Estadística

Autor: Triola, M. F.

Pearson, México, 2006, págs. 269-277, 280-286, 291-298

ISBN: 978-970-26-1287-2

Datos de catalogación bibliográfica

TRIOLA, MARIO F.

Estadística. Décima edición

PEARSON EDUCACIÓN, México, 2009

ISBN: 978-970-26-1287-2
Área: Matemáticas

Formato: 21 × 27 cm Páginas: 904

ISBN 9780321460929

Edición en español

Editor: Rubén Fuerte Rivera
e-mail: ruben.fuerte@pearsoned.com
Editor de desarrollo: Felipe Hernández Carrasco
Supervisor de producción: Juan José García Guzmán

Edición en inglés

Publisher: Greg Tobin
Executive Editor: Deirdre Lynch
Executive Project Manager: Christine O'Brien
Assistant Editor: Sara Oliver
Managing Editor: Ron Hampton
Senior Production Supervisor: Peggy McMahon
Senior Designer: Barbara T. Atkinson
Photo Researcher: Beth Anderson
Digital Assets Manager: Marianne Groth
Production Coordinator, Supplements: Emily Portwood
Media Producer: Cecilia Fleming
Software Development: Ted Hartman and Janet Wann
Marketing Manager: Phyllis Hubbard
Marketing Assistant: Celena Carr
Senior Author Support/Technology Specialist: Joe Vetere
Senior Prepress Supervisor: Caroline Fell
Rights and Permissions Advisor: Dana Weightman
Senior Manufacturing Buyer: Evelyn Beaton
Text and Cover Design: Leslie Haimes
Production Services, Composition and Illustration: Nesbitt Graphics

Cover Photo: Getty Images/Jean Louis Batt

DÉCIMA EDICIÓN, 2009

Cámara Nacional de la Industria Editorial Mexicana. Reg. Núm. 1031

ISBN 10: 970-26-1287-X
ISBN 13: 978-970-26-1287-2

PEARSON Educación ®

Impreso en México. *Printed in Mexico.*

1 2 3 4 5 6 7 8 9 0 - 11 10 09 08

b. Utilice la corrección por continuidad y calcule la probabilidad de seleccionar al azar a alguien con una puntuación de CI mayor que 105.
c. Compare los resultados de los incisos *a*) y *b*).

29. Normalización de calificaciones de un examen. Una profesora de estadística aplica un examen y descubre que las calificaciones se distribuyen normalmente, con una media de 25 y una desviación estándar de 5. Ella planea normalizar las calificaciones.
a. Si las normaliza sumando 50 a cada calificación, ¿cuál es la nueva media? ¿Cuál es la nueva desviación estándar?
b. ¿Será justo normalizarlas sumando 50 a cada calificación? ¿Por qué?
c. Si las calificaciones se normalizan según el siguiente esquema (en vez de sumar 50), calcule los límites numéricos de cada calificación.
A: 10% superior
B: calificaciones por arriba del 70% inferior y por debajo del 10% superior
C: calificaciones por arriba del 30% inferior y por debajo del 30% superior
D: calificaciones por arriba del 10% inferior y por debajo del 70% superior
F: 10% inferior
d. ¿Cuál método de normalización de las calificaciones es más justo: sumar 50 a cada calificación o emplear el esquema dado en el inciso *c*)? Explique.

30. Pruebas SAT y ACT. Las calificaciones de mujeres en la prueba SAT-I se distribuyen de manera normal, con una media de 998 y una desviación estándar de 202. Las calificaciones de mujeres en la prueba ACT se distribuyen de manera normal, con una media de 20.9 y una desviación estándar de 4.6. Suponga que las dos pruebas emplean escalas distintas para medir la misma habilidad.
a. Si una mujer obtiene una calificación en la prueba SAT que corresponde al percentil 67, calcule su calificación real en la prueba SAT y su calificación equivalente en la prueba ACT.
b. Si una mujer obtiene una calificación de 1220 en la prueba SAT, calcule su calificación equivalente en la prueba ACT.

6-4 Distribuciones y estimadores muestrales

Concepto clave El principal objetivo de esta sección es comprender el concepto de una *distribución muestral de un estadístico*, que es la distribución de todos los valores de ese estadístico cuando se obtienen todas las muestras posibles del mismo tamaño de la misma población. En específico, analizaremos la distribución muestral de la proporción y la distribución muestral de la media. También veremos que algunos estadísticos (como la proporción y la media) son útiles para estimar valores de parámetros poblacionales, mientras que otros estadísticos (como la mediana) no son buenos estimadores de los parámetros poblacionales.

En una encuesta realizada por el Education and Resources Institute, se eligieron al azar 750 estudiantes universitarios, y el 64% (o 0.64) de ellos dijeron que tenían al menos una tarjeta de crédito. Esta encuesta incluye únicamente a 750 participantes de una población de aproximadamente 15 millones de estudiantes universitarios. Sabemos que los estadísticos muestrales varían de manera natural de una muestra a otra, de manera que otra muestra de 750 estudiantes universitarios tal vez genere una proporción muestral diferente del 64% encontrado en la primera encuesta. En este contexto, debemos considerar a la proporción muestral como una variable aleatoria.

Dado que la muestra de 750 estudiantes universitarios representa un porcentaje tan pequeño de la población (0.005%), ¿realmente podemos esperar que la proporción muestral sea un estimado razonable de la proporción real de todos los estudiantes universitarios que tienen tarjetas de crédito? ¡Sí! Los profesionales de la estadística, al ser individuos tan inteligentes, han diseñado métodos para utilizar resultados muestrales y estimar parámetros poblacionales con bastante exactitud. ¿Como lo hacen? Su enfoque se basa en la comprensión del comportamiento de la estadística. Al entender la *distribución* de una proporción muestral, los especialistas en estadística pueden determinar qué tan exacta puede ser una proporción muestral individual. Además, ellos comprenden la distribución de medias muestrales. En general, conocen claramente el concepto de la *distribución muestral de un estadístico.*

¿Prevalecen los niños o las niñas en las familias?

El autor de este libro, sus hermanos y sus sobrinos suman un total de 11 hombres y sólo una mujer. ¿Será este ejemplo un fenómeno en el que un género particular prevalece en una familia? Este tema fue estudiado examinando una muestra aleatoria de 8770 hogares de Estados Unidos. Los resultados se reportaron en la revista *Chance*, en el artículo "Does Having Boys or Girls Run in the Family?", escrito por Joseph Rodgers y Debby Doughty. Parte de su análisis implica el uso de la distribución de probabilidad binomial. Ellos concluyeron que no encontraron "evidencias contundentes de que un sexo prevalezca más en la familia".

Definición

La **distribución muestral de un estadístico** (como una proporción muestral o una media muestral) es la distribución de todos los valores del estadístico cuando se obtienen todas las muestras posibles del mismo tamaño n de la misma población. (La distribución muestral de un estadístico generalmente se representa como la distribución de probabilidad en el formato de tabla, histograma de probabilidad o fórmula).

Distribución muestral de proporciones

La definición anterior se puede aplicar al estadístico específico de una proporción muestral.

Definición

La **distribución muestral de la proporción** es la distribución de probabilidad de proporciones muestrales, donde todas las muestras tienen el mismo tamaño muestral n y provienen de la misma población.

Comprenderemos mejor el importante concepto de una distribución muestral de la proporción si consideramos algunos ejemplos específicos.

EJEMPLO Distribución muestral de la proporción de niñas en dos nacimientos Cuando se eligen dos nacimientos al azar, el espacio muestral es hh, hm, mh, mm. Esos cuatro resultados igualmente probables sugieren que la probabilidad de 0 niñas es de 0.25, la probabilidad de una niña es de 0.50 y la probabilidad de 2 niñas es de 0.25. La imagen que se muestra más adelante indica la distribución de probabilidad del *número* de niñas, seguida por dos formatos diferentes (tabla y gráfica) que describen la distribución muestral de la *proporción* de niñas.

Una distribución muestral también se puede expresar como una fórmula (véase el ejercicio 15), además de los formatos de tabla y gráfica, o se puede describir de otra manera, como la siguiente: "La distribución muestral de la media de la muestra es una distribución normal con $\mu = 100$ y $\sigma = 15$". En esta sección, generalmente describimos una distribución muestral utilizando una tabla que lista los valores del estadístico de muestra, junto con sus probabilidades correspondientes. En capítulos posteriores usaremos algunas de las otras descripciones.

Aunque por lo general las encuestas implican tamaños muestrales de alrededor de 1000 a 2000, y tamaños poblacionales de hasta millones, el siguiente ejemplo implica una población de sólo tres valores, de manera que es fácil listar cada muestra posible.

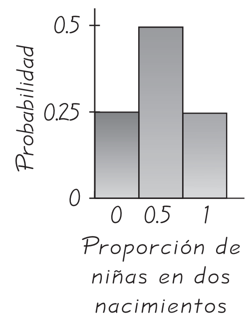
Número de niñas en 2 nacimientos *x*	*P(x)*
0	0.25
1	0.50
2	0.25

Distribución muestral de la proporción de niñas en 2 nacimientos

Tabla

Histograma de probabilidad

Proporción de niñas en 2 nacimientos	*Probabilidad*
0	0.25
0.5	0.50
1	0.25



EJEMPLO Distribución muestral de proporciones Un mariscal de campo lanzó 1 intercepción en su primer juego, 2 intercepciones en su segundo juego, 5 intercepciones en su tercer juego y después se retiró. Considere la *población* consistente en los valores 1, 2, 5. Observe que dos de los valores (1 y 5) son impares, de manera que la proporción de números impares en la población es 2/3.

a. Liste todas las muestras diferentes posibles de tamaño $n = 2$ seleccionadas con reemplazo. (Posteriormente explicaremos por qué el muestreo *con reemplazo* es tan importante). Para cada muestra, calcule la proporción de números *impares*. Utilice una tabla para representar la distribución muestral de la proporción de números impares.

b. Calcule la media de la distribución muestral para la proporción de números impares.

c. Para la población de 1, 2, 5, la proporción de números impares es 2/3. ¿La media de la distribución muestral de la proporción de números impares también es igual a 2/3? ¿Las proporciones muestrales coinciden con el valor de la proporción poblacional? Es decir, ¿las proporciones de la muestra tienen una media igual a la proporción poblacional?

continúa

Profetas de las ganancias

Muchos libros y programas de computadoras aseguran ser útiles para predecir números ganadores de la lotería. Algunos utilizan la idea de que ciertos números están “rezagados” (y deberían seleccionarse), ya que no han salido con frecuencia; otros creen que algunos números están “fríos” (y deben evitarse), ya que no han salido con frecuencia; y aún existen otros que utilizan la astrología, la numerología o los sueños. Como la selección de las combinaciones de números ganadores de la lotería son sucesos independientes, tales suposiciones son inútiles. Un método válido es el de elegir números que sean “raros”, en el sentido de que no son seleccionados por otras personas, de manera que si usted gana, no se verá obligado a compartir sus ganancias con otros. Por esta razón, la combinación de 1, 2, 3, 4, 5 y 6 es inadecuada, ya que muchos individuos la utilizan, mientras que la combinación 12, 17, 18, 33, 40, 46 es mucho mejor, al menos hasta la publicación de este libro.

Tabla 6-2 Distribución muestral de proporciones de números impares

Muestra	Proporción de números impares	Probabilidad
1, 1	1	1/9
1, 2	0.5	1/9
1, 5	1	1/9
2, 1	0.5	1/9
2, 2	0	1/9
2, 5	0.5	1/9
5, 1	1	1/9
5, 2	0.5	1/9
5, 5	1	1/9

SOLUCIÓN

a. En la tabla 6-2 se listan las nueve muestras diferentes posibles de tamaño $n = 2$, obtenidas con reemplazo de la población de 1, 2, 5. Esta tabla también contiene el número de valores muestrales que son números impares, incluyendo sus probabilidades. (Como existen 9 muestras igualmente probables, cada muestra tiene una probabilidad de 1/9). La tabla 6-3, que simplemente es una versión condensada de la tabla 6-2, representa de forma concisa la distribución muestral de la proporción de números impares.

b. La tabla 6-3 es una distribución de probabilidad, de manera que podemos calcular su media utilizando la fórmula 5-1 de la sección 5-2. La media de 2/3 se obtiene de la siguiente manera:

$$\mu = \Sigma[x \cdot P(x)] = (0 \cdot 1/9) + (0.5 \cdot 4/9) + (1 \cdot 4/9) = 6/9 = 2/3$$

c. La media de la distribución muestral de proporciones es 2/3, y 2/3 de los números en la población son impares. Esto no es una coincidencia. En general, la distribución muestral de proporciones tendrá una media que es igual a la proporción poblacional. A esto nos referimos cuando decimos que las proporciones de la muestra "coinciden" con la proporción poblacional.

INTERPRETACIÓN En el caso de seleccionar dos valores (con reemplazo) de la población de 1, 2, 5, ya identificamos la distribución muestral (tabla 6-3). También encontramos que la media de la distribución muestral es 2/3, que es igual a la proporción de números impares en la población. Por lo tanto, las proporciones muestrales tienden a coincidir con la proporción poblacional, en vez de tender sistemáticamente a subestimar o sobreestimar ese valor.

Tabla 6-3 Versión condensada de la tabla 6-2

Proporción de números impares	Probabilidad
0	1/9
0.5	4/9
1	4/9

El ejemplo anterior incluye una proporción bastante pequeña, de manera que ahora consideraremos los géneros de los senadores en el CVII (centésimo séptimo) Congreso. Como sólo existen 100 miembros [13 mujeres (M) y 87 hombres (H)], podemos listar la población completa:

H	**M**	H	H	**M**	H	H	H	H	H	H	H	**M**	H	H	H	H	H	H	H
H	H	H	H	H	H	H	H	H	H	H	H	**M**	**M**	H	H	H	H	H	H
H	H	H	**M**	H	H	H	H	**M**	H	H	H	H	H	H	H	H	H	H	H
M	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	**M**	**M**	**M**
H	H	H	**M**	H	**M**	H	H	H	H	H	H	H	H	H	H	H	H	H	H

La proporción poblacional de senadoras es $p = 13/100 = 0.13$. Por lo general, no conocemos a todos los miembros de la población, por lo que debemos estimarla a partir de una muestra. Con el propósito de estudiar el comportamiento de las proporciones muestrales, listamos unas pocas muestras de tamaño $n = 10$ e indicamos la proporción correspondiente de mujeres.

Muestra 1: H **M** H H **M** H H H H H $\rightarrow$ la proporción muestral es 0.2
Muestra 2: H **M** H H H H H H H H $\rightarrow$ la proporción muestral es 0.1
Muestra 3: H H H H H H **M** H H H $\rightarrow$ la proporción muestral es 0.1
Muestra 4: H H H H H H H H H H $\rightarrow$ la proporción muestral es 0
Muestra 5: H H H H H H H H **M** H $\rightarrow$ la proporción muestral es 0.1

Tabla 6-4 Resultados de 100 muestras

Proporción de senadoras	Frecuencia
0.0	26
0.1	41
0.2	24
0.3	7
0.4	1
0.5	1
Media:	0.119
Desviación estándar:	0.100

Preferimos no listar las 100,000,000,000,000,000,000 muestras diferentes posibles. En vez de ello, el autor seleccionó al azar 95 muestras adicionales, antes de parar de rodar las llantas de su automóvil. Si combinamos las 95 muestras adicionales con las cinco listadas antes, obtenemos las 100 muestras que se resumen en la tabla 6-4.

En la tabla 6-4 podemos ver que la media de las 100 proporciones muestrales es 0.119, pero si deseamos incluir todas las demás muestras posibles de tamaño 10, la media de las proporciones muestrales sería igual a 0.13, que es el valor de la proporción poblacional. La figura 6-17 presenta la distribución de las 100 proporciones de muestras resumidas en la tabla 6-14. La forma de esa distribución es bastante cercana a la que se habría obtenido con todas las muestras posibles de tamaño 10. Podemos observar que la distribución que se ilustra en la figura 6-17 tiene cierto sesgo hacia la derecha, pero con un poco de alargamiento podría aproximarse a una distribución normal. En la figura 6-18 mostramos los resultados obtenidos de 10,000 muestras, de tamaño 50, seleccionadas al azar y con reemplazo, de la lista anterior de 100 géneros. La figura 6-18 sugiere enfáticamente que la distribución se aproxima a la forma de campana que caracteriza a una distribución normal. Por consiguiente, los resultados de la tabla 6-4 y de la figura 6-18 sugieren lo siguiente.

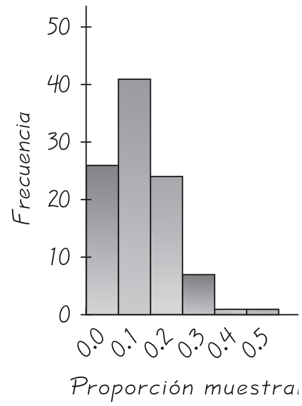
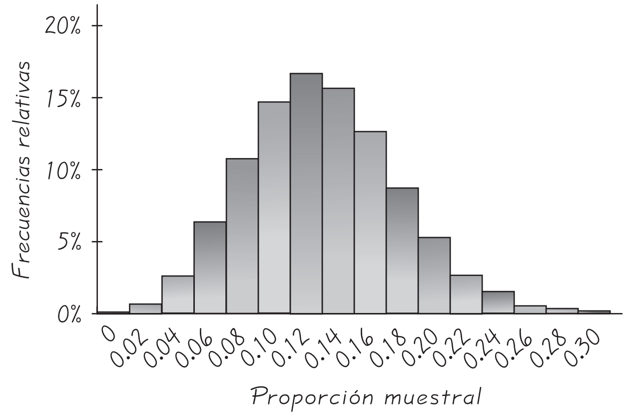


Figura 6-17

100 proporciones muestrales con $n = 10$ en cada muestra

Figura 6-18

10,000 proporciones muestrales con $n = 50$ en cada muestra



Propiedades de la distribución de proporciones muestrales

- **Las proporciones muestrales tienden a coincidir con el valor de la proporción poblacional. (Es decir, todas las proporciones muestrales posibles tienen una media igual a la proporción poblacional).**
- **En ciertas condiciones, la distribución de la proporción muestral puede aproximarse por medio de una distribución normal.**

Distribución muestral de la media

Ahora consideremos la distribución muestral de la media.

Definición

La **distribución muestral de la media** es la distribución de medias muestrales, donde todas las medias tienen el mismo tamaño muestral n y se obtienen de la misma población. (La distribución muestral de la media generalmente se representa como una distribución de probabilidad en formato de tabla, histograma de probabilidad o fórmula).

Nuevamente, en vez de usar conceptos demasiado abstractos, usaremos una población pequeña para ilustrar las propiedades importantes de esta distribución.

EJEMPLO Distribución muestral de la media Una *población* consiste en los valores 1, 2 y 5. Observe que la media de esta población es $\mu = 8/3$.

a. Liste todas las muestras posibles (con reemplazo) de tamaño $n = 2$, así como las medias muestrales y sus probabilidades individuales.

b. Calcule la media de la distribución muestral.

c. La media poblacional es 8/3. ¿Las medias muestrales coinciden con el valor de la media poblacional?

SOLUCIÓN

a. En la tabla 6-5 se listan las nueve muestras diferentes posibles de tamaño $n = 2$, obtenidas con reemplazo de la población 1, 2 y 5. La tabla 6-5 también presenta las medias muestrales e incluye sus probabilidades. (Como hay 9 muestras igualmente probables, cada muestra tiene una probabilidad de 1/9). La tabla 6-6, que es una versión condensada de la tabla 6-5, representa de forma concisa la distribución muestral de las medias muestrales.

b. La tabla 6-6 es una distribución de probabilidad, de manera que podemos calcular su media con la fórmula 5-1 de la sección 5-2. La media de 8/3 se obtiene de la siguiente manera:

$$\mu = \Sigma[x \cdot P(x)] = (1.0 \cdot 1/9) + (1.5 \cdot 2/9) + \cdots + (5.0 \cdot 1/9)$$
$$= 24/9 = 8/3$$

c. La media de la distribución muestral de proporciones es 8/3, y la media de la población también es 8/3. Nuevamente, no se trata de una coincidencia. En general, la distribución de medias muestrales tiene una media igual a la media poblacional. Por lo tanto, las medias muestrales tienden a coincidir con la media poblacional, en vez de ser sistemáticamente demasiado bajas o demasiado altas.

INTERPRETACIÓN En el caso de seleccionar dos valores (con reemplazo) de la población 1, 2 y 5, ya hemos identificado la distribución muestral (tabla 6-6). También observamos que la media de la distribución muestral es 8/3, que es igual a la media poblacional. Por consiguiente, las medias muestrales tienden a coincidir con la media poblacional.

Tabla 6-5 Distribución muestral de $\bar{x}$

Muestra	Media $\bar{x}$	Probabilidad
1, 1	1.0	1/9
1, 2	1.5	1/9
1, 5	3.0	1/9
2, 1	1.5	1/9
2, 2	2.0	1/9
2, 5	3.5	1/9
5, 1	3.0	1/9
5, 2	3.5	1/9
5, 5	5.0	1/9

↓

Tabla 6-6 Versión condensada de la tabla 6-5

$\bar{x}$	Probabilidad
1.0	1/9
1.5	2/9
2.0	1/9
3.0	2/9
3.5	2/9
5.0	1/9

Del ejemplo anterior observamos que la media de todas las medias muestrales posibles es igual a la media de la población original, $\mu = 8/3$. Podemos generalizar esto como una propiedad de las medias muestrales: para un tamaño muestral fijo, la media de todas las medias muestrales posibles es igual a la media de la población. Revisaremos esta importante propiedad en la siguiente sección.

Ahora hagamos una observación obvia, pero importante: *las medias muestrales varían*. Vea la tabla 6-5 y observe que las medias muestrales son diferentes. La primera media de muestra es 1.0, la segunda media de muestra es 1.5, etcétera. Esto nos conduce a la siguiente definición.

Definición

El valor de un estadístico, como la media muestral $\bar{x}$, depende de los valores particulares incluidos en la muestra y generalmente varía de una muestra a otra. Esta variabilidad de un estadístico se denomina **variabilidad de muestreo**.

En el capítulo 2 estudiamos las características importantes de un conjunto de datos: centro, variación, distribución, valores extremos y patrón temporal. Al examinar las muestras en la tabla 6-5, ya identificamos una propiedad que describe el comportamiento de las medias muestrales: la media de las medias muestrales es igual a la media de la población. Esta propiedad enfatiza la característica central; investigaremos otras características en la siguiente sección. Veremos que al incrementarse el tamaño de la muestra, la distribución muestral de medias muestrales tiende a convertirse en una *distribución normal*. En consecuencia, la distribución normal asume una importancia que va más allá de las aplicaciones ilustradas en la sección 6-3. La distribución normal se utilizará en muchos casos en los que deseamos emplear una media muestral $\bar{x}$ con el propósito de hacer alguna inferencia acerca de una media poblacional μ.

¿Cuáles estadísticos son buenos estimadores de parámetros?

En el capítulo 7 estudiaremos métodos formales para el uso de estadísticos muestrales con el fin de hacer estimaciones de los valores de parámetros de población. Algunos estadísticos funcionan mucho mejor que otros, y podemos juzgar su valor examinando sus distribuciones muestrales, como en el siguiente ejemplo.

EJEMPLO Distribuciones muestrales Una *población* consiste en los valores 1, 2 y 5. Si seleccionamos al azar muestras de tamaño 2 con reemplazo, existen nueve distintas muestras posibles, que se listan en la tabla 6-7. Como las nueve muestras distintas son igualmente posibles, cada muestra tiene una probabilidad de 1/9.

a. Para cada muestra calcule la media, la mediana, el rango, la varianza, la desviación estándar y la proporción de valores muestrales impares. (Para cada estadístico esto generará nueve valores que, cuando se asocien con nueve probabilidades de 1/9 cada una, se combinarán para formar una *distribución muestral* del estadístico).

b. Para cada estadístico, calcule la media de los resultados del inciso *a*).

Tabla 6-7 Distribuciones muestrales de estadísticos (para muestras de tamaño 2, obtenidas con reemplazo de la población 1, 2, 5)

Muestra	Media $\bar{x}$	Mediana	Rango	Varianza s^2	Desviación estándar s	Proporción de números impares	Probabilidad
1, 1	1.0	1.0	0	0.0	0.000	1	1/9
1, 2	1.5	1.5	1	0.5	0.707	0.5	1/9
1, 5	3.0	3.0	4	8.0	2.828	1	1/9
2, 1	1.5	1.5	1	0.5	0.707	0.5	1/9
2, 2	2.0	2.0	0	0.0	0.000	0	1/9
2, 5	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 1	3.0	3.0	4	8.0	2.828	1	1/9
5, 2	3.5	3.5	3	4.5	2.121	0.5	1/9
5, 5	5.0	5.0	0	0.0	0.000	1	1/9
Media de valores de los estadísticos	8/3	8/3	16/9	26/9	1.3	2/3	
Parámetro poblacional	8/3	2	4	26/9	1.7	2/3	
¿Coincide el estadístico muestral con el parámetro poblacional?	Sí	No	No	Sí	No	Sí	

c. Compare las medias del inciso *b*) con los parámetros de población correspondientes, después determine si cada estadístico coincide con el valor del parámetro poblacional. Por ejemplo, las medias muestrales tienden a centrarse alrededor del valor de la media poblacional, que es 8/3, de manera que las medias muestrales coinciden con el valor de la media poblacional.

SOLUCIÓN

a. Vea la tabla 6-7, que incluye los estadísticos individuales para cada muestra.

b. Las medias de los estadísticos muestrales aparecen casi al final de la tabla 6-7. La media de las medias muestrales es 8/3, la media de las medianas muestrales es 8/3, y así sucesivamente.

c. El renglón inferior de la tabla 6-7 está basado en una comparación de los parámetros poblacionales y los resultados de los estadísticos muestrales. Por ejemplo, la media poblacional de 1, 2, 5 es $\mu = 8/3$, y las medias muestrales "coinciden" con el valor de 8/3, ya que la media de las medias muestrales también es 8/3.

INTERPRETACIÓN Con base en los resultados de la tabla 6-7, podemos observar que cuando se utiliza un estadístico muestral para estimar un parámetro de población, algunos estadísticos son buenos en el sentido de que coinciden con el parámetro poblacional y, por lo tanto, tienden a producir buenos resultados. Estadísticos como éstos se denominan *estimadores sin sesgo*. Otros estadísticos no son tan buenos (ya que son *estimadores sesgados*). He aquí un resumen.

- **Estadísticos que coinciden con los parámetros poblacionales:** media, varianza, proporción
- **Estadísticos que no coinciden con los parámetros poblacionales:** mediana, rango, desviación estándar

Aun cuando la desviación estándar muestral no coincide con la desviación estándar poblacional, el sesgo es relativamente pequeño en muestras grandes, de manera que con frecuencia se utiliza *s* para estimar σ. En consecuencia, las medias, proporciones, varianzas y desviaciones estándar serán consideradas temas importantes en los siguientes capítulos, en tanto que la mediana y el rango se utilizarán en pocas ocasiones.

¿Por qué se hace el muestreo *con* reemplazo? Para muestras pequeñas como las que hemos considerado hasta ahora en esta sección, el muestreo *sin reemplazo* tiene la ventaja práctica de evitar una duplicación inútil, siempre que se selecciona el mismo elemento más de una vez. Sin embargo, estamos particularmente interesados en el muestreo *con reemplazo* por las siguientes razones: **1.** Cuando se selecciona una muestra relativamente pequeña de una población grande, no hay gran diferencia si realizamos el muestreo con reemplazo o sin él. **2.** El muestreo con reemplazo da como resultado sucesos independientes que no se ven afectados por resultados previos, y los sucesos independientes son más fáciles de analizar y derivan en fórmulas más simples. Por lo tanto, nos enfocamos en el comportamiento de muestras seleccionadas aleatoriamente *con reemplazo*. Muchos de los procedimientos estadísticos que se analizan en los siguientes capítulos se basan en el supuesto de que el muestreo se hizo con reemplazo.

El aspecto más importante de esta sección es analizar el concepto de distribución muestral de un estadístico. Considere el objetivo de tratar de calcular la temperatura corporal media de todos los adultos. Puesto que la población es demasiado grande, no

b. Calcule la media de la distribución muestral.
c. ¿Es la media de la distribución muestral [del inciso *b*)] igual a la proporción poblacional de defectos? ¿Coincide *siempre* la media de la distribución muestral de proporciones con la proporción poblacional?

13. Clasificación de competidores olímpicos de triatlón. Mujeres estadounidenses compitieron en el triatlón de los Juegos Olímpicos realizados en Atenas, y al final ocuparon los lugares 3, 9 y 23. Suponga que se seleccionan al azar muestras de tamaño 2 con reemplazo.
a. Utilice una tabla para describir la distribución muestral de las medias muestrales.
b. Puesto que los datos se refieren a lugares de clasificación, ¿realmente tiene sentido identificar la distribución muestral de las medias muestrales?

14. Mediana y lunas de Júpiter. Júpiter tiene 4 lunas grandes y 12 lunas pequeñas. Los tiempos de órbita de las 4 lunas grandes son los siguientes (en días): 1.8 (Io), 3.6 (Europa), 7.2 (Ganímedes) y 16.7 (Calixto). Suponga que se seleccionan al azar dos de estos valores con reemplazo.
a. Después de identificar las 16 diferentes muestras posibles, calcule la mediana de cada una y luego utilice una tabla para describir la distribución muestral de las medianas.
b. Calcule la media de la distribución muestral.
c. ¿Es la media de la distribución muestral [del inciso *b*)] igual a la mediana de la población? ¿Es la mediana un estimador sin sesgo de la mediana poblacional?

6-4 MÁS ALLÁ DE LO BÁSICO

15. Uso de una fórmula para describir una distribución muestral. Vea el primer ejemplo de esta sección, el cual incluye una tabla y una gráfica que describen la distribución muestral de las proporciones de niñas en dos nacimientos. Considere la fórmula que aparece abajo y evalúela utilizando proporciones muestrales x de 0, 0.5 y 1. Con base en los resultados, ¿la fórmula describe la distribución muestral? ¿Por qué?

$$P(x) = \frac{1}{2(2 - 2x)!(2x)!} \quad \text{donde } x = 0, 0.5, 1$$

16. Desviación media absoluta. La población de 1, 2 y 5 se utilizó para elaborar la tabla 6-7. Identifique la distribución muestral de la desviación media absoluta (definida en la sección 3-3), después determine si la desviación media absoluta de una muestra es un buen estadístico para estimar la desviación media absoluta de la población.

6-5 El teorema del límite central

Concepto clave En la sección 6-4 se incluyó cierto análisis de la distribución muestral de $\bar{x}$, y en esta sección se describen procedimientos para utilizarla en situaciones muy reales y prácticas. Los procedimientos de esta sección constituyen la base para la estimación de parámetros poblacionales y la prueba de hipótesis, temas que se analizarán con profundidad en los siguientes capítulos. Cuando se selecciona una muestra aleatoria simple de una población con media μ y desviación estándar σ, es esencial conocer los siguientes principios:

1. Si $n > 30$, entonces las medias muestrales tienen una distribución que se puede aproximar por medio de una distribución normal, con una media μ y una desviación estándar $\sigma/\sqrt{n}$. (Éste es el lineamiento que suele utilizarse, independientemente de la distribución de la población original).

2. Si $n \leq 30$ y la población original tiene una distribución normal, entonces las medias muestrales tienen una distribución normal con una media μ y una desviación estándar $\sigma/\sqrt{n}$.
3. Si $n \leq 30$, pero la población original no tiene una distribución normal, entonces no se aplican los métodos de esta sección.

Trate de conservar la siguiente idea en mente: cuando tomamos muestras de una población, deseamos conocer el comportamiento de las medias muestrales. El *teorema del límite central* nos dice que si el tamaño de una muestra es lo suficientemente grande, la distribución de las medias muestrales se puede aproximar por medio de una *distribución normal*, aun cuando la población original no esté distribuida de forma normal. Aunque hablamos de un "teorema", no incluimos pruebas rigurosas, sino que nos enfocamos en los *conceptos* y en su aplicación. He aquí los puntos clave que conforman una base importante para los siguientes capítulos.

El teorema del límite central y la distribución muestral de $\bar{x}$

Dado que:

1. La variable aleatoria x tiene una distribución (que puede o no ser normal) con media μ y desviación estándar σ.
2. Todas las muestras aleatorias del mismo tamaño n se seleccionan de la población. (Las muestras se seleccionan de manera que todas las muestras posibles de tamaño n tengan la misma probabilidad de ser seleccionadas).

Conclusiones:

1. Conforme el tamaño de la muestra aumenta, la distribución de las medias muestrales $\bar{x}$ se aproximará a una distribución *normal*.
2. La media de todas las medias muestrales es la media poblacional μ. (Es decir, la distribución normal de la conclusión 1 tiene una media μ).
3. La desviación estándar de todas las medias muestrales es $\sigma/\sqrt{n}$. (es decir, la distribución normal de la conclusión 1 tiene una desviación estándar $\sigma/\sqrt{n}$.)

Reglas prácticas de uso común

1. Si la población original no está distribuida normalmente, la siguiente es una directriz común: para muestras de tamaño n mayores que 30, la distribución de las medias muestrales puede aproximarse razonablemente bien por medio de una distribución normal. (Existen excepciones, como las poblaciones con distribuciones muy diferentes a la normal, que requieren tamaños de muestra mucho más grandes que 30, aunque tales excepciones son relativamente raras). La aproximación mejora conforme el tamaño muestral n se incrementa.
2. Si la población original se distribuye normalmente, entonces las medias muestrales estarán distribuidas normalmente para *cualquier* tamaño de muestra n (no sólo los valores de n mayores que 30).

El teorema del límite central implica dos distribuciones diferentes: la distribución de la población original y la distribución de las medias muestrales. Igual que en capítulos anteriores, utilizamos los símbolos μ y σ para denotar la media y la desviación estándar de la población original, pero ahora necesitamos nuevas notaciones para la media y la desviación estándar de la distribución de las medias muestrales.

Notación para la distribución muestral de $\overline{x}$

Si se seleccionan todas las muestras aleatorias posibles de tamaño n de una población con media μ y desviación estándar σ, la media de las medias muestrales se denota con $\mu_{\overline{x}}$, de manera que

$$\mu_{\overline{x}} = \mu$$

Asimismo, la desviación estándar de las medias muestrales se denota con $\sigma_{\overline{x}}$, de manera que

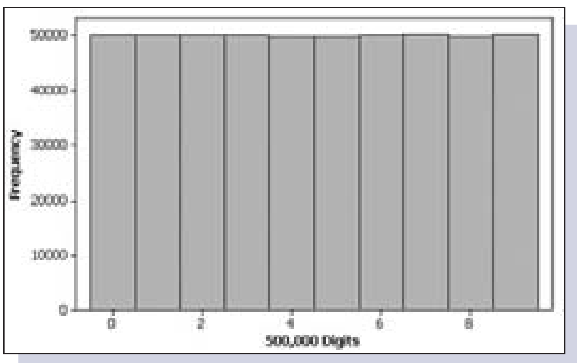
$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

$\sigma_{\overline{x}}$, suele denominarse el **error estándar de la media.**

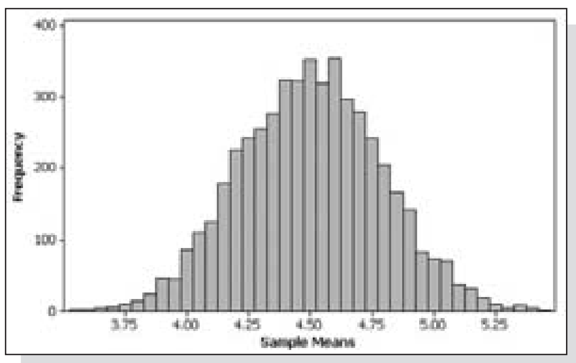
EJEMPLO **Simulación con dígitos aleatorios** Con frecuencia las computadoras se utilizan para generar dígitos aleatorios de números telefónicos para realizar encuestas. (Por ejemplo, el Pew Research Center genera aleatoriamente los últimos dos dígitos de números telefónicos para evitar un "sesgo de lista"). Los dígitos 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 se generan de forma que todos son igualmente probables. La imagen de Minitab que aparece a continuación incluye el histograma de 500,000 dígitos generados. Observe que la distribución tiene una apariencia uniforme, como esperábamos.

Ahora, agrupamos los 500,000 dígitos en 5000 muestras, cada una con $n = 100$ valores. Calculamos la media de cada muestra y mostramos el histograma de las 5000 medias muestrales. Observe el siguiente efecto sorprendente: **¡Aun cuando los 500,000 dígitos originales tienen una distribución *uniforme*, la distribución de las 500 medias muestrales es aproximadamente *normal*!** Un fenómeno verdaderamente fascinante e intrigante de la estadística es el hecho de que al obtener muestras de cualquier distribución podamos crear una distribución de medias muestrales que es normal o al menos aproximadamente normal.

Minitab



Minitab



Aplicación del teorema del límite central

Muchos problemas prácticos importantes se resuelven mediante el teorema del límite central. Cuando trabaje con este tipo de problemas, recuerde que si el tamaño de la muestra es mayor que 30, o si la población original se distribuye normalmente, debe tratar la distribución de medias muestrales como si fuera una distribución normal con media μ y desviación estándar $\sigma/\sqrt{n}$.

En el siguiente ejemplo, el inciso *a*) incluye un valor *individual*, pero el inciso *b*) incluye la media de una *muestra* de 20 hombres, por lo que debemos usar el teorema del límite central al trabajar con la variable aleatoria $\bar{x}$. Estudie este ejemplo con atención para comprender la diferencia significativa entre los procedimientos utilizados en los incisos *a*) y *b*). Observe cómo este ejemplo ilustra el siguiente procedimiento de trabajo:

- **Cuando trabaje con un valor *individual* de una población distribuida normalmente, utilice los métodos de la sección 6-3. Use $z = \dfrac{x - \mu}{\sigma}$.**
- **Cuando trabaje con una media de alguna *muestra* (o grupo), asegúrese de utilizar el valor de $\sigma/\sqrt{n}$ para la desviación estándar de las medias muestrales. Use $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$.**

PROBLEMA 6 DEL CAPÍTULO

EJEMPLO Seguridad de taxis acuáticos En el problema del capítulo señalamos que algunos pasajeros murieron cuando un taxi acuático se hundió en el Inner Harbor de Baltimore. Los hombres suelen ser más pesados que las mujeres y los niños; por lo tanto, supongamos que al cargar un taxi acuático la situación extrema es aquella en la que todos los pasajeros son hombres. En concordancia con los datos de la National Health and Nutrition Examination Survey, suponga que los pesos de los hombres se distribuyen normalmente, con una media de 172 libras y una desviación estándar de 29 libras.

a. Calcule la probabilidad de que, si se selecciona un hombre al azar, su peso sea mayor que 175 libras.

b. Calcule la probabilidad de que 20 hombres elegidos al azar tengan una media mayor que 175 libras (de manera que su peso total exceda la capacidad segura de 3500 libras).

SOLUCIÓN

a. *Enfoque: Utilice los métodos presentados en la sección 6-3* (porque estamos trabajando con un valor individual de una población distribuida normalmente). Buscamos el área de la región sombreada en la figura 6-19a). Si utilizamos la tabla A-2, convertimos el peso de 175 a su puntuación z correspondiente:

$$z = \frac{x - \mu}{\sigma} = \frac{175 - 172}{29} = 0.10$$

Nos remitimos a la tabla A-2 y utilizamos $z = 0.10$ para encontrar que el área acumulativa a la izquierda de 175 libras es 0.5398. Por lo tanto, la región sombreada es $1 - 0.5398 = 0.4602$. La probabilidad de que un hombre elegido al azar pese más de 175 libras es de 0.4602. (Si se utiliza una calculadora o un programa de cómputo en vez de la tabla A-2, el resultado más exacto es 0.4588 en lugar de 0.4602).

continúa

b. *Enfoque: Utilice el teorema del límite central* (porque estamos trabajando con la *media de una muestra* de 20 hombres y no con un solo hombre). Aun cuando el tamaño de la muestra no es mayor que 30, utilizamos una distribución normal por la siguiente razón: la población original de hombres tiene una distribución normal, de manera que las muestras de *cualquier* tamaño producirán medias distribuidas normalmente. Puesto que estamos trabajando con una distribución de medias muestrales, debemos utilizar los parámetros $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$, que se evalúan de la siguiente manera:

$$\mu_{\bar{x}} = \mu = 172$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{20}} = 6.4845971$$

He aquí un punto realmente importante: debemos utilizar la desviación estándar calculada de 6.4845971 y no la desviación estándar original de 29 (porque estamos trabajando con la distribución de medias muestrales cuya desviación estándar es 6.4845971, y no con la distribución de pesos individuales cuya desviación estándar es 29). Necesitamos calcular el área sombreada que se destaca en la figura 6-1b). Si usamos la tabla A-2, obtenemos la puntuación z relevante, que se calcula de la siguiente manera:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{175 - 172}{\frac{29}{\sqrt{20}}} = \frac{3}{6.4845971} = 0.46$$

Si nos remitimos a la tabla A-2, encontramos que $z = 0.46$ corresponde a una área izquierda acumulativa de 0.6772, de manera que la región sombreada es $1 - 0.6772 = 0.3228$. La probabilidad de que 20 hombres tengan un peso medio mayor que 175 libras es de 0.3228. (Si se utiliza una calculadora o un programa de cómputo, el resultado es 0.3218 en vez de 0.3228).

INTERPRETACIÓN Existe una probabilidad de 0.4602 de que un hombre pese más de 175 libras, y una probabilidad de 0.3228 de que 20 hombres tengan un peso medio mayor de 175 libras. Dado que la capacidad segura del taxi

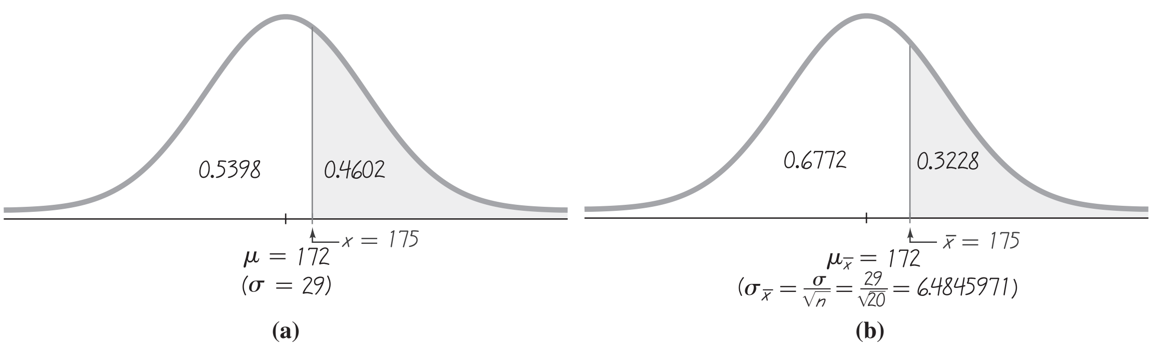


Figura 6-19 **Pesos de hombres**

a) Distribución de pesos individuales de hombres; *b*) Distribución de medias muestrales acuático es de 3500 libras, es muy probable (con una probabilidad de 0.3228) que se sobrecargue si se transporta a 20 hombres elegidos al azar. Dado que ya han muerto 21 personas, y dada la alta probabilidad de sobrecarga, lo más pertinente sería limitar el número de pasajeros a menos de 20. La capacidad de 20 pasajeros no es suficiente.

Los cálculos que se utilizaron aquí son exactamente iguales a los cálculos que utilizan los ingenieros al diseñar teleféricos, elevadores, escaleras eléctricas, aviones y otros aparatos que transportan personas.

El *efecto placebo*

Durante mucho tiempo se ha creído que los placebos realmente ayudan a algunos pacientes. De hecho, algunos estudios serios han demostrado que cuando se administra un placebo (un tratamiento sin valor medicinal), muchos sujetos de prueba manifiestan cierta mejoría. Los estimados de las tasas de mejoría van por lo común de una tercera a dos terceras partes de los pacientes. Sin embargo, un estudio más reciente sugiere que los placebos no tienen efecto real. Un artículo en la *New England Journal of Medicine* (vol. 334, núm. 21) se basó en la investigación de 114 ensayos médicos durante 50 años. Los autores del artículo concluyeron que los placebos parecen tener algún efecto sólo en aliviar el dolor, pero no en otras condiciones físicas. Ellos concluyen que, excepto en ensayos clínicos, el uso de placebos “no puede recomendarse”.

Interpretación de resultados

El siguiente ejemplo ilustra otra aplicación del teorema del límite central; examine con atención la conclusión a la que se llega. Este ejemplo ilustra el tipo de pensamiento que es fundamental para el importante procedimiento de prueba de hipótesis (que se estudia en el capítulo 8). Este ejemplo ilustra la regla del suceso infrecuente de la estadística inferencial, presentado inicialmente en la sección 4-1.

Regla del suceso infrecuente

Si, bajo cierto supuesto, la probabilidad de un suceso particular observado es pequeña de manera excepcional, concluimos que el supuesto probablemente no es correcto.

EJEMPLO Temperaturas corporales Suponga que la población de temperaturas corporales humanas tiene una media de 98.6°F, como suele creerse. También suponga que la desviación estándar de la población es 0.62°F (según datos de investigadores de la Universidad de Maryland). Si se selecciona al azar una muestra de tamaño $n = 106$, calcule la probabilidad de obtener una media de 98.2°F o menor. (En realidad se obtuvo el valor de 98.2°F; vea las temperaturas de media noche del día 2 en el conjunto de datos 2 del apéndice B).

SOLUCIÓN

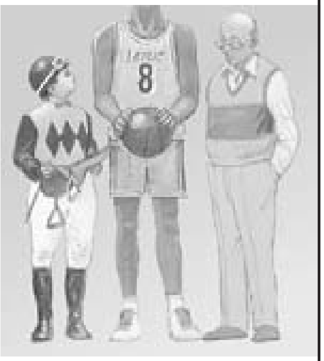
No tenemos información sobre la distribución de la población, pero, como el tamaño de la muestra $n = 106$ excede a 30, utilizamos el teorema del límite central y concluimos que la distribución de medias muestrales es normal, con estos parámetros:

$$\mu_{\bar{x}} = \mu = 98.6 \qquad \text{(por suposición)}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.62}{\sqrt{106}} = 0.0602197$$

La figura 6-20 destaca el área sombreada (observe la pequeña cola izquierda de la gráfica), correspondiente a la probabilidad que buscamos. Una vez encontrados los parámetros que se aplican a la distribución de la figura 6-20, ahora

continúa



El teorema del límite central borroso

En *The Cartoon Guide to Statistics*, los autores Gonick y Smith describen el teorema del límite central borroso de la siguiente manera: "Los datos que se ven influidos por efectos aleatorios muy pequeños y sin relación entre sí se distribuyen aproximadamente de manera normal. Esto explica por qué la normalidad está en todos lados: en las fluctuaciones del mercado de acciones, en los pesos de estudiantes, en los promedios anuales de temperatura y en las calificaciones del SAT. Todos son el resultado de muchos efectos diferentes". La estatura de las personas, por ejemplo, es el resultado de factores hereditarios, factores ambientales, nutrición, cuidado de la salud, región geográfica y otras influencias que, cuando se combinan, producen valores distribuidos de forma normal.

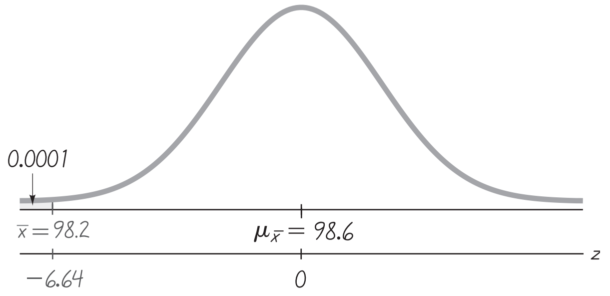


Figura 6-20

Distribución de temperaturas corporales medias, para muestras de tamaño $n = 106$

podemos encontrar el área sombreada utilizando los mismos procedimientos desarrollados en la sección 6-3. En la tabla A-2 primero encontramos la puntuación z:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{98.20 - 98.6}{0.0602197} = -6.64$$

Si nos remitimos a la tabla A-2, encontramos que $z = -6.64$ no aparece, pero para los valores de z que están por debajo de -3.49 utilizamos una área de 0.0001 para el área izquierda acumulativa hasta $z = -3.49$. Por lo tanto, concluimos que la región sombreada de la figura 6-20 es 0.0001. (Si se utiliza una calculadora TI-83/84 Plus o un programa de cómputo, el área de la región sombreada se acerca a 0.00000000002, pero incluso estos resultados son sólo aproximaciones. Con seguridad podemos reportar que la probabilidad es muy baja, menor que 0.0001).

INTERPRETACIÓN Los resultados demuestran que si la media de nuestra temperatura corporal es en realidad 98.6°F, entonces existe una probabilidad sumamente baja de obtener una media de muestra de 98.2°F o menor cuando se seleccionan 106 sujetos al azar. Los investigadores de la Universidad de Maryland obtuvieron una media muestral como ésta, ante lo cual existen dos explicaciones posibles: o la media de la población es realmente de 98.6°F y su muestra representa un suceso aleatorio extremadamente infrecuente, o en realidad la media poblacional es menor que 98.6°F y su muestra es típica. Como la probabilidad es tan baja, parece más razonable concluir que la media poblacional es menor que 98.6°F. Éste es el tipo de razonamiento que se usa en las *pruebas de hipótesis*, que se estudiarán en el capítulo 8. Por ahora, debemos enfocarnos en el uso del teorema del límite central para calcular la probabilidad de 0.0001, pero debemos señalar que este teorema se usará posteriormente para explicar algunos conceptos muy importantes en estadística.

Corrección para una población finita

Al aplicar el teorema del límite central, el uso de $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ supone que la población tiene un número infinito de miembros. Cuando hacemos un muestreo con reemplazo (es decir, cada elemento seleccionado se reincorpora a la muestra antes de hacer la siguiente selección), la población es efectivamente infinita. Aunque muchas aplicaciones realistas implican un muestreo sin reemplazo, estas muestras sucesivas dependen de resultados previos. En la fabricación, los inspectores de

6-5 MÁS ALLÁ DE LO BÁSICO

21. Diseño de asientos. Usted necesita construir una banca en la que se sentarán 18 jugadores universitarios de fútbol americano y debe determinar primero la longitud de la banca. Los hombres tienen anchuras de cadera que se distribuyen normalmente, con una media de 14.4 in y una desviación estándar de 1.0 in.

a. ¿Cuál será la longitud mínima de la banca si usted busca una probabilidad de 0.975 de que se ajuste a las anchuras de cadera combinadas de 18 hombres seleccionados al azar?

b. ¿Por qué sería incorrecto utilizar realmente el resultado del inciso *a*) como longitud de la banca?

22. Corrección para una población finita. El club Boston Women necesita un elevador limitado a 8 pasajeros. El club tiene 120 miembros mujeres con pesos que se aproximan a una distribución normal, con una media de 143 lb y una desviación estándar de 29 lb. (*Sugerencia:* Véase la explicación del factor de corrección para una población finita).

a. Si se seleccionan al azar 8 miembros diferentes, calcule la probabilidad de que su peso total no exceda la capacidad máxima de 1300 lb.

b. Si buscamos una probabilidad de 0.99 de que el elevador no se sobrecargue siempre que se seleccione al azar a 8 miembros como pasajeros, ¿cuál debe ser el peso máximo permitido?

23. Parámetros de población Una *población* consiste en los valores: 2, 3, 6, 8, 11, 18.

a. Calcule μ y σ.

b. Liste todas las muestras de tamaño $n = 2$ que pueden obtenerse con reemplazo.

c. Calcule la población de todos los valores de $\bar{x}$ al obtener la media de cada muestra del inciso *b*).

d. Calcule la media $\mu_{\bar{x}}$ y la desviación estándar $\sigma_{\bar{x}}$ para la población de medias muestrales obtenidas en el inciso *c*).

e. Verifique que

$$\mu_{\bar{x}} = \mu \quad \text{y} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

6-6 La distribución normal como aproximación de la distribución binomial

Concepto clave En esta sección se presenta un método para utilizar una distribución normal como aproximación de una distribución de probabilidad binomial. Si se satisfacen las condiciones $nq \geq 5$ entonces las probabilidades de una distribución de probabilidad binomial se pueden aproximar bastante bien utilizando una distribución normal con media $\mu = np$ y desviación estándar $\sigma = \sqrt{npq}$. Puesto que una distribución de probabilidad binomial generalmente usa sólo números enteros para la variable aleatoria x, mientras que la aproximación normal es continua, debemos utilizar una "corrección por continuidad", con un número entero x representado por el intervalo de $x - 0.5$ a $x + 0.5$. *Nota importante:* En vez de utilizar una distribución normal como aproximación de una distribución de probabilidad binomial, la mayoría de las aplicaciones prácticas de la distribución binomial se pueden manejar con un programa de cómputo o una calculadora, pero esta sección expone el importante principio de que una distribución binomial se puede aproximar por medio de una distribución normal, y este principio se utilizará en capítulos posteriores.

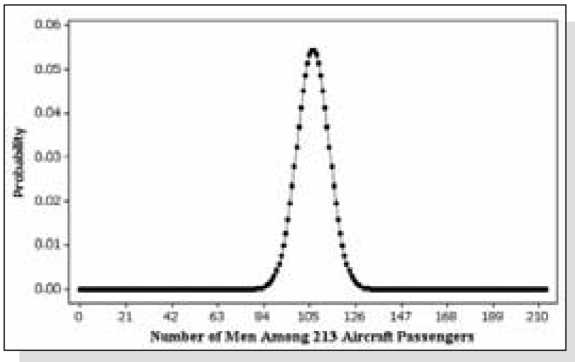
Considere la carga de un Boeing 767-300 de American Airlines que lleva 213 pasajeros. Al analizar la carga que puede transportar con seguridad, debemos con-

Voltaire vence a la lotería

En 1729, el filósofo Voltaire se hizo rico al diseñar un esquema para vencer a la lotería de París. El gobierno organizó una lotería para reembolsar bonos municipales que habían perdido cierto valor. La ciudad aportó grandes cantidades de dinero, con el efecto neto de que el valor total de los premios fuera mayor que el costo de todos los boletos. Voltaire organizó un grupo y compró todos los boletos de la lotería mensual y ganó durante más de un año. Por otro lado, un participante de la lotería del estado de Nueva York trató de ganar una parte de un premio excepcionalmente grande, que había crecido gracias a la falta de ganadores previos. Él quería extender un cheque por $6,135,756, que cubriera todas las combinaciones, pero el estado se rehusó con el argumento de que esto cambiaría la naturaleza de la lotería.

siderar el peso de los pasajeros. Sabemos que un hombre típico pesa alrededor de 30 libras más que una mujer típica, de manera que el número de pasajeros varones es un tema importante. Podemos utilizar la distribución de probabilidad binomial con $n = 213$, $p = 0.5$ y $q = 0.5$ (suponiendo que los hombres y las mujeres son igualmente probables). Observe que la pantalla de Minitab que aparece abajo incluye una gráfica de la probabilidad para cada número de pasajeros varones, desde 0 hasta 213, y observe que la gráfica tiene la apariencia de una distribución normal aun cuando los puntos graficados provienen de una distribución binomial. Esta gráfica sugiere que podemos utilizar una distribución normal para aproximar la distribución binomial.

Minitab



La distribución normal como aproximación de la distribución binomial

Si una distribución de probabilidad binomial satisface los requisitos $np \geq 5$ y $nq \geq 5$, entonces la distribución de probabilidad binomial puede aproximarse con una distribución normal con media $\mu = np$ y desviación estándar $\sigma = \sqrt{npq}$, así como con el número entero discreto x ajustado con una *corrección por continuidad*, de manera que x está representada por el intervalo de $x - 0.5$ a $x + 0.5$.

Cuando se utiliza una distribución normal como aproximación de una distribución binomial se sigue el siguiente procedimiento:

Procedimiento para el uso de una distribución normal como aproximación de una distribución binomial

1. Establezca que la distribución normal es una aproximación adecuada de la distribución binomial, verificando que $np \geq 5$ y $nq \geq 5$. (Si no se satisfacen ambas condiciones, entonces debe utilizar un programa de cómputo, una calculadora, la tabla A-1 o la fórmula de probabilidad binomial).
2. Obtenga los valores de los parámetros μ y σ calculando $\mu = np$ y $\sigma = \sqrt{npq}$.
3. Identifique el valor discreto x (el número de éxitos). Cambie el valor *discreto* x reemplazándolo con el *intervalo* de $x - 0.5$ a $x + 0.5$ (para mayor explicación,

véase el apartado "Correcciones por continuidad", más adelante en esta sección). Dibuje una curva normal e indique los valores de μ, σ y $x - 0.5$ o $x + 0.5$, según sea apropiado.

4. Modifique x reemplazándola por $x - 0.5$ o $x + 5$, según sea apropiado.
5. Utilice $x - 0.5$ o $x + 5$ (según sea apropiado) en vez de x, calcule el área correspondiente a la probabilidad deseada encontrando primero la puntuación z: $z = (x - \mu)/\sigma$. Ahora use esa puntuación z para encontrar el área a la izquierda del valor ajustado de x. Ahora esa área puede emplearse para identificar el área correspondiente a la probabilidad deseada.

Ilustraremos este procedimiento de aproximación normal con el siguiente ejemplo.

EJEMPLO **Carga de pasajeros en un Boeing 767-300** Un avión Boeing 767-300 de American Airlines tiene 213 asientos. Cuando se llena con pasajeros, equipaje, carga y combustible, el piloto debe verificar que el peso neto no rebase el límite máximo permitido, y el peso debe distribuirse adecuadamente para que el equilibrio de la aeronave permanezca dentro de los límites de seguridad aceptables. En vez de pesar a cada pasajero, sus pesos se estiman según las reglas de la Federal Aviation Administration. En realidad, sabemos que los hombres tienen un peso medio de 172 lb y que las mujeres tienen un peso medio de 143 lb, de manera que un número desproporcionadamente mayor de hombres podría provocar una situación insegura de sobrepeso. Suponga que, si hay por lo menos 122 hombres en una lista de 213 pasajeros, la carga debe ajustarse de alguna manera. Suponiendo que los pasajeros se registran al azar, que los hombres y las mujeres son igualmente probables, y que la aeronave está llena de adultos, calcule la probabilidad de que en un Boeing 767-300 con 213 pasajeros haya al menos 122 hombres.

SOLUCIÓN El problema implica una distribución binomial con un número fijo de ensayos ($n = 213$) que se supone son independientes, dos categorías (hombre, mujer) de resultados para cada ensayo, y la probabilidad de un hombre ($p = 0.5$) que se supone permanece constante de un ensayo a otro. Los cálculos con la fórmula de probabilidad binomial no son prácticos porque tendríamos que aplicarla 92 veces (una para cada valor de x desde 122 hasta 213, inclusive). En vez de ello, utilizamos el método de los cinco pasos para aproximar la distribución binomial con la distribución normal.

Paso 1: Verificación requerida: Primero debemos verificar que es razonable aproximar la distribución binomial con la distribución normal, porque $np \geq 5$ y $nq \geq 5$. Con $n = 213$, $p = 0.5$ y $q = 1 - p = 0.5$, verificamos las condiciones requeridas como sigue:

$$np = 213 \cdot 0.5 = 106.5 \quad \text{(Por lo tanto } np \geq 5.)$$
$$nq = 213 \cdot 0.5 = 106.5 \quad \text{(Por lo tanto } nq \geq 5.)$$

Paso 2: Ahora procedemos a calcular los valores de μ y σ, necesarios para la distribución normal. Obtenemos lo siguiente:

$$\mu = np = 213 \cdot 0.5 = 106.5$$
$$\sigma = \sqrt{npq} = \sqrt{213 \cdot 0.5 \cdot 0.5} = 7.2972598$$

continúa

¿Es seguro el paracaidismo?

De las más de 100,000 personas que realizan cerca de 2.25 millones de saltos en paracaídas, aproximadamente 30 mueren cada año. En comparación, un año típico incluye alrededor de 200 muertes en el buceo, 7000 ahogamientos, 900 muertes en bicicletas, 800 muertes por relámpagos y 1150 muertes por picaduras de abeja. Desde luego, estas cifras no significan necesariamente que el paracaidismo sea más seguro que andar en bicicleta o que la natación. En una comparación justa deben incluirse las tasas de mortalidad, no sólo el número total de fallecimientos. El autor, con gran osadía, realizó dos saltos en paracaídas, pero desistió después de no caer dentro de la amplia zona de aterrizaje, en ambas ocasiones. Él también ha volado un ala delta, un globo aerostático y en el dirigible Goodyear.

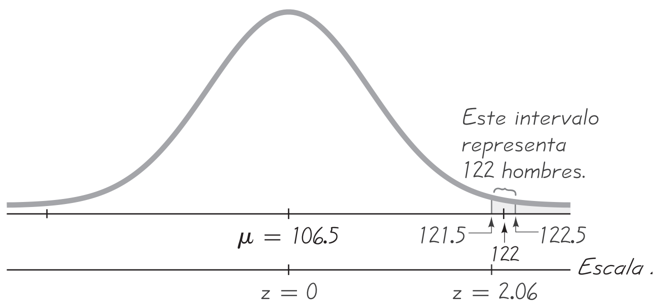


Figura 6-21 **Búsqueda de la probabilidad de "al menos" 122 hombres, entre 213 pasajeros**

Paso 3: Buscamos la probabilidad de "al menos 122 hombres" y el valor discreto de 122 se ajusta utilizando la corrección por continuidad de la siguiente manera: represente $x = 122$ por medio de la banda vertical limitada por 121.5 y 122.5.

Paso 4: Puesto que buscamos la probabilidad de *al menos* 122 hombres, queremos el área que representa el número entero discreto de 122 (la región limitada por 121.5 y 122.5), así como también el área a la derecha, como se indica en la figura 6-21.

Paso 5: Ahora podemos proceder a la búsqueda del área sombreada de la figura 6-21, utilizando los mismos métodos que se emplearon en la sección 6-3. Para usar la tabla A-2 de la distribución normal estándar, primero debemos transformar 121.5 a una puntuación z, después usar la tabla para encontrar el área a la izquierda de 121.5, que posteriormente se resta de 1. La puntuación z se obtiene como sigue:

$$z = \frac{x - \mu}{\sigma} = \frac{121.5 - 106.5}{7.2972598} = 2.06$$

Al emplear la tabla A-2, encontramos que $z = 2.06$ corresponde a una área de 0.9803, de manera que la región sombreada es $1 - 0.9803 = 0.0197$. Si se utiliza una calculadora TI-83/84 Plus o un programa de cómputo, se obtiene el resultado más exacto de 0.0199.

INTERPRETACIÓN Existe una probabilidad de 0.0197 de obtener al menos 122 hombres entre 213 pasajeros. Como esa probabilidad es demasiado baja, sabemos que una lista de 200 pasajeros pocas veces incluirá al menos 122 hombres, por lo que no será necesario ajustar la carga del avión con mucha frecuencia.

Correcciones por continuidad

El procedimiento que implica el uso de la distribución normal como aproximación de la distribución binomial incluye un ajuste en el que cambiamos un número entero discreto por un intervalo que está 0.5 por abajo y 0.5 por arriba del número discreto. Este paso en particular, denominado *corrección por continuidad*, suele ser difícil de comprender, por lo que ahora lo explicaremos con mayor detalle.

Definición

Cuando empleamos la distribución normal (que es una distribución de probabilidad *continua*) como una aproximación de la distribución binomial (que es *discreta*), se realiza una **corrección por continuidad** a un número entero discreto x en la distribución binomial, representando el valor único x en el *intervalo* de $x - 0.5$ a $x + 0.5$ (es decir, sumando y restando 0.5).

Las siguientes sugerencias prácticas le ayudarán a utilizar las correcciones por continuidad apropiadamente.

Procedimiento para correcciones por continuidad

1. Cuando use la distribución normal como aproximación de la distribución binomial, *siempre* aplique la corrección por continuidad. (Es necesario porque estamos utilizando la distribución normal *continua* para aproximar la distribución binomial *discreta*).
2. Para emplear la corrección por continuidad, primero identifique el número entero discreto x relevante al problema de probabilidad binomial. Por ejemplo, si usted está intentando calcular la probabilidad de obtener al menos 122 hombres entre 213 personas seleccionadas al azar, el número entero discreto relevante sería $x = 122$. Primero enfoque su atención en el valor x e ignore temporalmente si busca al menos x, más que x, menos que x o alguna otra condición.
3. Dibuje una distribución normal centrada alrededor de μ, después dibuje una *franja vertical* alrededor de x. Marque el lado izquierdo de la franja con el número igual a $x + 0.5$, y marque el lado derecho con el número igual a $x + 0.5$. Por ejemplo, para $x = 122$, dibuje una franja desde 121.5 hasta 122.5. *Considere el área completa de la franja para representar la probabilidad del número entero discreto x.*
4. Ahora determine si el valor de x debe incluirse en la probabilidad que busca. (Por ejemplo, "al menos x" incluye a x, pero "más que x" no la incluye). Después, determine si busca la probabilidad de al menos x, a lo sumo x, más que x, menos que x o exactamente x. Sombree el área a la derecha o izquierda de la franja, según sea apropiado; también sombree el interior de la franja *si y sólo si* x debe incluirse. Esta región total sombreada corresponde a la probabilidad buscada.

Para ver cómo resulta este procedimiento en las correcciones por continuidad, observe los casos comunes ilustrados en la figura 6-22. Esos casos corresponden a las aseveraciones de la siguiente lista.

Aseveración	Área
Al menos 122 (incluye 122 y números mayores)	A la *derecha* de 121.5
Más que 122 (no incluye 122)	A la *derecha* de 122.5
A lo sumo 122 (incluye 122 y números menores)	A la *izquierda* de 122.5
Menos que 122 (no incluye 122)	A la *izquierda* de 121.5
Exactamente 122	Entre 121.5 y 122.5

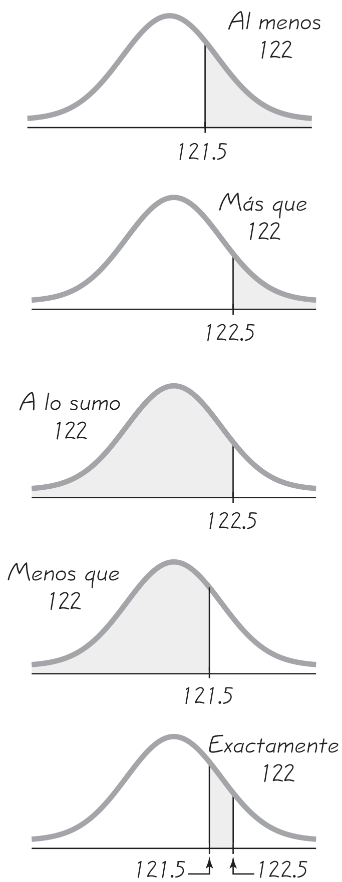


Figura 6-22

Uso de correcciones por continuidad

EJEMPLO Uso de Internet Una encuesta reciente reveló que, de 2013 adultos elegidos al azar, 1358 (o el 67.5%) afirmaron ser usuarios de Internet (según datos del Pew Research Center). Si la proporción de todos los adultos que utilizan Internet es en realidad de 2/3, calcule la probabilidad de que una muestra aleatoria de 2013 adultos produzca exactamente 1358 usuarios de Internet.

SOLUCIÓN Tenemos $n = 2013$ sujetos de encuesta independientes y $x = 1358$ de ellos son usuarios de Internet; suponemos que la proporción de la población es $p = 2/3$, de lo que se deduce que $q = 1/3$. Utilizaremos una distribución normal para aproximar la distribución binomial.

Paso 1: Primero verificamos los requisitos para determinar si es posible la aproximación normal:

$$np = 2013 \cdot 2/3 = 1342 \quad \text{(Por lo tanto } np \geq 5.)$$
$$nq = 2013 \cdot 1/3 = 671 \quad \text{(Por lo tanto } nq \geq 5.)$$

Paso 2: Ahora procedemos a calcular los valores de μ y σ, necesarios para la distribución normal. Obtenemos lo siguiente:

$$\mu = np = 2013 \cdot 2/3 = 1342$$
$$\sigma = \sqrt{npq} = \sqrt{2013 \cdot (2/3) \cdot (1/3)} = 21.150256$$

Paso 3: Dibujamos la curva normal de la figura 6-23. La región sombreada de la figura representa la probabilidad de obtener exactamente 1358 usuarios de Internet. Si utilizamos la corrección por continuidad, representamos $x = 1358$ por medio de la región ubicada entre 1357.5 y 1358.5.

Paso 4: He aquí el método empleado para calcular la región sombreada de la figura 6-23. Primero calcule el área total a la izquierda de 1358.5, después obtenga el área total a la izquierda de 1357.5, luego calcule la *diferencia* entre ambas áreas. Comencemos con el área total a la izquierda de 1358.5. Si deseamos utilizar la tabla A-2, primero debemos obtener la puntuación z que corresponde a 1358.5. Obtenemos

$$z = \frac{1358.5 - 1342}{21.150256} = 0.78$$

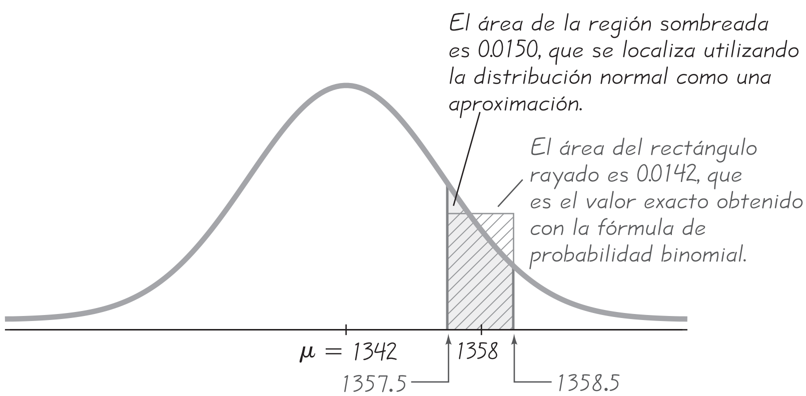


Figura 6-23 **Uso de la corrección por continuidad**

Usamos la tabla A-2 para encontrar que $z = 0.78$ corresponde a una probabilidad de 0.7823, que es el área total a la izquierda de 1358.5. Ahora, procedemos a obtener el área a la izquierda de 1357.5, calculando primero la puntuación z correspondiente a 1357.5:

$$z = \frac{1357.5 - 1342}{21.150256} = 0.73$$

Si utilizamos la tabla A-2, encontramos que $z = 0.73$ corresponde a una probabilidad de 0.7673, la cual es el área total a la izquierda de 1357.5. El área sombreada es $0.7823 - 0.7673 = 0.0150$. (Con un recurso tecnológico obtenemos 0.0142).

INTERPRETACIÓN Si suponemos que 2/3 de todos los adultos utilizan Internet, la probabilidad de obtener exactamente 1358 usuarios de Internet entre 2013 personas elegidas al azar es de 0.0150. Esta probabilidad nos indica que si la proporción de usuarios de Internet en la población es de 2/3, entonces es sumamente improbable que obtengamos exactamente 1358 usuarios de Internet al encuestar a 2013 personas. En realidad, cuando se encuesta a 2013 individuos, la probabilidad de *cualquier* número específico de usuarios de Internet es muy pequeña.

Si resolvemos el ejemplo anterior por medio de STATDISK, Minitab o una calculadora, obtenemos un resultado de 0.0142, pero el método de aproximación normal dio por resultado un valor de 0.0150. La discrepancia de 0.0008 proviene de dos factores: **1.** el uso de la distribución normal da como resultado un valor *aproximado* que corresponde al área de la región sombreada en la figura 6-23, mientras que el área correcta exacta es un rectángulo centrado por arriba de 1358 (la figura 6-23 ilustra esta discrepancia); **2.** el uso de la tabla A-2 nos obligó a calcular uno de un número limitado de valores basados en una puntuación z redondeada. El área del rectángulo es 0.0142, pero el área de la región sombreada aproximada es 0.0150.

Interpretación de los resultados

En realidad, cuando utilizamos una distribución normal como aproximación de la distribución binomial, nuestra meta no es simplemente calcular un número de probabilidad. A menudo necesitamos hacer algún *juicio* con base en el valor de probabilidad. Por ejemplo, suponga que el reportero de un periódico ve los datos muestrales del ejemplo anterior y, después de observar que el 67.5% de los adultos encuestados utilizaban Internet, escribe el encabezado "Más de 2/3 de los adultos utilizan Internet". Los datos muestrales no justifican ese encabezado por la siguiente razón: si la verdadera proporción poblacional *es igual a* 2/3 (en vez de ser mayor que 2/3), existe una elevada probabilidad (0.2327) de obtener al menos 1358 usuarios de Internet entre los 2013 adultos encuestados. (El área ubicada a la izquierda de 1357.5 es 0.7673, de manera que la probabilidad de obtener al menos 1358 usuarios de Internet es $1 - 0.7673 = 0.2327$). Es decir, con una proporción poblacional igual a 2/3, el resultado de 1358 usuarios de Internet *no es excepcionalmente elevado*. Este tipo de conclusiones tal vez parezcan un poco confusas en este momento, pero en los siguientes capítulos se presentarán métodos sistemáticos que las harán más fáciles. Por ahora, debemos comprender que las bajas probabilidades corresponden a sucesos con pocas posibilidades, mientras que las altas probabilidades corresponden a sucesos posibles. El valor de probabilidad de 0.05 suele utilizarse como punto de corte para distinguir entre sucesos improbables y sucesos probables. El siguiente criterio (de la sección 5-2) describe la aplicación de las probabilidades para distinguir resultados que pueden ocurrir fácilmente por azar de aquellos que son muy poco comunes.

Uso de las probabilidades para determinar cuando los resultados son poco comunes

- **Extremadamente alto:** x éxitos en n ensayos es un número *extremadamente alto* de éxitos si P(x o más) es muy pequeña (como 0.05 o menos).
- **Extremadamente bajo:** x éxitos en n ensayos es un número *extremadamente bajo* de éxitos si P(x o menos) es muy pequeña (como 0.05 o menos).

El papel de la aproximación normal

En realidad, casi todas las aplicaciones prácticas de la distribución de probabilidad binomial ahora se pueden trabajar bien con un programa de cómputo o una calculadora TI-83/84 Plus. En esta sección se presentan métodos para manejar casos en los que no se puede utilizar un programa de cómputo y, algo más importante, también ilustra el principio de que, en las circunstancias apropiadas, la distribución de probabilidad binomial puede aproximarse por medio de una distribución normal. Los capítulos posteriores incluyen procedimientos basados en el uso de una distribución normal como aproximación de una distribución binomial, de manera que esta sección establece las bases de esos importantes procedimientos.

6-6 DESTREZAS Y CONCEPTOS BÁSICOS

Conocimientos estadísticos y pensamiento crítico

1. **Distribución de proporciones muestrales.** Considere un estudio en el que obtenemos registros de los siguientes 50 bebés que nacen, luego calcule la proporción de niñas en esta muestra. Suponga que este estudio se repite muchas veces y que se utilizan las proporciones muestrales para construir un histograma. ¿Cuál sería la forma del histograma?

2. **Corrección por continuidad.** La prueba Wechsler se utiliza para medir puntuaciones de CI y está diseñada de tal manera que la media es 100 y la desviación estándar es 16. Se sabe que las puntuaciones de CI tienen una distribución normal. Suponga que deseamos calcular la probabilidad de que una persona elegida al azar tenga un CI igual a 107. ¿Cuál sería la corrección por continuidad y cómo se aplicaría para calcular esa probabilidad?

3. **Distribución de proporciones muestrales.** La planta Newport Bottling fabrica botellas de bebidas de cola que se empacan en grupos de seis. La probabilidad de una botella defectuosa es de 0.001. ¿Podemos aproximar la distribución de defectos en los empaques de seis como una distribución normal? ¿Por qué?

4. **Interpretación de probabilidad binomial.** En la prueba de un método de selección del género, 80 parejas reciben un tratamiento diseñado para incrementar la probabilidad de que un bebé sea niña. De 80 bebés nacidos, 47 fueron niñas. Si el método de selección del género no tiene efecto, la probabilidad de obtener exactamente 47 niñas es de 0.0264, y la probabilidad de obtener 47 niñas o más es de 0.0728. ¿Cual probabilidad se debe usar para evaluar la eficacia del método de selección del género? ¿Parece que el método es efectivo?