

Unidad 7

- Nociones de muestreo.

Capítulo

6

NOCIONES DE MUESTREO

6.1 INTRODUCCIÓN

En lo ya estudiado se ha hecho mención de los conceptos de población, de muestra y de inferencia estadística. Como la inferencia estadística respecto de los parámetros poblacionales con base en estadística de muestras es el fin primordial de este texto, son oportunos otros ejemplos para ilustrar dichos conceptos y otros con éstos relacionados. Por ejemplo, muchos han hablado de suprimir la pena de muerte; cabe preguntarse si sus opiniones a este respecto las comparten los estudiantes que están recibiendo educación universitaria, de quienes se presume mayor sensibilidad a las cuestiones importantes que la sociedad enfrenta. Acaso interese saber el porcentaje real de estudiantes universitarios que no estén de acuerdo con la pena de muerte. Se podría obtener este porcentaje al interrogar a todos los estudiantes si están en contra de la pena de muerte y registrar su respuesta. Pero el procedimiento exigiría mucho tiempo y probablemente sería impracticable. Para superar tal dificultad, se podría hacer la pregunta solamente a una porción del conjunto de estudiantes y tratar luego de deducir la actitud de todos los estudiantes con base en las respuestas de dicha porción de estudiantes.

El siguiente es otro ejemplo de inferencia estadística. La porción de estudiantes interrogados se llama técnicamente una *muestra* y el total de estudiantes, una *población*. Un procedimiento esencial de la inferencia estadística es tomar una muestra representativa de la población, y luego, con base en la información obtenida de la muestra,

hacer una afirmación útil acerca de la característica de la población. Para lograr este objetivo, como ya se señaló, la muestra debe ser *representativa*, entendida como tal una muestra que refleje las características fundamentales de la población. La representatividad está ligada a su vez a la aleatoriedad en la selección de la muestra. Esta aleatoriedad puede darse de distintas formas o a distinto nivel, dependiendo de lo que el investigador conozca acerca de la población y de la estructura misma de ésta. Tal circunstancia da lugar a la consideración de distintos procedimientos de selección de la muestra (*métodos de muestreo*), los cuales tienen como fin fundamental proporcionarnos distintas alternativas para que con cierto grado de confianza la muestra escogida sea efectivamente representativa.

6.2 DISEÑO DE MUESTREO

Antes de entrar en la discusión de las distintas formas como puede diseñarse el proceso de selección de la muestra, revisemos una vez más algunos conceptos ya tratados con el fin de que podamos precisar otros nuevos relacionados con los ya mencionados.

En primer lugar, recordemos que una variable aleatoria es una regla (función) que la asocia a cada resultado del experimento un único número real. Este conjunto de números corresponde esencialmente a la población en el sentido estadístico aunque es justo reconocer que esta concepción de población difiere de la idea que tenemos del término cuando se emplea en otras instancias.

Con el propósito de aclarar ideas, supongamos que se toma en consideración el conjunto de todos los estudiantes de la universidad. De los miembros de este conjunto pueden interesarnos distintos *atributos* como son: el puntaje obtenido por cada estudiante a su ingreso a la universidad, la cantidad de dinero gastada por cada uno de ellos en un año académico o la situación académica del estudiante en cuanto aprobado o aplazado. Los dos primeros atributos son *cuantitativos* por naturaleza, pero el tercero es *cualitativo*. A este tercer atributo le podemos dar carácter numérico al asignarle, por ejemplo, el valor 1 al aprobado y el valor 0 al aplazado.

También es importante notar que el conjunto de estudiantes (que podríamos llamar *población objetivo*) permanece inalterable a través de los distintos atributos considerados. Precisamente, el papel fundamental que desempeña la variable aleatoria es el de señalarnos cuál es el atributo de interés tomado en cuenta al estudiar una población objetivo dada.

Como ya fue comentado, cada valor de la variable aleatoria se identifica con un evento del espacio muestral y así a cada uno de estos valores le podemos asociar un valor de probabilidad para obtener lo que se llamó antes distribución de la variable, conocida también como *distribución de la población*. Así por ejemplo, decimos que los puntajes de ingreso de los estudiantes (población) tienen distribución o están distribuidos normalmente. Dentro de este contexto, usted escuchará y se acostumbrará a utilizar expresiones tales como, "la población tiene distribución normal", "la población tiene distribución de Poisson" o "la población tiene distribución binomial", etc.

Hay que distinguir entre población finita y población infinita. Se dice que una población es *finita* si el número de sus elementos corresponde a un número natural determinado, no importa que sea muy grande. Este es el caso que se da, por ejemplo, si consideramos las ventas individuales de todas las sociedades comerciales de los

Estados Unidos, los salarios recibidos por cada uno y todos los miembros de un sindicato, los títulos recibidos por cada uno y todos los estudiantes de un país.

En cambio, una población *infinita* es la que se puede poner en correspondencia con el conjunto de los números naturales (*infinita discreta*) o no es posible ponerla en correspondencia con ningún subconjunto de los naturales (*infinita continua*). Ejemplos de poblaciones con esta particularidad podrían ser el número de llamadas que llegan al conmutador de una gran empresa durante una semana (*infinita discreta*) o el tiempo de vida de una bombilla que se ha encendido hasta el término de su duración (*infinita continua*).

Con la finalidad de precisar aún más la idea de población infinita, suponga que un investigador se interese por los pesos de los cerdos en el pasado, en el presente y en el futuro. Si quisiera representar la colección de todos los cerditos que han nacido o van a nacer, se vería ante una verdadera "montaña" de cerditos. Si tratara de considerar los pesos de éstos, se las vería con millones de millones de pesos, la mayoría de los cuales no pueden siquiera determinarse, pues hay innumerables cerdos que aún están por nacer y habrá otros cerdos en el pasado que tal vez no se pesaron. Para tratar una población como ésta, los distintos pesos se relacionan con sus respectivas frecuencias relativas (probabilidad) que llevados al histograma y al tomar aproximaciones en libras se podría obtener una representación como se muestra en la figura 6.1, la cual correspondería a una distribución de probabilidad discreta.

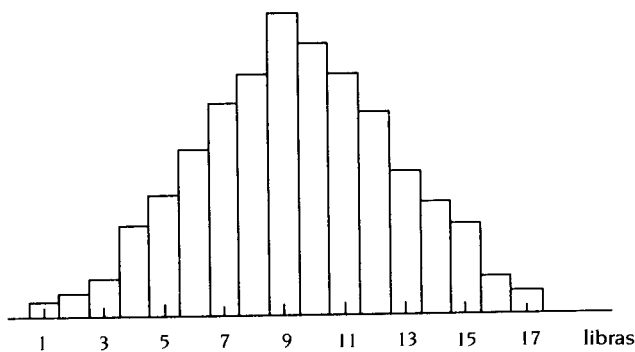


Figura 6.1 Pesos de los cerdos al nacer.

Si los pesos no se aproximan en libras o a una subdivisión de las mismas, éstos serían tratados como una variable continua y la población quedaría representada como una distribución continua, como se señala en la figura 6.2.

Cuando la población es infinita es imposible tomar una muestra tan grande como la población; realmente es posible, pero por lo general no tiene sentido tomar muestras tan grandes como una población finita. Así que, ya sea la población finita o infinita, la razón de ser de la estadística en lo que a la parte inferencial se refiere es la de proporcionar procedimientos mediante los cuales partiendo de la información de las muestras llegar a conclusiones referentes a la población. Para evitar resultados sin validez o sin lógica, es necesario estudiar la teoría del muestreo.

Estamos ahora en condiciones de entender el porqué de la necesidad de diseñar un muestreo.

Se entiende por *diseño de muestreo* el plan que se llevará a cabo para escoger la muestra, de tal forma que exista un convencimiento bien fundado de que la muestra

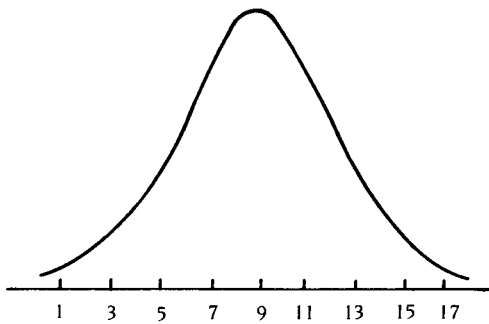


Figura 6.2 Pesos de los cerdos al nacer.

sea representativa. Para evaluar un diseño de muestreo generalmente se toman en cuenta dos criterios que son: la *fiabilidad* y la *efectividad*. La fiabilidad está ligada al denominado *error de muestreo* y la efectividad al costo del muestreo.

Una cuestión parece evidente. Cuando se procede por muestreo, se espera que existan errores. Un *error de muestreo* es la diferencia entre el valor de una estadística obtenido mediante los datos muestrales y el valor del correspondiente parámetro de la población debido a variaciones fortuitas en la selección de las unidades. La varianza de esta estadística (estimador) determina la fiabilidad del muestreo. Cuanto mayor sea la varianza del caso, menor es la *fiabilidad* del resultado de la muestra. Es bueno señalar que se presenta otro tipo de error que no tiene nada que ver con el muestreo; nos referimos al *error no muestral* o *sistemático*, el cual es el resultado de instrumentos de medición incorrectos, cuestionarios mal elaborados, etc. Las mediciones incorrectas no son resultados de la selección de la muestra; pueden darse aun al tomar todos los elementos de la población.

Un diseño de muestreo se considera *efectivo* si se obtiene el mismo grado de fiabilidad al menor costo posible. Un diseño de muestreo se considera más efectivo que otro si el primero tiene menor costo que el segundo dentro del mismo grado de fiabilidad.

En la recolección de los datos que van a servir de soporte para el estudio de la población se deben tener en cuenta dos aspectos básicos. El primero es si un mismo dato puede darse o no, tantas veces como sea posible en la muestra; de esta manera se origina lo que se llama *muestreo con repetición* y *muestreo sin repetición*. Lo segundo es lo referente a la manera como son recolectados los datos y hablamos así de *muestreo no probabilístico* y *muestreo probabilístico*.

6.2.1 Muestreo con repetición

En el capítulo cuatro se consideró el caso de muestras con repetición. Al retomar el tema decimos que se presenta *el muestreo con repetición cuando cada unidad, que forma parte de la muestra, una vez analizada se retorna a la población*. Esta es la forma más simple para describir un muestreo con repetición. Sin embargo, si se tiene en cuenta que lo estudiado por la estadística no son los objetos en sí, sino los valores de la variable considerada, podemos concebir el muestreo con repetición como aquel caso en donde existen "muchísimas" unidades que ofrecen la misma respuesta para la variable y que por tanto, al registrar un valor determinado quedan aún "infinitud" de unidades que darán la misma respuesta. Interpretado el muestreo con repetición

de esta forma, queda claro que para poblaciones infinitas el muestreo debe darse necesariamente con repetición. Ahora bien, si la población es pequeña, el muestreo con repetición sólo tiene sentido, si cada unidad una vez observada, se retorna físicamente a la población. El muestreo con repetición es fundamental en el estudio de problemas estadísticos mediante simulación.

6.2.2 Muestreo sin repetición

Es aquel en que cada unidad que forma parte de la muestra, una vez observada, no se retorna a la población. Esta circunstancia adquiere sentido sólo en poblaciones finitas.

Por lo expuesto, solamente en poblaciones finitas existe alternativa para considerar un muestreo con o sin repetición. Cuando en una población finita se hace un muestreo con repetición, las técnicas estadísticas para el caso de poblaciones infinitas son aplicables en estas circunstancias. Esto es, al realizar un muestreo de este tipo a partir de una población finita, "técnicamente" se considera a la misma como si fuera infinita.

6.2.3 Muestreo intencional, por criterio o subjetivo

Este procedimiento de escogencia de la muestra se basa en la experiencia personal que tenga el investigador, quien por su propio criterio decide qué muestra es representativa. Debido al factor subjetivo en la selección de la muestra, ésta puede resultar influida por las preferencias o tendencias del seleccionador. Este tipo de muestreo carece de bases teóricas que nos permitan evaluar su efectividad y fiabilidad. No obstante, es empleado con bastante regularidad en estudios estadísticos de auditoría por muestreo.

6.2.4 Muestreo por cuotas

En este caso el agente visitador o enumerador recoge información de personas o familias en número proporcional al de las que cumplen determinadas condiciones en la población y puede elegirlos a su arbitrio dentro de grupos establecidos de sexo, edad y ciertos niveles socioeconómicos. Así por ejemplo, se fija que el 18% de la muestra estará constituido por mujeres menores de 30 años que tengan educación secundaria y trabajen en determinado sector de la economía. En este muestreo se presenta también la subjetividad, pero a la vez puede ser asimilado a un tipo de muestreo que comentaremos más adelante, llamado **muestreo estratificado**.

6.2.5 Muestreo sin norma o circunstancial

En este tipo de muestreo la muestra se toma según lo que salga bien, por razones de comodidad, circunstancia y aun de capricho. Cuando la población es homogénea, la representatividad de la muestra puede ser satisfactoria, como cuando se trata de cierto tipo de bienes, como licores, por ejemplo. A los tipos de muestreo antes citados se les denomina *no probabilísticos*.

6.2.6 Muestreo probabilístico – Muestreo aleatorio simple

El muestreo probabilístico está caracterizado porque de antemano puede calcularse la probabilidad de obtener cada una de las muestras que sea posible seleccionar.

Dentro de esta clase de muestreo encontramos el denominado *muestreo aleatorio simple* (M.A.S.), el cual está caracterizado porque cada una de las muestras de tamaño n tiene la misma probabilidad de ser seleccionada. Esto es, si se trata de una población de tamaño N y de ella se extraen muestras no ordenadas sin repetición de tamaño n , entonces cada una de las $\binom{N}{n}$ muestras posibles que podemos formar tiene una probabilidad $\frac{1}{\binom{N}{n}}$ de ser escogida.

Así por ejemplo, supongamos que cierta compañía tiene en su nómina 30 vendedores ($N = 30$), tres de los cuales pertenecen al mismo grupo, llamémoslos A, B, C . Si se toma una muestra aleatoria simple sin repetición de tamaño tres ($n = 3$), la probabilidad de que se tome cualquier combinación de tres vendedores es

$$\frac{1}{\binom{N}{n}} = \frac{1}{\binom{30}{3}} = \frac{1}{\frac{30!}{(27!)(3!)}} = \frac{1 \times 2 \times 3}{28 \times 29 \times 30} = \frac{1}{4,060}$$

Así que la probabilidad de que los vendedores A, B, C entren en la muestra debe ser también $\frac{1}{4,060}$.

Este problema también puede ser resuelto recurriendo al principio de multiplicación de la siguiente forma:

Número de formas como puede escogerse el primer vendedor = $n_1 = 30$.
 Número de formas como puede escogerse el segundo vendedor = $n_2 = 29$.
 Número de formas como puede escogerse el tercer vendedor = $n_3 = 28$.
 Número de formas como pueden escogerse conjuntamente tres vendedores cualesquiera de los 30 = $n_1 \times n_2 \times n_3 = 30 \times 29 \times 28 = 24,360$.

Por otra parte se tiene,

Número de formas como puede escogerse el primer vendedor entre $A, B, C = 3$.
 Número de formas como puede escogerse el segundo vendedor = 2.
 Número de formas como puede escogerse el tercer vendedor = 1.
 Número de formas como puede escogerse los tres vendedores = $3 \times 2 \times 1 = 6$.

Así que la probabilidad pedida es $\frac{6}{24,360} = \frac{1}{4,060}$

Al utilizar un procedimiento más refinado de la probabilidad, el problema citado sería resuelto así:

Sean los eventos,

E_1 : uno de los vendedores, A, B, C es escogido en el primer intento
 E_2 : uno de los dos restantes es escogido en el segundo intento
 E_3 : el tercero es escogido en el tercer intento

Por lo anterior se tiene que $E_1 \cap E_2 \cap E_3$ representa el evento "los tres vendedores A, B, C son escogidos en los tres intentos".

Debemos calcular $P[E_1 \cap E_2 \cap E_3]$

Por el teorema de multiplicación se tiene, $P[E_1 \cap E_2 \cap E_3] = P[E_1] P[E_2|E_1]$

$$P[E_3|E_1 \cap E_2] = \frac{3}{30} \times \frac{2}{29} \times \frac{1}{28} = \frac{1}{4,060}$$

Para llevar a cabo un muestreo aleatorio simple se requiere que la muestra sea escogida siguiendo algún procedimiento que garantice aleatoriedad en la selección de cada una de las unidades que conformarán la muestra. Este procedimiento puede realizarse de varias formas, pero siempre partiendo del supuesto de que las unidades u objetos que constituyen la población son numerados. Estos números asignados a las unidades se anotan en fichas o tarjetas por separado, se colocan en una bolsa, se revuelven y luego se sacan una por una hasta completar n números, de acuerdo con el tamaño de muestra escogido. Si se trata de muestreo con repetición, cada ficha o papeleta deberá ser retornada a la bolsa antes de escoger la siguiente. En el caso de que sea sin sustitución, no se retorna.

El procedimiento puede simplificarse al utilizar una tabla de números aleatorios. El apéndice I es una de estas tablas. La construcción de una similar puede darse por distintos procedimientos; desde el más simple, como es el caso del lanzamiento de una moneda regular (al trabajar con sistema binario) o recurriendo a mecanismos más complejos que garanticen la aleatoriedad, tales como:

- Escoger ciertos dígitos comprendidos en el desarrollo decimal de 20 cifras de los logaritmos de los números (Fisher y Yates, 1943).
- Dividir un disco en diez sectores circulares iguales, cada uno marcado con una cifra diferente desde 0 hasta 9 que luego se hace girar (Kendall, Babington Smith, 1940).
- Registrar los números de las loterías (Royo y Ferrer, 1954).

La tabla de números aleatorios más ampliamente conocida es la construida en 1955 por la Rand Corporation, la cual comprende un millón de dígitos obtenidos mediante dispositivos electrónicos.

A los anteriores métodos, caracterizados por su sofisticación, podemos adicionar uno más sencillo de llevar a cabo, el cual consiste en colocar en una bolsa 10 fichas numeradas con el 0, 1, 2, ..., 9 y luego, mediante la extracción de las fichas con repetición, formar la tabla.

A continuación explicaremos el proceso para escoger aleatoriamente $n = 10$ unidades de una población formada por $N = 500$, por ejemplo, a partir de la tabla I:

En primer lugar debe localizarse un punto de partida aleatorio en la tabla. Esto puede hacerse de varias formas; una de ellas consiste en alejar la vista de la tabla mientras se deja caer un lápiz de punta. El punto de partida aleatorio es el dígito más próximo al sitio en donde el lápiz tocó la página.

Siguiendo este procedimiento, supongamos que el punto de partida quedó determinado por la intersección de la fila 22 y la columna 9 de la primera página de la tabla I, el dígito en este punto es 9.

Dado que se tienen 500 unidades para elegir, debemos utilizar números aleatorios comprendidos entre 1 y 500. El primer número de tres dígitos al empezar en el punto de partida aleatorio es 982, un número que no puede utilizarse (el máximo es 500). Se recorre entonces la tabla hacia abajo y encontramos el 498. Este número sí nos sirve y así se tomará la unidad identificada con el número 498 a la cual se le tomará el valor correspondiente a la variable que se esté estudiando. Al continuar con el proceso encontramos los números 421, 291 (aquí termina la página). Pasamos a la siguiente conservando la columna 9a., y encontramos los números 462, 669 (que no sirve), 737 (que tampoco sirve), 139, 177, 947, 546, 072, 914, 360, 978, 062, 300.

Una vez que hemos llegado a este punto se han completado los números de las 10 unidades que serán seleccionadas. Tales unidades son las identificadas con los números: 498, 421, 291, 462, 139, 177, 072, 360, 062 y 300. Escogidas éstas, se registran los valores de la variable que se esté estudiando. Dichos valores serán los que servirán de base para llevar a cabo el posible proceso inferencial.

A estas alturas, algunas precisiones son necesarias. En primer lugar, si al hacer el recorrido hacia abajo en la tabla llegamos al final y aún no hemos completado el número de unidades que debemos escoger, podemos pasar a la columna en donde comience el segundo bloque de dígitos (en nuestro ejemplo sería la columna 12) o también podría repetirse la experiencia de dejar caer el lápiz. Si la tabla se agota y todavía no se ha completado el número de unidades, deje caer el lápiz de nuevo o utilice otra tabla (es lo más aconsejado). En una tabla de números aleatorios para la escogencia de los números (aleatorios) puede moverse en cualquier sentido (vertical, horizontal, diagonalmente, etc). En segundo lugar si el muestreo se hace sin repetición, al encontrar un número repetido éste debe descartarse. Pero si es con repetición un mismo número se tendrá en cuenta tantas veces como aparezca, hasta completar la muestra.

A continuación ilustramos el procedimiento, además de otros aspectos tratados anteriormente, de cómo obtener una muestra aleatoria de determinado tamaño a partir de una población dada.

Supongamos que en la actualidad hay 100 estudiantes de la universidad que están tomando un primer curso de estadística. Esta será nuestra población objetivo y su tamaño es $N = 100$. En circunstancias como éstas podemos numerar los nombres de los estudiantes, de 1 a 100. Al hacer esta numeración (y esto debe tenerse siempre presente), por lo antes expuesto, la selección aleatoria mediante el uso de una tabla requiere de la consideración de números de tres dígitos y sólo uno de ellos presenta esta cantidad de dígitos. Entonces, con el fin de tomar en cuenta números con sólo dos dígitos, comenzamos la numeración desde 00 hasta 99. El atributo que vamos a considerar es el número de letras que tiene el nombre del estudiante escogido. De esta forma queda definida la variable aleatoria, X = número de letras del nombre del estudiante escogido.

Nos proponemos escoger aleatoriamente ocho nombres.

Ahora nos remitimos a la tabla I y suponemos que la punta del lápiz se ubicó en la fila 12 y columna 33 de la página 424. El número aleatorio de dos dígitos que ocupa esa posición es el 74. Esto nos indica que debemos comenzar con el nombre o estudiante que se encuentre reseñado con ese número. Si es "Adriana", entonces escribimos $x_1 = 7$; al seguir hacia abajo en la tabla encontramos el 67; digamos que corresponde a "Gregorio"; entonces $x_2 = 8$. En resumen, admitamos que ocurrió la siguiente secuencia de nombres de acuerdo con el número dado en la tabla:

Número aleatorio	Nombre	Número de letras (X)
74	Adriana	7
67	Gregorio	8
04	Mónica	6
54	Francisco	9
09	Rafael	6
69	Marlén	6
01	Martha	6

De esta forma se obtiene una muestra aleatoria simple extraída de la población de los 100 estudiantes que toman el primer curso de estadística.

Muchos de los computadores más comunes y aun calculadoras de bolsillo tienen capacidad para generar números aleatorios y los investigadores los utilizan como alternativa de las tablas impresas de números aleatorios. Sin embargo, los números "aleatorios" suministrados por la mayoría de los computadores son de hecho, "*números pseudoaleatorios*", debido a que son resultados de una fórmula determinística. Algunos estudiosos del tema, como George S. Fishman, afirman que dichos números sirven satisfactoriamente para muchos fines prácticos.

Finalmente, debemos decir que el procedimiento aleatorio simple puede utilizarse si las unidades de muestreo pueden identificarse fácilmente y si la población es homogénea y pequeña. Pero el método es costoso y lleva tiempo, especialmente si la población es grande, puesto que es necesario numerar todos los elementos. En realidad, muchos datos de las ciencias sociales no pueden recolectarse efectivamente por los procedimientos que acaban de describirse. Además, si los elementos próximos entre sí son más homogéneos que los que están apartados, una muestra aleatoria simple puede no ser representativa de la población. Por ejemplo, al estudiar los ingresos familiares en una gran ciudad, el muestreo aleatorio simple podría seleccionar, como es de pensarlo, un número relativamente elevado de familias con ingresos similares, lo cual resultaría posiblemente en un error de muestreo relativamente considerable y en una varianza elevada y por consiguiente en un bajo grado de fiabilidad para la estimación. En tales circunstancias, otros diseños de muestreo aleatorio serían más apropiados y son los que a menudo se emplean. Estos tipos de muestreo se llaman *restringidos*, en contraposición al que hemos venido considerando que se denomina *irrestringido*.

6.2.7 Muestreo sistemático

Uno de los diseños de muestreo *restringido* más utilizado es el *muestreo sistemático*. Consiste este método en particionar la población en n (tamaño de la muestra) grupos de tamaño m . Del primer grupo se toma al azar un elemento, digamos que es el numerado con k . Del segundo se toma el elemento con el número $k + m$, y así sucesivamente hasta completar la muestra.

Así por ejemplo, se va a tomar una muestra de tamaño 5 de una población de tamaño 20. Entonces las 20 unidades se particionan en 5 grupos de 4 elementos cada uno. El primero lo forman las unidades 1, 2, 3, 4; el segundo las unidades 5, 6, 7 y 8; y así sucesivamente. Ahora, del primer grupo seleccionamos aleatoriamente una unidad, digamos que es la identificada con el número 3; entonces del segundo se tomará la identificada con el número 7; del tercero el 11; del cuarto el 15 y del quinto el 19.

Algunos tratadistas recomiendan el muestreo sistemático siempre que se den las circunstancias puesto que consideran que debido a la mecánica de este muestreo hay mayor posibilidad de que la muestra resulte repartida en toda la población.

El muestreo sistemático se puede emplear fácilmente cuando se dispone previamente de una lista de las unidades de la población como en una guía telefónica. A veces las muestras tomadas para control de calidad estadístico son muestras sistemáticas, como es el caso cuando se obtiene una muestra al tomar un artículo de la corriente de producción usual a intervalos regulares.

El procedimiento del muestreo aleatorio sistemático selecciona una muestra más representativa que el muestreo aleatorio simple si los elementos cercanos de la población se asemejan más entre sí de lo que se parecen los que quedan distantes.

Si se va a estimar el ingreso promedio familiar en Los Ángeles, por ejemplo, las residencias de Beverly Hills y las de Watts estarán en proporción y por tanto igualmente representadas, si se utiliza el muestreo sistemático, mientras que el muestreo aleatorio simple podría seleccionar un número relativamente grande o pequeño de residencias de cada comunidad en que la situación económica de los residentes tiende a ser parecida. En el último caso, se tendría un mayor error de muestreo. Como la muestra sistemática no está sujeta a este tipo de error, tiende evidentemente a ser más representativa.

Por otra parte, la muestra sistemática es menos representativa que la muestra aleatoria simple en situaciones en que existe *periodicidad* oculta en la población, la cual es un movimiento cíclico o periódico de los datos con la longitud del ciclo con aproximación al valor m , tamaño de cada grupo. Por ejemplo, si se va a estimar la venta media diaria de tickets en una gran compañía aérea, una muestra que consista en las ventas de cada séptimo día no sería representativa de las ventas diarias de la compañía. La muestra incluirá ventas, por ejemplo, de todos los viernes o de todos los sábados y presentará un aspecto errado debido a la configuración semanal repetitiva de las ventas de billetes.

Por lo demás, el muestreo sistemático tiene la desventaja de numerar o de ordenar los elementos de una gran población, lo cual podría ser físicamente imposible si la población abarca todo un país o una zona geográfica considerable.

6.2.8 Muestreo estratificado

Otro plan de muestreo restringido que puede resultar más efectivo que el muestreo aleatorio simple es el *muestreo aleatorio estratificado*, procedimiento este que exige tener conocimiento previo de la población. Consiste en particionar la población en subpoblaciones al agrupar en ellas los elementos más parecidos entre sí. Cada subpoblación se llama *estrato* y en cada una de ellas se lleva a cabo un muestreo aleatorio simple para escoger la muestra. La muestra global se obtiene al combinar las submuestras de todos los estratos.

El reparto del tamaño de la muestra en los distintos estratos se llama *afijación*. Cuando la muestra se reparte por partes iguales en los estratos la afijación se dice *igual* o *uniforme*. Si se hace proporcional al número de elementos de cada estrato, tenemos entonces la *afijación proporcional*. Si se hace proporcional al número de elementos y a la varianza de cada estrato, se tiene en cambio la *afijación óptima*.

El muestreo por estratos es el más efectivo cuando se trata de poblaciones heterogéneas, tales como datos de desempleo (que varían de industria a industria y de ocupación a ocupación), ventas al por menor (que difieren entre las distintas regiones geográficas), o las actitudes de los consumidores respecto de los nuevos modelos de automóviles (en las que influyen factores tales como el sexo, la edad, y la categoría de los ingresos). Al hacerse 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. Así se puede controlar la proporción de cada estrato en la muestra global y al no dejarla al azar queda asegurado el carácter representativo de la muestra.

Si la varianza de la característica observada de cada estrato es menor que la de toda la población, como es lo usual debido a la mayor uniformidad dentro del estrato, resultará aumentada la fiabilidad para un tamaño de muestra dado o bien la efectividad para un grado dado de fiabilidad. Cuanto mayor sea la reducción de la varianza dentro del grupo tanto más grande será el aumento debido a la estratificación.

El aumento de la fiabilidad y de la efectividad puede incrementarse al clasificar aún más los estratos en subestratos, si ello fuera posible. Por ejemplo, al hacer una encuesta sobre una elección se pueden tomar como estratos los partidos. Puesto que las preferencias electorales están influidas a menudo por factores tales como el sexo, la edad, la raza, la religión y la formación académica, se podría subdividir por los estratos de acuerdo con estos factores. Dicho procedimiento de doble clasificación se denomina *doble estratificación*.

Las muestras resultantes se consideran muestras probabilísticas independientemente de como se haya estratificado la población. Para definir los estratos se pueden emplear datos anteriores, la intuición, un criterio sólido o bien, resultados preliminares procedentes de otros estudios. El muestreo por estratos es por consiguiente una combinación de submuestras de los estratos, los cuales son muestras aleatorias simples o sistemáticas.

En cuanto tales, todo elemento disponible de cada estrato tiene igual probabilidad de ser seleccionado y ésta será la situación, aun en el caso en que la muestra no sea proporcional, en el cual las probabilidades de ser seleccionado cada elemento individual de la población no son iguales.

6.2.9 Muestreo por conglomerados

El muestreo por conglomerados es un esquema en el cual se eligen los individuos por grupos llamados *conglomerados*. Cada conglomerado que resulte en la muestra se revisa total o parcialmente. Así, por ejemplo, si se tienen 100 cajas de manzanas y de éstas nos interesa, por ejemplo, estudiar su sabor. Lo indicado sería seleccionar aleatoriamente cierto número de cajas y de éstas escoger las manzanas para nuestro estudio. En este caso cada caja constituye un *conglomerado*. En el plan de muestreo por conglomerados los mejores resultados se obtienen cuando las diferencias entre éstos, respecto de la característica de interés, es lo más pequeña posible y las diferencias entre los elementos individuales dentro de cada conglomerado se hace tan grande como sea posible. Al tomar en cuenta esta particularidad del muestreo por conglomerados podemos decir que es totalmente opuesto al muestreo estratificado. Lo ideal sería que cada conglomerado constituyese una miniatura de la población y así un solo conglomerado sería una muestra satisfactoria. Pero en la práctica este requisito es difícil de llenar; no obstante es un principio importante que se debe tener en cuenta cuando se planea un muestreo por conglomerados.

Con frecuencia los conglomerados se llaman *unidades de muestreo primario*. Si todos los elementos o unidades elementales de los conglomerados 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 denomina *muestreo en dos etapas o muestreo bietápico*. Si intervienen más de dos etapas en la obtención de la muestra global, se dice que es un *muestreo polietápico*. Por ejemplo, en una encuesta sobre las actitudes de los estudiantes de enseñanza superior de un país, se toman como conglomerados o unidades de muestreo primarias las escuelas superiores y las universidades. Esta es la primera etapa; la segunda podría ser la selección de divisiones dentro de cada una de las escuelas superiores y universidades escogidas. Por último, en la tercera etapa, se toman estudiantes de todas las divisiones seleccionadas para constituir la muestra. Desde luego, los métodos de muestreo aleatorio se emplean en cada etapa. El objetivo del muestreo por conglomerados es el estudio de las características de los elementos individuales o unidades elementales, si bien se eligen inicialmente las unidades de muestreo primarias.

La ventaja principal del muestreo por conglomerados consiste en la gran reducción de costos para un grado dado de fiabilidad. Por ejemplo, el costo de tomar una muestra aleatoria simple de 5,000 mujeres en la población de amas de casa de un país sería tremendo. Si primero se tomara una muestra de 50 municipalidades, y luego se tomase una submuestra de cada una de las 50 municipalidades ya escogidas, la muestra 5,000 podría obtenerse a mucho menor costo del que se necesita para una muestra aleatoria simple de ese tamaño. A esto se debe que cuando se hacen muestreos de una población distribuida geográficamente se tomen como unidades primarias de muestreo las regiones geográficas, procedimiento que se conoce como *muestreo por zonas*.

El muestreo por conglomerados se utiliza a menudo en el control estadístico de la calidad. 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 comprobadas. O bien se puede tomar una muestra de cada lote para su comprobación, lo cual sería un muestreo en dos etapas.

La desventaja del muestreo por conglomerados consiste en su relativa ausencia de fiabilidad para un tamaño de muestra dado. La varianza de una muestra por conglomerados tiende a ser mayor que la de una muestra aleatoria tomada por cualquier otro de los procedimientos antes vistos. Pero esta desventaja no tiene que llevar necesariamente a la conclusión de que el muestreo por conglomerados no pueda ser más efectivo que cualesquiera otros planes de muestreo. Si el costo de seleccionar una unidad elemental se reduce mucho, como ocurre con el muestreo por conglomerados, la misma cuantía de gastos permitiría seleccionar una muestra mucho más grande. Como veremos luego, una muestra mayor tiene por resultado un error de muestreo menor. Es posible seleccionar una muestra por conglomerados lo suficientemente grande para lograr mayor fiabilidad al mismo costo que con cualquier otro muestreo aleatorio. Como suelen decir los estadísticos, la consideración del costo lleva con frecuencia al muestreo por conglomerados.

Ejercicios

6.1

1. Defina los siguientes términos de la manera más precisa posible:

- | | |
|------------------------------|-------------------------------|
| a) Muestreo aleatorio simple | b) Muestreo sistemático |
| c) Muestreo estratificado | d) Muestreo por conglomerados |

2. Explique la diferencia que existe en los siguientes términos:

- Fiabilidad del muestreo y efectividad del muestreo
- Muestreo uniforme y muestreo proporcional

3. Determine la población y proponga el diseño de muestreo más adecuado para estudiar las siguientes situaciones.

- En una industria que produce etiquetas para marcas de camisas existen 500 máquinas que elaboran la etiqueta. De estas 500 máquinas, 250 son operadas manualmente; 150 son semiautomáticas y 100 son totalmente automáticas.
- Una industria está considerando la revisión del sistema vacacional de sus empleados y desea determinar la proporción de empleados que apoyan la nueva propuesta. La industria cuenta con 50 plantas separadas, localizadas en todo el país.
- Un investigador de mercados desea hacer una encuesta de las actitudes de las amas de casa respecto de un nuevo aparato doméstico en cierta zona de la ciudad.

4. Se desea tomar una muestra consistente en el 10% de los alumnos inscritos en una universidad. Describa el procedimiento a seguir si la muestra es seleccionada por los métodos siguientes:
- Utilizando una tabla de números aleatorios
 - Utilizando bolas colocadas en una urna
 - Utilizando muestreo sistemático
5. Se va a hacer una encuesta para determinar las preferencias de los electores en una ciudad. Diseñe un plan de muestreo para que la muestra sea representativa de la población de todos los votantes.
6. Defina los conceptos siguientes:
- Muestreo por zonas
 - Muestreo aleatorio
 - Estratificación doble
 - Afijación óptima
7. Presente un caso en donde sea posible realizar un muestreo:
- Por criterio
 - Circunstancial
8. Compare las ventajas y desventajas de los siguientes pares de diseño de muestreo:
- Muestreo aleatorio simple y muestreo sistemático.
 - Muestreo estratificado y muestreo por conglomerados.
9. Se va a seleccionar una muestra de 50 trabajadores entre los miembros de un sindicato local que consta de 8,000 trabajadores. Describa el procedimiento para lograr tal muestra al utilizar la tabla de números aleatorios, apéndice I.
10. Una gran compañía tiene 300,000 empleados, cuya distribución por edades es:

Edad en años	Porcentaje
25 o menos	15
26-35	30
36-45	25
46-55	20
56 o más	10

Se desea una muestra del 2% de todos los empleados. Diseñe un plan de muestreo de manera que cada grupo de edad quede representado proporcionalmente.

11. La gerencia de una revista desea hacer una encuesta sobre la actitud de sus lectores respecto de cierta característica de la revista. Hay 500,000 suscriptores y el cuestionario va a ser enviado al 10% de ellos. ¿Cómo se toma esa muestra si se utiliza la técnica del muestreo sistemático?
12. Se hace una encuesta sobre la actitud de los maestros de escuela, hacia cierta cuestión en un país. Describa el procedimiento para tomar una muestra en tres etapas.
13. ¿En qué caso el muestreo circunstancial resulta efectivo?
14. ¿Cuál es la principal falla que se le señala al muestreo por criterio?
15. En un estudio estadístico puede resultar que sea necesario combinar el muestreo probabilístico con el no probabilístico. Exponga un caso en el que el estudio se inicie con un muestreo no probabilístico y se culmine con uno probabilístico.