F A C U L T A D D E E S T U D I O S S U P E R I O R E S

A R A G Ó N

SISTEMA UNIVERSIDAD ABIERTA

ECONOMÍA

MATERIAL DIDÁCTICO

ESTADÍSTICA II

2009-II

El Sistema Universidad Abierta ha cumplido 35 años de contribuir a la innovación de los procesos de aprendizaje en la Universidad Nacional Autónoma de México y en todo el país. En 1972, a iniciativa del Dr. Pablo González Casanova se creó el Sistema Universidad Abierta enfocado al aprendizaje y en las necesidades a satisfacer del estudiante, permitiéndole a éste integrar su educación a las exigencias prácticas de la vida tanto cotidiana como profesional.

La educación abierta y a distancia es una forma de organización y políticas que tienden a la flexibilización en cuanto a tiempos, plazos y formas de interacción entre estudiante y asesor.

La participación de estudiante y asesor en la construcción del conocimiento es en base a la corresponsabilidad de ambos protagonistas, especialmente del primero.

En este sentido, la División del Sistema Universidad Abierta y Educación Continua tiene la responsabilidad de poner al alcance de la mano todos los elementos necesarios para la consecución de los objetivos de aprendizaje. Un elemento básico de este proceso lo constituye el material didáctico en torno al cual giran las fortalezas del sistema abierto.

Los materiales didácticos más que una antología de lecturas, es una estrategia de trabajo diferente para garantizar su uso adecuado. En este caso, los materiales didácticos son autoadministrables, es decir, cuentan con los elementos suficientes para que el estudiante por sí mismo pueda comprender los objetivos de aprendizaje, desarrollar las actividades que le permitan alcanzarlos y contar con los elementos de evaluación y autoevaluación en el momento en que deben realizar sus exámenes.

Complementando lo anterior, la labor del asesor es potenciar la utilidad de estos materiales didácticos para hacer que los elementos básicos que se encuentran en ellos sean ampliados y profundizados a través de la discusión no sólo con un estudiante en particular sino con el total de participantes en cada asignatura.

El material didáctico y las sesiones de asesoría personalizada o grupal, a distancia o presencial son espacios de análisis donde el estudiante es activo promotor de su aprendizaje y no un pasivo oyente.

En este orden de ideas se cuenta con material didáctico de cuidadosa selección de lecturas que abarca los variados temas del programa de estudio e incluye de manera clara los objetivos y actividades para conseguirlos, asimismo, se encuentra en este material didáctico los elementos para medir el avance del aprendizaje.

Por otro lado el continuo avance tecnológico permite ofrecerte el material didáctico en Internet, accediendo a la plataforma “SUAragón en línea”, lo que permite consultarlo desde cualquier lugar y momento, así como interactuar con tus asesores y compañeros por medio del foro de discusión y recibir información propia de tus asignaturas.

Como toda actividad universitaria es un material que está sujeto a la crítica bajo la premisa de que todo es perfectible. Dado el vertiginoso avance de la ciencia en esta era del conocimiento, se considera también que es una obra temporal constantemente sujeta a revisión y modificación para mantenerla a tono con los cambios que el estudio de la Economía imponen.

Finalmente, la División del SUA Aragón destaca el esfuerzo que significó hacer llegar a sus manos este material didáctico. Para lograrlo se conjugaron muchos esfuerzos tanto académicos como prácticos por parte de los autores en un trabajo pionero en la más joven de las Facultades de la Universidad Nacional Autónoma de México.

Estadística II

ÍNDICE

INTRODUCCIÓN

Los métodos estadísticos se definen frecuentemente como métodos para tomar decisiones frente a hechos inciertos. El resultado de un experimento es, generalmente, incierto pero se espera que si el experimento se repite cierto número de veces puede que seamos capaces de construir un modelo probabilístico para el experimento y de tomar decisiones respecto al proceso experimental por medio de tal modelo.

La experiencia indica que muchas operaciones o experimentos reiterativos se comportan como si se produjeran bajo circunstancias esencialmente estables. Los juegos de azar, tales como arrojar una moneda al aire o lanzar un dado, presentan esta propiedad. Muchos experimentos y operaciones en las diversas ramas de la ciencia y de la industria se comportan también así.

Debido a la naturaleza de los datos y modelos estadísticos, es lógico que la probabilidad sea la herramienta fundamental en la teoría estadística. El estadístico considera la probabilidad como una idealización de la proporción de veces que un determinado resultado se presenta en las pruebas repetidas de un experimento; en consecuencia, un modelo de probabilidad es el tipo de modelo matemático elegido por él.

La idea de un modelo matemático para ayudar a resolver los problemas de la vida real es muy conocida en las diversas ciencias. Por ejemplo, un físico que esta estudiando el movimiento de un proyectil supone a menudo que las simples leyes de la mecánica proporciona un modelo satisfactorio a pesar de la complejidad del problema real. Para un trabajo más refinado introduce un modelo más complicado. Puesto que un modelo es sólo una idealización de la situación real, las conclusiones que se deduzcan de el únicamente son dignas de confianza en la medida en que el modelo elegido sea una aproximación suficientemente buena a la situación que se estudia. Por lo tanto, en cualquier problema dado es fundamental conocer bien el campo de aplicación para saber qué modelos son probablemente los realistas.

En el proceso de resolver un problema de la vida real por medio de la estadística pueden considerarse tres pasos. Primero, elección de un modelo matemático. Segundo, comprobación de la realidad del mismo. Tercero, obtención de las conclusiones adecuadas de este modelo para resolver el problema propuesto.

Alan Greenspan, ex presidente de la Reserva Federal de los Estados Unidos, conoce y entiende la importancia de las herramientas y técnicas estadísticas para proporcionar información precisa y oportuna que sirva para hacer declaraciones públicas con la fuerza de movilizar mercados bursátiles globales e influir en la política.

Como estudiante de economía y negocios, requerirás conocimientos básicos y habilidad para organizar, analizar y transformar datos, así como presentar la información. En esta materia, aprenderás las técnicas y métodos estadísticos básicos que mejorarán tu destreza para tomar buenas decisiones personales y de negocios.

NOMBRE DE LA ASIGNATURA:	Estadística II
LICENCIATURA:	Economía
SEMESTRE:	Cuarto
CICLO Y ÁREA A LA QUE PERTENECE:	Formación Teórica
CARÁCTER:	Obligatoria
NUMERO DE CRÉDITOS:	6 Créditos

Objetivo general del curso

Analizar estadísticamente trabajos de investigación económica, usando los diferentes elementos y herramientas de la estadística.

Criterios para la evaluación

El alumno de la asignatura de Estadística II, debe leer los materiales didácticos correspondientes a cada uno de los temas del programa de la asignatura (antología), para que al final resuelva un cuestionario por unidad temática el cual le permitirá apreciar desde su perspectiva el conocimiento y comprensión de los mismos.

Este cuestionario es una modalidad dentro del proceso de evaluación del aprendizaje del alumno, que el maestro puede tomar como parte de su proceso de evaluación además de los requerimientos individuales que bajo el principio de la libertad de cátedra le otorga la Universidad Nacional Autónoma de México, como lo son exámenes parciales por unidad, examen final, trabajos de investigación, controles de lectura, ensayos, etc.

INTRODUCCIÓN

En esta unidad iniciamos el estudio del muestreo, herramienta para inferir algo sobre una población. Después se construye una distribución de la media de la muestra para entender la forma como las medias muestrales tienden acumularse en torno a la media de la población. Además se demuestra que, para cualquier población, la forma de esta distribución de muestreo tiende a seguir la distribución de probabilidad normal cuando es lo suficientemente grande, es decir, que la distribución de frecuencias de la muestra representa de modo satisfactorio a la población muestreada.

ACTIVIDADES DE APRENDIZAJE

1.1. Realiza la lectura del material que se presenta en esta unidad.
1.2. Elabora un cuadro comparativo de los tipos de muestreos,
1.3. Elabora un resumen de la metodología para el muestreo.
1.4. Elabora un resumen de las distribuciones muéstrales.
1.5. Elabora un cuadro sinóptico de la estimación, tamaño de muestra y errores de muestreo y no muestreo.

Objetivos particulares

Mencionar las diferentes herramientas estadísticas que se utilizan en la elaboración de estudios por muestreo.

CONTENIDOS

1.1 Introducción al muestreo.
1.2 Diferentes tipos de muestreo.
1.3 Metodología para el estudio por muestreo.
1.4 Distribuciones muéstrales.
1.5 Estimación y tamaño de muestra.
1.6 Errores de muestreo y no muestreo.
1.7 Desarrollo de investigación económico-administrativa aplicando la metodología del estudio por muestreo.

Fichas bibliográficas de los documentos

Documento	Ficha
1.A.	LIND, Marchal y Wathen, **Estadística aplicada a los negocios y a la economía** 13ª ed., Editorial McGraw Hill, México 2008 Págs. 260-292
1.B	HOEL G. Paul y Jessen J. Raymond **Estadística básica para negocios y economía** Editorial C.E.C.S.A, México 1982 Págs. 145 - 171

1.A.	LIND, Marchal y Wathen, **Estadística aplicada a los negocios y a la economía** 13ª ed., Editorial Mc Graw Hill, México 2008 Págs. 260-292

[…]

8

Métodos de muestreo y el teorema del límite central.

OBJETIVOS:

Al concluir el capitulo, será capaz de:

1. Explicar la razón por la que una muestra es con frecuencia la única forma viable para conocer algo sobre una población.

2. Describir métodos para seleccionar una muestra.

3. Definir y construir una distribución muestral de la media de la muestra.

4. Comprender y explicar el teorema del *límite central.*

5. Aplicar el teorema del límite central para calcular probabilidades de seleccionar posibles medias muestrales de una población específica.

El informe anual de Nike indica que el estadounidense promedio compra 6.5 pares de zapatos deportivos al año. Suponga que la desviación estándar de la población es de 2.1 y que se analizará una muestra de 81 clientes el siguiente año. ¿Cuál es el error estándar de la media en este experimento? (Véase el objetivo 5 y el ejercicio 45)

Introducción

De los capítulos 2 a 4 se hizo hincapié en las técnicas para describir datos. Con el fin de ilustrar dichas técnicas, se organizaron los precios de 60 vehículos vendidos el mes pasado en Whitner Autoplex en una distribución de frecuencias para calcular las diversas medidas de ubicación y dispersión. Dichas medidas, como la media y la desviación estándar, describen el precio de venta habitual y la dispersión de los precios de venta. En estos capítulos se destacó la descripción de la condición de los datos: se describió algo que ya había sucedido.

El capítulo 5 comienza a establecer el fundamento de la inferencia estadística con el estudio de la probabilidad. Recuerde que, en la inferencia estadística, el objetivo es determinar algo sobre una *población* a partir sólo de una *muestra*. La población es todo el grupo de individuos u objetos en estudio, y la muestra es una parte o subconjunto de dicha población. El capítulo 6 amplía los conceptos de probabilidad al describir tres distribuciones de probabilidad discreta: binomial, hipergeométrica y de Poisson. El capítulo 7 describe la distribución de probabilidad uniforme y la distribución de probabilidad normal. Ambas son distribuciones continuas. Las distribuciones de probabilidad abarcan todos los posibles resultados de un experimento, así como la probabilidad asociada con cada resultado. Mediante las distribuciones de probabilidad se evaluó la probabilidad de que ocurra algo en el futuro.

Este capítulo inicia el estudio del muestreo, herramienta para inferir algo sobre una población. Primero se analizan los métodos para seleccionar una muestra de una población. Después se construye una distribución de la media de la muestra para entender la forma como las medias muestrales tienden a acumularse en torno a la media de la población. Por último, se demuestra que, para cualquier población, la forma de esta distribución de muestreo tiende a seguir la distribución de probabilidad normal.

Métodos de muestreo

Ya se mencionó en el capítulo 1 que el propósito de la estadística inferencial consiste en determinar algo sobre una población a partir de una muestra. Una muestra es una porción o parte de la población de interés. En muchos casos, el muestreo resulta más accesible que el estudio de toda la población. En esta sección se explican las razones principales para muestrear y, enseguida, diversos métodos para elegir una muestra.

Razones para muestrear

Cuando se estudian las características de una población, existen diversas razones prácticas para preferir la selección de porciones o muestras de una población para observar y medir. He aquí algunas razones para muestrear:

1. **Establecer contacto con toda la población requeriría mucho tiempo**. Un candidato para un puesto federal quizá desee determinar las posibilidades que tiene de resultar electo. Una encuesta de muestreo en la que se utiliza el personal y las entrevistas de campo convencionales de una empresa especializada en encuestas tardaría de uno o dos días. Con el mismo personal y los mismos entrevistadores, y laborando siete días a la semana, se requerirían 200 años para ponerse en contacto con toda la población en edad de votar. Aunque fuera posible reunir a un numeroso equipo de encuestadores, quizá no valdría la pena entrar en contacto con todos los votantes.

2. **El costo de estudiar todos los elementos de una población resultaría prohibitivo**. Las organizaciones que realizan encuestas de opinión pública y pruebas entre consumidores, como Gallup Polls y Roper ASW, normalmente entran en contacto con menos de 2 000 de las casi 60 millones de familias en Estados Unidos. Una organización que entrevista a consumidores en panel cobra cerca de $40 000 por enviar muestras por correo y tabular las respuestas con el fin de probar un producto (como un cereal para el desayuno, alimento para gato o algún perfume). La misma prueba del producto con los 60 millones de familias tendría un costo de aproximadamente $1 000 000 000.

Estadística en acción

Con el importante papel que desempeña la estadística inferencial en todas las ramas de la ciencia, es ya una necesidad la disponibilidad de fuentes copiosas de números aleatorios. En 1927 se publicó el primer libro de números aleatorios, con 41 600 dígitos aleatorios, generados por L. Tippett. En 1938, R. A. Fisher y E. Yates publicaron 15 000 dígitos aleatorios, generados con dos barajas. En 1955, RAND Corporation publicó un millón de dígitos aleatorios, generados por pulsos de frecuencia aleatorios de una ruleta electrónica. Para 1970, las aplicaciones del muestreo requerían miles de millones de números aleatorios. Desde entonces se han creado métodos para generar, con ayuda de computadoras, dígitos "casi" aleatorios, por lo que se les llama seudoaleatorios. Aun es motivo de debate la pregunta acerca de si un programa de computadora sirve para generar números aleatorios que de verdad sean aleatorios.

3. **Es imposible verificar de manera física todos los elementos de la población.** Algunas poblaciones son infinitas. Sería imposible verificar toda el agua del lago Erie en lo que se refiere a niveles de bacterias, así que se eligen muestras en diversos lugares. Las poblaciones de peces, aves, serpientes o mosquitos son grandes, y se desplazan, nacen y mueren continuamente. En lugar de intentar contar todos los patos que hay en Canadá o todos los peces del lago Pontchartrain, se hacen aproximaciones mediante diversas técnicas: se cuentan todos los patos que hay en un estanque, capturados al azar, se revisan las cestas de los cazadores o se colocan redes en lugares predeterminados en el lago.

4. **Algunas pruebas son de naturaleza destructiva.** Si los catadores de vino de Sutter Home Winery, California, se bebieran todo el vino para evaluar la vendimia, acabarían con la cosecha y no quedaría nada disponible para la venta. En el área de producción industrial las placas de acero cables y productos similares deben contar con una resistencia mínima a la tensión. Para cerciorarse de que el producto satisface la norma mínima, el departamento de control de calidad elige una muestra de la producción actual. Cada pieza se somete a tensión hasta que se rompe y se registra el punto de ruptura (medido en libras por pulgada cuadrada). Es obvio que si se sometieran todos los cables o todas las placas a pruebas de resistencia a la tensión no habría productos disponibles para vender u utilizar. Por la misma razón, Kodak selecciona sólo una muestra de película fotográfica y la somete a pruebas para determinar la calidad de todos los rollos que se producen; y sólo unas cuantas semillas se someten a pruebas de germinación en Burpee, antes de la temporada de siembra.

5. **Los resultados de la muestra son adecuados.** Aunque se contara con recursos suficientes, es difícil que la precisión de una muestra de 100% —toda la población— resulte esencial en la mayoría de los problemas. Por ejemplo, el gobierno federal utiliza una muestra de tiendas de comestibles distribuidas en Estados Unidos para determinar el índice mensual de precios de los alimentos. Los precios del pan, frijol, leche y otros productos de primera necesidad se incluyen en el índice. Resulta poco probable que la inclusión de todas las tiendas de comestibles de Estados Unidos influya significativamente en el índice, pues los precios de la leche, el pan y otros productos de primera necesidad no varían más de unos cuantos centavos de una cadena de tiendas a otra.

Muestreo aleatorio simple

El tipo de muestreo más común es el muestreo aleatorio simple.

MUESTREO ALEATORIO SIMPLE Muestra seleccionada de manera que cada elemento o individuo de la población tenga las mismas posibilidades de que se le incluya.

Para ejemplificar el muestreo aleatorio simple y la selección, suponga que una población consta de 845 empleados de Nitra Industries. Se va a elegir una muestra de 52 empleados de dicha población. Una forma de asegurarse de que todos los empleados de la población tienen las mismas posibilidades de que se les elija consiste en escribir primero el nombre de cada empleado en un papel y depositarlos todos en una caja. Después de mezclarlos, se efectúa la primera selección tomando un papel de la caja sin mirarlo. Se repite este proceso hasta terminar de elegir la muestra de 52 empleados.

Un método más conveniente de seleccionar una muestra aleatoria consiste en utilizar un número de identificación por cada empleado y una **tabla de números aleatorios** como la del apéndice B.6. Como su nombre lo indica, estos números se generaron mediante un proceso aleatorio (en este caso, con una computadora).

Una tabla de números aleatorios es una forma eficiente de seleccionar a los miembros de una muestra.

La probabilidad de 0, 1, 2,..., 9 es la misma para cada dígito de un número. Por consiguiente, la probabilidad de que se seleccione el empleado 011 es la misma que para los empleados 722 o 382. Al emplear números aleatorios para seleccionar empleados, se elimina la influencia o sesgo del proceso de selección.

En la siguiente ilustración aparece parte de una tabla de números aleatorios. Para seleccionar una muestra de empleados, elija primero un punto de partida en la tabla; cualquier punto sirve. Ahora suponga que el reloj marca las 3:04. Puede observar la tercera columna y enseguida desplazarse hacia abajo hasta el cuarto conjunto de números. El número es 03759. Como sólo hay 845 empleados, utilizará los tres primeros dígitos de un número aleatorio de cinco dígitos. Por tanto, 037 es el número del primer empleado que se convertirá en miembro de la muestra. Otra forma de elegir el punto de partida consiste en cerrar los ojos y señalar un número de la tabla. Para continuar, puede desplazarse en cualquier sentido. Suponga que lo hace hacia la derecha. Los primeros tres dígitos del número a la derecha de 03759 son 447, el número del siguiente empleado seleccionado para integrar la muestra. El siguiente número de tres dígitos a la derecha es 961. Omita 961, pues sólo hay 845 empleados. Continúe hacia la derecha y seleccione al empleado 784; después el 189 y así en lo sucesivo.

50525	57454	28455	68226	34656	38884	39018
72507	53380	53827	42486	54465	71819	91199
34986	74297	00144	38676	89967	98869	39744
68851	27305	03759	44723	96108	78489	18910
06738	62879	03910	17350	49169	03850	18910
11448	10734	05837	24397	10420	16712	94496

Segundo Empleado

Cuarto empleado

Punto de partida

Tercer Empleado

La mayoría de los paquetes de software contienen una rutina para seleccionar una muestra aleatoria simple. En el siguiente ejemplo se emplea el sistema Excel para elegir una muestra aleatoria.

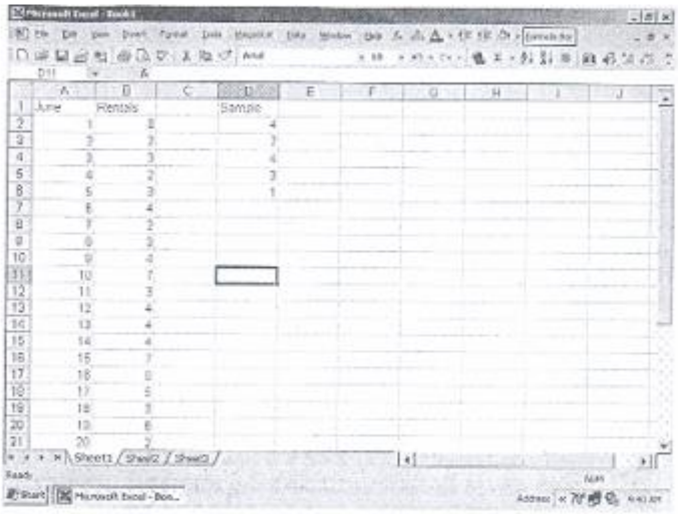
Jane y Joe Millar administran el Foxtrot Inn, una pensión donde dan alojamiento y desayuno, localizada en Tryon, Carolina del Norte. Se rentan ocho habitaciones en esta pensión. A continuación aparece el número de estas ocho habitaciones rentadas diariamente durante junio de 2006. Utilice Excel para seleccionar una muestra de cinco noches de junio.

Junio	Habitaciones en renta	Junio	Habitaciones en renta	Junio	Habitaciones en renta
1	0	11	3	21	3
2	2	12	4	22	2
3	3	13	4	23	3
4	2	14	4	24	6
5	3	15	7	25	0
6	4	16	0	26	4
7	2	17	5	27	1
8	3	18	3	28	1
9	4	19	6	29	3
10	7	20	2	30	3

Excel seleccionará la muestra aleatoria y arrojará los resultados. En la primera fecha muestreada había cuatro habitaciones rentadas de las ocho. En la segunda fecha muestreada de junio, se rentaron siete de las ocho habitaciones. La información aparece en la columna D de la hoja de cálculo de Excel. Los pasos en Excel se incluyen en la sección Comandos de software, al final del capítulo. El sistema Excel lleva a cabo el muestreo *con* reemplazo. Esto significa que tal vez el mismo día aparezca más de una vez en una muestra.

Estadística en acción

¿Es discriminación sacar ventaja del físico? Antes de contestar, considere un artículo reciente que apareció en *Personnel Journal*. Estos hallazgos indican que los hombres y mujeres atractivos ganan alrededor de 5% más que los que tienen una apariencia promedio, quienes, a su vez, ganan 5% más que sus compañeros poco agraciados. Esto se aplica tanto en hombres como en mujeres. También es cierto en el caso de gran variedad de ocupaciones, desde la construcción hasta la reparación de automóviles y los empleos de telemarketing, ocupaciones para las que, según se cree, la apariencia no es importante.



Muestreo aleatorio sistemático

El procedimiento de muestreo aleatorio simple resulta complicado en algunos estudios. 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 encontró 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 que 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.

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.

En la selección del primer recibo emplee el muestreo aleatorio simple. Por ejemplo, seleccionará 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 el hecho de 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 media de rentabilidad) gastan en publicidad la mayor parte del dinero ganado en ventas 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 agrupan 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 pocas probabilidades de que se les seleccione (0.13). Podría no seleccionar ninguna de las empresas que aparecen en los estratos 1 o 5 *sencillamente 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 x 50) empresa del estrato 1; 5 (0.10 x 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.

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.

Tabla 8.1 Numero seleccionado para una muestra aleatoria estratificada proporcional.

Estrato	Probabilidad (recuperación de capital)	Número de empresas	Frecuencia relativa	Numero 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

*0.02 de 50 = 1, 0.10 de 50 = 5, etcétera.

Muestreo por conglomerados

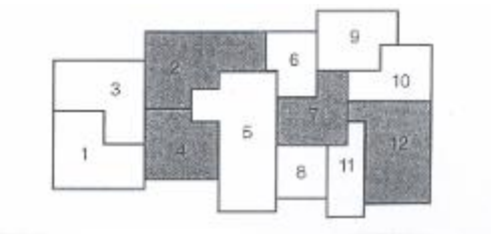
Otro tipo común de muestreo es el **muestreo por conglomerados**. Éste se emplea a menudo para reducir el costo de muestrear una población dispersa en cierta área geográfica.

MUESTREO ACUMULADO Una 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. Seria mejor aplicar el muestreo por conglomerados y subdividir el estado en pequeñas unidades: condados o regiones. Con frecuencia se es 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 entrevistarse con ellos (observe que se trata de una combinación de un muestreo por conglomerados y un muestreo aleatorio simple).

El estudio de los métodos de muestreo de las secciones anteriores no incluye todos los métodos de muestreo disponibles para el investigador. Si usted emprendiera un proyecto de investigación importante de marketing, finanzas, contabilidad u otras áreas necesitaría consultar libros dedicados exclusivamente a la teoría del muestreo y al diseño de muestras. Muchos métodos más de muestreo



"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 conocidas. Esta 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 la media de la muestra 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 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.

El siguiente ejemplo aclara el concepto de error de muestreo.

Revise el anterior de la página 263, 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 2006. Determine a 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?

Durante el mes se rentaron un total de 94 habitaciones. Así, la media de las unidades rentadas por noche es de 3.13. Esta es la media de la población. Este valor se designa con la letra griega μ.

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

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 para 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.

En este caso, con una población de 30 valores y muestras de 5 valores, existe una gran cantidad de posibles muestras, 142 506, para ser exactos. Para calcular este se aplica la fórmula de las combinaciones 5.10, de la página 168. 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 posibles muestras 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.

Distribución muestral de la media

Ahora que aparece 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 organiza 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.**

DISTRIBUCIÓN MUESTRAL DE LA MEDIA Distribución de probabilidad de todas las posibles medias de las muestras de un determinado tamaño muestra de la población.

Las medias muestrales varían de muestra en muestra.

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 empleado.

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 para 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?

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 muestra 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 posibles muestras, que se calcularon con la fórmula (5.10) de la página 168.

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

Aquí, 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 todos las posibles muestras 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, 8	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 para *n* = 2

Media muestral	Numero 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 $\mu_{\bar{X}}$. La μ recuerda que se trata de un valor poblacional, pues tomó en cuenta todas las muestras posibles. El subíndice $\bar{X}$ indica que se trata de a distribución muestral de la media.

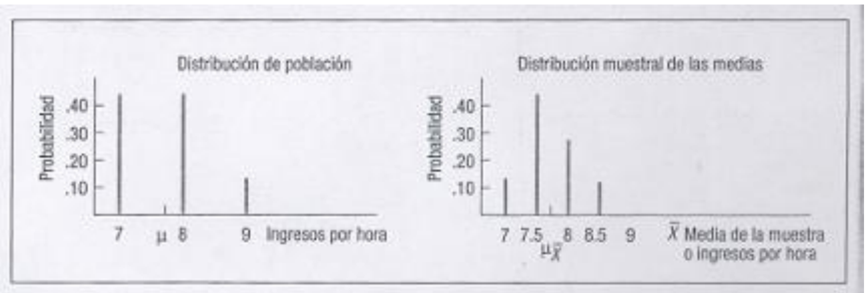
$$\mu_{\bar{X}} = \frac{\text{suma de todas las medias muestrales}}{\text{total de muestras}} = \frac{\$7.00 + \$7.50 + \ldots + \$8.50}{21} = \frac{\$162}{21} = \$7.71$$

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: $\mu = \mu_{\bar{X}}$

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 $700 a $850, mientras que los valores de población varían de $700 a $900. 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.



GRAFICA 8.1 Distribución de los valores de población y distribución muestral de medias

En resumen, tome todas las posibles muestras aleatorias de una población y calcule un estadístico muestral (la media de los ingresos percibidos) para 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.

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

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 resalta la importancia del tamaño de una muestra en relación con la distribución muestral de la media.

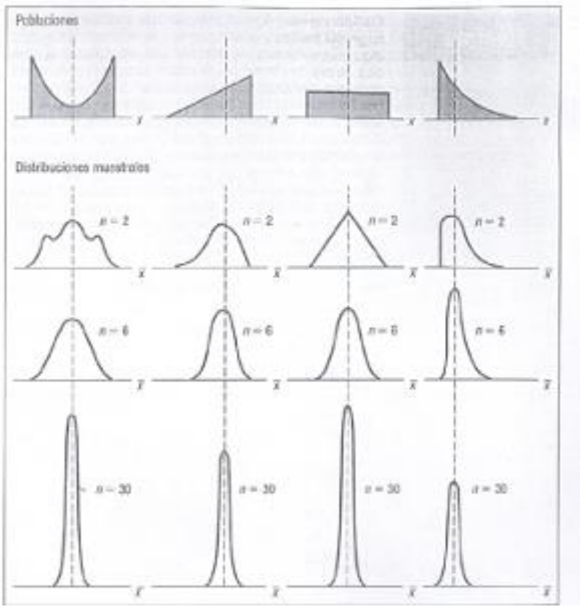
Teorema del límite central

En esta sección se estudia el **teorema del límite central**. Su aplicación a la distribución muestral de medias, en la sección anterior, permite utilizar la distribución de probabilidad normal para crear intervalos de confianza para la media poblacional (que se describe en el capítulo 9) y llevar a cabo pruebas de hipótesis (descritas en el capitulo 10). El teorema del límite central hace hincapié en que, en el caso de muestras aleatorias grandes, la forma de la distribución muestral de la media se aproxima a la distribución de probabilidad normal. La aproximación es más exacta en el caso de muestras grandes que en el de muestras pequeñas. Esta es una de las conclusiones más útiles de la estadística. Permite razonar sobre la distribución de las medias muestrales sin ninguna información acerca de la forma de la distribución de población de la que se toma la muestra. En otras palabras, el teorema del límite central se cumple en el caso de todas las distribuciones.

En seguida aparece el enunciado formal del teorema del límite central.

TEOREMA DEL LÍMITE CENTRAL Si todas las muestras de un tamaño en particular se seleccionan de cualquier población, la distribución muestral de la media se aproxima a una distribución normal. Esta aproximación mejora con muestras más grandes

Si la población obedece a una distribución normal, entonces, en el caso de cualquier tamaño de muestra, la distribución muestral de las medias también será de naturaleza normal. Si la distribución poblacional es simétrica (pero no normal), se verá que la forma normal de la distribución muestral de las medias se presenta con muestras tan pequeñas como 10. Por otra parte, sise comienza con una distribución sesgada o con colas gruesas, quizá se requieran muestras de 30 o más para observar la característica de normalidad. Este concepto se resume en la gráfica 8.2 para diversas formas de población. Observe la convergencia hacia una distribución normal sin importar la forma de la distribución de población. La mayoría de los especialistas en estadística consideran que una muestra de 30 o mayor es lo bastante grande para aplicar el teorema del limite central.

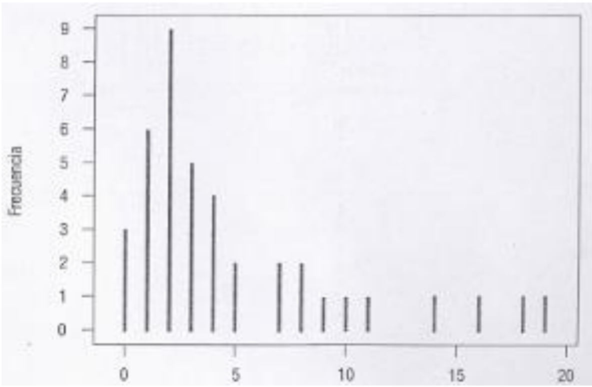


La idea de que la distribución muestral de las medias de una población que no es normal converge hacia la normalidad se ilustra en las gráficas 8.3, 8.4 y 8.5. En breve se analiza este ejemplo con más detalles, pero la gráfica 8 es la gráfica de una distribución de probabilidad discreta con sesgo positivo. Hay varias posibles muestras de 5 que puede seleccionar de esta población. Suponga que selecciona al azar 25 muestras de tamaño 5 cada una y calcula la media de cada muestra. Estos resultados se muestran en la gráfica 8.4. Observe que la forma de la distribución muestral de las medias cambió la forma de la población original aunque sólo seleccionó 25 de las diversas posibles muestras. En otras palabras, eligió 25 muestras al azar de tamaño 5 de una población positivamente sesgada, y encontró que la distribución muestral de las medias cambió en lo que se refiere a la forma de la población. A medida que toma muestras más grandes, es decir, *n* = 20 en lugar de *n* = 5, la distribución muestral de las medias se aproximará a la distribución normal. La gráfica 8.5 muestra los resultados de 25 muestras aleatorias de 20 observaciones cada una tomadas de la misma población. Note la clara tendencia hacia la distribución de probabilidad normal. Esta es la esencia del teorema del límite central. El siguiente ejemplo pondrá de relieve esta condición.

Ejemplo. Ed Spence dio inicio a su negocio de engranes hace 20 años. El negocio creció a lo largo del tiempo y ahora cuenta con 40 empleados. Spence Sprockets, Inc., encara algunas decisiones importantes relacionadas con la atención médica de sus empleados. Antes de tomar una decisión definitiva sobre el programa de atención médica que va a comprar, Ed decide formar un comité de cinco empleados. Se pedirá al comité que estudie el tema del cuidado de la salud y haga alguna recomendación sobre el plan que mejor convenga a los empleados. Ed cree que el punto de vista de los empleados más recientes en relación con el cuidado de la salud difiere de los empleados con más experiencia. Si Ed selecciona al azar este comité, ¿qué puede esperar en términos del promedio de años que llevan con Spence Sprockets los miembros del comité? ¿Cuál es la forma de la distribución de años de experiencia de todos los empleados (la población) en comparación con la forma de la distribución muestral de las medias? Los tiempos de servicio (redondeados al año inmediato) de los 40 empleados que actualmente están en nómina en Spence Sprockets, Inc., son los siguientes:

11	4	18	2	1	2	0	2	2	4
3	4	1	2	2	3	3	19	8	3
7	1	0	2	7	0	4	5	1	14
16	8	9	1	1	2	5	10	2	3

La gráfica 8.3 muestra la distribución de los años de experiencia de la población de 40 empleados actuales. La distribución de tiempos de servicio tiene un sesgo positivo, pues unos cuantos empleados han laborado en Spence Sprockets por un periodo extenso. En específico, seis empleados han laborado en la compañía 10 años o más. Sin embargo, como el negocio creció, el número de empleados se incrementó en los últimos cinco años. De los 40 empleados, 18 han laborado en la compañía dos años o menos.



GRAFICA 8.3 Tiempo de servicio en Spencer Sprockets, Inc., de los empleados.

Considere el primero de los problemas de Ed Spence. A él le gustaría formar un comité con el objeto de que estudien la cuestión del cuidado de cobertura de gastos médicos más adecuada para la mayoría de los trabajadores. ¿Cómo elegiría al comité? Si lo selecciona al azar, ¿qué tiempo medio de servicio de quienes forman parte del comité?

Para comenzar, Ed anota el tiempo de servicio de cada uno de los 40 empleados en papeles y los coloca en una gorra de béisbol. Después los revuelve y selecciona al azar cinco de ellos Los tiempos de servicio de estos cinco empleados son: 1, 90, 19 y 14 años. Por tanto, el tiempo medio de servicio de estos cinco empleados muestreados es de 8.60 años. ¿Cómo se compara este resultado con la media de la población? En este momento, Ed no conoce la media de la población, aunque el número de empleados de la población es de sólo 40, así que decide calcular la media del tiempo de servicio de *todos* sus empleados. Esta es de 4.8 años, que se determina al sumar los tiempos de servicio de todos los empleados y dividir el total entre 40.

$$\mu = \frac{11+4+18+\ldots+2+30}{40} = 4.80$$

La diferencia entre la media de la muestra (X) y la media de la población (μ) recibe el nombre de **error de muestreo.** En otras palabras, la diferencia de 3.80 años entre la media poblacional de 4.80 y la media muestral de 8.60 es el error de muestreo. Éste se debe al azar. Por consiguiente, si Ed selecciona a estos cinco empleados para formar el comité, el tiempo medio de servicio de éstos seria mayor que el de la media de la población.

¿Qué sucedería si Ed colocara de nuevo los papeles en la gorra y tomara otra muestra? ¿Esperaría que la media de esta segunda muestra fuera exactamente la misma que la anterior? Suponga que seleccione otra muestra de cinco empleados y encuentra que los tiempos de servicio de esta muestra son de 7, 4, 4, 1 y 3. La media muestral es de 3.80 años. El resultado de seleccionar 25 muestras de cinco empleados cada una se muestra en la tabla 8.5 y en la gráfica 8.4. En realidad hay 658 008 posibles muestras de 5 tomas de la población de 40 empleados, las cuales se determinan con la fórmula de las combinaciones (5.10) con 40 objetos tomados de 5 en 5. Observe la diferencia de forma de las distribuciones poblacional y muestral de medias. La población de tiempos de servicio de los empleados (gráfica 8.3) tiene un sesgo positivo, y la distribución de estas 25 medias muestrales no refleja el mismo sesgo positivo. También existe una diferencia en el rango de las medias muestrales en comparación con el rango de la población. La población varía de 0 a 19 años, mientras que las medias muestrales varían de 1.6 a 8.6 años.

TABLA 8.5 Veinticinco muestras aleatorias de cinco empleados

Muestra de identificación	Datos de la muestra					Media muestral
A	1	9	0	19	14	8.6
B	7	4	4	1	3	3.8
C	8	19	8	2	1	7.6
D	4	18	2	0	11	7.0
E	4	2	4	7	18	7.0
F	1	2	0	3	2	1.6
G	2	3	2	0	2	1.8
H	11	2	9	2	4	5.6
I	9	0	4	2	7	4.4
J	1	1	1	11	1	3.0
K	2	0	0	10	2	2.8
L	0	2	3	2	16	4.6
M	2	3	1	1	1	1.6
N	3	7	3	4	3	4.0
O	1	2	3	1	4	2.2
P	19	0	1	3	8	6.2
Q	5	1	7	14	9	7.2
R	5	4	2	3	4	3.6
S	14	5	2	2	5	5.6
T	2	1	1	4	7	3.0
U	3	7	1	2	1	2.8
V	0	1	5	1	2	1.8
W	0	3	19	4	2	5.6
X	4	2	3	4	0	2.6
Y	1	1	2	3	2	1.8

GRAFICA 8.4 Histograma de tiempos de servicio medios para 25 muestras de cinco empleados.

La tabla 8.6 contiene los resultados de seleccionar 25 muestras de 20 emplea dos cada una y el cálculo de las medias muestrales. Estas medias muestrales aparecen en la gráfica 8.5. Compare la forma de esta distribución con la población (gráfica 8.3) y con la distribución muestral de medias si la muestra es de $n = 5$ (gráfica 8.4). Observe dos importantes características:

TABLA 8.6 Muestras aleatorias y medias muestrales de 25 muestras de 20 empleados de Spence Sprockets, Inc.

Numero de Muestra	Datos de la muestra (tiempo de servicio)																				Media muestral
A	3	8	3	0	2	1	2	3	11	5	1	3	4	2	7	1	1	2	4	16	3.95
B	2	3	8	2	1	5	2	0	3	1	0	7	1	4	3	11	4	4	3	1	3.25
C	14	5	0	3	2	14	11	9	2	2	1	2	19	1	0	1	4	2	19	8	5.95
D	9	2	1	1	4	10	0	8	4	3	2	1	0	8	1	14	5	10	1	3	4.35
E	18	1	2	2	4	3	2	8	2	1	0	19	4	19	0	1	4	0	3	14	5.35
F	10	4	4	18	3	3	1	0	0	2	2	4	7	10	2	0	3	4	2	1	4.00
G	5	7	11	8	11	18	1	1	16	2	2	16	2	3	2	16	2	2	2	4	6.55

H	3	0	2	0	5	4	5	3	8	3	2	5	1	1	2	9	8	3	16	5	4.25
I	0	0	18	2	1	7	4	1	3	0	3	2	11	7	2	8	5	1	2	3	4.00
J	2	7	2	4	1	3	3	2	5	10	0	1	1	2	9	3	2	19	3	2	4.05
K	7	4	5	3	3	0	18	2	0	4	2	7	2	7	4	2	10	1	1	2	4.20
L	0	3	10	5	9	2	1	4	1	2	1	8	18	1	4	3	3	2	0	4	4.05
M	4	1	2	1	7	3	9	14	8	19	4	4	1	2	0	3	1	2	1	2	4.40
N	3	16	1	2	4	4	4	2	1	5	2	3	5	3	4	7	16	1	11	1	4.75
O	2	19	2	0	2	2	16	2	3	11	9	2	8	0	8	2	7	3	2	2	5.10
P	2	18	16	5	2	2	19	0	1	2	11	4	2	2	1	4	2	0	4	3	5.00
Q	3	2	3	11	10	1	1	5	19	16	7	10	3	1	1	1	2	2	3	1	5.10
R	2	3	1	2	7	4	3	19	9	2	2	1	1	2	2	2	1	8	0	2	3.65
S	2	14	19	1	19	2	8	4	2	2	14	2	8	16	4	7	2	9	0	7	7.10
T	0	1	3	3	2	2	3	1	1	0	3	2	3	5	2	10	14	4	2	0	3.05
U	1	0	1	2	16	1	1	2	5	1	4	1	2	2	2	2	2	8	9	3	3.25
V	1	9	4	4	2	8	7	1	14	18	1	5	10	11	19	0	3	7	2	11	6.85
W	8	1	9	19	3	19	0	5	2	1	5	3	3	4	1	5	3	1	8	7	5.35
X	4	2	0	3	1	16	1	11	3	3	2	18	2	0	1	5	0	7	2	5	4.30
Y	1	2	1	2	0	2	7	2	4	8	19	2	5	3	3	0	19	2	1	18	5.05

1. La forma de la distribución muestral de las medidas es diferente a la de la población. En la gráfica 8.3, la distribución de empleados tiene un sesgo positivo. No obstante, conforme selecciona muestras aleatorias de la población, cambia la forma de la distribución muestral de las medias. A medida que incrementa el tamaño de la muestra, la distribución muestral de las medias se aproxima a la distribución de probabilidad normal. Este hecho se ilustra con el teorema del límite central.

GRÁFICA 8.5 Histograma del tiempo medio de servicio de 25 muestras de 20 empleados

2. Hay menos dispersión en la distribución muestral de las medias que en la distribución de la población. En la población, los periodos de servicio variaron de 0 a 19 años. Cuando seleccionó muestras de tamaño las medias de las muestras variaron de 1.6 a 8.6 años, y cuando seleccionó muestras de 20, las medias variaron de 3.05 a 7.10 años.

También puede comparar la media de las medias de la muestra con la media de la población. La media de las 25 muestras de los 20 empleados de la tabla 8.6 es de 4.676 años.

$$\mu_{\bar{X}} = \frac{3.95 + 3.25 + \cdots + 4.30 + 5.05}{25} = 4.676$$

Emplee el símbolo $\mu_{\bar{X}}$ para identificar la media de la distribución muestral de las medias. El subíndice recuerda que la distribución se refiere a la media muestral. Se lee *mu subíndice X barra*. Observe que la media de las medias muéstrales, 4.676 años, se encuentra muy próxima a la media de la población de 4.80.

¿Qué concluye de este ejemplo? El teorema del límite central indica que, sin importar la forma de la distribución de población, la distribución muestral de la media se aproximará a la distribución de probabilidad normal. Cuanto mayor sea el número de observaciones en cada muestra, más evidente será la convergencia. El ejemplo de Spence Sprockets, Inc., demuestra el mecanismo del teorema del límite central. Comenzó con una población con sesgo positivo (gráfica 8.3). Después seleccionó 25 muestras aleatorias de 5 observaciones; calculó la medida de cada muestra y, por último, organizó las 25 medias de muestra en una gráfica (gráfica 8.4). Observó un cambio en la forma de la distribución muestral de las medias respecto de la propia de la población. El desplazamiento va de una distribución con sesgo positivo a una que tiene la forma de la distribución de probabilidad normal.

Para aclarar más los efectos del teorema del límite central, incremente el número de observaciones en cada muestra de 5 a 20. Seleccione 25 muestras de 20 observaciones cada una y calcule la media de cada muestra. Por último, organice estas medias muéstrales en una gráfica (gráfica 8.5). La forma del histograma de la gráfica 8.5 se desplaza claramente hacia la distribución de probabilidad normal.

En el capítulo 6, la gráfica 6.4 muestra diversas distribuciones binomiales con una proporción de éxitos de 0.1 0,- lo cual es otra demostración del teorema del limite central. Observe que, conforme se incrementa de 7 a 12 y de 20 a 40, el perfil de las distribuciones de probabilidad se desplaza para acercarse cada vez más a una distribución de probabilidad normal. La gráfica 8.5 de la página 279 también muestra la convergencia hacia la normalidad conforme *n* se incremente. Esto confirma de nuevo el hecho de que, conforme se incluyen más observaciones de la muestra de cualquier distribución poblacional, la forma de la distribución muestral de las medias se aproximará cada vez más a la distribución normal.

El teorema del límite central mismo (lea de nuevo la definición de la página 274) no dice nada sobre la dispersión de la distribución muestral de medias ni sobre la comparación entre la media de la distribución muestral de medias y la media de la población. Sin embargo, en el ejemplo de Spence Sprockets hay menor dispersión en la distribución de la media muestral que en la distribución de población, lo que indica la diferencia en el rango de la población y en el rango de las medias muéstrales. Observe que la media de las medias de las muestras se encuentra cerca de la media de la población. Se puede demostrar que la media de la distribución muestral es la media poblacional, es decir, que $\mu_{\bar{X}} = \mu$, y si la desviación estándar de la población es σ, la desviación estándar de las medias muéstrales es $\mu \sqrt{n}$ en la que *n* es el número de observaciones de cada muestra. Entonces, $\sigma \sqrt{n}$ es el **error estándar de la media**. En realidad, el nombre completo es desviación estándar de la distribución muestral de medias.

ERROR ESTÁNDAR DE LA MEDIA $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{N}}$ **[8.1]**

Esta sección permite importantes conclusiones.

1.- La media de la distribución muestral de las medias será exactamente igual a la media poblacional si selecciona todas las muestras posibles del mismo tamaño de una población dada. Es decir,

$$\mu = \mu_{\bar{X}}$$

Aunque no seleccione todas las muestras, es de esperar que la media de la distribución muestral de medias se aproxime a la media poblacional.

2. Habrá menos dispersión en la distribución muestra de las medias que en la población. Si la desviación estándar de la población es σ, la desviación estándar de la distribución muestral demedias es $\sigma / \sqrt{n}$. Note que, cuando se incrementa el tamaño de la muestra, disminuye el error estándar de la media.

<u>Uso de la distribución muestral de las medias</u>

El análisis anterior reviste importancia, pues la mayoría de las decisiones tomadas en los negocios tiene como fundamento los resultados de un muestreo. He aquí algunos ejemplos.

1.- Arm and Hammer Company desea cerciorarse de que su detergente para lavandería contiene realmente 100 onzas líquidas, como indica la etiqueta. Los registros de los procesos de llenado indican que la cantidad media por recipiente es de 100 onzas líquidas y que la desviación estándar es de 2 onzas liquidas. A las diez de la mañana el técnico de calidad realiza la verificación de 40 recipientes y encuentra que la cantidad media por recipiente es de 99.8 onzas liquidas. ¿Debe interrumpir el proceso de llenado o el error de muestreo es razonable?

2. A.C. Nielsen Company proporciona información a las empresas que se anuncian en televisión. Las investigaciones anteriores indican que, en promedio los adultos estadounidenses ven televisión 6.0 horas al día. La desviación estándar es de 1.5 horas. Para una muestra de 50 adultos que viven en el área de Greater de Boston, ¿sería razonable seleccionar al azar una muestra y encontrar que en promedio ven un promedio de 6.5 horas al día?

3. Haughton Elevator Company pretende formular especificaciones relacionadas con el número de personas que pueden desplazarse en un elevador nuevo de gran capacidad. Suponga que el peso medio de un adulto es de 160 libras, y que la desviación estándar es de 15 libras. Ahora bien, la distribución de pesos no sigue una distribución de probabilidad normal. Tiene un sesgo positivo. ¿Cuál es la probabilidad de que, en una muestra de 30 adultos, el peso medio sea de 170 o más libras?

En cada una de estas situaciones hay una población de la cual existe determinada información. Se toma una muestra de esta población y se quiere saber si el error de muestreo, es decir, la diferencia entre el parámetro de población y la muestra estadística, se debe al azar.

De acuerdo con los conceptos analizados en la sección anterior, es posible calcular la probabilidad de que la media de una muestra se encuentre dentro de cierto margen. La distribución de muestreo seguirá la distribución de probabilidad normal con dos condiciones:

1. Cuando se sabe que las muestras se toman de poblaciones regidas por la distribución normal. En este caso, el tamaño de la muestra no constituye un factor.

2. Cuando se desconoce la forma de la distribución de población o se sabe que no es normal, pero la muestra contiene por lo menos 30 observaciones. En este caso, el teorema del límite central garantiza que la distribución muestral de las medias sigue una distribución normal.

Aplique la fórmula (7.5) del capítulo anterior para convertir cualquier distribución normal en una distribución normal estándar. A este hecho también se le denomina valor z. Así, se emplea la tabla estándar normal del apéndice B.1 para determinar la probabilidad de seleccionar una observación que caerá dentro de un intervalo específico. La fórmula para determinar un valor *z* es:

$$z = \frac{x-\mu}{\sigma}$$

En esta fórmula, X es el valor de la variable aleatoria; μ es la media de la población y σ es la desviación estándar de la población.

Sin embargo, la mayor parte de las decisiones de negocios se refiere a una muestra, no a una sola observación. Así, lo importante es la distribución de X, la media muestral, en lugar de X, el valor de una observación. Este es el primer cambio en la fórmula (7.5). El segundo consiste en emplear el error estándar de la media de *n* observaciones en lugar de la desviación estándar de la población. Es decir, se usa $\sigma/\sqrt{n}$ en el denominador en vez de σ. Por consiguiente, para determinar la probabilidad de una media muestral con rango especificado, primero apliqué la fórmula para determinar el valor z correspondiente. Después consulte el apéndice B.1 para localizar la probabilidad.

CALCULO DEL VALOR z DE X CUANDO LA DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN Z= $\frac{X-\mu}{\sigma/\sqrt{n}}$

El siguiente ejemplo muestra la aplicación.

Ejemplo. El departamento de control de calidad de Cola, Inc., conserva registros sobre la cantidad de bebida de cola en su botella gigante. La cantidad real de bebida en cada botella es de primordial importancia, pero varía en una mínima cantidad de botella en botella. Cola, Inc., no desea llenar botellas con menos líquido del debido, pues tendría problemas en lo que se refiere a la confiabilidad de la etiqueta. Por otra parte, no puede colocar líquido de en las botellas porque regalaría bebida, lo cual reduciría sus utilidades. Los registros indican que la cantidad de bebida de cola tiene una distribución de probabilidad normal. La cantidad media por botella es de 31.2 onzas, y la desviación estándar de la población, de 0.4 onzas. Hoy, a las 8 de la mañana, el técnico de calidad selecciono al azar 16 botellas de la línea de llenado. La cantidad mediada bebida en las botellas es de 31.38 onzas. ¿Es un resultado poco probable? ¿Es probable que el proceso permita colocar demasiada bebida en las botellas? En otras palabras, ¿Es poco común el error da muestreo de 0.18 onzas?

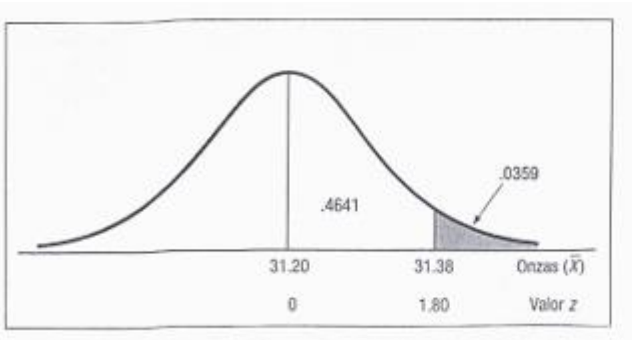
Utilice los resultados de la sección anterior para determinar la probabilidad de seleccionar una muestra de 16 (*n*) botellas de una población normal con una media de 31.2 (μ) onzas y una desviación estándar de la población de 0.4 (σ) onzas, y encontrar que la media muestral es de 31.38 ($\bar{x}$). Aplique la fórmula (8.2) para determinar el valor de z.

$$Z= \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{31.38-31.20}{0.4/\sqrt{16}} \quad =1.80$$

El numerador de esta ecuación, $\bar{x}$ - μ = 31.38-31.20 = .18, es el error muestral. El denominador, $\sigma/\sqrt{n}$ = 0.4 / $\sqrt{16}$ 0.1, es el error estándar de la distribución muestral de la medida. Así, los valores z expresan el error muestral en unidades estándar; en otras palabras, el error estándar.

Después, calcule la probabilidad de un valor z mayor que 1.80. En el apéndice B.1 localice la probabilidad correspondiente a un valor z de 1.80. Este valor es de 0.4641. La probabilidad de un valor z mayor que 1.80 es de 0.0359, que se calcula con la resta 0.5000 - 0.46

¿Qué concluye? No es probable —menos de 4% de probabilidad— que seleccione una muestra de 16 observaciones de una población normal con una mediada 31.2 onzas y una desviación estándar poblacional de 0.4 onzas, y determine que la media de la muestra es igual o mayor que 31.38 onzas. La conclusión es que en el proceso se vierte demasiada bebida de cola en las botellas. El técnico de control de calidad debe entrevistarse con el supervisor de producción para sugerir la reducción de la cantidad de bebida en cada botella. La información se resume en la gráfica 8.6.



GRAFICA 8.6 Distribución muestral de la cantidad media de bebida de cola en una botella gigante

Resumen del capítulo

I. Hay muchas razones para realizar el muestreo de una población.

A. Los resultados de una muestra permiten calcular adecuadamente el valor del parámetro poblacional, con la cual se ahorra tiempo y dinero.

B. Entrar en contacto con todos los miembros de la población consume demasiado tiempo.

C. Resulta imposible verificar y localizar a todas los miembros de la población.

D. El costo de estudiar a todos los elementos de la población resulta prohibitivo.

E. En una prueba con frecuencia se destruye el elemento de la muestra y no se puede regresar a la población.

II. En una muestra sin sesgo, todos los miembros de la población tienen una posibilidad de ser seleccionados para la muestra. Existen diversos métodos de muestreo de probabilidad.

A. En una muestra aleatoria simple, todos los miembros de la población tienen la misma posibilidad de ser seleccionados para la muestra.

B. En una muestra sistemática, se selecciona un punto de partida aleatorio y después se selecciona cada *k*-ésimo elemento subsiguiente de la población para formar la muestra.

C. En una muestra estratificada, la población se divide en varios grupos, a los que se denomina *estratos*, y enseguida se selecciona una muestra aleatoria de cada estrato.

D. En el muestreo por conglomerados, la población se divide en unidades primarias; después se toman las muestras de las unidades primarias.

III. El error de muestreo es la diferencia entre un parámetro poblacional y un estadístico de la muestra.

IV. La distribución muestral de las medias es una distribución de probabilidad de todas las posibles medias muéstrales del mismo tamaño de muestra.

A. Para un tamaño de muestra dado, la media de todas las posibles medias muéstrales tomadas de una población es igual a la media de la población.

B. Existe una menor variación en la distribución de las medias muéstrales que en la distribución de la población.

C. El error estándar de la media mide la variación de la distribución muestral de las medias. El error estándar se calcula de la siguiente manera:

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

D. Si la población se rige por una distribución normal, la distribución muestral de las medias también se regirá por la distribución normal para muestras de cualquier tamaño. Suponga que conoce la desviación estándar de la población. Para determinar la probabilidad de que una media muestral caiga dentro de determinada región, se aplica la fórmula

Clave de pronunciación

SÍMBOLO	SIGNIFICADO	PRONUNCIACIÓN
$\mu_{\bar{x}}$	Media de la distribución muestral de las medias	Mu subíndice X barra
$\sigma_{\bar{x}}$	Error estándar de la población de las medias de las muestras	Sigma subíndice X barra

1.B	HOEL G. Paul y Jessen J. Raymond **Estadística básica para negocios y economía** Editorial C.E.C.S.A, México 1982 Págs. 145 - 171

5

CONTENIDO

1. Introducción
2. Muestreo aleatorio
3. Esperanza
4. Muestreo de una población pequeña
5. La distribución de muestreo de $\bar{x}$
6. La distribución de $\bar{x}$ para una x normal
7. La distribución de $\bar{x}$ para una x no normal

8. **La variancia de muestra para una población pequeña**
9. **Ilustraciones de repaso**

Estadística II

MUESTREO

1. INTRODUCCIÓN

Los capítulos precedentes trataban acerca de la descripción de distribuciones de frecuencia obtenidas en las muestras de alguna población y de la construcción de distribuciones de probabilidad que puedan representar a la población. Se ha supuesto que la distribución de la población puede aproximarse en forma considerable si la muestra es lo suficientemente grande y, por lo tanto, que la distribución de frecuencias de la muestra representa de modo satisfactorio a la población muestreada. Por ejemplo, si un fabricante de camisas toma una muestra grande de medidas del cuello de hombres adultos, supone que tales medidas en la muestra tienen una distribución muy parecida a la distribución de las medidas en la población total de los hombres adultos. Si esto no fuera así, la muestra no serviría de base para las inferencias estadísticas respecto a la población que se muestrea.

El problema de cómo extraer muestras de modo que sean válidas las inferencias estadísticas hechas a partir de los datos suministrados por ellas, se discutió muy brevemente en el Cap. 2. Ahí se dijo que el muestreo llamado aleatorio era aceptable desde este punto de vista. Las propiedades del muestreo aleatorio que le confieren esta cualidad de servir como base para inferencias válidas se discutirá ahora y se darán métodos eficientes para la obtención de muestras aleatorias. Después de estos preliminares, se estudiará el problema general de la dependencia de la exactitud de una muestra respecto al tamaño de la misma. El estudio comprenderá resultados teóricos y varios experimentos de muestreo. El propósito de los experimentos es el de demostrar la variedad de muestreos y hacer más plausibles los resultados teóricos.

2. MUESTREO ALEATORIO

El muestreo aleatorio se definió en el Cap. 2 como un procedimiento de muestreo en el cual todo miembro de una población tiene la misma oportunidad de ser escogido. En términos de probabilidad, esto implica que la probabilidad de que un cierto miembro cualquiera sea escogido es igual a 1/*N*, donde *N* es el número total de individuos de que consta la población. En forma general, si la muestra contiene *r* individuos, en cuyo caso se dice que la muestra es de tamaño *r*, se dice que el muestreo es aleatorio si cualquier combinación de *r* individuos de la población tiene igual oportunidad de resultar elegida.

Aunque el muestreo aleatorio se presentó como aquel que tiene el mérito de producir muestras cuyas distribuciones representan en pequeña escala a la distribución de la población, existen razones más importantes para su aplicación. La más importante es que el muestreo aleatorio conduce a modelos probabilísticos para las distribuciones. Puesto que las conclusiones que deben derivarse de poblaciones por medio de muestras se basan en probabilidades, las muestras deben seleccionarse en tal forma que puedan aplicarse las leyes de probabilidad a ellas. Con objeto de ver por qué esto es así, considérese una vez más el problema discutido en el capítulo anterior referente a un muestreo de 400 votantes. En este punto se suponía que podían tratarse 400 votantes como 400 intentos independientes de un experimento para el cual la probabilidad de éxito en un solo intento es $p = 0.6$. Esta suposición permitía que se tratara el problema como un problema de distribución binomial. Así pues permitía también el cálculo de probabilidad de varios resultados posibles por medio de la distribución binomial y su aproximación normal.

Deben recalcarse las dos características importantes necesarias para permitir la aplicación de la distribución binomial a problemas prácticos como éste; estas características son la independencia de los intentos y la probabilidad constante en todos ellos. Para una población finita de *N* individuos la probabilidad de que un individuo elegido al azar esté a favor de cierto candidato no permanece constante al ir eligiendo los elementos de la muestra; sin embargo, si *N* es un número muy grande comparado con el número de elementos de la muestra la proporción de individuos favorables al candidato varía muy poco al ir tomando la muestra. Así, si las muestras se toman de modo que en cada etapa del muestreo cada miembro restante de la población tenga la misma oportunidad de ser elegido, las condiciones para un modelo binomial se cumplirán satisfactoriamente. Así, el muestreo aleatorio permitirá el uso de un modelo de distribución binomial para problemas tales, siempre que *N* sea grande. Si *N* no es suficientemente grande para justificar este empleo del modelo, se utilizará un modelo probabilístico más sofisticado; sin embargo, ello no se discutirá aquí.

Si se hubiese empleado otro tipo de muestreo para obtener la muestra de 400 votantes, no habría seguridad de que un modelo de distribución binomial fuese válido para calcular las probabilidades. Así, por ejemplo, si un investigador de la encuesta que nos ocupa empezase su recolección de datos al comienzo de una calle residencial y tomara las primeras 400 casas encontradas a lo largo de la calle podría encontrarse con un grupo de nivel económico uniforme, lo que podría implicar que la proporción de votantes a favor del candidato fuese en ese grupo considerablemente mayor o menor que la proporción en la ciudad entera. No se pueden, en general, hacer afirmaciones probabilísticas válidas al respecto de los resultados de otros métodos de muestreo. Por esta razón es que los estadígrafos insisten en que los muestreos sean aleatorios. Por el momento, sólo se discutirá el muestreo aleatorio, pero en un capítulo posterior se introducirán y describirán modificaciones más elaboradas del muestreo aleatorio.

Las muestras aleatorias pueden obtenerse con una técnica de juegos de azar. Así, por ejemplo, si hubiesen 10 000 votantes en una ciudad y se tratara de extraer una muestra de 400, bastaría escribir el nombre de cada votante en una tira de papel, mezclar estas tiras en un recipiente y tomar de ahí 400 tiras, una vez que han sido completamente mezcladas después de cada extracción individual.

Un método menos pesado y más económico de elegir una muestra al azar es el de emplear una tabla de números al azar. Esta tabla puede construirse escribiendo los dígitos de 0 a 9 en diez tiras de papel, mezclarlas completamente entre cada extracción, sustituyendo en cada caso la tira extraída, y tomando el dígito extraído en cada vez. La sucesión resultante de dígitos aleatorios podría usarse para construir una lista de números aleatorios, de, por ejemplo, cinco dígitos cada uno. La tabla II del apéndice contiene tales números aleatorios; sin embargo, éstos fueron obtenidos por un método más refinado que el de tiras de papel. Supóngase ahora que se tenían exactamente 10 000 votantes registrados en el distrito muestreado y que cada votante tiene un número, de 0000 a 9999. Entonces, para obtener una muestra de 400 nombres, sólo sería necesario seleccionar 400 juegos de números de cuatro dígitos de la Tabla II. Puesto que estos números ocurren en grupos de cinco, se pueden seleccionar solamente los primeros cuatro dígitos de un grupo para obtener un nombre. Hay cincuenta renglones de números para cada columna; luego serían suficientes ocho columnas. Por la forma en que se obtienen los números al azar, se infiere que todo número de cuatro dígitos tiene la misma probabilidad de formarse en un lugar específico de la tabla; por lo tanto, se elegirán los números sistemáticamente, leyendo una columna hacia abajo o salteados en la tabla. Podría bien suceder, naturalmente, que uno o más individuos se seleccionaran más de una vez cuando se usan números aleatorios, debido a que estos números se han formado independientemente y, por lo tanto, un número de cuatro dígitos particular puede ocurrir más de una vez en la tabla. En este caso se elegirán unos cuantos números extra para suplir las repeticiones.

Si hubiese 7 500 votantes registrados en la ciudad, entonces se descartaría cualquier número de cuatro dígitos obtenido de la tabla y mayor que 7 499, debido a que no existirían votantes asociados con estos números. Sin embargo, si hubiera sólo 2 000 votantes, no sería adecuado descartar todos los números de cuatro dígitos sobre 1999, porque entonces el ochenta por ciento de esos números serían descartados. En lugar de esto, se podrían asociar los dígitos pares 2, 4, 6, 8 con el dígito 0 y los dígitos nones 3, 5, 7, 9 con el dígito 1 en el primer dígito de los números de cinco dígitos en la Tabla II y entonces tomar los números de cuatro dígitos obtenidos. Con un poco de ingenio es posible usar números aleatorios para escoger una muestra aleatoria de cualquier población finita.

3. ESPERANZA

Como la mayoría de las inferencias estadísticas que se harán acerca de la población por medio de muestras, serán a propósito de la media y de la variancia, es tiempo de empezar el estudio de la confiabilidad en la media de muestra $\bar{x}$ y en la variancia de muestra s^2 como estimaciones de los correspondientes valores de los parámetros de la población μ y σ^2. Este capítulo se dedicará principalmente al estudio de las propiedades de la media de muestra, porque es el más importante de estos dos parámetros y más fácil de tratar que la variancia de muestra.

El problema de determinar la exactitud de una estimación por muestreo, de un parámetro de una población, puede dividirse en dos partes. Una es determinar si la estimación está sesgada. Si por ejemplo se estima la estatura de los hombres adultos por medio de una muestra tomada entre estudiantes universitarios, el valor de $\bar{x}$ tendería a sobrepasar el valor de μ, por que los estudiantes universitarios tienden a ser más altos que el resto de la población adulta de mayor edad. Una estimación que tiende a ser una subestimación o una sobreestimación diremos que está sesgada. Un poco más adelante daremos una definición más precisa. La segunda parte del problema es la de determinar la precisión de la estimación. Si una estimación no está sesgada, entonces precisión y exactitud significan lo mismo; pero si la estimación está sesgada, entonces puede muy bien ser bastante precisa y, por ser sesgada, será al mismo tiempo inexacta. La precisión se refiere al hecho de que muestreos repetidos dan valores muy cercanos entre sí.

El problema de determinar si una estimación está sesgada o no, se resuelve fácilmente por medio de un concepto muy importante en probabilidad llamado el valor esperado o esperanza, de una variable aleatoria. En relación a esto, considérese un juego en el cual se lanzan tres monedas y se recibe un dólar por cada águila que caiga. ¿Cuánto dinero se esperaría obtener si se pudiera jugar este juego una vez? De los resultados logrados en el capítulo anterior, las probabilidades asociadas con conseguir 0, 1, 2 y 3 águilas al lanzar tres monedas son $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}\ y\ \frac{1}{8}$, respectivamente. De manera intuitiva, se esperaría obtener 0 dólares $\frac{1}{8}$ del tiempo, 1 dólar $\frac{3}{8}$ del tiempo, 2 dólares $\frac{3}{8}$ del tiempo y 3 dólares $\frac{1}{8}$ del tiempo, si se pudiera jugar un número suficiente de veces. Por tanto, se esperaría obtener la cantidad

$$0\cdot\tfrac{1}{8}+1\cdot\tfrac{3}{8}+2\cdot\tfrac{3}{8}+3\cdot\tfrac{1}{8}=1\tfrac{1}{2}.$$

Esta cantidad ($1.50) es lo que comúnmente se llama la cantidad esperada que se ganaría si se jugara una vez. Este ejemplo es una ilustración del concepto general del valor esperado que sigue.

Supóngase que la variable aleatoria x debe tomar uno de los valores $x_1, x_2, .., x_k$ y que las probabilidades asociadas a esos valores son $P\{x_1\}, P\{x_2\}, ..., P\{x_k\}$, donde $\sum_{i=1}^{k} P\{x_i\} = 1$. Entonces, el *valor esperado* de esta variable aleatoria se define como la cantidad

(1) $$\sum[x] = \sum_{i=1}^{k} x_i P\{x_i\}.$$

En la ilustración anterior la variable aleatoria x era la cantidad de dinero que se ganaría al lanzar tres monedas, los valores posibles eran 0,1, 2 y 3, y las probabilidades correspondientes eran $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}\ y\ \frac{1}{8}$.

Debe hacerse notar que el valor esperado de una variable aleatoria no es necesariamente un valor posible para esa variable aleatoria. Por tanto, no hay nada paradójico en afirmar que el valor esperado del número total de accidentes automovilísticos que un conductor de automóvil tiene un periodo de cinco años es, por ejemplo, de 2.3. El valor esperado representa meramente un promedio, y comparando (1) con la fórmula (2) del Cap. 4, se ve que $E[x]$ no es otra cosa sino la media μ de la distribución de probabilidad de la variable aleatoria x. Ya que el valor esperado de una variable aleatoria es lo mismo que su media, se podría pensar que no tiene ningún caso introducir una nueva terminología; sin embargo, la esperanza va un paso más adelante. Suponga que el juego anterior se altera de manera que se gane $g(x_i)$ en lugar de x_i cuando se obtiene un valor de x_i. Por ejemplo, $g(x)$ puede ser la función $g(x)=x^2$. Entonces el valor esperado del juego estaría dado por la fórmula

(2) $$E[g(x)]=\sum_{i=1}^{k} g(x_i)P\{x_i\}.$$

De este modo, es posible hablar del valor esperado de cualquier función de una variable aleatoria en lugar de pensar en términos de la variable aleatoria en sí. Como un ejemplo, sea $g(x)=x^2$ en el ejemplo anterior, entonces

$$E[g(x)]=0\cdot\tfrac{1}{8}+1\cdot\tfrac{3}{8}+4\cdot\tfrac{3}{8}+9\cdot\tfrac{1}{8}=3.$$

Así, se podría esperar obtener $3 si las ganancias son los cuadrados del número de águilas obtenidas y se puede jugar una sola vez.

De la definición (2) se sigue que el operador valor esperado E tiene las siguientes propiedades:

(3) $$E[g(x)+c]=E[g(x)]+c$$

y

(4) $$E[cg(x)]=cE[g(x)].$$

La primera de estas propiedades se puede comprobar si se recuerda que $\sum_{i=1}^{k} P\{x_i\}=1$ y aplicando (2) para obtener

$$\begin{aligned}\sum_{i=1}^{k}[g(x_i)+c]P\{x_i\}&=\sum_{i=1}^{k} g(x_i)P\{x_i\}+c\sum_{i=1}^{k} P\{x_i\}\\&=E[g(x)]+c.\end{aligned}$$

La segunda propiedad se comprueba de manera similar. Así

$$\sum_{i=1}^{k}[cg(x_i)]P\{x_i\}=c\sum_{i=1}^{k} g(x_i)P\{x_i\}=cE[g(x)].$$

Una tercera propiedad que requiere una cantidad considerablemente mayor de antecedentes de probabilidad es el hecho de que el valor esperado de la suma de dos variables aleatorias es igual a la suma de sus valores esperados. Esto es cierto si las dos variables aleatorias corresponden a experimentos independientes o están íntimamente relacionadas en el mismo experimento. Así, suponga que se cambia el juego anterior permitiendo que se juegue por una bonificación de cinco dólares si se obtienen tres águilas. El premio será ganado si se lanza una sola moneda y sale una águila. Si x denota en total a ganar en el juego normal y y la ganancia por bonificación, entonces el total que se puede esperar está dado por $E[x+y]=E[x]+E[y]$. De una manera más general, para cualesquier dos variables aleatorias x y y y dos funciones cualesquiera de esas dos variables, es cierto que

(5) $$E[g(x)+h(y)]=E[g(x)]+E[h(y)]$$

En el ejemplo anterior los cálculos serían los siguientes:

$$E[x]=0\cdot\tfrac{1}{8}+1\cdot\tfrac{3}{8}+2\cdot\tfrac{3}{8}+3\cdot\tfrac{1}{8}=\tfrac{3}{2}$$
$$E[y]=5\cdot\tfrac{1}{8}\cdot\tfrac{1}{2}=\tfrac{5}{16}.$$

Por tanto,

$$E[x+y]=\tfrac{3}{2}+\tfrac{5}{16}=\tfrac{29}{16}.$$

Si se le pidiera a un individuo que pagara por jugar este último juego, no debería pagar más de 29/16 dólares por jugar una vez. El operador del juego sería bastante tonto si cobrara menos de 29/16 dólares por juego. Si el precio para jugar fuera exactamente de 29/16 dólares, lo cual no se puede pagar en efectivo real, entonces este juego sería un *juego justo.*

A pesar de que la discusión anterior ha sido orientada hacia juegos y la cantidad de dinero a ganar, la aplicación del operador del valor esperado en problemas de muestreo se dirigirá hacia la determinación de si una estimación está o no sesgada. Se usará también en relación a algunos de los problemas de toma de decisiones de un capítulo posterior.

4. MUESTREO DE UNA POBLACIÓN PEQUEÑA

Suponga que una población consta de N individuos y que se debe escoger una muestra aleatoria de tamaño n de ella. Sea X la variable aleatoria asociada con esta población y sean $X_1, X_2, ..., X_N$ los valores de esta variable aleatoria para los miembros de esta población. Entonces, la media de la distribución de la población está dada por

(6) $$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i.$$

Esto es lo mismo que la Fórmula (5) del Cap. 2, sustituyendo $\bar{x}$ por μ.

Sean $x_1, x_2, ..., x_n$ los valores de esta variable aleatoria correspondiente a una muestra aleatoria de tamaño n. Entonces la media de esta muestra está dada por

(7) $$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j.$$

Como primer paso en la determinación de la exactitud de $\bar{x}$ como una estimación de μ, se calculará el valor esperado de $\bar{x}$. De las propiedades (4) y (5) se sigue que

(8) $$E[\bar{x}] = \frac{1}{n}\sum_{j=1}^{n} E[x_j].$$

Para poder evaluar esta suma es necesario calcular el valor de $E[x_j]$ con $j = 1,2,...,n$. Considérese primero el cálculo de $E[x_1]$ donde x_1 es el primer valor de muestra a ser obtenido. Ya que x_1 puede asumir cualquiera de los N valores de la población $X_1, X_2, ..., X_N$ y que bajo el muestreo aleatorio cada uno de estos valores tiene la misma probabilidad, a saber $1/N$, de ser seleccionado, se sigue de la definición (1) y de la fórmula (6) que

$$E[x_1] = \sum_{i=1}^{N} X_i \frac{1}{N} = \mu.$$

A continuación, considérese el cálculo de $E[x_2]$, donde x_2 es el segundo valor de muestra a ser obtenido. La variable x_2 también puede tomar cualquiera de los valores $X_1, X_2, ..., X_N$, pero sólo cuando esté coordinada con el valor ya tomado por x_1. Así, x_2 puede tomar el valor de X_1 sólo si x_1 no lo ha hecho. La probabilidad de que x_2 tome X_1 está, por tanto, dada multiplicando la probabilidad de que x_1 no lo tomara por la probabilidad condicional de que x_2 lo hiciera si x_1 no lo tomó. En el muestreo aleatorio, donde cada miembro de la población tiene la misma oportunidad de ser escogido en cada fase, esta probabilidad está dada por

$$\frac{N-1}{N}\cdot\frac{1}{N-1} = \frac{1}{N}.$$

Esta misma probabilidad puede ser aplicada a cada uno de los valores restantes $X_2, ..., X_N$; por lo tanto, de la definición (1) se sigue que

$$E[x_2] = \sum_{i=1}^{N} X_i \frac{N-1}{N}\frac{1}{N-1} = \mu.$$

Para que x_3 tome el valor de X_1 es necesario que x_1 y x_2 no lo tomen, por tanto, con el mismo razonamiento anterior, se sigue que

$$E[x_3] = \sum_{i=1}^{N} X_i \frac{N-1}{N}\frac{N-2}{N-1}\frac{1}{N-2} = \mu.$$

Una continuación de este tipo de razonamiento dará por resultado que $E[x_j] = \mu$ para cada valor de j y, por tanto, de (8) que

(9) $$E[\bar{x}] = \frac{1}{n}\sum_{j=1}^{n} \mu = \mu.$$

Una estimación de un parámetro que tenga la propiedad de que su valor esperado sea igual al valor del parámetro se dice que es una *estimación insesgada* del parámetro. En vista del resultado dado por (9) se sigue que la media de la muestra $\bar{x}$ es una estimación insesgada de la media μ de la población. Está propiedad no requiere ningún conocimiento de la distribución de la población y es válida para cualquier muestra de cualquier tamaño y cualquier tamaño de población, siempre que $n \le N$. Todo lo que se ha usado en esta derivación es muestreo aleatorio y sus propiedades.

Considérese ahora el problema de si la variancia s^2 es una estimación insesgada de la variancia de población. En el capítulo anterior se definió la variancia de población sustituyendo la frecuencia relativa de muestra f_i/n que correspondía al valor observado x_i por la probabilidad $P(x_i)$. Esto suponía que el tamaño de la muestra podría agrandarse tanto como se quisiera y que f_i/n se aproximaría al valor $P(x_i)$ a medida que n fuera cada vez mayor. En una población pequeña, sin embargo, no es posible usar un valor límite al crecer n para definir la variancia de la población. Si $X_1, X_2, ..., X_N$ denotan los valores de la variable aleatoria X para los N miembros de la población, la variancia de población que será denotada por S^2, puede ser definida por la fórmula

$$S^2 = \sum_{i=1}^{N} \frac{(X_i - \mu)^2}{N-1} \tag{10}$$

Esta definición para población finita corresponde a la definición de la variancia de muestra s^2, al igual que μ dada por (6) corresponde a $\bar{x}$ dada por (7). Si se extrae una muestra de tamaño n de esta población, donde $n < N$, se puede mostrar, usando las propiedades del operador E del valor esperado, que

$$E[s^2] = S^2.$$

Así, la variancia de muestra s^2 es una estimación insesgada de la variancia de población S^2. Esta propiedad de s^2 es una de las razones por las cuales se dividió $\sum (x_i - \bar{x})^2 f_i$ entre $n-1$, en lugar de entre n, para definir la variancia de muestra. Esta propiedad es cierta para todos los tamaños de muestra y de población. S^2 difiere un poco de σ^2 porque si cada miembro de la población tuviera un valor diferente de la variable aleatoria X de manera que no hubiera repeticiones en los valores $X_1, X_2, ..., X_N$, entonces la definición de σ^2 dada por la fórmula (4) en el Cap. 4 estaría reducida a

$$\sigma^2 = \sum_{i=1}^{N} (X_i - \mu)^2 P\{X_i\}.$$

Pero si se supone que cada miembro de la población tiene asignada la misma probabilidad, $P\{X_i\} = 1/N$; así, de la Fórmula (10) se infiere que para una población finita de tamaño N

$$\sigma^2 = \frac{N-1}{N} s^2.$$

Cuando N es muy grande, el factor corrector $(N-1)/N$ es trivial y puede ser ignorado. Aún más, en experimentos de tipo binomial que incluyen realizaciones independientes repetidas, la población puede se0r considerada de tamaño infinito. De aquí en adelante se ignorará la diferencia entre S^2 y σ^2 que aparece en poblaciones pequeñas y se denotará a la variancia de población siempre por σ^2. En general, se puede decir que rara vez los problemas tratan con poblaciones tan pequeñas que valga la pena considerar la diferencia.

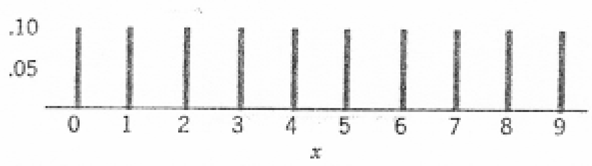


FIG. 1. Distribución de probabilidad uniforme sobre los dígitos

5. DISTRIBUCION DE MUESTREO DE $\bar{x}$

En las secciones precedentes se ha completado un aspecto de la determinación de la exactitud de $\bar{x}$ y s^2 cómo estimaciones de μ y σ^2 llegando al resultado de que tanto $\bar{x}$ como s^2 son estimaciones insesgadas de los valores de los correspondientes parámetros de población. La segunda parte de la investigación es determinar la precisión de esas estimaciones. En vista de que esto implica determinar cómo varían las estimaciones de las muestras alrededor del valor del parámetro estimado, es necesario estudiar la distribución de tales estimaciones. En este capítulo el estudio se restringirá al de la distribución de $\bar{x}$ bajo varias hipótesis en relación a la distribución de la variable aleatoria x.

La primera variable aleatoria que será estudiada es la variable x que toma los valores 0, 1, 2,..., 9 con igual probabilidad. La distribución de x está mostrada geométricamente en la Fig. 1. En vista de que el muestreo aleatorio de esta distribución, que recibe el nombre de distribución discreta uniforme, corresponde a muestrear dígitos de la tabla de los números aleatorios dada en la Tabla II, es fácil escoger muestras aleatorias para este caso.

Primero, se tomarán 25 muestras de tamaño $n = 1$ para observar la variabilidad del muestreo que se da en ese mínimo tamaño posible de la muestra. Esto se hará empezando con el primer dígito registrado en la Tabla II y procediendo hacia debajo de la primera columna hasta que se tengan 25 dígitos. Este muestreo produjo los siguientes valores.

Número de la muestra	x	Número de la muestra	x	Número de la muestra	x	Número de la muestra	x	Número de la muestra	x
1	0	6	2	11	8	16	9	21	3
2	3	7	6	12	7	17	7	22	4
3	1	8	3	13	2	18	5	23	5
4	3	9	0	14	8	19	9	24	1
5	6	10	4	15	3	20	1	25	6

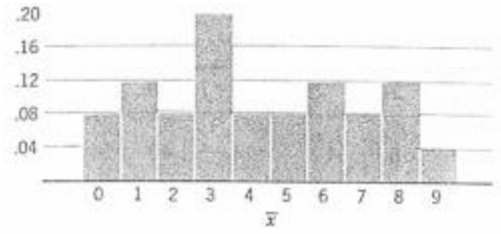


FIG. 2. Distribución de muestreo de $\bar{x}$ con $n = 1$

Contando las frecuencias f para cada valor de x y calculando las frecuencias relativas dividiendo entre 25 se tiene la siguiente tabla:

x_i	0	1	2	3	4	5	6	7	8	9
f_i	2	3	2	5	2	2	3	2	3	1
$\frac{f_i}{25}$	.08	.12	.08	.20	.08	.08	.12	.08	.12	.04

Esta distribución de muestreo está formada geométricamente en la Fig. 2 en la forma de un histograma. El eje de las x está marcado x porque la distribución de $\bar{x}$ es la que está siendo estudiada y las observaciones individuales pueden ser tratadas como medias basadas en muestras de tamaño uno cada una.

Para poder observar cómo cambia la distribución de $\bar{x}$ a medida que n aumenta, se efectuó el siguiente muestreo y se escogió $n = 5$. Esto se hizo calculando las medias de conjuntos consecutivos de cinco dígitos empezando con la última columna de tales conjuntos en la Tabla II y siguiendo hasta que se habían obtenido 25 medias. Este muestreo produjo los siguientes valores:

Número de la muestra	$\bar{x}$	Número de la muestra	$\bar{x}$	Número de la muestra	$\bar{x}$	Número de la muestra	$\bar{x}$	Número de la muestra	$\bar{x}$
1	4.8	6	5.0	11	4.2	16	5.2	21	5.6
2	4.6	7	3.8	12	4.8	17	3.2	22	3.4
3	2.8	8	5.0	13	5.2	18	5.6	23	3.6
4	2.0	9	5.0	14	2.0	19	4.4	24	3.8
5	6.2	10	1.4	15	6.2	20	4.6	25	4.4

Los resultados fueron clasificados en la siguiente tabla de frecuencia:

$\bar{x}_i$	0	1	2	3	4	5	6	7	8	9
f_i	0	1	2	3	6	10	3	0	0	0
$\frac{fi}{25}$	0	0.04	.08	.12	.24	.40	.12	0	0	0

El histograma de esta distribución está mostrado en la Fig. 3.

Finalmente, se llevó a cabo un experimento de muestreo escogiendo n igual a 10 promediando conjuntos de diez dígitos en la segunda columna de la Tabla II hasta que se obtuvieron 25 medias. Los resultados de este muestreo produjeron la siguiente tabla de frecuencia de valores clasificados en donde los valores que caían en los límites de dos intervalos fueron divididos entre los intervalos vecinos.

$\bar{x}_i$	0	1	2	3	4	5	6	7	8	9
f_i	0	0	0	4	10	8	3	0	0	0

$\frac{fi}{25}$	0	0	0	.16	.40	.32	.12	0	0	0

El histograma de esta distribución está mostrado en la Fig. 4.

Los experimentos anteriores dan una indicación de cuánta variación hay en una media de muestra $\bar{x}$ como función del tamaño de muestra n usada para determinar la media. Cuando $n = 10$ las medias de las muestras se concentran especialmente alrededor de la media de la población μ, que en la Fig. 1 se ve tiene el valor 4.5.

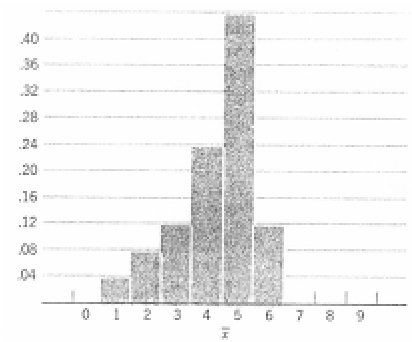


FIG. 3. Distribución de muestreo de x con $n = 5$

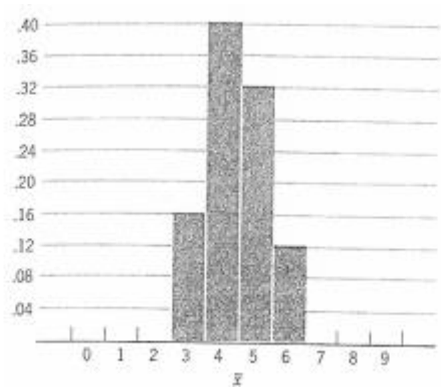


FIG. 4. Distribución de muestra de $\bar{x}$ con $n = 10$

Dado que la variancia o la desviación estándar se usan para medir la variabilidad, bastará con estudiar la variancia de estas distribuciones de muestreo. Los cálculos basados en las tres tablas anteriores de distribución de frecuencia correspondientes a los tres experimentos de muestreo dieron los valores empíricos $s_1{}^2 = 7.6, s_5{}^2 = 1.7$ y $s_{10}{}^2 = 9$, respectivamente. A pesar de que estas variancias dan una indicación del grado de concentración de las medias de la muestra en relación a la media de la población, están aún muy sujetas a la variabilidad del muestro para ser muy útiles; por tanto, lo que se necesita es un conocimiento de los valores a los que estas variancias se aproximarían si se continuaran los experimentos de muestreo indefinidamente en cada caso en lugar de interrumpir el experimento después que se obtienen 25 medias. La respuesta a esta pregunta está dada por una fórmula teórica de la variancia de $\bar{x}$, que puede ser obtenida mediante el uso adecuado del operador del valor esperado y sus propiedades. En vista de que la derivación es algo complicada, sólo presentamos el resultado final, a saber, la fórmula

(11)
$$\sigma_{\bar{x}}{}^2 = \frac{\sigma^2}{n}.$$

Esta es la fórmula deseada que mide la variabilidad de las medias de las muestras en función del tamaño n de la muestra. Es válida para el muestreo aleatorio del tipo que se ha usado en estos experimentos, el cual puede continuar indefinidamente y, por tanto, corresponde a tratar la población como si fuera infinita. Se necesita modificar esta fórmula si el muestreo se toma de una población finita pequeña. Esta modificación será presentada en una sección posterior.

Según la fórmula (11), las variancias de las tres distribuciones que corresponden a los tres experimentos anteriores de muestreo deben estar en la proporción 1:1/5:1/10, porque los valores de n en esos experimentos fueron 1, 5 y 10, respectivamente. En vista de que esas distribuciones, tal y como están representadas por las Figs. 2, 3 y 4, son sólo estimaciones de muestra de las distribuciones teóricas correspondientes, basándose solamente en muestras de 25, es poco razonable esperar que las variancias de las muestras de las tres distribuciones satisfagan estas proporciones teóricas con mucha exactitud. Sin embargo, las variancias de las muestras de $s_1{}^2 = 7.6, s_5{}^2 = 1.7$, y $s_{10}{}^2 = .9$ obtenidas para esas tres distribuciones, y que están en las razones aproximadas 1:22: .12, sí se aproximan a las razones teóricas 1: .20: .10. Si los experimentos de muestreo hubieran estado basados en, por ejemplo, 100 valores de la muestra de $\bar{x}$ cada uno en lugar de 25, las tres variancias de muestra se hubieran conformado con mucha mayor exactitud a las razones teóricas.*

6. DISTRIBUCIÓN DE $\bar{x}$ PARA UNA x NORMAL

La sección anterior estuvo dedicada al estudio de la variabilidad de $\bar{x}$ en el muestreo de una distribución uniforme sobre los dígitos 0, 1,…,9. En esta sección se adoptará un punto de vista más realista al muestrear a partir de una distribución normal. En vista de que muchas variables aleatorias de la vida real tienen distribuciones normales, al menos aproximadamente, los resultados basados en el muestreo de una distribución normal deben ser muy útiles.

Cualquier método para medir es de exactitud limitada; por tanto, sólo es posible construir distribuciones aproximadas para variables continuas. Para simplificar el experimento, se usará una aproximación un poco burda a la distribución de la variable normal estándar. Esta aproximación fue obtenida usando la Tabla IV para calcular el porcentaje del área bajo la curva normal mostrado en la Fig. 12 del Cap. 4, que corresponde a intervalos de una unidad centrando el intervalo medio en el origen. Los porcentajes resultantes escritos como fracciones decimales están mostrados en la Fig. 5, donde se ve el histograma de esta distribución normal aproximada.

Incidentalmente, los porcentajes obtenidos de la Fig. 5 al combinar los dos porcentajes últimos a cada lado, dan los porcentajes que son usados frecuentemente por los instructores que "califican con una curva" para determinar el porcentaje de calificaciones A, B, C, D y F.

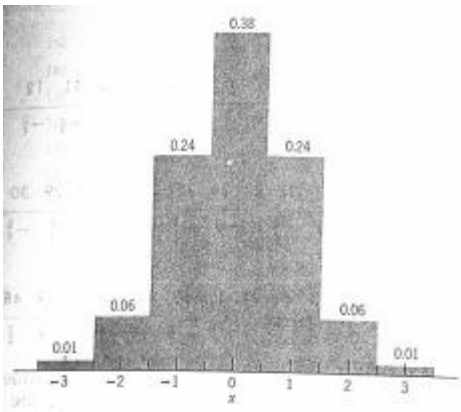


FIG. 5. Una aproximación a la distribución normal estándar

Se tomaron muestras de tamaño 4 de esta distribución discreta por medio de los números aleatorios encontrados en la Tabla 2. Primero, se construyó una forma de tabulación del tipo mostrado en la Tabla 1. Por este procedimiento, todos los pares de dígitos aleatorios fueron divididos en siete grupos, según las proporciones mostradas en la Fig. 5, y asociados con las marcas de clase de la Fig. 5. Por ejemplo, $x=-3$ se asigna al par 00, que es uno por ciento de todos los pares, y el valor $x=-2$ se asigna a los pares desde 01 a 06 inclusive, lo cual incluye el 6% de todos los pares de números aleatorios. Cuatro de esos pares de números aleatorios se leen en la Tabla II y se registran en el intervalo de clase apropiado para formar un experimento. Los resultados de los tres primeros experimentos de este tipo son mostrados en la Tabla 1. Este experimento fue repetido 50 veces. A continuación se calculó la media para cada experimento y se registró en la última fila de la Tabla 1, marcada $\bar{x}$. Las medias de la muestra de los 50 experimentos que fueron llevados a cabo son mostrados en la Tabla 2.

TABLA 1

x	Números aleatorios	1	2	3	…
-3	00				
-2	01-06		/	/	
-1	07-30	/		///	
0	31-68	//	/		
1	69-92	/	/		
2	93-98		/		
3	99				
$\bar{x}$		0	$\frac{1}{4}$	$-\frac{2}{4}$	…

* Una razón es un cociente. Decir que a, b, c están en la proporción 3:4:5 es decir que a/b=3/4 y b/c=4/5. (N. del T.)

TABLA 2

Núm. del experimento	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Valor de $\bar{x}$	0	$\frac{1}{4}$	$-\frac{2}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{2}{4}$	0	$\frac{1}{4}$	$-\frac{3}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	$-\frac{2}{4}$	$-\frac{2}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$	0	$-\frac{2}{4}$	0

19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
$-\frac{2}{4}$	$-\frac{2}{4}$	0	$-\frac{1}{4}$	$\frac{1}{4}$	0	$-\frac{1}{4}$	$\frac{3}{4}$	$\frac{2}{4}$	$-\frac{2}{4}$	$\frac{1}{4}$	$-\frac{2}{4}$	$-\frac{1}{4}$	0	$\frac{2}{4}$	0	$-\frac{1}{4}$	$\frac{2}{4}$	0	0	$-\frac{5}{4}$

40	41	42	43	44	45	46	47	48	49	50
$\frac{1}{4}$	0	$\frac{3}{4}$	$-\frac{3}{4}$	$\frac{5}{4}$	$-\frac{4}{4}$	0	0	$\frac{2}{4}$	0	$\frac{1}{4}$

Luego se tabularon los valores de $\bar{x}$ para obtener la tabla de frecuencias mostradas en la Tabla 3, en la cual el tercer renglón da en forma decimal los porcentajes de los correspondientes frecuencias absolutas.

Finalmente, con esos datos se trazó un histograma. Como la suma total de las áreas debe ser igual a uno, tal como sucede en todas las distribuciones de probabilidad, y como en esa tabla el intervalo de clase es de $\frac{1}{4}$, es necesario dibujar rectángulos cuyas alturas sean cuatro veces el valor respectivo dado $f/50$. El histograma que resulta, cuya área es la unidad, se muestra en la Fig. 6.

Comparando las Figs. 5 y 6, es evidente que las medias de muestras basadas en 4 mediciones, varían cada una aproximadamente la mitad de los que sucede en el caso de los valores de muestras de un elemento. Además, se ve que la distribución de $\bar{x}$ tiene una media cercana a 0, que es la media de la distribución de x. Por último, exceptuando a una irregularidad más o menos marcada en el intervalo centrado en $-\frac{1}{4}$, puede verse que la distribución de $\bar{x}$, excepto por la diferencia de dispersión, posee una distribución del mismo tipo aproximadamente normal que la distribución de x.

TABLA 3

$\bar{x}$	$-\frac{5}{4}$	$-\frac{4}{4}$	$-\frac{3}{4}$	$-\frac{2}{4}$	$-\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{5}{4}$
f	1	1	3	8	5	14	9	6	2	0	1
$f/50$	.02	.02	.06	.16	.10	.28	.18	.12	.04	.00	.02

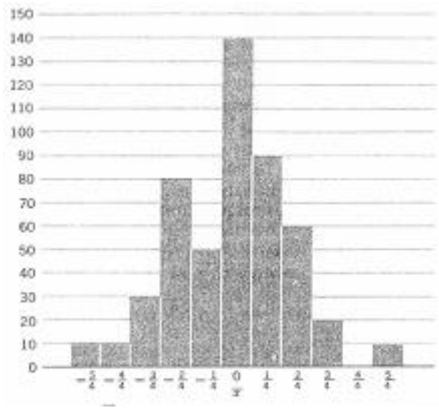


FIG. 6. Distribución de $\bar{x}$ para muestras de tamaño 4 de la distribución de la Fig. 5

Los cálculos de la media y la variancia de la distribución de $\bar{x}$ fueron hechos por medio de los valores de la Tabla 3. Estos cálculos dieron los valores -.035 y .23, respectivamente. De este modo, la media de la distribución $\bar{x}$ está muy cerca de la media de la distribución de x, en tanto que la variancia de la distribución de $\bar{x}$ es aproximadamente un cuarto de la variancia de la distribución de x, porque está última distribución se aproxima a una distribución normal con $\sigma = 1$. Estos valores están bastante cercanos a los valores esperados en las fórmulas (9) y (11).

Si se hubiera efectuado este experimento, digamos, 500 veces en vez de 50, habrían desaparecido las irregularidades tales como la que apareció en la Fig. 6 y se habrían hecho más patentes las propiedades que se discutieron respecto a la distribución de $\bar{x}$. En este caso, se habría visto que el histograma se podría ajustar perfectamente a una curva normal, que la media de la distribución de $\bar{x}$ se aproxima a cero y que el valor de la desviación típica de la distribución de $\bar{x}$ se aproxima mucho a un cuarto del valor de la variancia correspondiente de la distribución de x. Aún cuando, en este ejemplo, las muestras fueron tomadas de una distribución aproximadamente normal, tal como se muestra en la Fig. 5, en vez de hacerlo de una distribución exactamente normal, los resultados obtenidos son similares a los que se hubieran obtenido si la aproximación se hubiera mejorado, escogiendo un intervalo de clase muy pequeño.

Afortunadamente, no es necesario llevar a cabo experimentos repetidos de muestreo para llegar a la distribución de frecuencia teórica para $\bar{x}$. Usando las reglas de probabilidad y métodos matemáticos avanzados es posible derivar la ecuación de la curva que representa la distribución $\bar{x}$ cuando el muestreo es de una distribución exactamente normal, y no de una aproximación. Esto corresponde a lo que se ha hecho en el Cap. 4, para llegar a la distribución teórica para x binomial sin efectuar experimentos de muestreo. Resulta que $\bar{x}$ poseerá una distribución normal si x la posee, con la misma media que x, pero con una desviación estándar que es $1/\sqrt{n}$ veces la desviación estándar de x. Estos resultados matemáticos se expresan en la forma de un teorema.

(12) ***Teorema 1.*** *Si x posee una distribución normal cuya media es μ y cuya desviación estándar es σ, entonces la media de muestra $\bar{x}$ basada en muestras al azar de tamaño n, poseerá también una distribución normal cuya media será μ y cuya desviación estándar será $\sigma/\sqrt{n}$.*

La distribución $\bar{x}$ dada por este teorema frecuentemente recibe el nombre de distribución de muestreo de $\bar{x}$ debido a su conexión con experimentos repetidos de muestreo, aun cuando se deriva por métodos puramente matemáticos.

Este teorema es un teorema puramente matemático de distribuciones ideales que corresponden a curvas suaves: sin embargo, se puede esperar que las conclusiones sean válidas para poblaciones verdaderas, siempre y cuando la población sea grande y que el histograma de la población se pueda ajustar a una curva normal. Esto quiere decir que si se empieza con un número grande de muestras de tamaño n y se calcula el valor de $\bar{x}$ para cada muestra, entonces el histograma de los valores $\bar{x}$ se ajustará a la curva normal especificada en el teorema.

Los resultados del experimento de muestreo recién descrito están de acuerdo con este teorema. El histograma de la Fig. 6 presenta el aspecto típico de los histogramas que se obtienen al muestrear poblaciones normales y puesto que en este ejemplo $\sigma \doteq 1$ y $n = 4$ su media, -.035, y su desviación estándar $\sqrt{.23} = .48$, concuerdan con los valores teóricos de 0 y $\sigma/\sqrt{4} \doteq {}^{1}/_{2}$ dados por el teorema. Según este teorema, se puede concluir que las medias de las muestras de tamaño 4 obtenidas a partir de una población normal, poseen una dispersión con respecto a la media, cuyo valor es de solamente la mitad del valor que presenta la dispersión de las medias individuales de la población con respecto a su media, si medimos la dispersión por medio de la desviación estándar.

7. DISTRIBUCIÓN DE $\bar{x}$ PARA UNA x NO NORMAL

Supóngase ahora que la variable x no posee una distribución normal. En este caso, ¿qué puede decirse acerca de la distribución de $\bar{x}$? En vista de que las fórmulas (9) y (11) no dependían de la naturaleza de x, se sigue que la distribución de $\bar{x}$ tendrá una media μ y distribución estándar $\sigma/\sqrt{n}$. La única cuestión que queda entonces por resolver es la determinación de la naturaleza de la distribución. De los resultados del primer experimento de muestreo en donde x tenía una distribución uniforme mostrada en la Fig. 1, se puede observar que las distribuciones de $\bar{x}$ correspondientes a $n = 5$ y $n = 10$ también poseen distribuciones aproximadamente normales. Si se hubiera obtenido un conjunto de, por ejemplo, 250 en lugar de sólo 25 medias de muestras para construir las Figs. 3 y 4, los histogramas serían sin lugar a dudas mucho más regulares y simétricos, y se podría observar que las distribuciones son aproximadamente normales. Muchos estadígrafos han desarrollado experimentos de muestreo con diferentes clases de distribuciones no normales para x, para ver el efecto que tendría la falta de normalidad sobre la distribución de $\bar{x}$. El resultado sorprendente ha sido que si n es mayor que 25 la distribución de $\bar{x}$ parecerá normal, independientemente de la distribución de población que se elija para x. Hace algunos años un profesor de estadística avanzada retó a sus alumnos a que construyeran una distribución tan no normal como pudieran y les apostó que si él tomaba muestras de tamaño 25 de su población, la distribución resultante de $\bar{x}$ aparecería como una distribución normal. Una vez sentadas las bases del concurso se llevó a cabo el experimento y el profesor cenó esa noche a expensas de sus alumnos. Si la distribución de x no es muy distinta de la normal, la distribución de $\bar{x}$ aparecerá normal para valores de n no mayores que 5. Esta propiedad notable de $\bar{x}$ es de mucha importancia porque una gran parte de los problemas que se encuentran en la práctica, tratan con muestras suficientemente grandes para permitir la suposición de que $\bar{x}$ se encuentra distribuida normalmente, y permite, por lo tanto, el uso de los métodos de curvas normales ordinarias, para resolver problemas relacionados con medias sin preocuparse de la naturaleza de la distribución de la población original.

Un teorema matemático bien conocido, establece que bajo suposiciones poco restrictivas, la distribución de $\bar{x}$ tiende a una distribución normal al crecer el tamaño de la muestra, n .Este teorema, junto con los resultados de los experimentos de muestreo del tipo ya estudiado, se puede expresar en la siguiente forma:

(13) ***Teorema 2.*** *Si x posee una distribución con media μ y desviación estándar σ, entonces la media de muestra $\bar{x}$ basada en un muestreo aleatorio de tamaño n, tendrá una distribución que se aproxima a la distribución de una variable normal con media μ y desviación estándar $\sigma/\sqrt{n}$ al tender n a infinito.*

Este teorema permite calcular la probabilidad de que $\bar{x}$ esté en cualquier intervalo específico, si se conocen la media de la población y la variancia. Suponga que se sabe que $\mu = 68$ plg y que $\sigma = 3$ plg en la distribución de la estatura de hombres adultos. Si se toma una muestra aleatoria de tamaño $n = 100$, ¿cuál es la probabilidad de que el valor resultante de $\bar{x}$ difiera más de $\frac{1}{2}$ plg de la media de la población? Como se puede tratar a $\bar{x}$ como una variable normal con media $\mu_{\bar{x}} = 68$ y desviación estándar $\sigma_{\bar{x}} = 3/\sqrt{100} = .3$, es necesario calcular el área bajo una curva normal que tenga esta media y desviación estándar, área que se encuentre a la derecha de 68.5 y a la izquierda de 67.5. Por simetría, esto es igual al doble de la probabilidad de que $\bar{x}$ esté a la derecha de 68.5. Por tanto, basta calcular

$$z = \frac{68.5 - 68}{.3} = 1.67.$$

De la Tabla IV se sabe que esta probabilidad es .05; por lo tanto, la probabilidad de que $\bar{x}$ difiera de 68 en más de $\frac{1}{2}$ plg es igual a .10.

Los teoremas precedentes servirán de base para resolver varios tipos de problemas de inferencia estadística de los siguientes capítulos. Se trata de teoremas en extremo importantes y sumamente útiles.

8. VARIANCIA DE LA MUESTRA DE UNA POBLACIÓN PEQUEÑA

En la teoría anterior se supuso que los individuos eran seleccionados de una población por medio de números de muestreo aleatorio, o que la población era tan grande en relación al tamaño de la muestra, que la extracción de la muestra no causaba un efecto considerable en la composición de la población. Sin embargo, éste no es el caso en la mayoría de los problemas de muestreo de la vida real. La mayoría de los planes de muestreo no permiten que un individuo sea seleccionado dos veces en una muestra dada; por tanto, si la población no es grande en relación al tamaño de la muestra, la teoría anterior no será estrictamente correcta. La dificultad que surge en tales situaciones puede solucionarse al modificar la fórmula de la desviación estándar de $\bar{x}$.

Si N denota el tamaño de una población que es muestreada, y n denota el tamaño de la muestra tomada, entonces se puede demostrar que la fórmula $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ debe ser sustituida por la fórmula

(14)
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

Para saber qué efecto tiene el factor de corrección $\sqrt{(N-n)/(N-1)}$, considérense los tamaños de muestra y población en donde (a) $n = 5\%$ de N, (b) $n = 10\%$ de N, (c) $n = 20\%$ de N. En vista de que no tiene caso tomar muestras menores que 100, y de que $N-1$ diferirá de N en menos de uno por ciento, entonces, el factor de corrección anterior se puede expresar de la siguiente forma aproximada

$$\sqrt{\frac{N-n}{N}} = \sqrt{1-\frac{n}{N}}.$$

Los cálculos en los tres casos considerados aquí, han dado los siguientes valores para este factor de corrección: (a) .97, (b) .95, y (c) .89. De estos resultados se puede concluir con certeza que la fórmula original $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ estará equivocada en menos de un 10%, a no ser que la muestra constituya al menos el 20% de la población, y que, por tanto, no se debe uno preocupar mucho por el tamaño de la población, excepto que la muestra constituya por lo menos el 20% de la población. Desde un punto de vista más conservador, uno se debe preocupar solamente si la muestra constituye por lo menos el 10% de la población.

Se debe aplicar este mismo factor de corrección a la desviación estándar de una proporción dada por la fórmula (12) del Cap. 4 cuando el tamaño de la población es lo bastante pequeño para justificar su aplicación. Así, la Fórmula (12) del Cap. 4 quedará reemplazada por la fórmula

(15)
$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}}.$$

9. EJEMPLOS ILUSTRATIVOS DE REPASO

1. Se deberá tomar una muestra aleatoria de tamaño 100 de entre los empleados de una firma de negocios que tiene 10 sucursales con 400 empleados en cada sucursal. Explique cómo se debe usar la tabla de números aleatorios para obtener una muestra aleatoria de 100 (a) de toda la organización, (b) de manera que cada sucursal contribuya con una muestra de 10. (a) Como existen 4 000 empleados, se asociará cada empleado con uno de los números de cuatro dígitos desde 0000 hasta 3999. Para evitar descartar demasiados números aleatorios, se asociarán los números desde el 5 000 hasta el 8 999 con estos mismos individuos. Esto se hará restando 5 del primer dígito de los números en ese margen y descartando los números que empiecen con 4 o 9. (b) Haga lo mismo en relación a los números de tres dígitos en cada sucursal como se hizo con los números de cuatro dígitos para la totalidad de sucursales.

2. Suponga que $1/5$ es la probabilidad de que una propiedad sea vendida en el próximo año y que, si esto sucede, se obtendrá una ganancia de \$20 000; en el caso contrario, la pérdida será de \$4 000. (a) Calcule la ganancia esperada para esta probable venta. (b) ¿Qué valor de la probabilidad haría esta ganancia esperada igual a cero?

(a) E = \$ 20 000 $1/5$ - \$4 000 $4/5$ = 800. (b) Despejando p en la ecuación 20 000 p - 4 000 (1 - p) = 0 tenemos $p = 1/6$.

3. Suponga que la probabilidad de que una nueva medicina cure una cierta enfermedad es de $1/5$. Si se toma de muestra aleatoria de 30 pacientes con esta enfermedad y se les administra la medicina, (a) ¿cuál es el número de pacientes curados?, (b) ¿cuál es la variancia del número de pacientes curados? y

- (c) ¿cuáles serán estas dos cantidades si la muestra tomada es seleccionada de un solo hospital cuyo número total de pacientes con esta enfermedad es de 60?

(a) Sea x la cantidad de pacientes curados, $E[x] = np = 30 \cdot 1/5 = 6$ porque x es una variable binomial aleatoria.

(b) $\sigma^2 = npq = 30 \cdot 1/5 \cdot 4/5 = 24/5$,

(c) $E[x] = 6$ como en el caso anterior, pero

$$\sigma^2 = npq\frac{N-n}{N-1} = \frac{24}{5}\frac{60-30}{59} = \frac{24}{5}\frac{30}{59} = \frac{144}{59}.$$

Así, la variancia es aproximadamente $1/2$ mayor que en el caso anterior.

4. Sea x el peso en onzas de una lata de café de una cierta marca.

Si el peso medio es $\mu = 16.3$ onzas y la desviación estándar es $\sigma = .2$ oz, (a) calcule la probabilidad de que la media $\bar{x}$ de la muestra basada en una muestra de tamaño $n = 10$ sea menor que 16 oz. (b) ¿Cuál sería esta probabilidad para $n = 1$?

(a) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{10}} = 0.63, z = \frac{16-16.3}{.063} = -\frac{.3}{.063} = -5.$

La probabilidad es tan pequeña que no está registrada en las tablas.

(b) $z = \frac{16-16.3}{.2} = -\frac{.3}{.2} = -1.5. P\{z < -1.5\} = .07.$

Así aproximadamente el 7% de las latas pesará menos de una libra.

- **5.** Las x para una población con $N = 5$ elementos son 1, 2, 3, 4, 5. (a) Calcule μ, S^2 y σ^2 para esta población; (b) enumere todas las muestras posibles con $n = 3$ y demuestre que $E[\bar{x}]$ es idéntica a μ en (a). (c) Calcule $E\left[(\bar{x}-\mu)^2\right]$ y demuestre que es idéntica al valor de $\sigma_{\bar{x}}^2$ obtenido teóricamente, es decir,

$$\sigma_{\bar{x}}^2 = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n};$$

(d) demuestre que $E[s^2] = s^2$.

(a) Las X_i en este caso son $X_1 = 1$, $X_2 = 2$, $X_3 = 3$, $X_4 = 4$, y $X_5 = 5$, Por tanto, a partir de (6),

$$\mu = \frac{1}{N}\sum_{i=1}^{5} X_i = \frac{1}{5}(1+2+3+4+5) = 3.$$

De (10),

$$S^2 = \frac{\sum_{i=1}^{N}(X_i-\mu)^2}{N-1}$$

$$= \frac{1}{5-1}\left[(1-3)^2+(2-3)^2+(3-3)^2+(4-3)^2+(5-3)^2\right]$$

$$= \frac{2^2+1^2+0^2+1^2+2^2}{4} = \frac{10}{4} = 2.5.$$

Ya que $\sigma^2 = \frac{N-1}{N}S^2$, se tiene $\sigma^2 = \frac{4}{5}(2.5) = 2.$

(b) El número total de posibles muestras está dado por $\binom{5}{3}=10.$ Se encuentran en la siguiente tabla con sus x_j y $\bar{x}$ observadas.

Número de la muestra	Elementos de la muestra x_j	$\bar{x}$	s^2	$\bar{x}-\mu$	$(\bar{x}-\mu)^2$
1	1,2,3	$\frac{6}{3}$	$\frac{3}{3}$	$-\frac{3}{3}$	$\frac{9}{9}$
2	1,2,4	$\frac{7}{3}$	$\frac{7}{3}$	$-\frac{2}{3}$	$\frac{4}{9}$
3	1,2,5	$\frac{8}{3}$	$\frac{13}{3}$	$-\frac{1}{3}$	$\frac{1}{9}$
4	1,3,4	$\frac{8}{3}$	$\frac{7}{3}$	$-\frac{1}{3}$	$\frac{1}{9}$
5	1,3,5	$\frac{9}{3}$	$\frac{12}{3}$	0	0
6	1,4,5	$\frac{10}{3}$	$\frac{13}{3}$	$+\frac{1}{3}$	$\frac{1}{9}$
7	2,3,4	$\frac{9}{3}$	$\frac{3}{3}$	0	0
8	2,3,5	$\frac{10}{3}$	$\frac{7}{3}$	$+\frac{1}{3}$	$\frac{1}{9}$
9	2,4,5	$\frac{11}{3}$	$\frac{7}{3}$	$+\frac{2}{3}$	$\frac{4}{9}$
10	3,4,5	$\frac{12}{3}$	$\frac{3}{3}$	$+\frac{3}{3}$	$\frac{9}{9}$
Totales		$\frac{90}{3}$	$\frac{75}{3}$	0	$\frac{30}{9}$

En vista de que cada una de las 10 posibles muestras puede ser seleccionada en un muestreo aleatorio sin reemplazo con probabilidad de $1/10$, entonces

$$[\bar{x}] = \sum_{i=1}^{10}\bar{x}_i P\{\bar{X}_i\} = \frac{\sum \bar{x}_i}{10} = \frac{\frac{90}{3}}{10} = 3,$$

que es idéntica a μ.

(c) $\sigma_{\bar{x}}^{\ 2} = E\left[(\bar{x}-\mu)^2\right] = \frac{\sum(\bar{x}_i-\mu)^2}{10} = \frac{\frac{30}{9}}{10} = \frac{1}{3}.$

De la Fórmula (14),

$$\sigma_{\bar{x}}^{\ 2} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2}{3}\left(\frac{5-3}{5-1}\right) = \frac{2}{3}\cdot\frac{2}{4} = \frac{1}{3}$$

que es idéntico al resultado anterior.

(d) $E\left[s^2\right] = \frac{\sum s_i^{\ 2}}{10} = \frac{75}{30} = 2.5 = S^2.$

Por tanto, s^2 es una estimación insesgada de S^2.

GUÍA DE AUTOEVALUACIÓN

Preguntas abiertas

1. Roberto y María administran el Foxtrot Inn, una pensión donde dan alojamiento y desayuno, localizada en Tryon, Carolina del Norte. Se rentan ocho habitaciones en esta pensión. A continuación aparece el número de estas ocho habitaciones rentadas diariamente durante junio de 2006. Utiliza Excel para seleccionar una muestra de cinco noches de junio.

2. La siguiente lista incluye a los estudiantes que se matricularon en un curso de introducción a la estadística administrativa. Se elige al azar a tres estudiantes, a quienes se formulan varias preguntas relacionadas con el contenido del curso y el método de enseñanza.
a) Se escriben a mano los números 00 a 45 en papeletas y se colocan en un recipiente. Los tres números seleccionados son 31, 7 y 25 ¿Qué estudiantes se van a incluir en la muestra?
b) Ahora utilice la tabla de dígitos aleatorios, para seleccionar tu propia muestra
c) ¿Qué harías si localizara el número 59 en la tabla de números aleatorios?

3. Slim Industries cuenta con siete empleados de producción (a quienes se les considera la población). En la tabla siguiente se incluyen los ingresos por hora de cada empleado.

Empleado	Ingreso por hora	Empleado	Ingreso por hora
Juan	$7	Cristina	$7
Samuel	7	Arturo	8
Susana	8	César	9
Roberto	8		

a) ¿Cuál es la media de la población?
b) ¿Cuál es la distribución muestral de la media para muestras de tamaño 2?
c) ¿Cuál es la media de la distribución muestral de la media?
d) ¿Qué observaciones es posible hacer sobre la población y la distribución muestral de la media?

4. 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) Elabora una lista de todas las muestras posibles de 2 ejecutivos de la población y calcula las medias.
c) Organiza las medias en una distribución muestral
d) Compara la media poblacional y la media de las medias de las muestras.
e) Compara la dispersión en la población con la dispersión de la distribución muestral de la media.

5. Una población consta de los siguientes cuatro valores: 12, 12, 14 y 16.
a) Enumera todas las muestras de tamaño 2 y calcula la media de cada muestra
b) Calcula la media de la distribución muestral de la media y la media de población. Compara los dos valores.
c) Compara la dispersión en la población con la de las medias de las muestras.

Preguntas de falso-verdadero

1. Una población que se divide en subgrupos, denominados estratos y se selecciona al azar una muestra de cada estrato se le conoce como muestreo aleatorio sistemático.

2. El error de muestreo es la diferencia entre el estadístico de una muestra y el parámetro de la población correspondiente.

3. La distribución de probabilidad de todas las posibles medias de las muestras de un determinado tamaño muestra de la población se le conoce como distribución muestral de la media.

Preguntas de opción múltiple

1. A cada nuevo empleado se le proporciona un número de identificación. Los archivos del personal se ordenan en secuencia comenzando con el empleado número 0001. Para sondear a los empleados, primero se eligió el número 0153. Los números 0253, 0353, 0453 y así sucesivamente, se convierten en miembros de la muestra. Este tipo de muestreo recibe el nombre de:
a) Muestreo aleatorio simple

b) Muestreo sistemático
c) Muestreo aleatorio estratificado
d) Muestreo por conglomerados

2. Divides un barrio en cuadras. Enseguida seleccionas 12 cuadras al azar y concentras tú sondeo en esas 12 cuadras. Este tipo de muestreo se denomina:
a) Muestreo aleatorio simple
b) Muestreo sistemático
c) Muestreo aleatorio estratificado
d) Muestreo por conglomerados

3. El error de muestreo es:
a) Igual a la media poblacional
b) Un parámetro poblacional
c) Siempre positivo
d) La diferencia entre el estadístico de la muestra y el parámetro de la población.

BIBLIOGRAFÍA COMPLEMENTARIA

- CAMACHO Rosales Juan, **Estadística con SPSS para Window versión 12**, Editorial Alfaomega, 2006 pp.410.
- CRYER D. Jonathan y Miller B. Robert. **Statistics for Business" Data Analysis and Modeling**, Editorial Duxbury Press, Estados Unidos, 1994, pp. 883.
- WAYNE Daniel W. y Terrell C. James "**Business Statistics for Management and Economics**", 6a ed., Editorial Houghton Mifflin Company, Estados Unidos, 1992, pp.920.
- DES Raj., **Teoría del Muestreo**, Editorial Fondo de Cultura Económica, 1980, pp.305.
- LEE, Cheng F., **Statics for Business and Financial Economics Editorial D.C. Heath and Company**, Canada 1993, pp. 928.
- MENDENHALL, William, **Estadística para Administradores**, Editorial Grupo Editorial Iberoamérica, México, 1993 pp. 817
- MORSE B. Lawrence, **Statistics for Business and Economics**, Editorial Harper Collins, Estados Unidos 1993, pp.989.
- SALVATORE Dominick y Reagle Derrick, **Estadística y Econometría**, 2ª ed., Editorial McGraw Hill, México 2004, pp. 357.
- SPIEGEL R. Murray, Schiller John y Scrinivasan R. Alu. **Probabilidad y Estadística**, 2ª ed., Editorial McGraw Hill, México, 2003, pp. 416.