Muestreo aleatorio sistemático

En algunos estudios, el procedimiento de muestreo aleatorio simple resulta complicado. Por ejemplo, suponga que la división de ventas de Computer Graphic, Inc., necesita calcular rápidamente el ingreso medio en dólares por venta del mes pasado. La división confirmó que se registraron 2 000 ventas y se almacenaron en cajones de archivo, y se decidió seleccionar 100 recibos para calcular el ingreso medio en dólares. El muestreo aleatorio simple requiere la numeración de cada recibo antes de utilizar la tabla de números aleatorios para seleccionar los 100 recibos. Dicho proceso de numeración puede tardar mucho tiempo. En su lugar, es posible aplicar el **muestreo aleatorio sistemático**.

MUESTREO ALEATORIO SISTEMÁTICO

Se selecciona un punto aleatorio de inicio y posteriormente se elige cada *k*-ésimo miembro de la población.

Estadística en acción

Los métodos de muestreo aleatorio y sin sesgos son muy importantes para realizar inferencias estadísticas válidas. En 1936 se efectuó un sondeo de opinión para predecir el resultado de la carrera presidencial entre Franklin Roosevelt y Alfred Landon. Se enviaron diez millones de papeletas en forma de postales retornables gratuitas a domicilios tomados de directorios telefónicos y registros de automóviles. Se contestó una alta proporción de papeletas, con 59% en favor de Landon y 41% de Roosevelt. El día de la elección, Roosevelt ganó con 61% de los votos. Landon obtuvo 39%. Sin duda, a mediados de la década de 1930, la gente que tenía teléfono y automóvil no era representativa de los votantes estadounidenses.

Primero se calcula *k*, que es el resultado de dividir el tamaño de la población entre el tamaño de la muestra. En el caso de Computers Graphic, Inc., seleccione cada vigésimo recibo (2

000/100) de los cajones del archivo; al hacerlo evita el proceso de numeración. Si *k* no es un número entero, hay que redondearlo.

Para seleccionar el primer recibo emplee el muestreo aleatorio simple. Por ejemplo, seleccione un número de la tabla de números aleatorios entre 1 y *k*, en este caso, 20. Suponga que el número aleatorio resultó ser 18. Entonces, a partir del recibo 18, se seleccionará cada vigésimo recibo (18, 38, 58, etc.) como muestra.

Antes de aplicar el muestreo aleatorio sistemático, debe observar con cuidado el orden físico de la población. Cuando el orden físico se relaciona con la característica de la población, no debe aplicar el muestreo aleatorio sistemático. Por ejemplo, si los recibos se archivan en orden creciente de ventas, el muestreo aleatorio sistemático no garantiza una muestra aleatoria. Debe aplicar otros métodos de muestreo.

Muestreo aleatorio estratificado

Cuando una población se divide en grupos a partir de ciertas características, se aplica el **muestreo aleatorio estratificado** con el fin de garantizar que cada grupo se encuentre representado en la muestra. A los grupos también se les denomina **estratos**. Por ejemplo, los estudiantes universitarios se pueden agrupar en estudiantes de tiempo completo o de medio tiempo, por sexo, masculino o femenino, tradicionales o no tradicionales. Una vez definidos los estratos, se aplica el muestreo aleatorio simple en cada grupo o estrato con el fin de formar la muestra.

MUESTRA ALEATORIA ESTRATIFICADA

Una población se divide en subgrupos, denominados *estratos*, y se selecciona al azar una muestra de cada estrato.

Por ejemplo, puede estudiar los gastos en publicidad de las 352 empresas más grandes de Estados Unidos. Suponga que el objetivo del estudio consiste en determinar si las empresas con altos rendimientos sobre el capital (una medida de rentabilidad) gastan en publicidad la mayor parte del dinero ganado que las empresas con un registro de bajo rendimiento o déficit. Para asegurar que la muestra sea una representación imparcial de las 352 empresas, éstas se deben agrupar de acuerdo con su rendimiento porcentual sobre el capital. La tabla 8-1 incluye los estratos y las frecuencias relativas. Si aplicara el muestreo aleatorio simple, observe que las empresas del tercero y cuarto estratos tienen una probabilidad alta de que se les seleccione (0.87), mientras que las empresas de los demás estratos tienen menos (0.13). Podría no seleccionar ninguna de las empresas que aparecen en los estratos 1 o 5 *sencilla* mente por *azar*. No obstante, el muestreo aleatorio estratificado garantizará que por lo menos una empresa de los estratos 1 o 5 aparezca en la muestra. Considere una selección de 50 compañías para llevar a cabo un estudio minucioso. Entonces se seleccionará de forma aleatoria 1 (0.02 × 50) empresas del estrato 1; 5 (0.10 × 50), del estrato 2, etc. En este caso, el número de empresas en cada estrato es proporcional a la frecuencia relativa del estrato en la población. El muestreo estratificado ofrece la ventaja de que, en algunos casos, refleja con mayor fidelidad las características de la población que el muestreo aleatorio simple o el muestreo aleatorio sistemático.

TABLA 8-1: Número seleccionado de una muestra aleatoria estratificada proporcional

Estrato	Probabilidad (recuperación de capital)	Número de empresas	Frecuencia relativa	Número muestreado
1	30% y más	8	0.02	1*
2	De 20% a 30%	35	0.10	5*
3	De 10% a 20%	189	0.54	27
4	De 0% a 10%	115	0.33	16
5	Déficit	5	0.01	1
Total		352	1.00	50

Muestreo por conglomerados

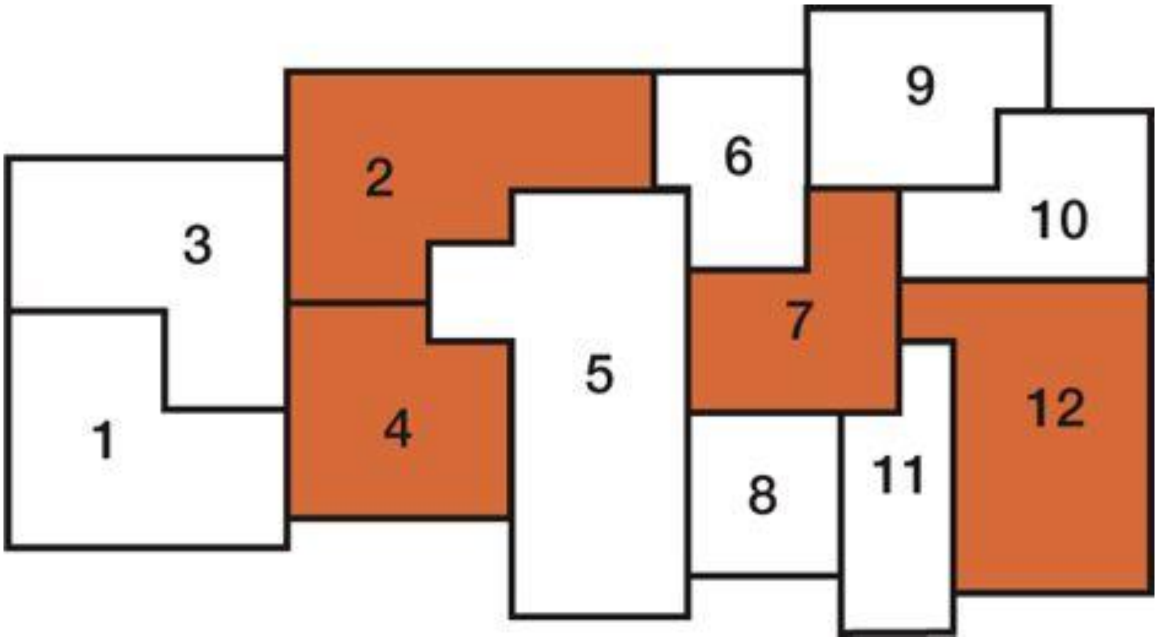
Otro tipo común de muestreo es el muestreo por conglomerados, que a menudo se emplea para reducir el costo de muestrear una población dispersa en cierta área geográfica.

MUESTREO POR CONGLOMERADOS

La población se divide en conglomerados a partir de los límites naturales geográficos o de otra clase. A continuación se seleccionan los conglomerados al azar y se toma una muestra de forma aleatoria con elementos de cada grupo.

Suponga que desea determinar la opinión de los residentes de algún estado con referencia a las políticas federales y estatales de protección ambiental. Seleccionar una muestra aleatoria de residentes y ponerse en contacto con cada persona requeriría mucho tiempo y resultaría muy costoso. Sería mejor aplicar el muestreo por conglomerados y subdividir el estado en pequeñas unidades: condados o regiones. Con frecuencia se les conoce como unidades primarias.

Suponga que dividió el estado en 12 unidades primarias, seleccionó al azar cuatro regiones, 2, 7, 4 y 12, y concentró su atención en estas unidades primarias. Usted puede tomar una muestra aleatoria de los residentes de cada una de estas regiones y entrevistarse con ellos (observe que se trata de una combinación de un muestreo por conglomerados y un muestreo aleatorio simple).



8.3: “Error” de muestreo

En la sección anterior se estudiaron métodos de muestreo útiles para seleccionar una muestra que constituya una representación imparcial o sin sesgos de la población. Es importante señalar que, en cada método, la selección de cualquier posible muestra de determinado tamaño de una población tiene una posibilidad o probabilidad conocida, que constituye otra forma de describir un método de muestreo sin sesgo.

Las muestras se emplean para determinar características de la población. Por ejemplo, con la media de una muestra se calcula la media de la población. No obstante, como la muestra forma parte o es una porción representativa de la población, es poco probable que su media sea *exactamente igual* a la media poblacional. Asimismo, es poco probable que la desviación estándar de la muestra sea *exactamente igual* a la desviación estándar de la población. Por lo tanto, puede esperar una diferencia entre un *estadístico de la muestra* y el *parámetro de la población* correspondiente. Esta diferencia recibe el nombre de **error de muestreo**.

ERROR DE MUESTREO

Diferencia entre el estadístico de una muestra y el parámetro de la población correspondiente.

Ejemplo

Revise el ejemplo anterior de la página 268, en el que estudió el número de habitaciones rentadas en Foxtrot Inn, en Tryon, Carolina del Norte. La población se refiere al número de habitaciones rentadas cada uno de los 30 días de junio de 2011. Determine la media de la población. Utilice Excel u otro software de estadística para seleccionar tres muestras aleatorias de cinco días. Calcule la media de cada muestra y compárela con la media poblacional. ¿Cuál es el error de muestreo en cada caso?

Solución

Durante el mes se rentaron un total de 94 habitaciones. Por lo tanto, la media de las unidades que se rentaron por noche es de 3.13. Ésta es la media de la población. Este valor se designa con la letra griega μ.

$$\mu = \frac{\sum X}{N} = \frac{0 + 2 + 3 + \cdots + 3}{30} = \frac{94}{30} = 3.13$$

La primera muestra aleatoria de cinco noches dio como resultado el siguiente número de habitaciones rentadas: 4, 7, 4, 3 y 1. La media de esta muestra de cinco noches es de 3.8 habitaciones, que se representa como $\overline{X}_1$. La barra sobre la *X* recuerda que se trata de una media muestral, y el subíndice 1 indica que se trata de la media de la primera muestra.

$$\overline{X}_1 = \frac{\sum X}{n} = \frac{4 + 7 + 4 + 3 + 1}{5} = \frac{19}{5} = 3.80$$

El error de muestreo de la primera muestra es la diferencia entre la media poblacional (3.13) y la media muestral (3.80). De ahí que el error muestral sea:

$$\overline{X}_1 - \mu = 3.80 - 3.13 = 0.67$$

La segunda muestra aleatoria de cinco días de la población de 30 días de junio arrojó el siguiente número de habitaciones rentadas: 3, 3, 2, 3 y 6. La media de estos cinco valores es de 3.4, que se calcula de la siguiente manera:

$$\overline{X}_2 = \frac{\sum X}{n} = \frac{3 + 3 + 2 + 3 + 6}{5} = 3.4$$

El error de muestreo es ($\overline{X}_2 - \mu = 3.4 - 3.13 = 0.27$).

En la tercera muestra aleatoria, la media fue de 1.8, y el error de muestro fue de −1.33.

Cada una de estas diferencias, 0.67, 0.27 y –1.33, representa el error de muestreo cometido al calcular la media de la población. A veces estos errores son valores positivos, lo cual indica que la media muestral sobreexcedió la media poblacional; otras veces son valores negativos, lo cual indica que la media muestral resultó inferior a la media poblacional.

Random Samples

	A	B	C	D	E	F	G	H
1	June	Rentals			Sample 1	Sample 2	Sample 3	
2	1	0			4	3	0	
3	2	2			7	3	0	
4	3	3			4	2	3	
5	4	2			3	3	3	
6	5	3			1	6	3	
7	6	4		Totals	19	17	9	
8	7	2		Sample means	3.8	3.4	1.8	
9	8	3						
10	9	4						
11	10	7						
12	11	3						
13	12	4						
14	13	4						
15	14	4						
16	15	7						

En este caso, con una población de 30 valores y muestras de 5 valores, existe una gran cantidad de muestras posibles, 142 506, para ser exactos. Para calcular este valor se aplica la fórmula de las combinaciones (5-10), de la página 174. Cada una de las 142 506 diferentes muestras cuenta con las mismas posibilidades de que se le seleccione. Cada muestra puede tener una media muestral diferente y, por consiguiente, un error de muestreo distinto. El valor del error de muestreo se basa en el valor particular de las 142 506 muestras posibles seleccionadas. Por consiguiente, los errores de muestreo son aleatorios y se presentan al azar. Si determinara la suma de estos errores de muestreo en una gran cantidad de muestras, el resultado se aproximaría mucho a cero. Sucede así porque la media de la muestra constituye un *estimador sin sesgo* de la media de la población.

Una media simple es una estimación sin sesgo de la media poblacional.

8.4: Distribución muestral de la media

Debido a que existe la posibilidad de que se presente un error de muestreo cuando se emplean los resultados del muestreo para aproximar un parámetro poblacional, ¿cómo hacer un pronóstico preciso relacionado con el posible éxito de un nuevo dentífrico u otro producto sobre la única base de los resultados del muestreo? ¿Cómo puede el departamento de control de calidad de una compañía de producción en serie enviar un cargamento de microchips a partir de una muestra de 10 chips? ¿Cómo pueden las organizaciones electorales de CNN-*USA Today* o ABC *News-Washington Post* hacer pronósticos precisos sobre la elección presidencial con base en una muestra de 1 200 electores registrados de una población de cerca de 90 millones? Para responder estas preguntas, primero hay que precisar el concepto de *distribución muestral de la media*.

Las medias muestrales del ejemplo anterior varían de una muestra a la siguiente. La media de la primera muestra de 5 días fue de 3.80 habitaciones, y la media de la segunda muestra fue de 3.40 habitaciones. La media poblacional fue de 3.13 habitaciones. Si se organizan las medias de todas las muestras posibles de 5 días en una distribución de probabilidad, el resultado recibe el nombre de **distribución muestral de la media**.

Las medias muestrales varían de muestra en muestra.

DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Distribución de probabilidad de todas las posibles medias de las muestras de un determinado tamaño muestral de la población.

El siguiente ejemplo ilustra la construcción de una distribución muestral de la media.

Ejemplo

Tartus Industries cuenta con siete empleados de producción (a quienes se les considera la población). En la tabla 8-2 se incluyen los ingresos por hora de cada uno de ellos.

TABLA 8-2: Ingresos por hora de empleados de producción en Tartus Industries

Empleado	Ingresos por hora	Empleado	Ingresos por hora
Joe	$7	Jan	$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

1. ¿Cuál es la media de la población?

2. ¿Cuál es la distribución muestral de la media de muestras de tamaño 2?

3. ¿Cuál es la media de la distribución muestral de la media?

4. ¿Qué observaciones es posible hacer sobre la población y la distribución muestral de la media?

Solución

He aquí las respuestas.

1. La media de la población es de $7.71, que se determina de la siguiente manera:

$$\mu = \frac{\sum X}{n} = \frac{\$7 + \$7 + \$8 + \$8 + \$7 + \$8 + \$9}{7} = \$\,7.71$$

Identifique la media de la población por medio de la letra griega μ. En los capítulos 1, 3 y 4 se convino en identificar los parámetros poblacionales con letras griegas.

2. Para obtener la distribución muestral de la media se seleccionó, sin reemplazos de la población, todas las muestras posibles de tamaño 2 y se calcularon las medias de cada muestra. Hay 21 muestras posibles, que se calcularon con la fórmula (5-10) de la página 174.

$$_NC_n = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

donde $N = 7$ es el número de elementos de la población, y $n = 2$, el número de elementos de la muestra.

En la tabla 8-3 se ilustran las 21 medias muestrales de todas las muestras posibles de tamaño 2 que pueden tomarse de la población. Estas 21 muestras se utilizan para construir una distribución de probabilidad, que es la distribución muestral de la media, la cual se resume en la tabla 8-4.

TABLA 8-3: Medias muestrales de todas las muestras posibles de 2 empleados

Muestra	Empleados	Ingresos por hora	Suma	Media	Muestra	Empleados	Ingresos por hora	Suma	Media
1	Joe, Sam	\$7, \$7	\$14	\$7.00	12	Sue, Bob	\$8, \$8	\$16	\$8.00
2	Joe, Sue	7, 8	15	7.50	13	Sue, Jan	8, 7	15	7.50
3	Joe, Bob	7, 8	15	7.50	14	Sue, Art	8, 8	16	8.00
4	Joe, Jan	7, 7	14	7.00	15	Sue, Ted	8, 9	17	8.50
5	Joe, Art	7, 8	15	7.50	16	Bob, Jan	8, 7	15	7.50
6	Joe, Ted	7, 9	16	8.00	17	Bob, Art	8, 8	16	8.00
7	Sam, Sue	7, 8	15	7.50	18	Bob, Ted	8, 9	17	8.50
8	Sam, Bob	7, 8	15	7.50	19	Jan, Art	7, 8	15	7.50
9	Sam, Jan	7, 7	14	7.00	20	Jan, Ted	7, 9	16	8.00
10	Sam, Art	7, 8	15	7.50	21	Art, Ted	8, 9	17	8.50
11	Sam, Ted	7, 9	16	8.00					

TABLA 8-4: Distribución muestral de la media con *n* = 2

Media muestral	Número de medias	Probabilidad
\$7.00	3	.1429
7.50	9	.4285
8.00	6	.2857
8.50	3	.1429
	21	1.0000

3. La media de la distribución muestral de la media se obtiene al sumar las medias muestrales y dividir la suma entre el número de muestras. La media de todas las medias muestrales se representa mediante $\mu_{\overline{X}}$. La μ recuerda que se trata de un valor poblacional, pues tomó en cuenta todas las muestras posibles. El subíndice $\overline{X}$ indica que se trata de la distribución muestral de la media.

$$\mu_{\overline{X}} = \frac{\text{Suma de todas las medias muestrales}}{\text{Total de muestras}} = \frac{\$\,7.00 + \$\,7.50 + \cdots + \$\,8.50}{21}$$

$$= \frac{\$\,162}{21} = \$\,7.71$$

La media de la población es igual a la media de las medias muestrales.

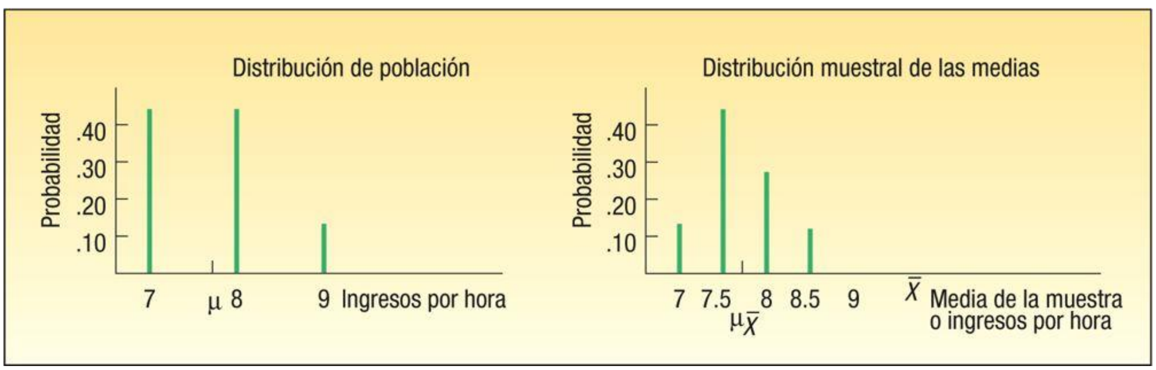
4. Consulte la gráfica 8-1, donde aparecen las dos distribuciones poblacionales y la distribución muestral de la media. Caben las siguientes observaciones:

a) La media de la distribución muestral de la media ($7.71) es igual a la media de la población: $\mu = \mu_{\overline{X}}$.

b) La dispersión de la distribución muestral de las medias es menor que la dispersión de los valores de población. La media de las muestras varía de $7.00 a $8.50, mientras que los valores de población varían de $7.00 a $9.00. Observe que, conforme se incrementa el tamaño de la muestra, se reduce la dispersión de la distribución muestral de las medias.

c) La forma de la distribución muestral de la media y la forma de la distribución de frecuencias de los valores de población son diferentes. La distribución muestral de las medias tiende a adoptar más forma de campana y a aproximarse a la distribución de probabilidad normal.

GRÁFICA 8-1



Distribución de los valores de población y distribución muestral de las medias.

En resumen, tome todas las posibles muestras aleatorias de una población y calcule un estadístico muestral (la media de los ingresos percibidos) de cada una. Este ejemplo ilustra las importantes relaciones entre la distribución poblacional y la distribución muestral de la media:

1. La media de las medias de las muestras es exactamente igual a la media de la población.

2. La dispersión de la distribución muestral de la media es más estrecha que la distribución poblacional.

3. La distribución muestral de la media suele tener forma de campana y se aproxima a la distribución de probabilidad normal.

Dada una distribución de probabilidad normal o de forma de campana, se aplican los conceptos del capítulo 7 para determinar la probabilidad de seleccionar una muestra con una media muestral específica. En la siguiente sección se resalta la importancia del tamaño de una muestra en relación con la distribución muestral de la media.

Autoevaluación 8-3

Los tiempos de servicio de los ejecutivos que laboran en Standard Chemicals son los siguientes:

Nombre	Años
Señor Snow	20
Señora Tolson	22
Señor Kraft	26
Señora Irwin	24
Señor Jones	28

a) De acuerdo con la fórmula de las combinaciones, ¿cuántas muestras de tamaño 2 son posibles?

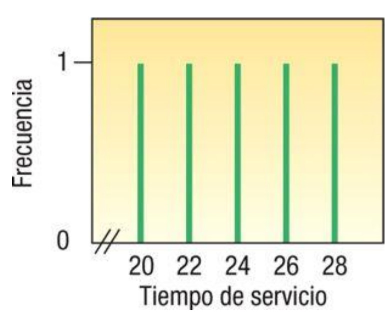
b) Elabore una lista de todas las muestras posibles de 2 ejecutivos de la población y calcule las medias.

c) Organice las medias en una distribución muestral.

d) Compare la media poblacional y la media de las medias de las muestras.

e) Compare la dispersión en la población con la dispersión de la distribución muestral de la media.

f) A continuación se muestra una gráfica con los valores de la población. ¿Tienen los valores de población una distribución normal (en forma de campana)?



g) ¿Comienza la distribución muestral de la media que se calculó en el inciso *c*) a indicar una tendencia a adoptar forma de campana?

Ejercicios

5. Una población consta de los siguientes cuatro valores: 12, 12, 14 y 16.

***a*)** Enumere todas las muestras de tamaño 2 y calcule la media de cada muestra.

***b*)** Calcule la media de la distribución muestral de la media y la media de la población. Compare los dos valores.

***c*)** Compare la dispersión en la población con la de las medias de las muestras.

6. Una población consta de los siguientes cinco valores: 2, 2, 4, 4 y 8.

***a*)** Enumere todas las muestras de tamaño 2 y calcule la media de cada muestra.

***b*)** Calcule la media de la distribución muestral de las medias y la media de la población. Compare los dos valores.

***c*)** Compare la dispersión en la población con la de las medias de las muestras.

7. Una población consta de los siguientes cinco valores: 12, 12, 14, 15 y 20.

***a*)** Enumere todas las muestras de tamaño 3 y calcule la media de cada muestra.

***b*)** Calcule la media de la distribución muestral de las medias y la media de la población. Compare los dos valores.

***c*)** Compare la dispersión de la población con la de las medias de las muestras.

8. Una población consta de los siguientes cinco valores: 0, 0, 1, 3 y 6.

***a*)** Enumere todas las muestras de tamaño 3 y calcule la media de cada muestra.

***b*)** Calcule la media de la distribución muestral de las medias y la media de la población. Compare los dos valores.

***c*)** Compare la dispersión de la población con la de las medias de las muestras.

9. El despacho de abogados Tybo and Associates consta de seis socios. En la siguiente tabla se incluye el número de casos que en realidad atendió cada socio en los tribunales durante el mes pasado.

Socio	Número de casos
Ruud	3
Wu	6
Sass	3
Flores	3
Wilhelms	0
Schueller	1

***a*)** ¿Cuántas muestras de 3 son posibles?

***b*)** Enumere todas las muestras posibles de 3 y calcule el número medio de casos en cada muestra.

***c*)** Compare la media de la distribución muestral de las medias con la de la media poblacional.

***d*)** En una gráfica similar a la 8-1, compare la dispersión en la población con la de las medias muestrales.

10. Mid-Motors Ford tiene cinco vendedores. Los cinco representantes de ventas y el número de automóviles que vendieron la semana pasada son los siguientes:

Representantes de ventas	Autos vendidos
Peter Hankish	8
Connie Stallter	6
Juan López	4
Ted Barnes	10
Peggy Chu	6

***a*)** ¿Cuántas muestras de tamaño 2 son posibles?

***b*)** Enumere todas las muestras posibles de tamaño 2 y calcule la media en cada muestra.

***c*)** Compare la media de la distribución muestral de la media con la de la media poblacional.

***d*)** En una gráfica similar a la 8-1, compare la dispersión de la población con la de la media de la muestra.

Lind, Douglas A. *Estadística aplicada a los negocios y la economía, 15th Edition*. McGraw-Hill Interamericana, 2012. VitalBook file.