ de la muestra para estimar μ, la media de la población, no podemos esperar que la media de la muestra sea exactamente igual a la media de la población. El valor absoluto de la diferencia entre la media de la muestra $\bar{x}$ y la media de la población μ es $|\bar{x} - \mu|$ y se llama *error muestral*. La razón práctica de que nos interese la distribución muestral de $\bar{x}$ es que la podemos usar para determinar información probabilística acerca del tamaño del error muestral. Para demostrar esta aplicación retomemos el problema de EAI.

Supongamos que el director de personal cree que la media de la muestra será un estimado aceptable de la media de la población si el promedio muestral dista menos de \$500 del promedio poblacional. En términos de probabilidad, lo que le preocupa en realidad al director es la siguiente pregunta: ¿Cuál es la probabilidad de que la media de la muestra que obtengamos de una muestra aleatoria simple de 30 gerentes de EAI esté dentro del intervalo de \$500 dólares alrededor de la media de la población?

Figura 7.6 DISTRIBUCIÓN DE $\bar{x}$ PARA EL SALARIO ANUAL PROMEDIO DE UNA MUESTRA ALEATORIA SIMPLE DE 30 GERENTES DE EAI



Como hemos identificado las propiedades de la distribución muestral de $\bar{x}$ (véase figura 7.6), usaremos esa información para contestar la pregunta. Veamos la distribución muestral de $\bar{x}$ que se vuelve a mostrar en la figura 7.7. El director de personal pregunta sobre la probabilidad de que la media de la muestra sea entre $51,300 y $52,300 dólares. Si el valor de la media de la muestra de $\bar{x}$ está en este intervalo, se aproximará a $500 dólares de la media de la población. La probabilidad correspondiente es el área de la distribución muestral que vemos en la figura 7.7. Como la distribución muestral es normal, con promedio de 51,800 y desviación estándar de 730.30, podemos usar la tabla de distribución normal estándar de probabilidades para determinar el área o probabilidad. Cuando $\bar{x} = 51{,}300$, tenemos

$$z = \frac{51{,}300 - 51{,}800}{730.30} = -.68$$

Consultamos la tabla de distribución normal estándar de probabilidades y vemos que el área entre $z = 0$ y $z = -.68$ es .2518. Las operaciones análogas con $\bar{x} = 52{,}300$ muestran que el área entre $z = 0$ y $z = +.68$ es .2518. Así, la probabilidad de que el valor de la media de la muestra esté entre 51,300 y 52,300 es $.2518 + .2518 = .5036$.

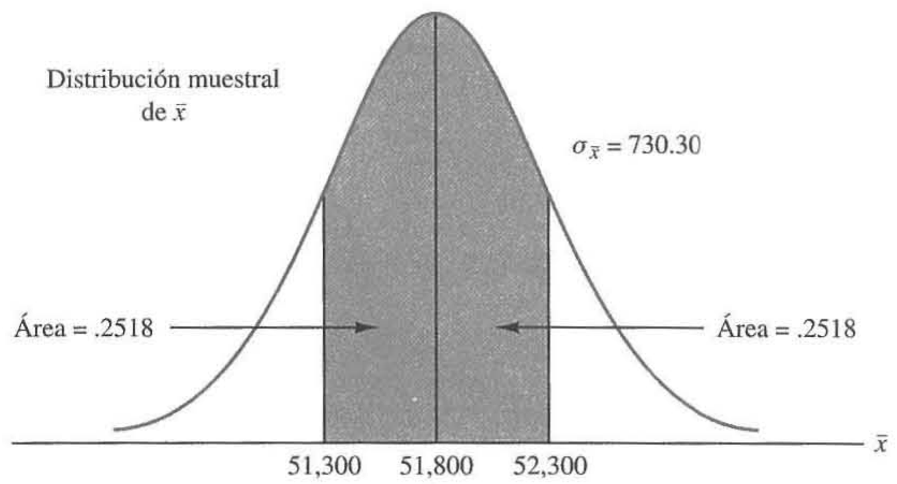
Los cálculos anteriores indican que una muestra aleatoria simple de 30 gerentes de EAI tiene una probabilidad de .5036 de dar como resultado una media de la muestra $\bar{x}$ que quede a $500 dólares o menos de la media de la población. Entonces, hay una probabilidad de que la media de la muestra quede a más de $500 dólares de la media de la población. En otras palabras, una muestra aleatoria simple de 30 gerentes de EAI tiene, aproximadamente una posibilidad de 50-50 de quedar dentro de los $500 dólares permisibles. Quizá deba considerarse una muestra de mayor tamaño. Investigaremos esa posibilidad con la relación entre el tamaño de la muestra y la distribución muestral de $\bar{x}$.

Se puede usar la distribución muestral de $\bar{x}$ para obtener información probabilística sobre la cercanía de $\bar{x}$ a la media poblacional μ.

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

Supongamos que en el problema de EAI seleccionamos una muestra aleatoria simple de 100 gerentes, en lugar de los 30 que consideramos inicialmente. Parece intuitivo que con el incremento en los datos asociados que tiene al mayor tamaño de la muestra, la media de la muestra basada en $n = 100$ sea un estimado mejor del promedio poblacional que el

Figura 7.7 LA PROBABILIDAD DE QUE UNA MEDIA DE MUESTRA QUEDE A $500 DÓLARES O MENOS DE LA MEDIA POBLACIONAL



promedio basado en una muestra de $n = 30$. Para ver el grado de mejoría, examinaremos la relación entre el tamaño de la muestra y la distribución muestral de $\bar{x}$.

Observemos primero que $E(\bar{x}) = \mu$ independientemente del tamaño de muestra. Así, la media de todos los valores posibles de $\bar{x}$ es igual a la media de la población, μ independientemente del tamaño de la muestra n. Sin embargo, vemos que el error estándar de la media, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, se relaciona con la raíz cuadrada del tamaño de la muestra. En forma específica, siempre que aumenta el tamaño de la muestra, disminuye el error estándar de la media, $\sigma_{\bar{x}}$. Con $n = 30$, el error estándar de la media del problema de EAI es 730.30. Sin embargo, al aumentar el tamaño de la muestra a $n = 100$, el error estándar de la media disminuye a

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

Las distribuciones muestrales de $\bar{x}$ con $n = 30$ y $n = 100$ se ven en la figura 7.8. Como la distribución muestral con $n = 100$ tiene un menor error estándar, los valores de $\bar{x}$ tienen menos variación y tienden a estar más cerca de la media de la población que los de $\bar{x}$ con $n = 30$.

Podemos usar la distribución muestral de $\bar{x}$ en el caso con $n = 100$ para calcular la probabilidad de que una muestra aleatoria simple de 100 gerentes de EAI dé como resultado una media de la muestra dentro de un intervalo de \$500 dólares alrededor de la media de la población. Como la distribución muestral es normal, con media de \$51,800 dólares y desviación estándar de 400, podemos emplear la tabla de la distribución normal estándar para determinar el área o probabilidad. Cuando $\bar{x} = 51,300$ (figura 7.9),

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

Al consultar esa tabla vemos que el área entre $z = 0$ y $z = -1.25$ es .3944. Con operaciones parecidas para $\bar{x} = 52,300$, vemos que la probabilidad de que el valor de la media de la muestra esté entre 51,300 y 52,300 es $.3944 + .3944 = .7888$. Así, al aumentar el tamaño de la muestra de 30 a 100 gerentes de EAI, hemos aumentado de .5036 a .7888 la probabilidad de obtener una media de la muestra dentro de un margen de \$500 dólares respecto a la media de la población.

Figura 7.8 COMPARACIÓN DE LAS DISTRIBUCIONES MUESTRALES DE $\bar{x}$ PARA MUESTRAS ALEATORIAS SIMPLES DE $n = 30$ Y $n = 100$ GERENTES DE EAI

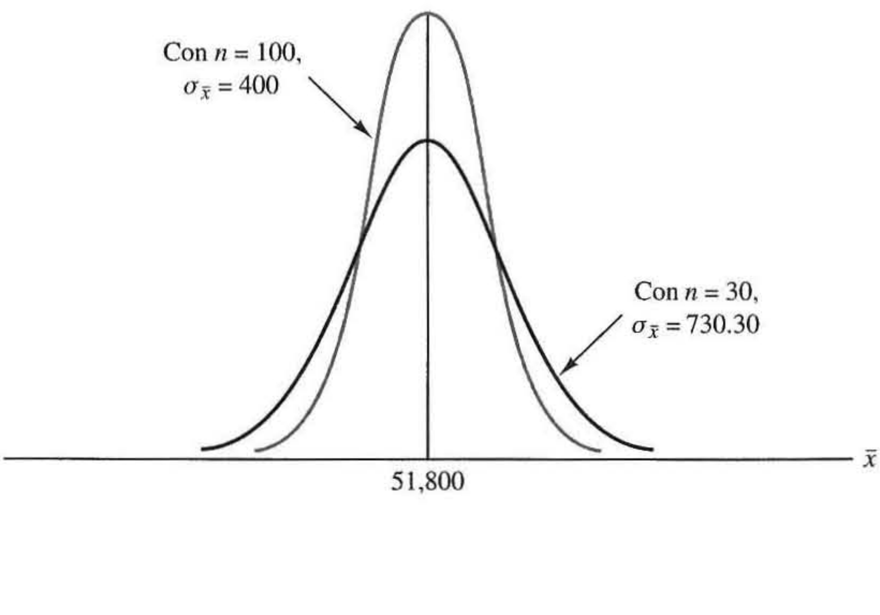
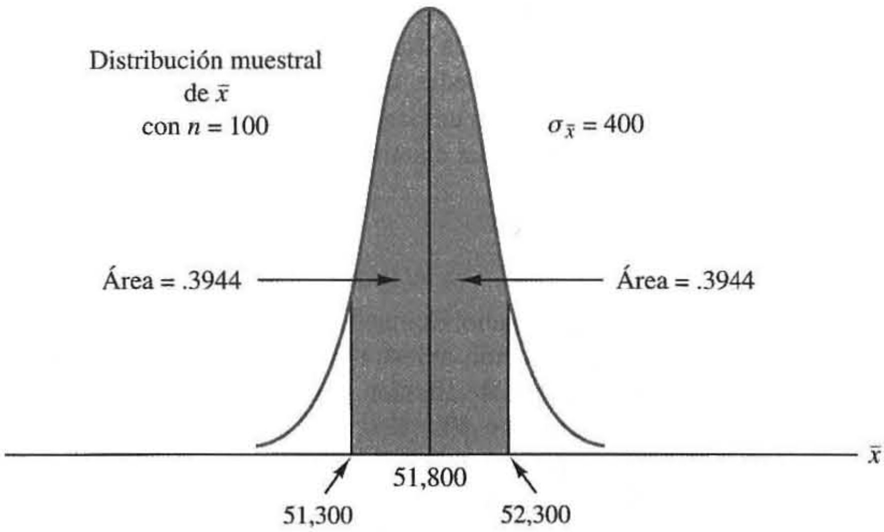


Figura 7.9 PROBABILIDAD DE QUE UNA MEDIA DE MUESTRA QUEDE A $500 DÓLARES O MENOS DE LA MEDIA POBLACIONAL CUANDO SE USA UNA MUESTRA ALEATORIA SIMPLE DE 100 GERENTES DE EAI



Lo importante de todo esto es que cuando se aumenta el tamaño de la muestra, disminuye el error estándar de la media. En consecuencia, con el mayor tamaño de muestra se obtendrá una mayor probabilidad de que la media de la muestra quede dentro de límites especificados respecto a la media de la población.

NOTAS y comentarios

1. Al presentar la distribución muestral de $\bar{x}$ para el problema de EAI, aprovechamos la ventaja de que se conocían la media de la población $\mu = 51{,}800$ y la desviación estándar de la población $\sigma = 4000$. Sin embargo, en el caso general no se conocen los valores de esos parámetros, que se necesitan para determinar la distribución muestral de $\bar{x}$. En el capítulo 8 mostraremos cómo se usan la media y la desviación estándar de la muestra, $\bar{x}$ y s, de una muestra aleatoria simple, cuando se desconocen μ y σ.
2. Para demostrar teóricamente el teorema del límite central se requieren observaciones independientes, o elementos, en la muestra. Esta condición se cumple en las poblaciones infinitas, y en las poblaciones finitas cuando el muestreo se efectúa con remplazo. Aunque ese teorema no menciona directamente el muestreo sin remplazo para poblaciones finitas, la práctica general de la estadística ha sido aplicar los resultados del teorema del límite central en este caso, cuando el tamaño de población es grande.

EJERCICIOS

MÉTODOS

20. Una población tiene 200 de media y 50 de desviación estándar. Se tomará una muestra aleatoria simple de tamaño 100, y se usará la media de la muestra $\bar{x}$ para estimar la media de la población.

- **a.** ¿Cuál es el valor esperado de $\bar{x}$?
- **b.** ¿Cuál es la desviación estándar de $\bar{x}$?
- **c.** Determine la distribución muestral de $\bar{x}$.
- **d.** ¿Qué indica la distribución muestral de $\bar{x}$?

21. ¿Qué papel importante desempeña el teorema del límite central cuando $\bar{x}$ se usa para estimar μ?

AUTOEXAMEN

22. Una población tiene una media de 200 y una desviación estándar de 50. Supongamos que se selecciona una muestra aleatoria simple de tamaño 100, y que se usa $\bar{x}$ para estimar μ.
  a. ¿Cuál es la probabilidad de que la media de la muestra quede dentro de ± 5 de la media de la población?
  b. ¿Cuál es la probabilidad de que la media de la muestra quede dentro de ± 10 de la media de la población?

23. Suponga que $\mu = 32$ y que la desviación estándar es $\sigma = 5$; también, que la población tiene 1000 elementos y que se usa una muestra aleatoria simple de 30 elementos para obtener información sobre esta población.
  a. ¿Cuál es el valor esperado de $\bar{x}$?
  b. ¿Cuál es la desviación estándar de $\bar{x}$?

24. Suponga que la desviación estándar de la población es $\sigma = 25$. Calcule el error estándar de la media, $\sigma_{\bar{x}}$, para tamaños de muestra de 50, 100, 150, y 200. ¿Qué se puede decir acerca del tamaño del error estándar de la media cuando aumenta el tamaño de la muestra?

25. Una muestra aleatoria simple de tamaño 50 se selecciona de una población con $\sigma = 10$. Calcule el valor del error estándar de la media en cada uno de los casos siguientes (si es necesario, aplique el factor de corrección de población finita).
  a. El tamaño de población es infinito.
  b. El tamaño de población es $N = 50{,}000$.
  c. El tamaño de población es $N = 5000$.
  d. El tamaño de población es $N = 500$.

26. Una población tiene 400 de media y 50 de desviación estándar. La distribución de sus probabilidades se desconoce.
  a. Un investigador empleará muestras aleatorias simples de 10, 20, 30 o 40 artículos para reunir datos sobre la población. ¿Con cuál de esas alternativas de tamaño de muestra podrá usar una distribución normal de probabilidades para describir la distribución muestral de $\bar{x}$? Explique su respuesta.
  b. Indique la distribución muestral de $\bar{x}$ para los casos en los que sea adecuada la distribución normal de probabilidades.

27. Una población tiene 100 de media con 16 de desviación estándar. ¿Cuál es la probabilidad de que una media de muestra quede a ± 2 o menos de la media de población, en cada uno de los siguientes tamaños muestrales?
  a. $n = 50$ **b.** $n = 100$ **c.** $n = 200$ **d.** $n = 400$
  e. ¿Cuál es la ventaja de un tamaño grande de muestra?

APLICACIONES

28. Retome de nuevo el problema de muestreo de EAI y suponga que la muestra aleatoria simple hubiera sido de 60 gerentes.
  a. Trace la distribución muestral de $\bar{x}$ cuando se usan muestras aleatorias simples de tamaño 60.
  b. ¿Qué sucede a la distribución muestral de $\bar{x}$ si se usan muestras aleatorias simples de tamaño 120?
  c. ¿Qué afirmación general se puede hacer sobre lo que sucede con la distribución muestral de $\bar{x}$ al aumentar el tamaño de la muestra? ¿Parece lógico? Explique su respuesta.

AUTOEXAMEN

29. En el problema de muestreo de EAI (véase la figura 7.7) demostramos que para $n = 30$, había la probabilidad de .5036 de obtener una media de la muestra a $\pm\$500$ dólares o menos de la media de la población.
  a. ¿Cuál es la probabilidad de que $\bar{x}$ quede a menos de \$500 dólares de la media de la población si se usa un tamaño de muestra de 60?
  b. Conteste el inciso *a* de nuevo pero ahora cuando el tamaño de la muestra es de 120.

30. El precio de la media por galón de gasolina regular vendida en Estados Unidos es de \$1.20 dólares (*The Energy Information Administration*, 3 de marzo de 1997). Suponga que el precio de la media de la población es $\mu = \$1.20$ dólares por galón, y que la desviación estándar de la población es $\sigma = .10$ dólar. También suponga que se selecciona una muestra aleatoria de gasolineras, y que se calcula un precio de la media de la muestra con los datos reunidos en esas gasolineras.
  a. Muestre la distribución de muestreo de la media de la muestra $\bar{x}$ para las 50 gasolineras.
  b. ¿Cuál es la probabilidad de que la muestra aleatoria simple produzca una media de la muestra a menos de \$.02 dólar de la media de la población?
  c. ¿Cuál es la probabilidad de que la muestra aleatoria simple produzca una media de la muestra a menos de \$.01 dólar de la media de la población?

31. Una encuesta efectuada por la Asociación Automovilística Estadounidense mostró que una familia de 4 miembros gasta \$215.60 dólares diarios, en promedio, en sus vacaciones. Suponga que \$215.60 dólares es el promedio poblacional de gastos diarios por familia de 4 y que \$85.00 dólares es la desviación estándar poblacional. Suponga que se selecciona una muestra aleatoria de 40 familias para ciertos estudios.

a. Determine la distribución de muestreo de $\bar{x}$ la media del gasto diario de una familia de 4.

b. ¿Cuál es la probabilidad de que la muestra aleatoria simple de 40 familias produzca una media de la muestra que quede a menos de \$20 dólares de la media de la población?

c. ¿Cuál es la probabilidad de que esa muestra produzca una media de la muestra que quede a menos de \$10 dólares de la media de la población?

32. El Programa de Pruebas Universitario de la Oficina Universitaria Americana reportó una calificación SAT de la media de la población de $\mu = 960$ (*The New York Times*, 1998 Almanac). Suponga que la desviación estándar de la población es $\sigma = 100$.

a. ¿Cuál es la probabilidad de que una muestra aleatoria de 75 estudiantes produzca una media de la muestra de calificación SAT que quede a menos de 10 de la media de población?

b. ¿Cuál es la probabilidad de que una muestra aleatoria de 75 estudiantes produzca una media de la muestra de calificación SAT que quede a menos de 20 de la media de población?

33. La media del sueldo anual de graduados en contabilidad, durante 1996 y 1997, fue de \$30,393 dólares (*U.S. News Online*, 28 de diciembre de 1997). Suponga que μ = \$30,393 dólares para la población de graduados en contabilidad, y que la desviación estándar es σ = \$2000 dólares.

a. ¿Cuál es la probabilidad de que una muestra aleatoria simple de graduados en contabilidad tenga su media dentro de ±\$250 dólares de la media de la población, si el tamaño de la muestra es 30, 50, 100, 200 y 400?

b. ¿Qué ventaja se tiene con una muestra mayor, al tratar de estimar una media de población?

34. En 1993 las mujeres tomaron un promedio de 8.5 semanas sin goce de sueldo en sus trabajos, después del nacimiento de su bebé (*U.S. News & World Report*, 27 de diciembre de 1993). Suponga que 8.5 semanas es la media de población y que 2.2 semanas es la desviación estándar de población.

a. ¿Cuál es la probabilidad de que una muestra aleatoria simple de 50 mujeres arroje una media de muestra de permiso sin goce de sueldo entre 7.5 y 9.5 semanas?

b. ¿Cuál es la probabilidad de que esa muestra tenga una media de entre 8 y 9 semanas?

35. Para estudiar la rapidez de crecimiento de cierta planta, un botánico usará una muestra aleatoria simple de 25 plantas para obtener datos. Después de analizar los datos, cree que el error estándar de la media es demasiado grande. ¿Qué tamaño de muestra aleatoria simple debe usar el botánico para reducir el error estándar a la mitad de su valor actual?

36. El precio promedio poblacional de una casa nueva monofamiliar es de \$166,500 dólares (*New One-Family Houses Sold*, Oficina del Censo de Estados Unidos, 1997). Suponga que la desviación estándar de la población es de \$42,000 y que se seleccionará una muestra de 100 casas monofamiliares nuevas.

a. Presente la distribución muestral del precio de la media de la muestra de casas, basada en la muestra de 100.

b. ¿Cuál es la probabilidad de que la media de la muestra de las 100 compras quede a \$10,000 dólares o menos de la media de la población?

c. Repita el inciso *b* para los valores de \$5000, \$2500 y \$1000 dólares.

d. Para estimar el precio de la media de población con ±\$2500 y ±\$1000 dólares de aproximación, ¿qué recomienda usted?

37. Una llenadora automática de latas de sopa tiene las siguientes características: $\mu = 15.9$ onzas y $\sigma = .5$ onzas.

a. Presente la distribución muestral de $\bar{x}$, que es la media de la muestra para 40 latas que el inspector de control de calidad selecciona al azar.

b. ¿Cuál es la probabilidad de que en una muestra de 40 latas la media sea mayor que 16 onzas?

38. Se informa en la revista Business Week que entre sus suscriptores, los que planean comprar un automóvil nuevo durante los próximos 12 meses pretenden gastar un promedio de \$27,100 dólares. (*Business Week*, Perfil del Suscriptor, 1996). Suponga que el precio del nuevo vehículo, para la población de suscriptores de *Business Week*, tiene una media de $\mu = \$27{,}100$ dólares y que su desviación estándar es $\sigma = \$5200$ dólares.

a. ¿Cuál es la probabilidad de que el precio de la media de la muestra del nuevo vehículo quede a \$1000 dólares o menos de la media de población, si la muestra es de 30 suscriptores?

b. ¿Cuál es la probabilidad de que ese precio quede a \$1000 dólares o menos de la media de población, si la muestra es de 50 suscriptores?

c. ¿Cuál es la probabilidad de que ese precio quede a \$1000 dólares o menos de la media de población, si la muestra es de 100 suscriptores?

d. ¿Recomendaría usted un tamaño muestral de 30, 50 o 100, si se desea tener una probabilidad mínima de .90 de que ese precio quede a \$1000 dólares o menos de la media de la población?

39. Para estimar la edad media de una población de 4000 empleados, se selecciona una muestra aleatoria simple de 40 empleados.

a. ¿Usaría usted el factor de corrección por población finita para calcular el error estándar de la media? Explique por qué.

b. Si la desviación estándar de la población es $\sigma = 8.2$ años, calcule el error estándar, aplicando y sin aplicar el factor de población por población finita. ¿Cuál es el criterio para no tomar en cuenta ese factor siempre que $n/N \leq .05$?

c. ¿Cuál es la probabilidad de que la edad promedio de una muestra de los empleados tenga una aproximación de ± 2 años a la media de la edad de la población?

40. Una biblioteca presta un promedio de $\mu = 320$ libros por día, con desviación estándar $\sigma = 75$ libras. Se tiene una muestra de 30 días de funcionamiento, y $\bar{x}$ es la cantidad de la media de la muestra de libros prestados en un día.

a. Presente la distribución muestral de $\bar{x}$.

b. ¿Cuál es la desviación estándar de $\bar{x}$?

c. ¿Cuál es la probabilidad de que la media de una muestra de 30 días sea entre 300 y 340 libros?

d. ¿Cuál es la probabilidad de que la media de una muestra sea de 325 libros o más prestados diariamente?

7.6 DISTRIBUCIÓN MUESTRAL DE $\bar{p}$

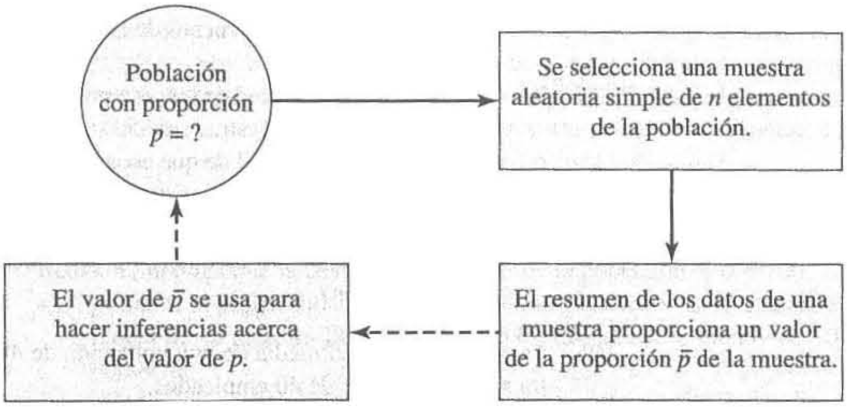
En muchos casos de los negocios y la economía se usa la proporción muestral $\bar{p}$ para hacer inferencias estadísticas sobre la proporción poblacional p. Este proceso se describe en la figura 7.10. Podemos predecir que en cada repetición del proceso obtendremos un valor distinto de la proporción de una muestra $\bar{p}$. La distribución de probabilidades de todos los valores posibles de esa proporción $\bar{p}$ se llama distribución de la proporción muestral $\bar{p}$.

Distribución muestral de $\bar{p}$

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

Para determinar lo cercano que está la proporción muestral $\bar{p}$ de la proporción poblacional p, necesitamos comprender las propiedades de la distribución muestral de $\bar{p}$: su valor esperado, su desviación estándar y la forma de su distribución.

Figura 7.10 PROCESO ESTADÍSTICO PARA USAR UNA PROPORCIÓN MUESTRAL PARA HACER INFERENCIAS ACERCA DE UNA PROPORCIÓN POBLACIONAL



Valor esperado de $\bar{p}$

El valor esperado de $\bar{p}$, la media de todos los valores posibles de $\bar{p}$, se puede expresar como sigue:

Valor esperado de $\bar{p}$

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

en donde

$E(\bar{p})$ = el valor esperado de la variable aleatoria $\bar{p}$
p = la proporción poblacional

La ecuación (7.4) indica que la media de todos los valores posibles de $\bar{p}$ es igual a la proporción p de la población. Recordamos que en la sección 7.1 se demostró que $p = .60$ para la población de EAI, siendo p la proporción de la población de gerentes que participaron en el programa de adiestramiento gerencial de esa empresa. Así, el valor esperado de $\bar{p}$ para el problema de muestreo de EAI es .60.

Desviación estándar de $\bar{p}$

La desviación estándar de $\bar{p}$ se llama *error estándar de la proporción*. Igual que en el caso de la media de muestra $\bar{x}$, la desviación estándar de $\bar{p}$ depende de si la población es finita o infinita. A continuación tenemos las dos ecuaciones de la desviación estándar de $\bar{p}$.

Desviación estándar $\bar{p}$ $q = 1 - p$

Población finita *Población infinita*

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}} \qquad \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{7.5}$$

Al comparar las dos ecuaciones (7.5) vemos que la única diferencia es el empleo del factor de corrección por población finita, $\sqrt{(N-n)/(N-1)}$.

Como en el caso de la media de muestra $\bar{x}$, vemos que la diferencia entre las ecuaciones para población finita e infinita se hace despreciable si el tamaño de la población finita es grande en comparación con el tamaño de la muestra. Seguiremos la misma regla general que recomendamos para la media de muestra. Esto es, si la población es finita y $n/N \leq .05$, usaremos $\sigma_{\bar{p}} = \sqrt{p(1-p)/n}$. Sin embargo, si la población es finita y si $n/N > .05$, se debe usar el factor de corrección por población finita, como lo indica la ecuación (7.5). De nuevo, a menos que se diga otra cosa específicamente, en el libro supondremos que el tamaño de la población es grande en relación con el tamaño de la muestra, y que no es necesario el factor de corrección por población finita.

En el caso del estudio de EAI, sabemos que la proporción de la población de gerentes que participó en el programa de adiestramiento gerencial es $p = .60$. Si $n/N = 30/2500 = .012$, podemos pasar por alto el factor de corrección por población finita al calcular la desviación estándar de $\bar{p}$. Para la muestra aleatoria simple de 30 gerentes, $\sigma_{\bar{p}}$ es

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.60(1-.60)}{30}} = .0894$$

Forma de la distribución muestral de $\bar{p}$

Ahora que conocemos el promedio y la desviación estándar de $\bar{p}$, vamos a conocer la forma de su distribución muestral. Al aplicar el teorema del límite central a $\bar{p}$ se obtiene el resultado siguiente.

La distribución muestral de $\bar{p}$ se puede aproximar con una distribución normal de probabilidades, siempre que el tamaño de muestra sea grande.

En el caso de $\bar{p}$, se puede considerar que el tamaño de muestra es grande cuando se cumplen las dos condiciones siguientes:

$$np \geq 5$$
$$n(1-p) \geq 5$$

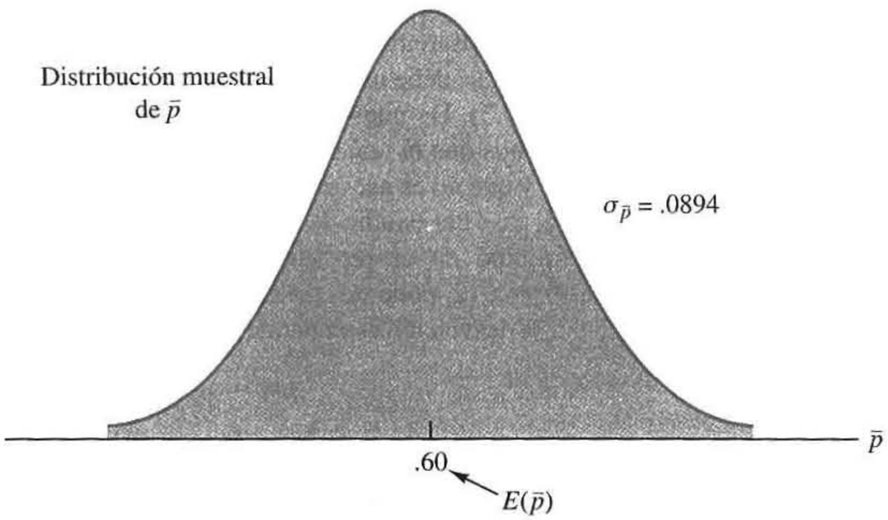
En el problema de muestreo de EAI conocemos que la proporción poblacional de gerentes que participaron en el programa de adiestramiento es $p = .60$. Con una muestra aleatoria simple de tamaño 30, $np = 30(.60) = 18$ y $n(1-p) = 30(.40) = 12$. De esta forma, la distribución muestral de $\bar{p}$ se puede aproximar con una curva de distribución normal de probabilidades como la de la figura 7.11.

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

Siempre que se selecciona una muestra aleatoria simple y que el valor de la proporción de la muestra $\bar{p}$ se usa para estimar el valor de la proporción poblacional p, podemos predecir que hay cierto error de muestreo. En este caso, el error de muestreo es el valor absoluto de la diferencia entre el de la proporción muestral $\bar{p}$ y el de la proporción poblacional p. El valor práctico de la distribución muestral de $\bar{p}$ es que se puede usar para proporcionar información probabilística acerca del error de muestreo.

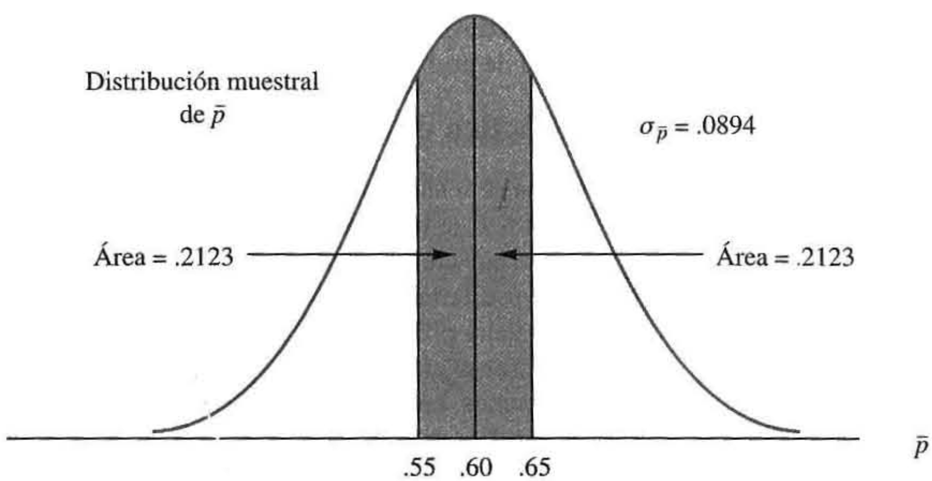
Supongamos, en el problema de EAI, que el director de personal desea conocer la probabilidad de obtener un valor de $\bar{p}$ que se acerque a .05 o más de la proporción poblacional de gerentes de EAI que participaron en el programa de adiestramiento. Esto es, ¿Cuál es la probabilidad de obtener una muestra con proporción muestral $\bar{p}$ entre .55 y

Figura 7.11 DISTRIBUCIÓN MUESTRAL DE $\bar{p}$ PARA LA PROPORCIÓN DE GERENTES DE EAI QUE PARTICIPARON EN EL PROGRAMA DE ADIESTRAMIENTO GERENCIAL



.65? El área de la figura 7.12 indica esa probabilidad. Usando el hecho de que la distribución muestral $\bar{p}$ se puede aproximar con una distribución normal de probabilidades con promedio .60 y desviación estándar $\sigma_{\bar{p}} = .0894$, vemos que la variable aleatoria normal estándar que corresponde a $\bar{p} = .55$ tiene un valor $z = (.55 - .60)/.0894 = -.56$. Al consultar la tabla de distribución normal estándar de probabilidades vemos que el área entre $z = -.56$ y $z = 0$ es .2123. De igual manera, cuando $\bar{p} = .65$ el área entre $z = 0$ y $z = .56$ es .2123. Entonces, la probabilidad de seleccionar una muestra que dé como resultado una proporción muestral más cercana que .05 a la proporción poblacional p es $.2123 + .2123 = .4246$.

Figura 7.12 DISTRIBUCIÓN MUESTRAL DE $\bar{p}$ PARA EL PROBLEMA DE MUESTREO DE EAI



Si deseamos aumentar el tamaño de muestra a $n = 100$, el error estándar de la proporción se transforma en

$$\sigma_{\bar{p}} = \sqrt{\frac{.60(1 - .60)}{100}} = .0490$$

Con un tamaño de muestra de 100 gerentes de EAI, se puede calcular ahora la probabilidad de que la proporción en esa muestra tenga un valor a .05 o menos de la proporción de población. Como la distribución muestral es aproximadamente normal, con media de .60 y desviación estándar de .0490, usamos la tabla de distribución normal estándar de probabilidades para determinar el área o probabilidad. Cuando $\bar{p} = .55$, tenemos que $z = (.55 - .60)/.0490 = -1.02$. En esa tabla vemos que el área entre $z = -1.02$ y $z = 0$ es .3461. Igualmente, en .65 el área entre $z = 0$ y $z = 1.02$ es .3461. Así, si el tamaño de la muestra aumenta de 30 a 100, la probabilidad de que la proporción muestral $\bar{p}$ esté dentro de .05 de la proporción poblacional p aumentará de $.3461 + .3461 = .6922$.

EJERCICIOS

MÉTODOS

41. Se selecciona una muestra aleatoria simple de tamaño 100, de una población con $p = .40$.
- **a.** ¿Cuál es el valor esperado de $\bar{p}$?
- **b.** ¿Cuál es la desviación estándar de $\bar{p}$?
- **c.** Describa la distribución muestral de $\bar{p}$.
- **d.** ¿Qué indica la distribución muestral de $\bar{p}$?

AUTOEXAMEN

42. La proporción de una población es de .40. Se tomará una muestra aleatoria simple de tamaño 200 y se usará la proporción $\bar{p}$ de la muestra para estimar la de la población.
- **a.** ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm.03$ o menos de la proporción poblacional?
- **b.** ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm.05$ o menos de la proporción poblacional?

43. Suponga que la proporción poblacional es de .55. Determine el error estándar de la proporción, $\sigma_{\bar{p}}$, para tamaños de muestra de 100, 200, 500, y 1000. ¿Qué puede decir acerca del error estándar de la proporción al aumentar el tamaño de la población?

44. La proporción de población es de .30. ¿Cuál es la probabilidad de que una proporción muestral esté a $\pm.04$ o menos de la proporción de población para cada uno de los tamaños siguientes de muestra?
- **a.** $n = 100$ **b.** $n = 200$ **c.** $n = 500$ **d.** $n = 1000$
- **e.** ¿Cuál es la ventaja de un mayor tamaño muestral?

APLICACIONES

AUTOEXAMEN

45. El presidente de Distribuidores Díaz, S. A., cree que el 30% de los pedidos a su empresa provienen de clientes nuevos. Se va a usar una muestra aleatoria simple de 100 pedidos para comprobar lo que dice, que $p = .30$.
- **a.** Suponga que el presidente está en lo correcto y que $p = .30$. ¿Cuál es la distribución muestral de $\bar{p}$ para este estudio?
- **b.** ¿Cuál es la probabilidad de que la proporción muestral de $\bar{p}$ esté entre .20 y .40?
- **c.** ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm.05$ o menos de la proporción poblacional $p = .30$?

46. La asociación Grocery Manufacturers of America informa que el 76% de los consumidores leen los ingredientes que aparecen en la etiqueta de los productos (*America by the Numbers*, 1993). Suponga que la proporción de población es $p = .76$, y que de la población se selecciona una muestra de 400 consumidores.

a. Describa la distribución de la proporción muestral $\bar{p}$, que es la proporción, en la muestra, de los consumidores que leen la lista de los ingredientes.

b. ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm.03$ o menos de la proporción poblacional?

c. Conteste el inciso *b* pero ahora con una muestra de 750 consumidores.

47. Louis Harris & Associates llevó a cabo una encuesta entre 403 altos ejecutivos, para conocer cómo evalúan la economía estadounidense durante los 12 meses venideros (*Business Week*, 16 de junio de 1997). Suponga que el 80% de toda la población de ejecutivos se muestra optimista acerca de esa economía. Sea $\bar{p}$ la proporción muestral de ejecutivos encuestados que se muestra optimista acerca de esa economía.

a. Describa la distribución muestral de $\bar{p}$ si la proporción poblacional es $p = .80$.

b. ¿Cuál es la probabilidad de que la proporción muestral, en la encuesta de Harris, quede a $\pm.02$ o menos de la proporción poblacional?

c. ¿Cuál es la probabilidad de que esa distancia sea $\pm.03$ o menor?

48. Si bien la mayoría de las personas cree que el desayuno es el alimento más importante del día, el 25% de los adultos no desayunan (*U.S. News & World Report*, 10 de noviembre de 1997). Suponga que la proporción poblacional es $p = .25$, y que $\bar{p}$ es la proporción poblacional de adultos que no desayunan, determinada con una muestra de 200 adultos.

a. Muestre la distribución muestral de $\bar{p}$.

b. ¿Cuál es la probabilidad de que la proporción muestral quede a $\pm.03$ o menos de la proporción poblacional?

c. ¿Cuál es esa probabilidad con $\pm.05$ o menos?

49. Determinado municipio tiene una tasa de desempleo de 9%. Una agencia estatal lleva a cabo una encuesta mensual de 800 individuos para vigilar la tasa de desempleo en el municipio.

a. Suponga que $p = .09$. ¿Cuál es la distribución muestral de $\bar{p}$ cuando se usa una muestra de tamaño 800?

b. ¿Cuál es la probabilidad de observar una proporción muestral $\bar{p}$ de al menos .08?

50. El Instituto de Investigación de Políticas Femeninas informó que hoy las mujeres forman el 37% de las personas sindicalizadas, porcentaje récord de todos los tiempos (*The Wall Street Journal*, 26 de julio de 1994). Suponga que la proporción poblacional de mujeres sindicalizadas es $p = .37$, y que se selecciona una muestra aleatoria simple de 1000 personas sindicalizadas.

a. Indique la distribución muestral de $\bar{p}$, la proporción de mujeres en la muestra.

b. ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm.03$ de la proporción poblacional?

c. Conteste el inciso *b*, pero ahora con una muestra aleatoria simple de 500.

51. En el problema de muestreo de EAI, ¿cuál es la probabilidad de que la proporción muestral de gerentes que terminaron el programa de adiestramiento, $\bar{p}$, esté a $\pm.05$ o menos de la proporción poblacional, $p = .60$? Emplee tamaños muestrales de 60 y de 120.

52. Suponga que el 15% de los artículos que se producen en una línea de ensamble son defectuosos, pero que el gerente de producción no se ha enterado. También suponga que el departamento de aseguramiento de la calidad prueba 50 piezas para determinar la calidad de la operación de armado. Sea $\bar{p}$ la proporción muestral de piezas defectuosas que encontró la prueba de aseguramiento de calidad.

a. Describa la distribución muestral de $\bar{p}$.

b. ¿Cuál es la probabilidad de que la proporción de la muestra esté a $\pm.03$ o menos de la proporción de piezas defectuosas en la población?

c. Si la prueba indica que $\bar{p} = .10$ o más de piezas defectuosas, la línea de ensamble se para y se investiga la causa de los defectos. ¿Cuál es la probabilidad de que la muestra de 50 piezas lleve a la conclusión de que debe pararse la línea de ensamble?

53. El Instituto de Mercadotecnia de Alimentos indica que el 17% de las familias gastan más de \$100 dólares por semana en abarrotes (*USA Today*, 21 de junio de 1994). Suponga que la proporción poblacional es $p = .17$ y que de ella se selecciona una muestra aleatoria simple de 800 familias.
   a. Describa la distribución muestral de $\bar{p}$, la proporción muestral de familias que gastan más de \$100 dólares por semana en abarrotes.
   b. ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm .02$ o menos de la proporción poblacional?
   c. Conteste el inciso *b*, pero ahora para una muestra de 1600 familias.

7.7 PROPIEDADES DE LOS ESTIMADORES PUNTUALES

En este capítulo hemos mostrado cómo se pueden usar los estadísticos de muestras: la media de la muestra $\bar{x}$, la desviación estándar de muestra s y la proporción muestral $\bar{p}$ como estimadores puntuales de sus correspondientes parámetros poblacionales, μ, σ, y p. Resulta intuitivamente atractivo, que cada uno de esos estadísticos de muestra sea el estimador puntual de su parámetro poblacional correspondiente. Sin embargo, antes de emplear algún estadístico de muestra como estimador puntual, se debe comprobar si tiene ciertas propiedades asociadas con los buenos estimadores puntuales. En esta sección describiremos las propiedades de los buenos estimadores puntuales, que son: insesgo, eficiencia y consistencia.

En vista de que se pueden emplear diversos estadísticos de muestra como estimadores puntuales de distintos parámetros poblacionales, usaremos la siguiente notación general en esta sección.

θ = el parámetro poblacional de interés

$\hat{\theta}$ = el estadístico de muestra o estimador puntual de θ

La notación θ es la letra griega theta, y la notación $\hat{\theta}$ se llama "theta con sombrero". En general, θ representa cualquier parámetro de población, como la media poblacional, desviación estándar poblacional, proporción poblacional, etc.; $\hat{\theta}$ el estadístico de muestra correspondiente, como la media de muestra, la desviación estándar de muestra y la proporción muestral.

Insesgo

Si el valor esperado del estadístico de muestra es igual al parámetro poblacional que se estima, se dice que ese estadístico es un *estimador insesgado* del parámetro poblacional. A continuación definiremos la propiedad de insesgo.

Insesgo

El estadístico de muestra $\hat{\theta}$ es un estimador insesgado del parámetro poblacional θ si

$$E(\hat{\theta}) = \theta \tag{7.6}$$

en donde

$E(\hat{\theta})$ = valor esperado del estadístico de muestra $\hat{\theta}$

Por consiguiente, el valor esperado o media, de todos los valores posibles de un estadístico de muestra insesgado es igual al parámetro de población que se estima.

La figura 7.13 describe los casos de estimadores puntuales insesgado y sesgado. En la ilustración del estimador insesgado, la media de la distribución muestral es igual al valor del parámetro poblacional. Los errores de muestreo se compensan en este caso, porque a veces el valor del estimador puntual $\hat{\theta}$ puede ser menor que θ y otras veces puede ser mayor. En el caso de un estimador sesgado, la media de la distribución muestral es menor que, o mayor que el valor del parámetro poblacional. En la figura 7.13(b), $E(\hat{\theta}) > \theta$; entonces,el estadístico de muestra tiene una gran probabilidad de sobrestimar el valor del parámetro poblacional. La cantidad de sesgo se indica en la figura.

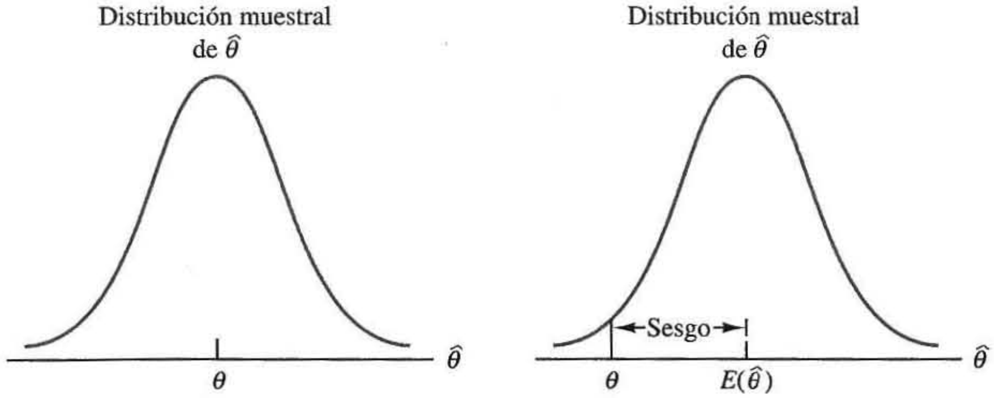
Al describir las distribuciones muestrales de la media y la proporción muestrales, dijimos que $E(\bar{x}) = \mu$ y $E(\bar{p}) = p$. Entonces, tanto $\bar{x}$ como $\bar{p}$ son estimadores insesgados de sus parámetros poblacionales correspondientes, μ y p.

Dejaremos hasta el capítulo 11 una descripción más detallada de la distribución de la desviación estándar de muestra s y de la varianza de muestra s^2. Sin embargo, se puede demostrar que $E(s^2) = \sigma^2$. Así llegamos a la conclusión de que la varianza de muestra s^2 es un estimador insesgado de la varianza poblacional σ^2. De hecho, cuando presentamos por primera vez las fórmulas de la varianza y la desviación estándar de muestra en el capítulo 3, en los denominadores apareció $n - 1$ y no n. La razón de este cambio es hacer que la varianza de muestra sea un estimador insesgado de la varianza poblacional. Si hubiéramos usado n en el denominador, la varianza muestral sería un estimador sesgado, que tendería a subestimar un poco la varianza poblacional.

Eficiencia

Suponga que se puede usar una muestra aleatoria simple de n elementos para obtener dos estimadores puntuales del mismo parámetro poblacional. En este caso, preferiríamos usar el estimador puntual con la menor desviación estándar, porque tiende a proporcionar estimados más cercanos al parámetro poblacional. Se dice que el estimador puntual con la menor desviación estándar tiene una mayor *eficiencia relativa* que el otro.

Figura 7.13 EJEMPLOS DE ESTIMADORES PUNTUALES INSESGADO Y SESGADO



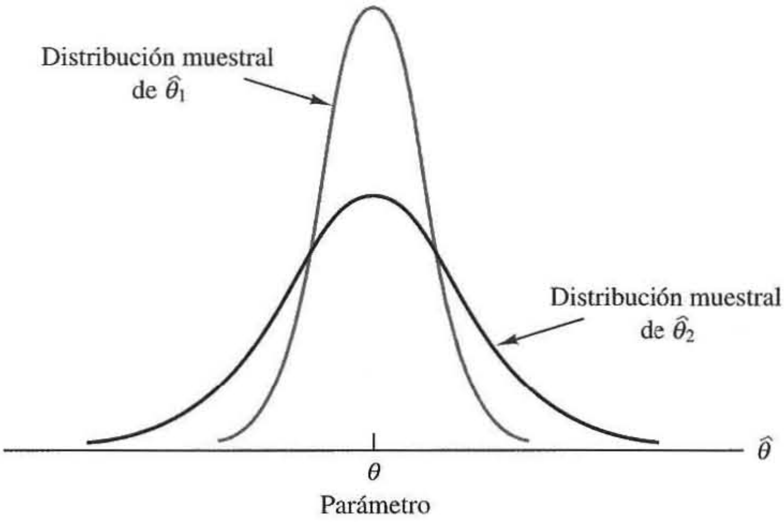
El parámetro θ está ubicado en la media de la distribución muestral; $E(\hat{\theta}) = \theta$

(a) Estimador insesgado

El parámetro θ no está ubicado en la media de la distribución muestral; $E(\hat{\theta}) \neq \theta$

(b) Estimador sesgado

Figura 7.14 DISTRIBUCIONES MUESTRALES DE DOS ESTIMADORES PUNTUALES INSESGADOS



Al tomar muestras de una población con distribución normal, la desviación estándar de la media de muestra es menor que la de la mediana muestral. Por consiguiente, la media de muestra es más eficiente que la mediana muestral.

La figura 7.14 muestra las distribuciones muestrales de dos estimadores puntuales insesgados, $\hat{\theta}_1$ y $\hat{\theta}_2$. Observe que la desviación estándar de $\hat{\theta}_1$ es menor que la de $\hat{\theta}_2$; Entonces, los valores de $\hat{\theta}_1$ tienen mayor posibilidad de estar cerca del parámetro θ que los valores de $\hat{\theta}_2$. Como la desviación estándar del estimador puntual $\hat{\theta}_1$ es menor que la de $\hat{\theta}_2$, $\hat{\theta}_1$ es relativamente más eficiente que $\hat{\theta}_2$ y será el estimador puntual preferido.

Consistencia

Una tercera propiedad asociada con los buenos estimadores puntuales es la consistencia. Hablando en términos generales, un estimador puntual es consistente si sus valores tienden a acercarse al parámetro de población conforme se incrementa el tamaño de la muestra. En otras palabras, un tamaño grande de muestra tiende a proporcionar un mejor estimador puntual que un tamaño pequeño. Observe que, para la media de muestra $\bar{x}$, demostramos que su desviación estándar es $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Como $\sigma_{\bar{x}}$ se relaciona con el tamaño de muestra, de tal manera que las muestras mayores dan menores valores de $\sigma_{\bar{x}}$, llegamos a la conclusión de que un tamaño de muestra mayor tiende a producir estimados puntuales más cercanos a la media de población μ. En este sentido podemos decir que la media de muestra $\bar{x}$ es un estimador consistente de la media de población μ. Con el mismo razonamiento también podemos llegar a la conclusión que la proporción muestral $\bar{p}$ es un estimador consistente de la proporción poblacional p.

NOTAS y comentarios

En el capítulo 3 dijimos que la media y la mediana son dos medidas de la tendencia central. En este capítulo sólo hablamos de la media. La razón es que en el muestreo de una población normal, donde la media y la mediana poblacionales son idénticas, el error estándar de la mediana es, aproximadamente, 25% mayor que el error estándar de la media. Recuerde que en el problema de EAI en el que $n = 30$, el error estándar de la media es $\sigma_{\bar{x}} = 730.30$. El error estándar de la mediana sería, aproximadamente, $1.25 \times (730.30) = 913$. En vista de lo anterior, la media de la muestra tendrá mayor probabilidad de estar dentro de una distancia especificada a la media de población.

7.8 OTROS MÉTODOS DE MUESTREO

En esta sección se presenta una introducción a métodos de muestreo distintos del aleatorio simple. En el capítulo 21 se verán más detalles acerca de esos métodos.

Hemos descrito el procedimiento de muestreo aleatorio simple, y las propiedades de las distribuciones muestrales de $\bar{x}$ y $\bar{p}$ cuando se usa ese muestreo. Sin embargo, el muestreo aleatorio simple no es el único método de muestreo con el que se cuenta. Existen otros, como el muestreo aleatorio estratificado, el de conglomerados y el sistemático, que son alternativas del muestreo aleatorio simple, y en algunos casos presentan ventajas sobre éste. En esta sección describiremos brevemente algunos métodos alternativos de muestreo. En el capítulo 21 se presenta una descripción más detallada de ellos.

Muestreo aleatorio estratificado

En este tipo de muestreo, primero se divide a la población en grupos de elementos llamados estratos, de tal manera que cada elemento en la población pertenece a uno y sólo a un estrato. La base de formación de los estratos, por ejemplo, por departamento, ubicación, edad, giro industrial, etc., queda a discreción de quien diseña la muestra. Sin embargo, los mejores resultados se obtienen cuando los elementos dentro de cada estrato son tan semejantes como sea posible. La figura 7.15 es un diagrama de una población dividida en *H* estratos.

Después de formar los estratos se toma una muestra aleatoria simple de cada uno. Se dispone de fórmulas para combinar los resultados para las muestras de estrato individual en un estimado del parámetro poblacional de interés. El valor del muestreo aleatorio estratificado depende de cuán homogéneos sean los elementos dentro de los estratos. Si son similares (homogeneidad), los estratos tendrán bajas varianzas. En este caso se pueden usar tamaños de muestra relativamente pequeños para obtener buenos estimados de las características de los estratos. Si los estratos son homogéneos, el procedimiento de muestreo aleatorio estratificado producirá resultados tan precisos como el muestreo aleatorio simple, pero con menor tamaño total de muestra.

Muestreo por conglomerados

En el *muestreo por conglomerados*, se divide primero a la población en conjuntos separados de elementos, llamados *conglomerados*. Cada elemento de la población pertenece a uno y sólo a un grupo (véase la figura 7.16). A continuación se toma una muestra aleatoria simple de los conglomerados. Todos los elementos dentro de cada conglomerado muestreado forman la muestra. El muestreo de conglomerados tiende a proporcionar los mejores resultados cuando los elementos de los conglomerados son heterogéneos (desiguales). En el caso ideal, cada conglomerado es una versión representativa, en pequeña escala, de toda la población. El valor del muestreo por conglomerados depende de cuán representativo sea cada conglomerado de la población total.

Figura 7.15 DIAGRAMA DEL MUESTREO ALEATORIO SIMPLE ESTRATIFICADO

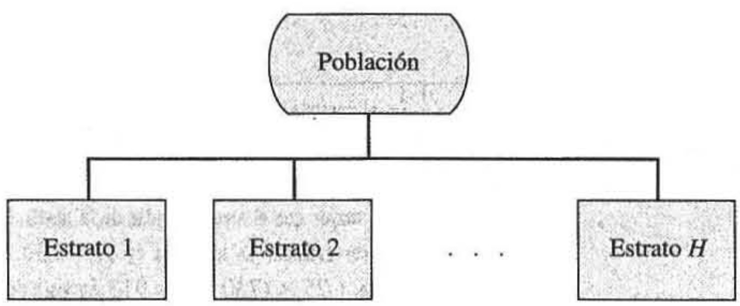
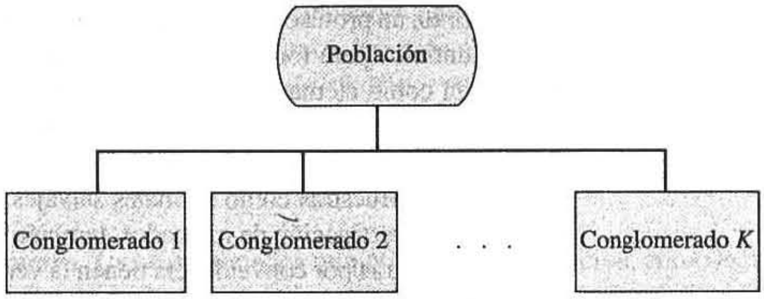


Figura 7.16 DIAGRAMA DEL MUESTREO POR CONGLOMERADOS



Si al respecto todos los conglomerados se asemejan entre sí, al muestrear una pequeña cantidad de conglomerados se obtendrán buenos estimados de los parámetros de población.

Una de las principales aplicaciones del muestreo por conglomerados es el muestreo de áreas, en los que los conglomerados son manzanas urbanas, u otras áreas bien definidas. Por lo general, el muestreo de conglomerados requiere un tamaño de muestra total mayor que el muestreo aleatorio simple o el muestreo aleatorio estratificado. Sin embargo, puede originar ahorros porque cuando se manda a un entrevistador a aplicar un cuestionario a un conglomerado muestreado (por ejemplo, una manzana urbana), se pueden obtener muchas observaciones muestrales en un tiempo relativamente corto. En consecuencia, se puede obtener un mayor tamaño de muestra con un costo bastante menor por elemento, y por ende, probablemente un costo total menor.

Muestreo sistemático

En algunos casos, en especial cuando hay grandes poblaciones, es tardada la selección de una muestra aleatoria simple cuando se determina primero un número aleatorio y después se cuenta o se busca en la lista de elementos de la población hasta encontrar el elemento correspondiente. Una alternativa al muestreo aleatorio simple es el *muestreo sistemático*. Por ejemplo, si se desea una muestra de tamaño 50 de una población con 5000 elementos, podríamos muestrear un elemento de cada 5000/50 = 100 en la población. Una muestra sistemática en este caso implica seleccionar al azar uno de los primeros 100 elementos de la lista de la población. Se identifican los demás elementos de la muestra comenzando por el primero obtenido al azar y a continuación seleccionando cada 100o. elemento. En efecto, se identifica la muestra de 50 recorriendo la población en forma sistemática, e identificando cada 100o.elemento después del primero que se seleccionó al azar. Por lo general será más fácil identificar la muestra de 50 de este modo que si se usara el muestreo aleatorio simple. Ya que el primer elemento se seleccionó de manera aleatoria, generalmente se asume un muestreo sistemático para tener las propiedades de una muestra aleatoria simple. Esta hipótesis se aplica en especial cuando la lista de los elementos de la población es una ordenación aleatoria de ellos.

Muestreo por conveniencia

Los métodos de muestreo que se han descrito se llaman técnicas de *muestreo probabilístico*. Los elementos seleccionados de la población tienen una probabilidad conocida de ser incluidos en la muestra. La ventaja del muestreo probabilístico es que la distribución del estadístico de muestra que se trate, por lo general, se puede identificar. Se pueden usar fórmulas como las del muestreo aleatorio simple, que se presentaron en este capítulo, para determinar las propiedades de la distribución muestral. Ésta se puede usar a continuación para establecer afirmaciones probabilísticas acerca de posibles errores de muestreo asociados con los resultados de la muestra.

El *muestreo por conveniencia* es una técnica de *muestreo no probabilístico*. Como su nombre lo indica, la muestra se identifica, principalmente, por conveniencia. Se incorporan elementos en la muestra sin probabilidades preespecificadas o conocidas de selección. Por ejemplo, un profesor que lleva a cabo una investigación universitaria puede usar alumnos voluntarios para formar una muestra, tan sólo porque dispone fácilmente de ellos y participan como elementos a un costo pequeño o nulo. Igualmente, un inspector puede muestrear un embarque de naranjas seleccionándolas, con mucho tedio y dificultad, entre varias cajas. Sería poco práctico marcar cada naranja y usar un método de probabilidad para el muestreo. Muestras como animales salvajes o plantas silvestres, o equipos de voluntarios para investigación de mercados, también son muestras por conveniencia.

Las muestras por conveniencia tienen la ventaja de la fácil selección y recolección de sus datos. Sin embargo, es imposible evaluar la "bondad" de la muestra en función de su representatividad de la población. Una muestra por conveniencia puede dar buenos resultados o no. No hay un procedimiento estadísticamente justificado que permita un análisis de probabilidades o inferencias acerca de la calidad de los resultados de esa muestra. A veces los investigadores aplican métodos estadísticos diseñados para muestras de probabilidad a una muestra por conveniencia, y dicen que ésta se puede manejar como si fuera una muestra aleatoria. Sin embargo no se puede sostener este argumento, y debemos tener mucho cuidado al interpretar los resultados de muestras por conveniencia, cuando se usan para hacer inferencias acerca de poblaciones.

Muestreo por juicio

Otra técnica más de muestreo no probabilístico es el *muestreo por juicio*. En este método, la persona más capaz en el tema del estudio selecciona a los individuos u otros elementos de la población que siente son los más representativos de esa población. Con frecuencia es una manera relativamente fácil de seleccionar una muestra. Por ejemplo, un reportero puede muestrear a dos o tres senadores, considerando que ellos reflejan la opinión general de todos los senadores. Sin embargo, la calidad de los resultados muestrales depende del juicio de la persona que seleccionó la muestra. De nuevo, se necesita tener gran cuidado al llegar a conclusiones basadas en muestras por juicio, para después hacer inferencias acerca de poblaciones.

NOTAS *y comentarios*

Recomendamos muestrear siguiendo uno de los métodos de muestreo probabilístico: el muestreo aleatorio simple, el aleatorio estratificado simple, el por conglomerados o el sistemático. Para esos métodos se dispone de fórmulas para evaluar la "bondad" de los resultados muestrales en lo referente a la aproximación de ellos a las características poblacionales que se estiman. Con el muestreo por conveniencia o el muestreo por juicio no se puede establecer una evaluación de la bondad. Por consiguiente, se debe tener gran cuidado al interpretar los resultados cuando se usaron métodos de muestreo no probabilístico para obtener información estadística.

RESUMEN

En este capítulo presentamos los conceptos del muestreo aleatorio simple y de las distribuciones muestrales. Demostramos cómo se puede seleccionar una muestra aleatoria simple, y cómo se pueden emplear los datos obtenidos con ella para desarrollar estimados puntuales de los parámetros de población. Como las diversas muestras aleatorias simples dan como resultado una diversidad de valores de los estimadores puntuales, estos estimadores, como $\bar{x}$ y $\bar{p}$ son variables independientes. La distribución de probabilidades de esas variables aleatorias se llama distribución muestral. En particular, describimos las distribuciones de la media de la muestra $\bar{x}$ y de la proporción muestral $\bar{p}$.

Al considerar las características de las distribuciones muestrales de $\bar{x}$ y $\bar{p}$, dijimos que $E(\bar{x}) = \mu$ y $E(\bar{p}) = p$. Después de desarrollar las fórmulas de la desviación estándar o del error estándar para esos estimadores, indicamos cómo el teorema del límite central es la base para usar una distribución normal de probabilidades y aproximar a esas distribuciones muestrales en el caso de muestra grande. Presentamos reglas fáciles para determinar cuándo se satisfacen las condiciones de tamaño de muestra grande. A continuación describimos tres propiedades de los estimadores puntuales: insesgo, eficiencia y consistencia. Describimos brevemente otros métodos de muestreo, como el muestreo aleatorio estratificado, el de por conglomerados y el de por conveniencia.

GLOSARIO

Parámetro Una característica numérica de una población, como la media de población μ, desviación estándar poblacional σ, proporción poblacional p, etcétera.

Muestreo aleatorio simple Para población finita: una muestra seleccionada de tal manera que cada muestra posible de tamaño n tiene la misma probabilidad de ser seleccionada. Para población infinita: una muestra seleccionada de tal manera que cada elemento proviene de la misma población y los elementos sucesivos se seleccionan en forma independiente.

Muestreo sin remplazo Una vez incluido en la muestra un elemento de la población, sale de ésta y ya no se puede seleccionar por segunda vez.

Muestreo con remplazo Al seleccionar cada elemento para la muestra, se regresa a la población. Un elemento que ya se seleccionó se puede volver a seleccionar, y en consecuencia puede aparecer más de una vez en la muestra.

Estadístico de muestra Es una característica de la muestra, como la media de la muestra $\bar{x}$, la desviación estándar de muestra s, la proporción muestral $\bar{p}$, etc. El valor del estadístico de muestra se usa para estimar el valor del parámetro poblacional.

Distribución muestral Una distribución de probabilidades que consta de todos los valores posibles de un estadístico de muestra.

Estimado puntual Un valor numérico único que se usa como estimado de un parámetro poblacional.

Estimador puntual El estadístico de muestra, como $\bar{x}$, s, y $\bar{p}$, que produce el estimado puntual del parámetro poblacional.

Factor de corrección por población finita El término $\sqrt{(N-n)/(N-1)}$ que se usa en las fórmulas de $\sigma_{\bar{x}}$ y $\sigma_{\bar{p}}$ cuando se selecciona una muestra de una población finita, no de una población infinita. La regla fácil que generalmente se acepta es no tomar en cuenta el factor de corrección por población finita siempre que $n/N \leq .05$.

Error estándar La desviación estándar de un estimador puntual.

Teorema del límite central Un teorema que permite usar la distribución normal de probabilidades para aproximar la distribución de $\bar{x}$ y $\bar{p}$ de muestra cuando el tamaño de la muestra es grande.

Insesgo Propiedad de un estimador puntual cuando su valor esperado es igual al parámetro poblacional que estima.

Eficiencia relativa Dados dos estimadores puntuales insesgados del mismo parámetro poblacional, el que tenga la menor varianza es el que tiene mayor eficiencia que el otro.

Consistencia Propiedad de un estimador puntual que se manifiesta cuando los tamaños mayores de muestra tienden a producir estimados puntuales más cercanos al parámetro poblacional.

Muestreo aleatorio simple estratificado Método para seleccionar una muestra en el que primero se divide a la población en estratos y a continuación se toma una muestra aleatoria simple de cada estrato.

Muestreo por conglomerados Método probabilístico de muestreo en el cual primero se divide la población en conglomerados y después se selecciona uno o más conglomerados para muestrearlos.

Muestreo sistemático Método para elegir una muestra seleccionando al azar uno de los primeros k elementos y a continuación cada k-ésimo elemento.

Muestreo por conveniencia Método no probabilístico en el que los elementos se seleccionan para la muestra con base en la conveniencia.

Muestreo por juicio Un método no probabilístico de muestreo en el que se seleccionan los elementos de la muestra basándose en el juicio de la persona que hace el estudio.

FÓRMULAS CLAVE

Valor esperado de $\bar{x}$

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

Desviación estándar de $\bar{x}$

Población finita *Población infinita*

$$\sigma_{\bar{x}} = \sqrt{\frac{N-n}{N-1}}\left(\frac{\sigma}{\sqrt{n}}\right) \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{7.2}$$

Valor esperado de $\bar{p}$

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

Desviación estándar de $\bar{p}$

Población finita *Población infinita*

$$\sigma_{\bar{p}} = \sqrt{\frac{N-n}{N-1}}\sqrt{\frac{p(1-p)}{n}} \qquad \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \tag{7.5}$$

EJERCICIOS COMPLEMENTARIOS

54. Supermercados Nacionales tiene 4800 tiendas al menudeo en todo el país. Al término de cada año se selecciona una muestra de 35 almacenes para inventarios físicos. Los resultados de las muestras de inventario se usan en declaraciones anuales al fisco. Suponga que las tiendas al menudeo están en una lista secuencial de computadora. Comenzando en la parte inferior de la segunda columna de números aleatorios en la tabla 7.1, sin tener en cuenta el primer dígito de cada grupo y usando números aleatorios de cuatro dígitos que comienzan con 8112, avance *hacia arriba* de la columna para identificar las primeras cinco tiendas que se incluirán en la muestra aleatoria simple.
55. Un estudio de *The Orlando Sentinel* en 1993 arrojó datos acerca de cuánto esperan los clientes en Burger King, McDonald's y Wendy. Suponga que la media poblacional del tiempo de espera, en un pedido de despacho en automóvil, es de cuatro minutos, y que la desviación estándar poblacional es de 1.5 minutos.
    a. Suponga que se seleccionará una muestra de clientes en automóvil y que se calculará el tiempo promedio de espera. Presente la distribución de la media de muestra.
    b. ¿Cuál es la probabilidad de que la media de muestra del tiempo de espera esté a $\pm .25$ minutos o menos de la media poblacional del tiempo de espera?

56. La media del tiempo de transporte al trabajo, para las personas de Chicago, es 31.5 minutos (1998 *Information Please Almanac*). Suponga que la media de población es $\mu = 31.5$ minutos y que la desviación estándar poblacional es $\sigma = 12$ minutos. Se selecciona una muestra de 50 residentes de Chicago.
   a. Muestre la distribución muestral de $\bar{x}$ que es la media del tiempo de transporte al trabajo para los 50 residentes de Chicago.
   b. ¿Cuál es la probabilidad de que la media de la muestra quede a ± 1 minuto o menos de la media poblacional?
   c. ¿Cuál es la probabilidad de que la media de la muestra quede a ± 3 minutos o menos de la media poblacional?

57. Un componente eléctrico se diseña para tener una vida media de servicio de 3000 horas, con desviación estándar de 800 horas. Un cliente compra un lote de 50 componentes; suponga que se considera que este lote es una muestra aleatoria simple de la población de componentes. ¿Cuál es la probabilidad de que la vida promedio del grupo de 50 componentes sea, cuando menos, de 2750 horas? ¿O al menos de 3200 horas?

58. La Oficina de Estadísticas del Trabajo, en Estados Unidos, informó que el salario promedio por hora, para personas con ocupaciones ejecutivas, administrativas y gerenciales, es de \$24.07 dólares (*The Wall Street Journal Alamanc* 1998). Suponga que la media de población es $\mu = \$24.07$ dólares y que la desviación estándar poblacional es $\sigma = \$4.80$ dólares. Se va a seleccionar una muestra de 120 individuos en niveles ejecutivos, administrativos y gerenciales.
   a. ¿Cuál es la probabilidad de que la media de muestra quede a $\pm\$0.50$ dólar o menos de la media poblacional?
   b. ¿Cuál es la probabilidad de que la media de muestra quede a $\pm\$1.00$ dólar o menos de la media poblacional?

59. Según la revista *USA Today* (11 de abril de 1995), la cantidad promedio de días que los agentes viajeros están en carretera, en un año, es 115. La desviación estándar es de 60 días por año. Suponga que estos resultados son válidos para la población, y que de ella se escoge una muestra de 50 agentes viajeros.
   a. ¿Qué valor tiene el error estándar del promedio?
   b. ¿Cuál es la probabilidad de que la media de muestra sea más de 115 días por año?
   c. ¿Cuál es la probabilidad de que la media de muestra quede a ± 5 días o menos de la media de población?
   d. ¿Cómo cambiaría la probabilidad obtenida en el inciso *c* si aumentara a 100 el tamaño de la muestra?

60. En una población de 5000 alumnos se selecciona una muestra aleatoria simple de 50 para estimar la media de la población de calificaciones.
   a. ¿Usaría usted el factor de corrección por población finita para calcular el error estándar de la media? Explique por qué.
   b. Si la desviación estándar de la población es $\sigma = .4$, determine el error estándar de la media, primero con y después sin el factor de corrección por población finita. ¿Cuál es la justificación de no tomar en cuenta este factor cuando $n/N \leq .05$?
   c. ¿Cuál es la probabilidad de que la media de las calificaciones en la muestra de 50 alumnos esté a $\pm.10$ o menos de la media poblacional de calificaciones?

61. Tres empresas tienen inventarios de distinto tamaño. La empresa A tiene una población de 2000 artículos, la B de 5000 artículos y la C de 10,000. La desviación estándar de la población del costo de los artículos es $\sigma = 144$. Un consultor en estadística recomienda que cada empresa tome una muestra de 50 artículos de su población para suministrar estimados de validez estadística del costo promedio por artículo. Los gerentes de la empresa pequeña dicen que como tiene la población más pequeña, debería ser posible obtener los datos con una muestra mucho menor que la necesaria para las empresas más grandes. Sin embargo, el consultor dice que para obtener el mismo error estándar, y en consecuencia la misma precisión en los resultados de la muestra, todas las empresas deben usar el mismo tamaño de muestra, independientemente del tamaño de población.
   a. Aplique el factor de corrección por población finita y calcule el error estándar de cada una de las empresas, si la muestra es de tamaño 50.
   b. ¿Cuál es la probabilidad de que para cada empresa la media de muestra $\bar{x}$ quede a ± 25 o menos de la media de población μ?

62. Un investigador informa los resultados de una encuesta diciendo que el error estándar de la media es de 20. La desviación estándar de población es de 500.
    a. ¿De qué tamaño fue la muestra que se usó en esta encuesta?
    b. ¿Cuál es la probabilidad de que el estimado quede a ± 25 o menos de la media de la población?

63. Un inspector de control de calidad supervisa periódicamente un proceso de producción. El inspector selecciona muestras aleatorias simples de 30 productos terminados y calcula la media de muestra de los pesos del producto $\bar{x}$. Si los resultados de las pruebas, durante un gran intervalo de tiempo, muestran que el 5% de los valores de $\bar{x}$ son mayores de 2.1 libras, y el 5% son menores de 1.9 libras, ¿cuáles son la media y la desviación estándar de la población que produce este proceso?

64. ¿Cuál es el factor más importante para los agentes viajeros que se hospedan en un hotel? Según *USA Today*, el 74% de los viajeros dicen que lo más importante es llegar a un cuarto en donde no se permite fumar (*USA Today*, 11 de abril de 1995). Suponga que la proporción poblacional es $p = .74$, y que se selecciona una muestra de 200 agentes viajeros.
    a. Describa la distribución muestral de $\bar{p}$, la proporción muestral de los agentes viajeros que dicen que el factor más importante en su estancia en un hotel es un cuarto en donde no se permite fumar.
    b. ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm .04$ o menos de la proporción poblacional?
    c. ¿Cuál es la probabilidad de que la proporción de la muestra esté a $\pm .02$ o menos de la proporción poblacional?

65. Una empresa de mercadotecnia lleva a cabo encuestas telefónicas con una tasa de respuesta histórica de 40%. ¿Cuál es la probabilidad de que en una nueva muestra de 400 números telefónicos cooperen al menos 150 personas, respondiendo a las preguntas? En otras palabras, ¿cuál es la probabilidad de que la proporción muestral sea, cuando menos, $150/400 = .375$?

66. Una corrida de producción no se acepta para su embarque a los clientes si una muestra de 100 artículos contiene 5% o más de piezas defectuosas. Si una corrida de producción tiene una proporción poblacional de piezas defectuosas de $p = .10$, ¿cuál es la probabilidad de que $\bar{p}$ sea cuando menos de .05?

67. La proporción de individuos asegurados por una empresa, que recibieron cuando menos una multa de tráfico durante un periodo de 5 años es de .15.
    a. Describa la distribución muestral de $\bar{p}$ si se usa una muestra de 150 personas aseguradas para estimar la proporción de quienes recibieron al menos una multa en ese periodo.
    b. ¿Cuál es la probabilidad de que la proporción muestral esté a $\pm .03$ o menos de la proporción de población?

68. Lori Jeffrey es una exitosa representante de ventas de una gran editorial de libros de texto universitarios. De los registros de llamadas a sus clientes el 25% de éstas resultan en la adopción del libro que ella sugiere. Considerando que sus llamadas durante un mes forman una muestra de todas las llamadas posibles de ventas, suponga que un análisis estadístico de los datos da como resultado un error estándar de población igual a .0625.
    a. ¿De qué tamaño fue la muestra que se usó en este análisis? Esto es, ¿cuántas llamadas hizo Lori durante el mes?
    b. Sea $\bar{p}$ la proporción muestral de adopciones de libro obtenida durante el mes. Describa la distribución muestral de $\bar{p}$.
    c. Con la distribución muestral de $\bar{p}$, calcule la probabilidad de que Lori logre adopciones de libro en 30% o más de sus llamadas durante un periodo de un mes.

APÉNDICE 7.1 EL VALOR ESPERADO Y LA DESVIACIÓN ESTÁNDAR DE $\bar{x}$

En este apéndice presentaremos la base matemática de las expresiones para $E(\bar{x})$, el valor esperado de $\bar{x}$ de acuerdo con la ecuación (7.1), y de $\sigma_{\bar{x}}$, la desviación estándar de $\bar{x}$ según la ecuación (7.2).

Valor esperado de $\bar{x}$

Suponga que una población tiene medio μ y varianza σ^2. Se selecciona una muestra aleatoria simple de tamaño n, y sus observaciones individuales se representan con x_1, $x_2, \ldots, x_n$. La media muestra se calcula como sigue:

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

Con muestras aleatorias simples repetidas de tamaño n, $\bar{x}$ es una variable aleatoria que asume distintos valores numéricos, dependiendo de los n artículos específicos que se hayan seleccionado. El valor esperado de la variable aleatoria $\bar{x}$ es la media de todos los valores posibles de $\bar{x}$.

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

Como $E(x_i) = \mu$, para toda x_i, podemos escribir

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

Esta ecuación indica que la media de todos los valores posibles de $\bar{x}$ es igual a la media poblacional μ. Esto es, que, $E(\bar{x}) = \mu$.

Desviación estándar de $\bar{x}$

De nuevo supongamos una población con media μ, varianza σ^2, y una media de muestra

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

Con muestras aleatorias simples repetidas de tamaño n, sabemos que $\bar{x}$ es una variable aleatoria que toma distintos valores numéricos, dependiendo de los elementos específicos n que se hayan seleccionado. A continuación deduciremos la ecuación de la desviación estándar de los valores de $\bar{x}$ que es, $\sigma_{\bar{x}}$, para el caso de una población infinita. La deducción de la ecuación de $\sigma_{\bar{x}}$ cuando la población es finita y el muestreo se lleva a cabo sin remplazo es más difícil y sale de los propósitos de este libro.

Regresando al caso de la población infinita, recordamos que una muestra aleatoria simple de una población infinita está formada por observaciones $x_1, x_2, \ldots, x_n$ que son independientes. Las dos ecuaciones siguientes son fórmulas generales de la varianza de variables aleatorias.

$$\text{Var}(ax) = a^2\,\text{Var}(x) \tag{7.7}$$

siendo a una constante y x una variable aleatoria. También

$$\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y) \tag{7.8}$$

en donde x y y son variables aleatorias *independientes*. Con las ecuaciones (7.7) y (7.8) podemos deducir la expresión de la varianza de la variable aleatoria $\bar{x}$ como sigue:

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

Aplicando la ecuación (7.7), en donde $1/n$ se considera como la constante, tenemos

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

$$= \left(\frac{1}{n}\right)^2 \text{Var}(x_1 + x_2 + \cdots + x_n)$$

En el caso de la población infinita, las variables aleatorias $x_1, x_2, \ldots, x_n$ son independientes. En consecuencia, la ecuación (7.8) nos permite escribir

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

Porque para toda x_i, $\text{Var}(x_i) = \sigma^2$, y entonces

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

Con n valores de σ^2 en esta ecuación,

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

Al sacar la raíz cuadrada llegamos a la ecuación de la desviación estándar de $\bar{x}$.

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