

Métodos prácticos de inferencia estadística

Gildaberto Bonilla

trillas 

QA
276.6
.B65
1991

CMR006952

Catalogación en la fuente

Bonilla, Gildaberto

Métodos prácticos de inferencia estadística. --
México : Trillas, 1991.

402 p. ; 23 cm.

Bibliografía: p. 397

Incluye índices

ISBN 968-24-3960-4

1. Estadística matemática. 2. Muestreo
(Estadística). 3. Diseño experimental. I. t.

LC- QA276.6'B6.5 D- 519.52'B243m

La presentación y disposición en conjunto de
MÉTODOS PRÁCTICOS DE INFERENCIA ESTADÍSTICA
son propiedad del editor. Ninguna parte de esta obra
puede ser reproducida o transmitida, mediante ningún sistema
o método, electrónico o mecánico (incluyendo el fotocopiado,
la grabación o cualquier sistema de recuperación y almacenamiento
de información), sin consentimiento por escrito del editor

Derechos reservados

© 1991, Editorial Trillas, S. A. de C. V.,
Av. Río Churubusco 385, Col. Pedro María Anaya,
C.P. 03340, México, D. F.

Miembro de la Cámara Nacional de la
Industria Editorial. Reg. núm. 158

Primera edición, 1988

Segunda edición, septiembre 1991

(primera publicada por Editorial Trillas, S. A. de C. V.)

ISBN 968-24-2931-4

Impreso en México

Printed in Mexico

Prólogo

Desde la aparición de mi primer libro, *Elementos de estadística descriptiva y probabilidad*, publicado por UCA Editores en El Salvador, muchos colegas que lo utilizan como texto en las distintas universidades de ese país me sugirieron escribir otro para facilitar la comprensión de la parte inferencial, con un enfoque lo más práctico posible. Este libro es la respuesta a tales peticiones. Espero no defraudarlos; he tenido presentes valiosas sugerencias, en especial las referentes a preparar un texto asequible a personas con poca cultura matemática. Utilizo la notación matemática solamente cuando es necesario, y evito por completo las demostraciones y deducciones. Explico cada tema de manera informal e intuitiva y lo acompaño de ejemplos prácticos para facilitar su comprensión e incrementar su utilidad en la toma de decisiones.

El enfoque anterior, aplicado a través de toda la obra, asegura el cumplimiento de mi objetivo principal al escribir este libro: lograr un texto introductorio a la estadística inferencial que fuera tanto práctico, claro y comprensible como interesante e informativo.

Métodos prácticos de inferencia estadística constituye una exposición práctica de los principales temas estadísticos para realizar investigaciones mediante muestreo en las áreas de economía, administración de empresas, educación, psicología y ciencias políticas.

En el capítulo 1 se estudian, desde los puntos de vista probabilístico y no probabilístico, hasta los principales diseños muestrales que se requieren para tomar muestras representativas.

En un orden secuencial y lógico, el capítulo 2 trata de las relaciones que guardan estadísticos y parámetros, con aplicación del método deductivo. Se parte de una población concreta y se calculan sus parámetros; luego, de esa población se extraen todas las muestras posibles de un determinado tamaño, con y sin reemplazamiento, para luego

calcular los estadísticos: media y varianza, los cuales a su vez se comparan con los parámetros y así se obtienen las relaciones entre ellos. El capítulo sienta las bases para la estimación estadística y prueba de hipótesis, las cuales se estudian en éste y en los capítulos 3 y 4.

Concretamente, en el capítulo 4 se destacan las vías más sencillas para probar las hipótesis. Su comprensión hace posible una perspectiva dinámica de la inferencia y sus variadas aplicaciones.

En el capítulo 5 se estudian varios procedimientos especiales para muestras pequeñas. En particular, procedimientos de estimación por intervalo y prueba de hipótesis sobre la media de una población, así como la prueba de la hipótesis para diferencias de las medias.

En el capítulo 6 se estudian los interesantes temas de la regresión y la correlación, comenzando con el análisis de dos variables para determinar la posible asociación entre ellas; se mide su relación funcional y la intensidad con que están relacionadas. El capítulo finaliza con el análisis simultáneo de dos variables.

En el capítulo 7 se tratan las pruebas no paramétricas; es decir, aquellas que no especifican las condiciones de los parámetros de la población de la cual se extrae la muestra. Una de las ventajas de las técnicas no paramétricas es su facilidad y rapidez de aplicación. Existe una gran cantidad de pruebas no paramétricas; en este libro sólo se estudian las de uso más generalizado.

A sugerencia del vicerrector de la Universidad Centroamericana "José Simeón Cañas", al final del libro se presenta una exposición sencilla de las principales pruebas no paramétricas; de gran utilidad especialmente para los estudiantes de psicología y ciencias de la educación.

A lo largo del libro se presentan ejemplos de aplicación y ejercicios elaborados para reafirmar y evaluar la comprensión de los estudiantes, así como para fomentar en éstos el diseño de proyectos de investigación.

Índice de contenido

Prólogo	5
Cap. 1. Diseños muestrales	9
1.1. Definiciones, 9. 1.2. Muestreos aleatorios, 12. 1.3. Muestreos no aleatorios, 20.	
Cap. 2. Distribución muestral de estadísticos	25
2.1. Distribución muestral de medias, 26. 2.2. Distribución muestral de proporciones, 58.	
Cap. 3. Estimación de parámetros	83
3.1. Definiciones, 83. 3.2. Estimación de la proporción poblacional, 94. 3.3. Determinación del tamaño adecuado de la muestra, 100.	
Cap. 4. Aspectos básicos de pruebas de hipótesis	111
4.1. Definición, 111. 4.2. Planteamiento de las hipótesis, 112. 4.3. Errores tipo I y tipo II, 112. 4.4. Pruebas unilaterales y bilaterales, 113. 4.5. Procedimiento para probar la hipótesis, 118. 4.6. Diferencia entre una proporción muestral y una poblacional, 127. 4.7. Prueba de hipótesis en una distribución muestral de diferencias de medias $\bar{X}_1 - \bar{X}_2$, 132. 4.8. Prueba de hipótesis para diferencias de proporciones, 141.	
Cap. 5. Muestreo pequeño	149
5.1. La distribución t de Student, 149. 5.2. Prueba de la hipótesis sobre la media poblacional, 158. 5.3. Prueba de la hipótesis para diferencias de dos medias, en muestreo pequeño, 163. 5.4. Distribución χ^2 cuadrada, 170.	

5.5. Cuadros de contingencia, 182. 5.6. Fórmulas eficientes para el cálculo de X^2 , 187. 5.7. Limitaciones de la prueba χ^2 cuadrada, 194. 5.8. Pruebas de homogeneidad, 195. 5.9. Ejercicios de refuerzo, 203.

Cap. 6. Regresión y correlación	221
6.1. Introducción, 221. 6.2. Regresión lineal simple, 222. 6.3. Regresión curvilínea, 228. 6.4. Desviación estándar de regresión, 241. 6.5. Coeficiente de correlación, 245. 6.6. Regresión y correlación para datos agrupados, 254. 6.7. Ejercicios, 261. 6.8. Regresión y correlación múltiples, 267. 6.9. Ejercicios, 285. 6.10. Ejercicios de refuerzo, 288. 6.11. Correlación por rangos, 305. 6.12. Coeficiente de correlación para datos nominales clasificados en tablas 2×2 , 316. 6.13. Coeficiente de correlación para datos nominales clasificados en tablas de contingencia mayores de 2×2 , 318. 6.14. Ejercicios, 323.	
Cap. 7. Pruebas no paramétricas	329
7.1. Prueba del signo, 330. 7.2. Prueba de rangos señalados y pares igualados de Wilcoxon, 338. 7.3. Prueba U de Mann-Whitney, 346. 7.4. Prueba H de Kruskal-Wallis, 353. 7.5. Prueba de la mediana, 358.	
Respuestas de los ejercicios	367
Apéndice de tablas	387
Bibliografía	397
Índice analítico	399

1

Diseños muestrales

1.1. DEFINICIONES

Los *diseños muestrales* son procedimientos para extraer muestras de poblaciones con el objeto de conocer sus características promedio. Es interesante saber cómo, a través del análisis de sólo una parte representativa de la población, se pueden inferir resultados válidos para el total de ella. Antes de estudiar los distintos diseños muestrales definiremos algunos conceptos básicos.

1.1.1. Conceptos básicos del muestreo

- *Población* es un agregado de unidades individuales, compuesto de personas o cosas que se hallan en una situación determinada. Las unidades individuales se llaman *unidades elementales*. Definir una población es determinar sus unidades elementales de acuerdo con el interés que se tiene respecto a alguna característica de aquélla. Tanto la definición de una población como la característica por observar de sus unidades elementales dependen de la naturaleza del problema. Por ejemplo, si el problema es “Camisas para personas adultas de El Salvador”, se trata de determinar la cantidad adecuada de producción de camisas de acuerdo con las diversas medidas. La población son todas las personas adultas de El Salvador. La característica de interés son las medidas del cuello de las personas adultas en dicho país. Veamos otro ejemplo: las cuotas diferenciadas para alumnos de la UES (Universidad de El Salvador). El problema por resolver es la asignación de las cuotas a los estudiantes. La población son los alum-

nos (o sus padres) de la UES; la característica de interés es el monto de los ingresos de dichos estudiantes.

Las poblaciones pueden ser infinitas o finitas. Una población *infinita* es la que contiene un número infinito de unidades elementales; por ejemplo, el conjunto de piezas que se obtienen en un proceso productivo; en el sentido de que se siguen produciendo indefinidamente. Otro ejemplo son todos los posibles resultados al lanzar una moneda sin cesar.

Una población es *finita* cuando tiene un número finito de unidades elementales. Por ejemplo, los estudiantes de una determinada universidad; el número de escuelas que existen en una determinada ciudad, el número de árboles de coco sembrados en una determinada parcela, etcétera.

El número de unidades elementales de una población se denota con la letra N .

- Una *muestra* es una parte de la población; por ejemplo, cuando se desea hacer un estudio relativo al rendimiento académico de los alumnos de cierta universidad, y para esto se toma sólo un grupo de estudiantes de la misma. Todos los estudiantes de ella son la población y el grupo escogido constituye la muestra. Es importante hacer notar que para hacer una investigación mediante el análisis de una muestra, ésta tiene que ser, necesariamente, *representativa*. La representatividad de la muestra implica que cada unidad de la población debe tener igual probabilidad de ser seleccionada. En estas condiciones, se dice que la muestra es aleatoria. La obtención de una muestra representativa es uno de los aspectos más importantes de la teoría estadística. Incluye preguntas como, ¿qué tan grande debe ser la muestra?, ¿qué tipo de datos deben ser recolectados?, ¿cómo se recogerán éstos? Estas preguntas serán contestadas más adelante. (El número de unidades elementales de una muestra se denota con la letra n).
- El *muestreo* es el proceso de seleccionar una parte del todo.
- La *medición* es la etapa intermedia que consiste en contar, medir y formular preguntas.
- La *estimación* es el proceso de hacer deducciones sobre la población total, a partir de la información de la muestra. El muestreo y la estimación conforman el diseño de la muestra.
- El *muestreo probabilístico* es un proceso de selección de muestras en el cual los elementos son escogidos por métodos aleatorios. Existen numerosas variaciones en el muestreo probabilístico, pero todas comparten un rasgo común: la selección de la muestra se realiza por procedimientos al azar y con probabilidades conocidas.
- El *muestreo no probabilístico* incluye todos los métodos en que las unidades elementales de la población no se seleccionan al azar.

- Las *listas* son un inventario de unidades de una población o subpoblación, que tiene una correspondencia directa, uno a uno, entre cada ítem listado y la unidad que representa. Un ejemplo de lista son todos los alumnos de la universidad, identificados cada uno por su nombre. Si las unidades de la lista son elementos de la población, se trata de una lista de elementos. A veces las unidades del listado pueden ser grupos de elementos ubicados en conglomerados naturales convenientes, identificables y concretos; por ejemplo, los departamentos de un edificio, grupos de escuelas, grupos de fábricas, etc. En estos casos, el inventario consiste de una lista de conglomerados.
- Las *unidades de muestreo* son aquellos elementos o grupos de elementos que forman la base de una selección de muestras. Pueden o no ser idénticas al listado de las unidades. Cuando se dispone de una lista completa de elementos de la población, en general, es más conveniente extraer una muestra directamente de la lista, en cuyo caso el muestreo sería lo mismo que el listado de unidades.
- La *estructura de muestreo* está constituida por los materiales y procedimientos utilizados para contabilizar la población cuando no se dispone de una lista completa de elementos. La estructura de muestreo consistirá, en general, de mapas, croquis, listas, fotografías aéreas e instrucciones sobre la forma como estos elementos deben utilizarse. Se trata básicamente del procedimiento operativo y de los materiales usados para esquematizar la población al diseñar la muestra.
- *Fiabilidad y efectividad* son generalmente los dos criterios para evaluar un diseño de muestreo. Es de esperar que en éste haya errores. Un error de muestreo es la diferencia entre el valor del estadístico (aproximado) y el del parámetro (exacto); se mide por su fiabilidad, o por la precisión del muestreo, la cual está relacionada con la varianza del estadístico muestral. A mayor varianza, menor fiabilidad en el resultado. El criterio de efectividad está asociado con el costo del muestreo. Un diseño de muestreo es más efectivo que otro si con él se obtiene el mismo grado de fiabilidad, con un costo menor.
- Un *evento* es el resultado de un experimento. Por ejemplo, al lanzar una moneda, los eventos posibles son “cara” y “cruz”.
- La *probabilidad* (P) es una medida de las veces que se espera que ocurra un evento que depende del azar.

Si el total de resultados posibles de un experimento son M , y éstos son mutuamente exclusivos e igualmente posibles, y si un evento particular e puede ocurrir de m maneras, entonces la probabilidad de la ocurrencia de e es: $P(e) = m/M$. Por ejemplo, la probabilidad de obtener “cara” al lanzar una moneda es $1/2$. Esta fracción también podemos interpretarla así: *esperamos* obtener “cara” una vez ($m = 1$) en un total de dos lanzamientos de la moneda ($M = 2$).

La probabilidad siempre será un número comprendido entre 0 y 1.

1.1.2. Diseño de muestreos alternativos

Los diseños de muestreo se dividen en dos grandes grupos: aleatorios y no aleatorios.

El muestreo *aleatorio* es un proceso de selección al azar; se denomina también muestreo *probabilístico*, ya que involucra las leyes de la probabilidad.

El muestreo *no aleatorio* es un proceso subjetivo de selección muestral, sin intervención del azar; es decir, una muestra no aleatoria se selecciona de acuerdo con el juicio experto del investigador.

1.2. MUESTREOS ALEATORIOS

1.2.1. Muestreo aleatorio simple

El muestreo aleatorio simple es un procedimiento de selección por el cual todos y cada uno de los elementos de la población tienen igual probabilidad de ser incluidos en la muestra. Entonces, si toda unidad de muestreo tiene la misma probabilidad de ser escogida, se sigue que *las muestras aleatorias de igual tamaño, tomadas de una población dada, tienen la misma probabilidad de ser tomadas.*

Partamos, por ejemplo, de la población de tres calificaciones: 2, 4 y 6. De esta población podemos tomar muestras de tamaño $n = 2$; con o sin reemplazo. *Con reemplazo* significa que se selecciona una unidad elemental y luego ésta se regresa a la población antes de tomar la siguiente; por tanto, cada unidad puede ser seleccionada más de una vez en la misma muestra. Es *sin reemplazo*, cuando, una vez escogida una unidad elemental, ésta ya no se regresa a la población, y por tanto, cada unidad puede aparecer sólo una vez.

Veamos el caso de muestreo con reemplazo, tomando todas las posibles muestras de $n = 2$ que se pueden tomar de la población: 2, 4 y 6. El número de muestras por tomar es $N^n = 3^2 = 9$.

Las muestras son:

2,2	2,4	2,6
4,2	4,4	4,6
6,2	5,4	6,6

Cada una de estas muestras tiene $1/9$ de probabilidad de ser escogida.

Ahora veamos el caso de muestreo sin reemplazo. Tomemos siempre muestras de tamaño $n = 2$. El número de muestras por tomar es el resultado del desarrollo del combinatorio

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

Al sustituir:

$$\binom{3}{2} = \frac{3!}{2!1!} = \frac{3 \times 2!}{2!1!} = 3$$

Las muestras son:

2,4

2,6

4,6

Cada una de estas muestras tiene $1/3$ de probabilidad de ser escogida.

¿Cómo escoger una muestra aleatoria simple? Uno de los métodos sencillos consiste en numerar todos los elementos de la población, escribir los números en tarjetas, fichas o bolas, poner luego en una caja o bolsa estos objetos numerados y mezclarlos completamente. Se determina entonces el tamaño n de la muestra y se sacan los objetos al azar, uno por uno, hasta tener el número deseado.

Otro método más técnico y confiable consiste en usar tablas de números aleatorios. En la mayor parte de los libros de estadística aparece este tipo de tablas; sin embargo, uno mismo puede elaborar las suyas, haciendo uso de una urna con 10 bolas –de ping pong, por ejemplo– numeradas del 0 al 9; se sacan las bolas con reemplazamiento, se anotan los números que salgan y se forman bloques de números con los dígitos que se desee. El proceso es eminentemente aleatorio, ya que cada dígito tiene la misma probabilidad de salir: $1/10$.

Veamos algunos ejemplos donde se utiliza la tabla 1.1 de números aleatorios.

Supóngase que tenemos una población de tamaño $N = 100$ y deseamos tomar una muestra de tamaño $n = 15$. Primero se numera la población de la siguiente manera: por ser $N = 100$ potencia de 10, la numeración de la población queda así: 00-01-02-03-04-05-06-07-08-09-10-11-12 ... 99. Una vez numerados los elementos de la población, se toman los primeros dos dígitos de los bloques de números de la tabla, con un comienzo aleatorio. Supongamos que el comienzo fue a partir del tercer bloque y quinto renglón; es decir:

58854

56144

01148

78742

.

.

.

Tabla 1.1. Números aleatorios*

55034	81217	90564	81943	11241	84512	12288	89862	00760	76159
25521	99536	43233	48786	49221	06960	31564	21458	88199	06312
85421	72744	97242	66383	00132	05661	96442	37388	57671	27916
61219	48390	47344	30413	39392	91365	56203	79204	05330	31196
20230	03147	58854	11650	28415	12821	58931	30508	65989	26675
95776	83206	56144	55953	89787	64426	08448	45707	80364	60262
07603	17344	01148	83300	96955	65027	31713	89013	79557	49755
00645	17459	78742	39005	36027	98807	72666	54484	68262	38827
62950	83162	61504	31557	80590	47893	72360	72720	08396	33674
79350	10276	81933	26347	08068	67816	06659	87917	74166	85519
48339	69834	59047	82175	92010	58446	69591	56205	95700	86211
05842	08439	79836	50957	32059	32910	15842	13918	41365	80115
25855	02209	07307	59942	71389	76159	11263	38787	61541	22606
25272	16152	82323	70718	98081	38631	91956	49909	76253	33970
73003	29058	17605	49298	47675	90445	68919	05676	23823	84892
81310	94430	22663	96584	38142	00146	17496	51115	61458	65790
10024	44713	59832	80721	63711	67882	25100	45245	55743	67618
84671	52806	89124	37691	20897	82339	22627	06142	05773	03547
29296	58162	21858	33732	94056	88806	54603	00384	66340	69232
51771	94074	70630	41286	90583	87680	13661	55627	23670	35109
42166	56251	60770	51672	36031	77273	85218	14812	90758	23677
78355	67041	22492	51522	31164	30450	27600	44428	96380	26772
09552	51347	33864	89018	73418	81538	77399	30448	97740	18158
15771	63127	34847	05660	06156	48970	55699	61818	91763	20821
13231	99058	93754	36730	44286	44326	15729	37500	47269	13333
50583	03570	38472	73236	67613	72780	78174	18718	99092	64114
99485	57330	10634	74905	90671	19643	69903	60950	17968	37217
54676	39524	73785	48864	69835	62798	65205	69187	05572	74741
99343	71549	10248	76036	31702	76868	88909	69574	27642	00336
35492	40231	34868	55356	12847	68093	52643	32732	67016	46784
98170	25384	03841	23920	47954	10359	70114	11177	63298	99903
02670	86165	56860	02592	01646	42200	79950	37764	82341	71952
63934	42879	81637	79952	07066	41625	96804	92388	88860	68580
56851	12778	24309	73660	84264	24668	16686	02239	66022	64133
05464	28892	14271	23778	88599	17081	33884	88783	39015	57118
15025	20237	63386	71122	06620	07415	94982	32324	79427	70387
95610	08030	81469	91066	88857	56583	01224	28097	19726	71465
09026	40378	05731	55128	74298	49196	31669	42605	30368	96424
81431	99955	52462	67667	97322	69808	21240	65921	12629	92896
21431	59335	58627	94822	65484	09641	41018	85100	16110	32077

*Tomada de Rand Corporation, *A million random digits with 100 000 normal deviates*, The Free Press., Glencoe, Ill., 1955.

Los números escogidos –siguiendo hacia abajo– son: 58-56-01-78-61-81-59-79-07-82-17-22-89-21 y 70. Obsérvese que después del 22 sigue el 59, el cual no se tomó, porque ya había sido seleccionado.

Supóngase ahora que la población tiene $N = 825$ elementos, de los cuales tomaremos una muestra de $n = 20$. Por no ser N potencia de 10 y constar de tres cifras, los elementos de la población se numeran así: 001-002-003-004-005-006-007-008-009-010-011-012-013 ... 825. Ya numerados esos elementos, se toman de la tabla números de tres cifras, sin pasar de 825. Supongamos que el comienzo aleatorio es el sexto bloque y el onceavo renglón; es decir:

58446
32910
76159

·
·
·

Los números seleccionados –siguiendo hacia abajo– son: 584-329-761-386-001-678-823-772-304-815-489-443-727-196-627, y así sucesivamente hasta completar la muestra. Obsérvese que después del número 386 sigue el 904, que no se tomó por ser mayor que 825; lo mismo se hizo con los números 888 y 876, mayores que el tamaño de la población.

Vamos a suponer, esta vez, que $N = 6\ 645$, y la muestra tomada es $n = 400$. Los elementos de la población quedan numerados así: 0001-0002-0003-0004 ... 3645. Supongamos que el comienzo aleatorio es el octavo bloque y el veinteavo renglón; es decir:

55627
14812
44428, etcétera.

Esta vez los números serán escogidos siguiendo hacia arriba: 5562-0038-0614-4534-5111-0567-4990-3878-1391-5620-5448-4570-3050-3738-2145; en este número termina el bloque octavo. Los siguientes se pueden escoger ya sea el séptimo o el doceavo bloque de números; si optamos por el séptimo, los números que siguen son: 1228-3156-5620-5893-0844, y así sucesivamente hasta tener toda la muestra.

1.2.2. Muestreo sistemático

Según este procedimiento, se obtiene una muestra tomando cada K -ésima unidad de la población, tras numerar las unidades elementales

de ésta o haberlas ordenado de alguna manera. La K representa un número entero, que es aproximadamente la “razón de muestreo” entre el tamaño de la población N y el tamaño de la muestra n ; es decir,

$$K = N/n$$

Así, si la población consta de 3 600 unidades y se desea tomar una muestra de tamaño $n = 400$, entonces la razón de muestreo

$$K = 3\,600/400 = 9,$$

y la muestra se obtiene tomando una unidad de cada 9 de la población. Para lograr la aleatoriedad, el procedimiento debe empezar al azar. Así, como en nuestro ejemplo $K = 9$, se escriben en 9 papelitos números del 1 al 9; luego se extrae uno al azar. Supongamos que salió el papelito con el número 5; ésta será la primera unidad que formará la muestra; el siguiente es $5 + K$, o sea, $5 + 9 = 14$; el tercero es $14 + 9 = 23$, y así sucesivamente hasta completar los 400 de la muestra.

1.2.3. Muestreo aleatorio estratificado

El proceso consiste en dividir la población en grupos llamados *estratos*. Dentro de cada estrato están los elementos situados de manera más homogénea con respecto a las características en estudio. Para cada estrato se toma una submuestra mediante el procedimiento aleatorio simple, y la muestra global se obtiene uniando las submuestras de todos los estratos.

La elección de los elementos de cada estrato puede ser proporcional o no.

El muestreo por estratos puede ser más efectivo si se trata de poblaciones heterogéneas. Al hacer la estratificación, las clases se establecen de modo que las unidades de muestreo tiendan a ser uniformes dentro de cada clase, y las clases tiendan a ser diferentes entre sí.

Si la desviación típica de la característica observada en cada estrato es menor que la de toda la población, como es lo usual, debido a la mayor uniformidad dentro del estrato, resultará que la fiabilidad aumenta para un tamaño de muestra dado, o bien la efectividad para un grado de fiabilidad.

El aumento de fiabilidad y efectividad se puede incrementar clasificando todavía más los estratos en subestratos, si ello fuese posible. Por ejemplo, al hacer una encuesta sobre la elección de un candidato político, se pueden tomar como estratos los partidos políticos. Dado que las preferencias electorales están influidas a menudo por factores

como el sexo, la edad, la profesión, la religión, etc., se podrían subdividir los estratos de acuerdo con dichos factores. Este procedimiento de doble clasificación se llama *estratificación cruzada*. El muestreo por estratos es una combinación de submuestras de los mismos.

Supóngase, por ejemplo, que se desea establecer un seguro de vida para los maestros que laboran en San Salvador. Para tal propósito se tomará una muestra que represente el 10% de la población de maestros de dicho país (24 000), cuya distribución por edades es:

<i>Edad (años)</i>	<i>Núm. de profesores</i>
25 o menos	2 880
26-35	7 680
36-45	6 240
46-55	4 800
56 o más	2 400
<i>Total</i>	24 000

El siguiente paso será un plan de muestreo de manera que cada grupo de edad quede representado proporcionalmente. El diseño apropiado es el muestreo aleatorio estratificado proporcional que se presenta en la tabla 1.2.

Tabla 1.2. Distribución por edades de los maestros de San Salvador y cálculos del porcentaje por estratos y unidades elementales de las submuestras.

<i>Estrato</i>	<i>Núm. de profesores</i>	<i>Porcentaje</i>	<i>Unidades por estrato (submuestras)</i>
25 o menos	2 880	12	288
26-35	7 680	32	768
36-45	6 240	26	624
46-55	4 800	20	480
56 o más	2 400	10	240
<i>Total</i>	24 000	100	$n = 2\,400$

El tamaño de la muestra es $24\,000 \times 0.10 = 2\,400$.

El porcentaje de los estratos se calculó así:

$$2\,880/24\,000 \times 100 = 12\%$$

para el estrato 25 años o menos; para el estrato 26-35 años:

$$7\,680/24\,000 \times 100 = 32\%,$$

y así sucesivamente.

Las unidades muestrales de cada estrato se obtuvieron multiplicando la proporción de cada estrato por el tamaño de la muestra, que es $n = 2\,400$. Por tanto, para el estrato 25 años o menos: $0.12 \times 2\,400 = 288$; para el estrato 26-35 años: $0.32 \times 2\,400 = 768$; y así sucesivamente. Éstas constituyen las submuestras.

Veamos cómo escoger las unidades elementales de cada estrato. Se puede hacer de dos maneras: 1) por muestreo aleatorio simple, utilizando tablas de números aleatorios, y 2) por muestreo aleatorio sistemático.

La solución por muestreo aleatorio simple, usando tablas de números aleatorios, sería la siguiente: para tomar la submuestra del estrato 25 años o menos, numeramos la subpoblación de este estrato que tiene $N_1 = 2\,880$, así: 0001-0002-0003... 2 880. Utilizamos la tabla 1.1 de números aleatorios y suponemos que el comienzo aleatorio está en la tercera columna (bloque) y tercer renglón (de números); es decir, comenzamos en 97242. A partir de este número se tomarán las primeras cuatro cifras, siguiendo hacia abajo, hacia arriba o en forma horizontal. Hagámoslo hacia abajo; es decir:

97242
47334
58854

·
·
·

Los números escogidos son los siguientes: 0114-0730-1 760-2 266-2 185-2 249-1 063-1 024-0384-1 427-0573. En este momento hemos terminado la tercera columna; luego continuamos en la cuarta columna, ya sea de abajo hacia arriba o de arriba hacia abajo. Nosotros vamos a continuar de abajo hacia arriba, así: 0259-2 392-0566-0658-2 634-1 165, después se puede pasar a la quinta columna, etc., hasta tomar toda la submuestra de ese estrato. Lo mismo haremos con el resto de estratos.

La solución por muestreo aleatorio sistemático se hace en la forma siguiente. Primero se numeran las subpoblaciones de los estratos. Por ejemplo, para el estrato 26-35 años se numeran así: 1, 2, 3, 4, 5 ... 7 680; luego se establece la razón de muestreo $K = N_2/n_2 = 7\,680/768 = 10$; en seguida, se numeran 10 papelitos del 1 al 10, luego se revuelven en una caja y se extrae uno. Supongamos que trae escrito el número 7; éste

constituye, por tanto, la primera unidad muestral; la siguiente es $7 + K$, o sea, $7 + 10 = 17$; la tercera $17 + 10 = 27$, y así sucesivamente, hasta completar los 768 de este estrato. Se hace lo mismo con el resto de los estratos.

1.2.4. Muestreo por conglomerados

El muestreo por conglomerados es lo contrario del muestreo por estratos. Consiste en seleccionar primero, al azar, grupos, llamados *conglomerados* de elementos individuales de la población, y en tomar luego todos los elementos o una submuestra de ellos dentro de cada conglomerado, para constituir así la muestra global.

Para lograr los mejores resultados con el procedimiento de muestreo anterior, las diferencias entre los conglomerados se hacen tan pequeñas como sea posible; en tanto que las diferencias entre los elementos individuales dentro de cada conglomerado se hacen tan grandes como sea posible. Lo ideal sería que cada conglomerado fuera una miniatura de toda la población; así un solo conglomerado sería una muestra satisfactoria.

Con frecuencia, los conglomerados se llaman *unidades de muestreo primarias*. Si todos los elementos o unidades elementales de los conglomerados seleccionados se incluyen en la muestra, el procedimiento se llama *muestreo de una etapa*. Si se saca una submuestra aleatoria de elementos de cada conglomerado seleccionado, se tiene lo que se llama *muestreo en dos etapas*. Si intervienen más de dos etapas en la obtención de la muestra total se dice que es un *muestreo de etapas múltiples* o *multietápico*. Por ejemplo, en una encuesta sobre actitudes de los estudiantes de la Universidad de El Salvador, se toman como conglomerados las distintas facultades (que constituyen unidades de muestreo primarias). Ésta es la primera etapa. La segunda podría ser la selección de departamentos dentro de cada facultad. Por último, en la tercera etapa, se toman estudiantes de todos los departamentos seleccionados para constituir la muestra global. Desde luego, los métodos de muestreo se emplean en cada etapa.

La ventaja principal del muestreo por conglomerados es la gran reducción de costos para un grado dado de fiabilidad. Por ejemplo, el costo de tomar una muestra aleatoria simple de 1 200 estudiantes de la Universidad de El Salvador, que tiene aproximadamente 35 000 estudiantes, sería elevado. Si primero se toma una muestra aleatoria de, por ejemplo, 10 departamentos, y luego se toma una submuestra de cada departamento de los ya escogidos, la muestra de $n = 1\,200$ se podría obtener a mucho menor costo que el que se necesitaría para una muestra aleatoria simple de ese tamaño. A esto se debe que, cuan-

do se hacen muestreos de una población distribuida geográficamente, se toman como unidades primarias de muestreo las regiones geográficas; este procedimiento se conoce como *muestreo por zonas*.

El muestreo por conglomerados se utiliza a menudo en control de calidad estadístico. Primero se seleccionan lotes o tandas de producción, al azar, como conglomerados; si el muestreo es de una etapa, todas las unidades de los lotes seleccionados pueden ser comprobados; o bien se puede tomar una submuestra de cada lote para su comprobación, lo cual constituye un muestreo en dos etapas.

1.3. MUESTREOS NO ALEATORIOS

1.3.1. Muestreo dirigido

Este tipo de muestreo también recibe el nombre de *muestreo adaptado*; consiste en seleccionar las unidades elementales de la población según el juicio de los encuestadores, quienes consideran que las unidades seleccionadas son representativas. En el muestreo dirigido se desconoce la probabilidad de que una unidad elemental sea elegida; en consecuencia, no se pueden construir intervalos de confianza, sólo se pueden hacer estimaciones puntuales. Tampoco se pueden aplicar los principios de la teoría del muestreo. Sin embargo, en muchas situaciones resulta más ventajoso echar mano del muestreo dirigido que del aleatorio. La selección aleatoria simple puede omitir los elementos más importantes, mientras que el muestreo dirigido con seguridad los incluirá en la muestra. Por ejemplo, un estudio de la industria de bebidas en El Salvador sería de dudosa calidad si no toma en cuenta La Constancia y la Embotelladora Tropical.

El muestreo dirigido, con un buen manejo, puede ser de utilidad práctica para resolver problemas comerciales cotidianos y tomar decisiones de política pública. Los ejecutivos, funcionarios y políticos, a menudo están presionados por el tiempo y las circunstancias y no pueden esperar por diseños probabilísticos. El muestreo dirigido es, entonces, el único método práctico, ya que permite disponer rápidamente de las estimaciones indispensables para llegar a decisiones sobre problemas urgentes. Es mejor decidir de esta manera que hacerlo sin ningún dato estadístico.

Cuando queremos estudiar algunas particularidades desconocidas de una población, algunas de cuyas características conocemos, podemos estratificar esa población de acuerdo con esas propiedades conocidas y seleccionar las unidades muestrales de cada estrato, basado en el juicio personal. Es aconsejable usar el muestreo dirigido para realizar estudios piloto o de sondeo. La confiabilidad de los resultados

muestrales en el muestreo dirigido depende de la calidad de los conocimientos o del juicio del investigador.

1.3.2. Muestreo por cuotas

El muestreo por cuotas es una técnica de uso corriente en encuestas de opinión pública. Al usar esta técnica, el entrevistador selecciona, de acuerdo con su criterio, un número determinado de individuos u objetos (*cuota*) de cada uno de varios sectores de la población. Por ejemplo, se le puede pedir que entreviste a 25 señoras de los mercados, 30 amas de casa, 20 estudiantes universitarios, 25 obreros de la construcción, etc. Otro ejemplo, para control de calidad, puede ser examinar cinco unidades cada 30 segundos en un proceso de producción.

1.3.3. Muestreo deliberado

Consiste en tomar un segmento o fracción de la población por su asequibilidad. Por ejemplo, una muestra obtenida de listas fácilmente disponibles, como las guías de teléfonos, constituye una muestra deliberada. Por la comodidad o facilidad en tomar la muestra se sacrifica cierto grado de representatividad de las características poblacionales; sin embargo, este tipo de muestreo puede ser de utilidad para estudios piloto o de sondeo.

1.4. EJERCICIOS

1. Se desea tomar una muestra aleatoria de tamaño $n = 975$ de la población estudiantil de la UCA (Universidad Centroamericana) –que vamos a suponer asciende a $N = 6\,500$ estudiantes– con el objeto de conocer su opinión respecto a las leyes tributarias aprobadas por la asamblea. Describa el procedimiento con cada uno de los siguientes métodos:
 - a) Muestreo aleatorio simple, usando tablas de números aleatorios.
 - b) Muestreo sistemático.
2. Un estudiante que concluyó sus estudios de psicología eligió para su trabajo de tesis el siguiente problema: *¿Cuáles son los factores fundamentales que incidieron durante el año escolar de 1986, en el bajo rendimiento escolar de los educandos del tercer grado de las instituciones educativas del núcleo 0-14 de la ciudad de Soyapango, en el aprendizaje de las cuatro operaciones matemáticas básicas?* Determine:

- a) La población objetivo.
 - b) El marco de referencia.
 - c) El diseño muestral más adecuado.
3. ¿Cómo usaría la tabla 1.1 de números aleatorios para tomar una muestra de $n = 10$ comunidades de las dañadas por el terremoto del 10 de octubre con el objeto de investigar las necesidades más urgentes de la población damnificada?
 4. Suponga que se desea realizar una encuesta de escolares del área metropolitana para obtener información sobre su salud, o sobre antecedentes de algún tema determinado. Explique la metodología que seguiría en esa encuesta, utilizando el muestreo por conglomerados en tres etapas (multietápico).
 5. Suponga que se desea hacer un estudio sobre la procedencia (zona de residencia) de los estudiantes de la UCA, con el objeto de estudiar la posibilidad de resolver el problema del transporte de sus hogares a la universidad. Explique como tomaría la muestra usando muestreo estratificado proporcional. Tome la muestra de tal manera que su tamaño sea 15% del total de la población estudiantil, la cual es $N = 6\,500$ estudiantes.
 6. Un sindicato de trabajadores va a seleccionar una muestra entre sus miembros para determinar la actitud general hacia un programa de actividades sociales puesto en marcha recientemente. Determine:
 - a) La población objetivo.
 - b) El marco de referencia.
 - c) El diseño muestral más adecuado.
 7. El total de maestros que trabajan en el nivel básico en El Salvador es aproximadamente 22 000. Su distribución por edades es:

<i>Edad (años)</i>	<i>Porcentaje</i>
25 o menos	15
26-35	30
36-45	25
46-55	20
56 o más	10

- Se desea hacer una investigación relativa a la adquisición de un seguro colectivo de vida. Para tal fin se quiere tomar una muestra que constituya el 5% de los 22 000 maestros del nivel básico. Diseñe un plan de muestreo de manera que cada grupo de edad quede representado proporcionalmente.
8. Al efectuar una auditoría de las cuentas de una empresa, es típico que el auditor seleccione una muestra de cuentas y verifique los saldos con los clientes. Explique cómo tomaría la muestra con el muestreo por cuotas.

9. El supervisor de la cadena de ferreterías Freund desea estimar la proporción de circuitos integrados defectuosos enviados a sus almacenes por determinado fabricante. Los circuitos vienen en paquetes de 25 cajas, cada una de las cuales contiene 10 de ellos. ¿Cómo tomaría una muestra con el muestreo por conglomerados?
- Observación.* Tanto los paquetes como las cajas pueden usarse como conglomerados.
10. A continuación se presenta la siguiente información sobre las concentraciones de azúcar en la sangre, extraída en ayunas a 60 individuos aparentemente normales:

Núm. del individuo	Concentración	Núm. del individuo	Concentración	Núm. del individuo	Concentración
1	91	21	92	41	106
2	94	22	117	42	85
3	115	23	98	43	93
4	85	24	89	44	103
5	89	25	105	45	119
6	107	26	101	46	90
7	94	27	81	47	82
8	105	28	108	48	90
9	94	29	94	49	113
10	103	30	104	50	104
11	104	31	107	51	97
12	105	32	94	52	101
13	88	33	101	53	90
14	107	34	95	54	88
15	90	35	80	55	108
16	95	36	194	56	95
17	104	37	94	57	100
18	93	38	102	58	103
19	109	39	89	59	108
20	87	40	98	60	85

Tome de esta población una muestra de $n = 15$, usando la tabla 1.1 de números aleatorios.