MUESTREO

Como ya hemos establecido con anterioridad, el principal objetivo de la estadística es hacer inferencias de una población a partir de los datos de una muestra. El muestreo, es el proceso por el cual se define la muestra de una población en estudio, y es el que nos proporciona las consideraciones a tomar de dicha muestra, es decir, cual será el tamaño de la muestra y como se genera el conjunto de elementos o datos muestrales.

Partiendo de que objeto de nuestro interés es la población. Normalmente, trabajar con la población es muy costoso y el tiempo que requiere recoger los datos puede resultar muy largo. Quizá por esta razón los censos de población se realizan cada 10 años.

Supongamos que nos interesa conocer los hábitos de lectura de un grupo de 30 estudiantes. En este caso no tenemos problemas para trabajar con la población, ya que contamos solamente con 30 personas. Sin embargo, las inferencias sobre sus hábitos estarán solo limitadas a este grupo de estudiantes. Si requerimos un estudio más amplio, por ejemplo para unos 300,000 estudiantes, entonces realizar el estudio en una población tan grande va a requerir que tomemos más tiempo y el costo se eleva considerablemente. En estos casos es que el muestreo puede ser más práctico y económico.

Por otro lado, si al obtener la muestra de la población solo la tomamos de los estudiantes del área de matemáticas, o si la muestra solo incluye a estudiantes de escuelas particulares, etc. En estos casos, los resultados no van a ser muy fiables ya que no incluirán los hábitos de lectura de estudiantes de otras disciplinas, condiciones económicas, sexo, etc.

Así, resulta que no todas las muestras que se pueden extraer de una población son útiles desde el punto de vista estadístico. Ahora comenzaremos el estudio, muy descriptivo y resumido, de diferentes técnicas de muestreo que permiten evitar (o al menos reducir) el impacto sobre el resultado final de los errores que acabamos de mencionar.

Pasos para seleccionar una muestra:

1. Definir el objetivo del estudio.
2. Definir la población objetivo.
3. Seleccionar un procedimiento.
4. Definir el tamaño de la muestra.
5. Seleccionar las unidades muéstrales.

Existen varias clases de muestreo, y se usan ó no de acuerdo al tipo de estudio, objetivo del estudio, la forma de las variables, características de la población, incluso hasta la cantidad de recursos disponibles. Sin embargo, solo tenemos dos métodos para seleccionar muestras de poblaciones: el muestreo determinista, no aleatorio, y el muestreo aleatorio (que incorpora el azar como recurso en el proceso de selección). Cuando este último cumple con la condición de que todos los elementos de la población tienen alguna oportunidad de ser escogidos en la muestra, si la probabilidad correspondiente a cada sujeto de la población es conocida de antemano, recibe el nombre de *muestreo probabilístico*. Una muestra obtenida por muestreo determinista puede basarse en la experiencia de alguien con la población. Algunas veces una muestra determinista se usa como guía o muestra tentativa para decidir cómo tomar una muestra aleatoria más adelante.

Muestreo probabilístico

Forman parte de este tipo de muestreo todos aquellos métodos para los que puede calcularse la probabilidad de extracción de cualquiera de las muestras posibles. Este conjunto de técnicas de muestreo es el más aconsejable, aunque en ocasiones no es posible optar por él. En este caso se habla de muestras probabilísticas, pues no es en rigor correcto hablar de *muestras representativas* dado que, al no conocer las características de la población, no es posible tener certeza de que tal característica se haya conseguido.

La herramienta del muestreo más importante para recoger los datos de la muestra es la encuesta. Algunas definiciones son necesarias,

a) El diseño de muestreo o *diseño de encuesta* especifica el método de obtención de la muestra.
b) Un *elemento* es un objeto del cual se toma una medición. Por ejemplo en el caso de la encuesta de ingreso-gasto de los hogares, ele elemento es la familia,
c) Las *unidades muestrales* son grupos de elementos de una población.

Para obtener una muestra aleatoria, de unidades muestrales, es necesario contar con el ***marco muestral***, que es una lista de unidades muestrales.

Entonces, la primera acción para realizar una encuesta por muestreo consiste en la identificación de las unidades de muestreo y la lista que contiene estas unidades.

Errores en el muestreo.

Las fuentes de error más comunes en el muestreo son:

a) Variación aleatoria. Por ejemplo para determinar el ingreso de los hogares de una comunidad, podemos seleccionar solamente a los estratos más altos, o más aún los niveles de ingreso medios que pueden pasar desapercibido y producir inferencias erróneas.
b) Especificación errónea o deficiente de la población. Este tipo de error lo produce generalmente un marco muestral erróneo. Por ejemplo en una encuesta de opinión sobre las elecciones podemos incluir en el marco muestral a personas que no van a votar el día de la elección, o como es una moda actual realizar las encuestas por teléfono, lo cual impide que participen electores que no tienen teléfono, o no lo pagaron, etc.
c) La no respuesta. Es común suponer que los elementos de la muestra que responden tienen comportamientos similares a los que no responden. Es común en las encuestas de opinión que los que no responden son quienes prefieren que las cosas se queden como están.

Cuando se trabaja con una muestra aleatoria se deben tener en cuenta dos aspectos principales:

- El método de selección
- El tamaño de la muestra

Muestreo simple aleatorio

Es la extracción de una muestra de una población finita, en el que el proceso de extracción es tal que garantiza que cada uno de los elementos de la población tiene la misma oportunidad de ser elegido en la muestra. Nótese que esto no es lo mismo que decir que todos los elementos de la población tienen igual probabilidad de ser elegidos. Esta condición garantiza sin embargo la equiprobabilidad. Muchas de las estimaciones resultantes de un muestreo de este tipo se dicen *insesgadas*. Esto significa en el caso particular (por ejemplo) de un porcentaje, lo siguiente: si en la población un determinado porcentaje de individuos presenta la característica A, la extracción aleatoria garantiza matemáticamente que, por término medio, se obtendrá el mismo porcentaje de datos muestrales con esa característica.

El muestreo aleatorio simple puede ser de tres tipos:

Sin reposición de los elementos, cada elemento extraído se descarta para la subsiguiente extracción. Por ejemplo, si se extrae una muestra de una "población" de bombillas para estimar la vida media de las bombillas que la integran, no será posible medir más que una vez la bombilla seleccionada.

Con reposición de los elementos las observaciones se realizan con reemplazamiento de los individuos, de forma que la población es idéntica en todas las extracciones. En poblaciones muy grandes, la probabilidad de repetir una extracción es tan pequeña que el muestreo puede considerarse sin reposición aunque, realmente, no lo sea.

Con reposición múltiple En poblaciones muy grandes, la probabilidad de repetir una extracción es tan pequeña que el muestreo puede considerarse sin reposición. Cada elemento extraído se descarta para la subsiguiente extracción.

El primer paso en el muestreo simple aleatorio es encontrar el tamaño de la muestra y posteriormente enumerar todos los individuos de la población. Una vez cubiertos estos pasos procedemos a seleccionar la muestra. Hay varias maneras de hacerlo, escribir números en papelitos y meterlos en una urna, posteriormente seleccionamos el número de papelitos que correspondan al tamaño de la muestra y al relacionar estos con la lista nos permitirá obtener la muestra correspondiente. El método más utilizado es utilizar las llamadas tablas de números aleatorios, que representan el procedimiento más habitual para obtener una muestra aleatoria simple dentro de una población finita, estos números pueden ser generados también por medios electrónicos, incluso si no tenemos algo más con los números del directorio telefónico.

Procedimiento de selección de la muestra.

1) Definir la población de estudio.
2) Asignar un número a cada individuo de la población
3) Determinar el tamaño de muestra óptimo o para el estudio.
4) Seleccionar la(s) muestra(s) de manera sistemática por medio de algún medio mecánico (Tablas de números aleatorios, bolas dentro de una bolsa, números aleatorios generados con una calculadora, etc.)
5) Y se eligen tantos individuos como sea necesario para completar el tamaño de muestra que necesitamos.

Para calcular la media poblacional μ y el total poblacional τ, basados en un muestreo aleatorio simple, utilizamos las siguientes formulas;

El cuadro siguiente ilustra una tabla de números aleatorios. La tabla completa se muestra el final.

TABLA DE NUMEROS ALEATORIOS

17841	**49597**	**92623**	**80005**	**11177**	**15145**	**46379**
84970	**47043**	**64048**	**06993**	**17369**	**70932**	**47950**
30524	**27250**	**73072**	**52654**	**33653**	**30422**	**22347**
56211	**27219**	**44652**	**09467**	**62848**	**82479**	**35068**
66110	**69181**	**13200**	**93239**	**25591**	**21248**	**06881**
.....	**.....**	**.....**	**.....**	**......**	**......**	**.......**
.....	**......**	**......**	**......**	**.......**	**......**	**.......**
72208	**67425**	**77273**	**35454**	**43798**	**89958**	**98485**
62663	**32726**	**14266**	**48467**	**36706**	**90411**	**84898**
99530	**11547**	**35629**	**86192**	**25909**	**97084**	**30951**
36626	**80491**	**21369**	**48285**	**59708**	**44408**	**75096**

Fuente. Elaboración propia

Ejemplo. En una compañía con 750 trabajadores se quiere obtener una muestra aleatoria de 15 elementos para un chequeo médico. Los trabajadores fueron numerados del 1 al 750 y mediante una tabla de números aleatorios se procedió a seleccionarlos. El punto de arranque en la tabla se fijó mediante la hora en ese momento, 2:04, por lo tanto se inició en la columna 3, renglón 4. Como los números de los trabajadores van desde 1 hasta 750 solo se toman en cuenta las primeras 3 cifras de cada número que se encuentren en ese rango. En seguida se muestra una parte de la tabla, con el primer y segundo seleccionado:

TABLA DE NUMEROS ALEATORIOS

17841	49597	92623	80005	11177	15145	46379
84970	47043	64048	06993	17369	70932	47950
30524	27250	73072	52654	33653	30422	22347
56211	**272**19	**446**52	**094**67	**628**48	82479	**350**68
66110	**691**81	**132**00	93239	**255**91	**212**48	**068**81
28710	**524**14	**558**93	**256**32	64856	51745	46855
38939	15777	66270	53052	05160	94786	81987
31297	00722	88300	21109	13124	96742	64968
34043	19959	77949	24510	93510	40492	81113
74996	32698	29430	58603	43879	7861	15870
36626	80491	21369	48285	59708	44408	75096

Fuente. Elaboración propia

Es decir, la muestra estaría integrada con los trabajadores, 272, 446, 094, 628, 350, 661, 691, 132, 255, 212, 068, 287, 524, 558 y 256.

Estimador de la media poblacional μ, para un muestreo aleatorio simple

$$\mu = \overline{y} = \sum_{i=1}^{n} \frac{y_i}{n}$$

Varianza estimada del estimador $\overline{y}$

$$\hat{\sigma}_{\overline{y}}^2 = \left(\frac{s^2}{n}\right)\left(\frac{N-n}{N}\right) \qquad \text{donde} \qquad s^2 = \sum_{i=1}^{n} \frac{(y_i - \overline{y})^2}{n-1} = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}$$

Límite para el error de estimación. $\overline{y} \pm 2\hat{\sigma}_{\overline{y}}$

A. *Estimador del total poblacional τ, para un muestreo aleatorio simple*

$$\hat{\tau} = N\overline{y} = \frac{N\sum_{i=1}^{n} y_i}{n}$$

Varianza estimada de $\hat{\tau}$

$$\hat{\sigma}_{\hat{\tau}}^2 = N^2\hat{\sigma}_{\overline{y}}^2 = N^2\left(\frac{s^2}{n}\right)\left(\frac{N-n}{N}\right)$$

Límite para el error de estimación

$$N\overline{y} \pm 2\hat{\sigma}_{\hat{\tau}}$$

Ejemplo. En la tabla aparecen los saldos correspondientes a las cuentas de una muestra de un laboratorio de análisis clínico de un hospital de tamaño $n = 8$. El laboratorio tiene un total de 1000 cuentas por cobrar

a) Estime el saldo promedio para las N cuentas y establezca una cota para el error de estimación.
b) Estime el total τ de los saldos de todas las cuentas y establezca una cota para el error de estimación.

No. Elemento	Adeudo y_i	y_i^2
1	30.2	912.04
2	14.5	210.25
3	33.5	1122.25
4	32	1024
5	17.5	306.25
6	10	100
7	23.4	547.56
8	27.5	756.25
Suma	188.6	4978.6

a) $\overline{y} = \frac{188.6}{8} = \$\ 23.58$

La cota del error $s^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} = \frac{\left(4978.6 - \frac{188.6^2}{8}\right)}{7} = \frac{532.36}{7} = \$\ 76.05$

La varianza estimada es entonces $\hat{\sigma}_{\overline{y}}^2 = \left(\frac{s^2}{n}\right)\left(\frac{N-n}{N}\right) = \left(\frac{76.05}{8}\right)\left(\frac{1000-8}{1000}\right) = 9.43$

Estimación del saldo promedio μ, y una cota de error; $23.58 \pm 2\sqrt{9.43} = \$23.58 \pm 6.14$

b) El total τ de los saldos de todas las cuentas $\hat{\tau} = N\overline{y} = 1000(23.575) = 23575$

Dado que la varianza estimada de $\hat{\tau}$ es $\hat{\sigma}^2_{\hat{\tau}} = N^2\hat{\sigma}^2_{\overline{y}}$ una estimación del total de los saldos. La cota de error que le corresponde es;

$$\hat{\tau} \pm 2\hat{\sigma}_{\hat{\tau}} = N\overline{y} \pm 2N\hat{\sigma}_{\overline{y}} = \$23{,}575 \pm 2(1000)\sqrt{9.43} = \$23{,}575 \pm 9{,}430$$

B. *Estimación de la proporción poblacional para una muestra aleatoria simple.*

Este análisis se realiza cuando estamos interesados en estimar la proporción (porcentaje) de la población que tiene una característica específica. Tiene un comportamiento binomial ya que una observación pertenece o no a la categoría estudiada.

Estimador de la proporción poblacional $\widehat{\mathbf{p}} = \frac{\mathbf{y}}{\mathbf{n}}$ donde y es en este caso el número total de los elementos de la muestra que tienen determinada característica.

Varianza estimada del estimador $\boldsymbol{\hat{\sigma}^2_{\hat{p}} = \left(\frac{\hat{p}\hat{q}}{n-1}\right)\left(\frac{N-n}{N}\right)}$ con $\boldsymbol{\hat{p} + \hat{q} = 1}$

Cota para el error de estimación. $\hat{\mathrm{p}} \pm 2\hat{\sigma}_{\hat{p}}$

Ejemplo:
Para el ejercicio anterior, supongamos que 2 de las 8 cuentas de la muestra tienen saldos vencidos. Estime el total de cuentas atrasadas y establezca una cota para el error de la estimación.

La proporción muestral esta dada por: $\hat{\mathrm{p}} = \frac{2}{8} = 0.25$

La cota del error de estimación es $\hat{\sigma}^2_{\hat{p}} = \left(\frac{0.25(0.75)}{8-1}\right)\left(\frac{1000-8}{1000}\right) = 0.026$ con $\hat{q} = 1 - 0.25 = 0.75$

Por lo tanto, se estima que el 25% (0.25) de las cuentas tienen saldo vencido, con una cota de error de

$$0.25 \pm 2\sqrt{0.026} = 0.25 \pm 0.32$$

Tamaño de la muestra

El número de observaciones necesarias para estimar una media y/o total poblacional con un límite para el error de estimación de magnitud B se obtiene de la siguiente manera, con un límite para el error de la estimación,

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \qquad donde \quad D = \frac{B^2}{4}$$

Ejemplo. La cantidad promedio de cuentas por cobrar de un hospital debe ser estimada, aunque no se cuenta con mucha información, se sabe que las cuentas tienen una varianza de $\sigma^2 = 625$. Existen $N = 1000$ cuentas abiertas, encuentre el tamaño de muestra necesario para estimar μ con un límite de para el error de estimación de $B = \$3$.

Se calcula entonces D:

$$D = \frac{B^2}{4} = \frac{3^2}{4} = 2.25$$

De esta manera el tamaño de la muestra es,

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} = \frac{1000 * 625}{(1000-1) * 2.25 + 625} = 217.56$$

Se necesitan tomar 218 observaciones para estimar la media de las cuentas por cobrar μ, con un límite para el error de estimación de $3 pesos

De manera similar para determinar el tamaño de la muestra para estimar el total poblacional τ

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \qquad donde \quad D = \frac{B^2}{4N^2}$$

Ejemplo. Un investigador está interesado en estimar la ganancia en peso total en un periodo de 4 semanas de $N = 1000$ pollitos alimentados con cierto alimento. Obviamente, pesar cada uno de los pollitos sería una tarea tediosa y demorada. Se debe determinar el número de pollitos que serán seleccionados para estimar τ con un límite para el error de estimación igual a 1000 gramos. Estudios previos dicen que la varianza poblacional σ^2 es de aproximadamente 36 gramos. Determine el tamaño de muestra requerido.

$$D = \frac{B^2}{4N^2} = \frac{1000^2}{4(1000)^2} = 0.25$$

De esta manera el tamaño de la muestra es,

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} = \frac{1000 * 36}{(1000-1) * 0.25 + 36} = 125.98$$

Se requiere pesar a 126 pollitos para estimar el total poblacional τ, la ganancia en peso en las 4 semanas de N=1000 pollitos, con un límite para el error de estimación igual a 1000 gramos.

Muestreo sistemático

Es el método donde se selecciona un punto aleatorio de inicio y posteriormente se elige cada $k\_ésimo$ miembro de la población. Es similar al muestreo simple aleatorio; sin embargo, no requiere contar con el marco muestral. Pudiera ser más económico y rápido, no obstante, tiene la desventaja de la periodicidad, es decir, al obtener las unidades o elementos muestrales de manera sistemática, se pueden realizar mediciones que obtienen estimaciones sesgadas.

El muestreo aleatorio sistemático, parte de encontrar una muestra a partir de un factor determinista. A partir de una selección al azar de un elemento de la población, seleccionamos los restantes al aplicar este factor de selección, el cual es proporcional al tamaño de la población.

$$k = \frac{N}{n}$$

Por ejemplo una empresa que esté interesada en estudiar el comportamiento de sus ventas, puede ordenar los totales de ventas diarias en una lista, si el factor es por ejemplo de 7, seleccionaría siempre el mismo día de la semana, lo cual podría ser un inconveniente ya que siempre será seleccionado el mismo día de la semana, por ejemplo, el lunes, presentando estimadores de ventas muy por debajo de lo real, o por el contrario seleccionar todos los viernes, donde la estimación queda sobre valorada.

El resultado de k se redondea al entero más cercano. Este procedimiento se hace más sencillo porque en lugar de extraer n números aleatorios sólo se extrae uno. Y porque es fácil si al igual que el muestreo aleatorio simple, se tienen enumerados todos los elementos de la población, o si de lo contrario no se tienen enumerados de todos modos se puede realizar pero se debe observar el orden físico de los elementos de la población. Cuando el orden físico de la población se relaciona con la característica de la población no se debe aplicar el muestreo aleatorio sistemático. El riesgo de este tipo de muestreo está en los casos en que se dan periodicidades en la población ya que al elegir a los miembros de la muestra con una periodicidad constante podemos introducir una homogeneidad que no se da en la población.

Procedimiento:

1) Definir la población de estudio.
2) Determinar el tamaño de muestra requerido.
3) Se calcula la muestra sistemática dividiendo la población entre el tamaño de la muestra.
4) El valor de *k* es el intervalo de selección que indica cada *k* de veces que un elemento de la población se integrará a la muestra (en el caso de no estar enumerados los elementos). Y también es el intervalo de selección del cual se escogerá un número aleatoriamente dentro de este intervalo (en caso de que los elementos estén enumerados), y de ahí se parte para seleccionar las muestras en los demás grupos o intervalos de selección.

Ejemplo:

Cuando los elementos no están enumerados. Una empresa armadora de autos va a probar el sistema de frenos, el departamento de ingeniera considera que una muestra de 50 autos de una población de 500 será suficiente para considerar si el lote se regresa a fabrica, $k = \frac{500}{50}, k = 10$. Ya los autos se encuentran estacionados en forma ordenada en un almacén, este intervalo de selección indica que cada 10 autos que contemos se integrará a la muestra. El primero en la muestra es el décimo auto estacionado, el segundo en la muestra es el vigésimo, el tercero el trigésimo y así de diez en diez hasta completar los 50 autos de la muestra.

Cuando los elementos están numerados. Si la población se compone de una cartera de clientes pre numerados; por ejemplo, $N = 800$ y se quiere extraer una muestra sistemática de *n=40,* se aplica la formula,

$$k = \frac{N}{n} = \frac{800}{40} \cong 20$$

De este intervalo selecciona un número aleatorio entre 1 y 20, y se incluye cada vigésimo elemento tras la primera selección de la muestra. Supongamos que el primer número seleccionado es 8. Los elementos que serían seleccionados para integrar la muestra,

28, 48, 68, 88, 108, 128, 148, 168, 188, 208, 228, 248, 268, 288, 308, 328, 348, 368, 388, 408, 428, 448, 468, 488, 508, 528, 548, 568, 588, 608, 628, 648, 668, 688, 708, 728, 748, 768 y 788.

Muestreo estratificado

Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos con respecto a alguna característica de las que se van a estudiar y tomar una muestra de cada uno de estos grupos. Posteriormente, si utilizamos muestreo simple aleatorio para extraer los elementos de la muestra de cada estrato, tendremos un ***Muestreo aleatorio estratificado***. Normalmente si no se especifica lo contrario se llamará solamente ***muestreo estratificado***.

Un estrato se define mediante algunas características comunes como son el sexo, la población, la edad, la profesión, etc. A diferencia del muestreo simple, el muestreo estratificado es útil porque garantiza que cada grupo esté representado en la muestra. El procedimiento de muestreo será más útil mientras más homogéneos sean los diferentes estratos.

Hay algunas ventajas al utilizar el muestreo estratificado,

- Además de estimar los parámetros de la población, nos permite realizar estimaciones de cada uno de los estratos. Por ejemplo es posible estimar el comportamiento de los consumidores en cada estrato para después estimar el comportamiento de la población.
- El muestreo estratificado nos puede garantizar una muestra más representativa. La estratificación reduce la probabilidad de obtener una muestra no representativa al asegurar que un número de elementos sean seleccionados de cada estrato.
- No es necesario disponer de la lista de toda la población sino de las sub poblaciones de orden superior extraídas.
- Existe una considerable reducción de costos. Es más fácil tomar una muestra de por ejemplo zonas o colonias específicas de una ciudad que muestrear toda la ciudad

La distribución de la muestra en función de los diferentes estratos, puede ser de diferentes tipos:

a) Afijación simple. A cada estrato le corresponde igual número de elementos muestrales. La desventaja es que favorece a los estratos más pequeños
b) Proporcional. Cada estrato se encuentra representado en la muestra en proporción exacta al tamaño de la población total.
c) Óptima. Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.

A cada uno de estos estratos se le asignaría una cuota que determinaría el número de miembros del mismo que compondrán la muestra. Dentro de cada estrato se suele usar, como ya lo indicamos antes, la técnica de muestreo aleatorio o sistemático, una de las técnicas de selección más usadas en la práctica.

Si N es el tamaño de la población y n el de la muestra. La población será dividida en L estratos de tamaño $l_1, l_2, l_3, \; ..l_L$

$$N = l_1 + \; l_2 + \; l_3 + \; ..l_L = \sum_{i=1}^{L} l_i$$

Y la muestra

$$n = n_1 + \; n_2 + \; n_3 + \; ..n_L = \sum_{i=1}^{L} n_i$$

Por ejemplo, para un estudio de opinión, puede resultar interesante estudiar por separado las opiniones de hombres y mujeres pues se estima que, dentro de cada uno de estos grupos, puede haber cierta homogeneidad. Así, si la población está compuesta de un 55% de mujeres y un 45% de hombres, se tomaría una muestra que contenga también esos mismos porcentajes de hombres y mujeres.

Una vez especificados los estratos, usamos el procedimiento mostrado en el muestreo simple aleatorio para seleccionar una muestra aleatoria para cada estrato. Usando el método de afijación (o asignación) proporcional, el tamaño de la muestra n se divide en cada uno de los estratos, de acuerdo a la siguiente regla,

$$n_i = n\left(\frac{N_i}{N}\right) \quad i = 1,2,3, \dots, L \qquad donde \;\; N_i \; es \; el \; número \; de \; elementos \; de \; estrato \;\; i$$

$$Por \; lo \; tanto \quad N = \sum_{i=1}^{L} N_i \;\; es \; el \; tamaño \; de \; la \; población.$$

En resumen, para seleccionar una muestra estratificada por afijación proporcional realizamos las siguientes acciones:

1) Definir la población en estudio
2) Determinar el tamaño de la muestra requerido
3) Establecer los estratos
4) Determinar la frecuencia relativa del muestreo de cada estrato por el tamaño de la muestra total, para obtener de cada estrato la cantidad de individuos que integrarán la muestra total.
5) Seleccionar y extraer de cada estrato la cantidad de elementos que formarán parte de la muestra, por el método de muestreo aleatorio simple.

Ejemplo. En un club de tenis, los 500 socios se reparten por edades en cuatro categorías: la 1ª con 200 socios, la 2ª con 175, la 3ª con 75 y la 4ª con 50. Se quiere seleccionar una muestra de 40 socios.

a) ¿Qué tipo de muestreo deberíamos realizar si queremos que estén representadas todas las edades?

b) ¿Cuántos individuos elegiríamos de cada categoría, si atendiéramos a razones de proporcionalidad?

Solución:

a) Deberíamos realizar un muestreo aleatorio estratificado.

b) Llamamos n1, n2, n3, n4 al número de individuos que tendríamos que seleccionar en cada categoría (1a, 2a, 3ª y 4ª , respectivamente). Entonces:

$$\frac{n_1}{200} = \frac{n_2}{175} = \frac{n_3}{75} = \frac{n_4}{50} = \frac{40}{500}$$

Así, debemos seleccionar $\frac{n_1}{200} = \frac{40}{500} \rightarrow n_1 = 200\left(\frac{4}{50}\right) = 16$ de la misma manera calculamos las demás categorías.

$$n_1 = 16; \quad n_2 = 175\left(\frac{4}{50}\right) = 14; \quad n_2 = 75\left(\frac{4}{50}\right) = 6; \; n_2 = 50\left(\frac{4}{50}\right) = 4$$

A. Estimación de la media poblacional para una muestra aleatoria estratificada.

Ahora podemos encontrar la *media y la varianza para cada estrato*, de la siguiente forma,

$$Media \;\; \overline{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \;\; y \; la \; varianza \; de \; cada \; estrato \;\; s_i^2 = \frac{\sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2}{n_i - 1} \quad i = 1,2,3, \dots \; L$$

$$El \; término \; y_{ij} \; es \; la \;\; j - ésima \; observación \; del \; estrato \; i$$

Para estimar la *media poblacional* para una muestra aleatoria estratificada.

$$Estimador \;\; \overline{y}_{est} = \frac{1}{N} \sum_{i=1}^{L} N_i \overline{y}_i$$

Y la varianza estimada del estimador es

$$\hat{\sigma}^2_{\overline{y}_{est}} = \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \left(\frac{s_i^2}{n_i}\right)\left(\frac{N_i - n_i}{N_i}\right)$$

Finalmente, la cota para el error de la estimación es $\overline{y}_{est} \pm 2\hat{\sigma}_{\overline{y}_{est}}$

Ejemplo: Si se tiene que seleccionar una muestra de 20 personas, de una comunidad de 500 habitantes, con el fin de hacerles una encuesta sobre la inversión en ahorros diaria. Los habitantes están repartidos en 5 comunidades, en donde el tamaño de cada estrato es:

58	144	147	94	40	26	135	9	2	16	129	42	5	150	22	126	149	69	109	19	51	3	4	39	11
114	116	79	50	146	104	87	33	83	126	71	68	53	41	122	62	6	144	8	149	111	98	31	146	2
70	5	36	55	148	141	81	144	112	99	36	107	104	145	95	43	95	73	39	52	30	131	140	88	60
52	118	110	33	144	15	25	58	76	29	49	108	67	34	88	38	129	4	101	72	105	144	59	132	51
137	106	41	113	39	139	128	55	17	16	105	116	96	45	86	71	96	129	94	118	40	68	9	9	16
131	35	68	69	61	42	35	9	116	108	2	145	80	27	121	13	116	94	49	121	11	47	62	64	103

Estrato	Colonia	Tamaño N_i		No de muestras por estrato	s_i^2
1	Mochitlán	100	$n_1 = 40\left(\frac{100}{500}\right)$	8	\$ 16.81
2	Quechultenango	150	$n_2 = 40\left(\frac{150}{500}\right)$	12	\$ 22.09
3	Juan R Escudero	50	$n_3 = 40\left(\frac{50}{500}\right)$	4	\$ 125.44
4	San Marcos	125	$n_4 = 40\left(\frac{125}{500}\right)$	10	\$ 45.2
5	Ayutla	75	$n_5 = 40\left(\frac{75}{500}\right)$	6	\$ 130.2
Total		500		40	

Para fines prácticos se considera como dato la varianza de cada estrato s_i^2

Los habitantes de cada comunidad están registrados y se les asignará un número, por ejemplo, en el estrato 1 hay 100 habitantes entonces se numerará de 001 a 100, en el estrato 2 hay 150 y se numerará de 001 a 150 y así sucesivamente se hará con los demás estratos. Y del tamaño de cada estrato se sacaran el número de muestras que se obtuvieron, por medio del método de muestreo aleatorio simple con la tabla de números aleatorios siguiente.

Los datos de cada comunidad se anexan al final.

Del estrato 1 (1 a 100) se tomarán las 8 muestras de la fila 1 de izquierda a derecha. Las muestras son: 58, 94, 40, 26, 9, 2, 16 y 42

Del estrato 2 (1 a 150) se tomarán las 12 muestras de la fila 2 de izquierda a derecha. Las muestras son: 114, 116, 79,50, 146, 104, 87, 33, 83, 126, 71 y 68

Del estrato 3 (1 a 50) se tomarán las 4 muestras de la fila 3 de izquierda a derecha. Las muestras son: 5, 36, 43 y 39

Del estrato 4 (1 a 125) se tomarán las 10 muestras de la fila 4 de izquierda a derecha. Las muestras son: 52, 118, 110, 33, 15, 25, 58, 76, 29 y 49

Del estrato 5 (1 a 75) se tomarán las 6 muestras de la fila 5 de izquierda a derecha. Las muestras son: 41, 39, 55, 17, 16 y 45

Estrato	Colonia	Tamaño N_i	Promedio de la muestra $\bar{y}$	No de muestras por estrato	s_i^2
1	Mochitlán	100	10.05	8	\$ 16.81
2	Quechultenango	150	24.8	12	\$ 22.09
3	Juan R Escudero	50	24.3	4	\$ 125.44
4	San Marcos	125	34.6	10	\$ 45.2
5	Ayutla	75	53.4	6	\$ 130.2
Total		500		40	

A partir de la tabla anterior, se estima la inversión en ahorros promedio.

$$\tilde{y}_{est} = \frac{1}{N}\sum_{i=1}^{L} N_i \bar{y}_i = \frac{1}{500}[100(10.05) + 150(24.8) + 50(24.3) + 125(34.6) + 75(53.4)]$$

$$= \frac{1}{500}(15270) \approx 29.0$$

Por lo tanto el ahorro promedio estimado que realizó la comunidad es \$ 28.54 y la varianza estimada es,

$$\hat{\sigma}^2_{\bar{y}_{est}} = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\left(\frac{s_i^2}{n_i}\right)\left(\frac{N_i-n_i}{N_i}\right) = \frac{1}{(500)^2}\left[(100)^2\left(\frac{100-8}{100}\right)\left(\frac{16.81}{8}\right) + (150)^2\left(\frac{150-12}{150}\right)\left(\frac{22.09}{12}\right) + (50)^2\left(\frac{50-4}{50}\right)\left(\frac{125.44}{4}\right) + (125)^2\left(\frac{125-10}{125}\right)\left(\frac{45.2}{10}\right) + (75)^2\left(\frac{75-6}{75}\right)\left(\frac{130.2}{6}\right)\right]$$

$$\hat{\sigma}^2_{\bar{y}_{est}} == \frac{1}{(500)^2}(19331.5 + 38105.25 + 72128 + 64975 + 112297.5) = 1.227$$

Los ahorros promedio con una cota de error de la estimación son entonces,

$$\bar{y}_{est} \pm 2\hat{\sigma}_{\bar{y}_{est}} = \$\,29 \pm 2\sqrt{1.227} = \$29 \pm 1.1077$$

B. *Estimador del total poblacional para una muestra aleatoria estratificada*

Si el objetivo de la encuesta es utilizar el muestreo estratificado para estimar el total poblacional τ procedemos de la siguiente forma,

Estimador del total poblacional $\hat{\tau} = N\bar{y}_{est}$ y la varianza estimada del estimador $\hat{\sigma}^2_{\hat{\tau}} = N^2\hat{\sigma}^2_{\bar{y}_{est}}$
La cota de error de la estimación $\hat{\tau} \pm 2\hat{\sigma}_{\hat{\tau}}$

Regresando a nuestro ejemplo, para estimar el ahorro total poblacional.

$$\hat{\tau} = 500(29) = 14500$$

Las cotas del error son,

$$\hat{\sigma}^2_{\hat{\tau}} = N^2\hat{\sigma}^2_{\bar{y}_{est}} = (500)^2(1.227) = 306925$$

y la cota $\hat{\tau} \pm 2\hat{\sigma}_{\hat{\tau}} = 14500 \pm 2\sqrt{306925} = \$14{,}500 \pm \$1{,}108$

C. *Estimación de la proporción poblacional para una muestra aleatoria estratificada.*

Estimador de la proporción poblacional $\hat{p}_{est} = \frac{1}{N}\sum_{i=1}^{L} N_i \hat{p}_i$

Varianza estimada del estimador $\boldsymbol{\hat{\sigma}^2_{\hat{p}_{est}} = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \left(\frac{\hat{p}_i\,\hat{q}_i}{n_i-1}\right)\left(\frac{N_i-n_i}{N_i}\right)}$ con $\boldsymbol{\hat{p}_i + \hat{q}_i = 1}$

Y las cotas para el error de la estimación $\hat{p}_{est} \pm 2\hat{\sigma}_{\hat{p}_{est}}$

Tamaño de muestra para estimar el promedio con asignación proporcional.

$$n = \frac{\sum_{i=1}^{k} n_i \tilde{s}_i^2}{\frac{NB^2}{k^2} + \frac{1}{N}\sum_{i=1}^{k} n_i \tilde{s}_i^2}$$

Dónde: B = error de estimación

k = Percentil que se halla en la tabla de la distribución normal y depende del nivel de confianza

EJEMPLO. Se desea estimar la nota promedio de los estudiantes de administración de empresas diurna y nocturna en una universidad. En la carrera diurna (estrato 1) hay 280 estudiantes y en la nocturna (estrato 2) hay 200 estudiantes. Determine el tamaño de muestra necesario para cumplir el objetivo con un error máximo de 0.15 y una confiabilidad del 95 por ciento.

Por un estudio realizado tiempo atrás se conocen las varianzas de las notas de administración diurna y nocturna, las que respectivamente son: 0.31 y 0.28.

Solución

Considerando que las varianzas son similares, se trabaja con muestreo estratificado con asignación proporcional. El error (B) es 0.15 y para una confiabilidad del 95 por ciento el valor correspondiente en la distribución normal es 1.96, entonces, k = 1.96:

$$n_1 = 280; \quad n_2 = 200; \quad N = 480; \quad \tilde{s}_1^2 = 0.31; \quad \tilde{s}_2^2 = 0.28$$

Para hallar el tamaño de muestra se utiliza la ecuación anterior

$$n = \frac{280(0.31) + 200(0.28)}{480\frac{0.15^2}{1.96^2} + \frac{1}{480}[280(0.31) + 200(0.28)]} = 45.93$$

El tamaño de la muestra es de 46 estudiantes. Esta muestra se reparte proporcionalmente al tamaño de los estratos, como se realizó con anterioridad

$$n_1 = 46\frac{280}{480} = 26.83; \quad n_2 = 46\frac{200}{480} = 19.17$$

Se deben seleccionar 27 estudiantes de administración de empresas diurna y 19 de la nocturna.

Tamaño de muestra para estimar el total con asignación proporcional

$$n = \frac{\sum_{i=1}^{k} n_i \tilde{s}_i^2}{\frac{B^2}{k^2 N} + \frac{1}{N}\sum_{i=1}^{k} n_i \tilde{s}_i^2}$$

EJEMPLO. Se desea hacer un estudio para estimar el consumo total de gasolina en una ciudad, halle el tamaño de muestra necesario para cumplir éste objetivo. Los vehículos se clasificaron en tres grupos o estratos, particulares (1), públicos (2) y oficiales (3). En la oficina de circulación y tránsito se obtuvo la siguiente información sobre los vehículos matriculados en la ciudad; vehículos particulares 7,627, públicos 2,392 y oficiales 534.

Solución

Como no se dispone de estudios similares, se toma una muestra piloto, con la cual se obtienen las siguientes varianzas sobre el consumo semanal en galones:

$$\tilde{s}_1^2 = 137.6; \quad \tilde{s}_2^2 = 138; \quad \tilde{s}_3^2 = 135.86$$

Asumiendo un error de estimación máximo de 15,000 galones, (B = 15,000), y una confiabilidad del 95 por ciento, el valor de k en la distribución normal es 1.96.

Considerando que las varianzas en los tres estratos son similares, se trabaja con muestreo estratificado con asignación proporcional. Para calcular el tamaño de la muestra se utiliza la ecuación anterior.

$$N_1 = 7{,}627 \quad N_2 = 2{,}392 \quad N_3 = 534 \quad N = 10{,}553$$

$$n = \frac{7627(137.6) + 2392(138) + 534(135.86)}{\frac{15{,}000^2}{1.96^2(10{,}553)} + \frac{1}{10553}[7627(137.6) + 2392(138) + 534(135.86)]} = 255$$

$$n_1 = 255\frac{7{,}627}{10{,}553} = 184; \quad n_2 = 255\frac{2{,}392}{10{,}553} = 58; \quad n_3 = 255\frac{534}{10{,}553} = 13$$

Para estimar el consumo total de gasolina con un error máximo de 15.000 galones/semana, se debe seleccionar una muestra de 255 autos repartida así: 184 autos particulares, 58 públicos y 13 oficiales.

Recuerde que si se desea, se puede disminuir el error máximo admisible, pero esto conlleva a un aumento en el tamaño de la muestra.

Tamaño de muestra para estimar la proporción con asignación proporcional

$$n = \frac{\sum_{k=1}^{L} n_k p_k q_k}{N\frac{B^2}{k^2} + \frac{1}{N}\sum_{k=1}^{L} n_k p_k q_k}$$

EJEMPLO. En vista de la recesión económica existente, una empresa textil pretende reducir el número de días laborables por semana a cuatro. Otra alternativa consiste en clausurar una de sus tres plantas y despedir a los trabajadores. Para tener una idea de la opinión de los trabajadores, el gerente de personal de la empresa desea seleccionar una muestra de empleados de las tres plantas para estimar la proporción de trabajadores que prefieren la reducción de la semana de trabajo, con un error de estimación máximo de 0.1.

La empresa emplea 150 personas en la planta 1.65 en la planta 2 y 40 en la 3. Se estima que cerca del 75 por ciento de los de la planta tres están a favor de la reducción de la semana de trabajo, mientras que en las otras plantas este porcentaje parece corresponder al 50 por ciento. Encuentre el tamaño de muestra y la asignación necesaria en cada estrato.

Solución

$$N_1 = 150 \quad N_2 = 65 \quad N_3 = 40 \quad N = 255$$

$$p_1 = 0.5 \quad p_2 = 0.5 \quad p_3 = 0.75 \quad B = 0.1$$

Por la diferencia en el tamaño de las plantas, se utiliza el muestreo estratificado con asignación proporcional.

Asumiendo un nivel de confianza del 95 por ciento, el valor correspondiente en la distribución normal es 1.96 (k=1.96).

Para determinar el tamaño de la muestra sustituimos en la ecuación,

$$n = \frac{150(0.5) + 65(0.5)(0.5) + 40(0.75)(0.25)}{255\frac{0.1^2}{1.96^2} + \frac{1}{255}[150(0.25) + 65(0.25) + 40(0.1875)]} = 67.76$$

$$n_1 = 68\frac{150}{255} = 40; \quad n_2 = 68\frac{65}{255} = 17.33; \quad n_3 = 68\frac{40}{255} = 10.67$$

Muestreo por conglomerados

Técnica similar al muestreo por estadios múltiples, se utiliza cuando la población se encuentra dividida, de manera natural, en grupos que se supone que contienen toda la variabilidad de la población, es decir, la representan fielmente respecto a la característica a elegir, pueden seleccionarse sólo algunos de estos grupos o *conglomerados* para la realización del estudio.

En una muestra de conglomerados, se divide *N* elementos de la población en varios grupos de tal manera que cada uno sea representativo de toda la población. Este procedimiento tiende a proporcionar mejores resultados cuando los elementos dentro de los conglomerados no son semejantes. Lo ideal es que cada conglomerado sea una representación, a pequeña escala, de la población. Se aplica en el muestreo de áreas, en la que los conglomerados son manzanas, ciudades, distritos electorales, países, etc. En este tipo de muestreo es imprescindible diferenciar entre unidad de análisis entendida como quiénes va a ser medidos y unidad muestral que se refiere al conglomerado a través del cual se logra el acceso a la unidad de análisis.

Procedimiento:

1) Dividir la población en conglomerados.
2) Crear una lista de todos ellos
3) Seleccionar al azar el número de conglomerados que desee.
4) Tomar una muestra aleatoria simple de uno de los elementos de cada conglomerado.

Dentro de los conglomerados seleccionados se ubicarán las unidades elementales, por ejemplo, las personas a encuestar, y podría aplicársele el instrumento de medición a todas las unidades, es decir, los miembros del conglomerado, o sólo se podría aplicar a algunos de ellos, seleccionados al azar. Este método tiene la ventaja de simplificar la recogida de información muestral.

Cuando, dentro de cada conglomerado seleccionado, se extraen algunos individuos para integrar la muestra, el diseño se llama *muestreo bietápico*.

Las ideas de estratos y conglomerados son, en cierto sentido, opuestas. El primer método funciona mejor cuanto más homogénea es la población respecto del estrato, aunque más diferentes son éstos entre sí. En el segundo, ocurre lo contrario. Los conglomerados deben presentar toda la variabilidad, aunque deben ser muy parecidos entre sí.

Ejemplo:

Si se va a realizar una encuesta sobre las políticas y leyes del municipio, se podría dividir el municipio en distritos, por ejemplo en 13 distritos, de esos tres se toma al azar el 4, 5, 9 y 11, y solo concentrándonos en estos distritos, tomamos una muestra aleatoria de habitantes de cada uno de esos distritos, para entrevistarlos.

A. *Estimación de la media poblacional en un muestreo por conglomerados*

Al usar muestreo por conglomerados, la media poblacional, se estima de la siguiente forma. En estos cálculos, n_i es el número de elementos del $\mathrm{i}-\mathrm{ésimo}$ conglomerado y t_i es el total de las mediciones del conglomerado.

$$\hat{\mu} = \frac{\sum_{i=1}^{m} t_i}{\sum_{i=1}^{m} n_i}$$

Varianza estimada del estimador

$$\hat{\sigma}^2_{\overline{y}_c} = \left(\frac{M-m}{Mm\overline{m}^2}\right)\left(\frac{\sum_{i=1}^{m}(t_i - \overline{y}_c n_i)^2}{m-1}\right)$$

Cotas para el error de la estimación $\overline{y}_c \pm 2\hat{\sigma}_{\overline{y}_c}$ donde $\overline{n} = \frac{1}{m}\sum_{i=1}^{m} n_i$ $\quad \overline{t} = \frac{1}{m}\sum_{i=1}^{m} t_i$

M es el número de conglomerados en la población y m es el número de conglomerados en la muestra.

B. Estimación del total poblacional en un muestreo por conglomerados

$$\hat{\tau} = \frac{M}{m}\sum_{i=1}^{m} t_i$$

Varianza estimada del estimador

$$\hat{\sigma}^2_{\hat{\tau}} = M^2\left(\frac{M-m}{Mm}\right)\left(\frac{\sum_{i=1}^{m}(t_i - \overline{t})^2}{m-1}\right)$$

Cotas para el error de la estimación $\hat{\tau} \pm 2\hat{\sigma}_{\hat{\tau}}$

C. Estimación de la proporción poblacional para un muestreo por conglomerados.

Estimador $\hat{p}_c = \frac{\sum_{i=1}^{m} a_i}{\sum_{i=1}^{m} n_i}$

Varianza estimada del estimador $\hat{\boldsymbol{\sigma}}^2_{\hat{\boldsymbol{p}}_c} = \left(\frac{M-m}{Mm\overline{m}^2}\right)\left(\frac{\sum_{i=1}^{m}(a_i - \overline{p}_c n_i)^2}{m-1}\right)$

Y las cotas para el error de la estimación $\hat{p}_c \pm 2\hat{\sigma}_{\hat{p}_c}$

ANEXO

TABLA DE NUMEROS ALEATORIOS

17841	49597	92623	80005	11177	15145	46379
84970	47043	64048	06993	17369	70932	47950
30524	27250	73072	52654	33653	30422	22347
56211	27219	44652	09467	62848	82479	35068
66110	69181	13200	93239	25591	21248	06881
28710	52414	55893	25632	64856	51745	46855
38939	15777	66270	53052	05160	94786	81987
31297	00722	88300	21109	13124	96742	64968
34043	19959	77949	24510	93510	40492	81113
74996	32698	29430	58603	43879	7861	15870
09224	49628	26353	25592	78113	27589	87512
16925	04512	74150	64475	04497	64977	30847
95370	15305	08474	58306	65393	86919	16478
74611	68568	45153	72541	14812	64511	20253
24791	18151	68084	01936	53838	48954	25322
22283	30815	82384	19084	41248	77855	22366
71898	26726	89650	31162	86245	00370	31000
95413	08931	96334	31263	45687	89601	25395
50790	90191	37070	59230	21080	86042	46441
86175	96384	63337	73013	34939	59945	62412
51485	90027	98827	43212	93302	64337	34026
24958	56475	29207	62272	41011	56041	37735
47249	28708	17767	20087	43020	20963	59504
62152	80266	99282	22863	81820	09317	74915
09135	46518	71377	14410	69712	65884	14366
43770	34210	35225	08830	65793	43288	22567
49358	18612	11688	52443	39456	65328	44806
67452	60795	63023	21400	02021	20485	09224
39578	43182	40366	02955	47485	54797	16874
49630	42256	95206	52914	15086	01292	24360
86158	26615	20228	14854	00161	64983	59471
81648	63523	82624	81928	54646	62114	36529
72208	67425	77273	35454	43798	89958	98485
62663	32726	14266	48467	36706	90411	84898
99530	11547	35629	86192	25909	97084	30951
36626	80491	21369	48285	59708	44408	75096

Fuente. Elaboración propia

Anexo 2. Datos ejercicio muestreo estratificado

	Datos muestreo estratificado comunidad Mochitlán						
Número	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro
1	8.9	26	8.0	51	12.7	76	7.6
2	11.4	27	8.0	52	15.8	77	6.0
3	9.8	28	9.7	53	17.7	78	7.6
4	11.9	29	8.9	54	8.5	79	10.7
5	6.8	30	13.7	55	12.7	80	6.7
6	9.4	31	8.9	56	16.9	81	6.0
7	10.1	32	8.3	57	12.9	82	4.3
8	4.7	33	10.5	58	12.7	83	11.4
9	7.2	34	16.6	59	11.1	84	9.9
10	11.1	35	15.9	60	11.9	85	5.8
11	11.5	36	9.7	61	13.5	86	6.1
12	9.1	37	8.9	62	7.9	87	10.0
13	9.2	38	12.1	63	9.0	88	11.4
14	8.8	39	16.9	64	6.0	89	4.8
15	17.7	40	6.5	65	7.4	90	6.0
16	12.1	41	11.3	66	7.2	91	9.2
17	15.1	42	15.2	67	14.8	92	8.3
18	8.6	43	17.7	68	7.6	93	9.1
19	7.9	44	20.0	69	8.4	94	7.3
20	10.3	45	13.5	70	6.8	95	6.2
21	15.2	46	13.5	71	8.2	96	5.1
22	7.2	47	15.8	72	6.8	97	6.0
23	8.9	48	9.6	73	8.4	98	8.1
24	9.7	49	11.5	74	7.6	99	4.9
25	10.1	50	20.1	75	14.3	100	6.8
Promedio	10.16						

	Datos muestreo estratificado comunidad Juan R Escudero								
Número	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro
1	13.0	11	20.8	21	14.4	31	28.6	41	20.1
2	31.1	12	24.4	22	16.0	32	27.3	42	17.0
3	18.9	13	14.3	23	19.5	33	27.3	43	15.3
4	15.9	14	17.3	24	18.9	34	21.8	44	17.5
5	15.6	15	18.6	25	16.0	35	18.2	45	16.0
6	16.0	16	23.2	26	29.2	36	38.7	46	23.3
7	21.8	17	14.8	27	38.7	37	23.3	47	21.5
8	15.7	18	31.9	28	21.8	38	22.9	48	13.0
9	25.7	19	17.2	29	16.8	39	27.7	49	23.0
10	27.3	20	17.2	30	47.6	40	11.4	50	36.1
Promedio	21.80								

Número	Datos muestreo estratificado comunidad Quechultenango						
	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro
1	29.2	39	32.3	77	20.7	115	16.9
2	20.4	40	24.1	78	43.0	116	17.7
3	24.6	41	18.3	79	37.8	117	15.4
4	33.9	42	23.8	80	22.4	118	16.9
5	25.8	43	20.7	81	15.4	119	37.8
6	20.0	44	20.7	82	27.2	120	17.1
7	28.9	45	19.0	83	39.5	121	37.8
8	42.8	46	23.8	84	29.1	122	33.7
9	25.8	47	33.9	85	18.5	123	30.5
10	22.4	48	18.5	86	19.0	124	16.9
11	25.8	49	17.1	87	28.9	125	19.0
12	18.7	50	19.0	88	19.0	126	20.7
13	39.5	51	22.4	89	20.4	127	15.4
14	56.3	52	18.8	90	22.0	128	23.1
15	28.9	53	37.8	91	22.4	129	25.5
16	18.6	54	28.9	92	36.1	130	24.1
17	36.1	55	28.9	93	18.6	131	19.0
18	32.9	56	15.4	94	22.3	132	21.6
19	24.1	57	36.1	95	32.3	133	30.5
20	33.9	58	32.4	96	34.6	134	24.1
21	27.6	59	22.4	97	20.7	135	20.3
22	24.6	60	23.8	98	32.3	136	27.6
23	13.8	61	21.6	99	19.0	137	24.1
24	32.3	62	18.1	100	22.2	138	23.9
25	20.4	63	32.5	101	27.2	139	22.0
26	28.9	64	29.2	102	35.6	140	20.4
27	30.6	65	27.6	103	19.0	141	39.7
28	22.4	66	25.8	104	22.1	142	24.2
29	32.3	67	22.4	105	45.9	143	21.1
30	25.8	68	36.0	106	27.6	144	18.5
31	36.8	69	18.5	107	43.0	145	22.3
32	17.0	70	32.5	108	17.1	146	24.6
33	13.5	71	15.4	109	30.8	147	25.1
34	22.4	72	19.3	110	27.2	148	25.4
35	27.2	73	19.0	111	33.9	149	21.9
36	27.5	74	20.5	112	25.5	150	19.3
37	22.3	75	18.7	113	19.0		
38	19.0	76	32.3	114	22.4		
Promedio	25.50						

Número	Datos muestreo estratificado comunidad San Marcos								
	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro
1	40.7	26	64.5	51	26.7	76	47.4	101	38.7
2	31.5	27	29.1	52	32.5	77	31.0	102	28.7
3	45.4	28	29.1	53	26.7	78	36.3	103	36.3
4	26.0	29	26.0	54	28.7	79	43.0	104	43.3
5	31.5	30	31.0	55	21.6	80	28.8	105	33.5
6	36.3	31	26.2	56	21.6	81	24.0	106	40.6
7	33.7	32	30.3	57	53.1	82	24.9	107	31.0
8	29.1	33	31.4	58	48.6	83	31.5	108	38.7
9	21.7	34	27.1	59	55.5	84	40.9	109	41.1
10	19.0	35	31.5	60	41.0	85	38.7	110	33.9
11	28.1	36	60.2	61	50.6	86	45.4	111	33.1
12	26.7	37	36.3	62	50.7	87	23.8	112	40.6
13	45.4	38	45.4	63	31.5	88	21.7	113	31.5
14	21.7	39	45.4	64	38.6	89	19.5	114	31.5
15	42.8	40	33.4	65	47.7	90	42.8	115	50.7
16	28.7	41	26.7	66	26.6	91	25.7	116	31.2
17	38.3	42	50.0	67	31.4	92	26.0	117	38.2
18	45.4	43	47.6	68	53.1	93	19.4	118	35.8
19	29.1	44	24.0	69	26.7	94	34.6	119	33.9
20	45.7	45	26.7	70	40.6	95	47.6	120	40.6
21	28.6	46	45.5	71	29.1	96	36.3	121	64.5
22	79.2	47	26.7	72	26.2	97	31.5	122	33.5
23	60.4	48	33.9	73	26.7	98	50.7	123	26.3
24	24.0	49	26.2	74	32.5	99	26.7	124	33.9
25	21.7	50	29.1	75	31.5	100	38.7	125	33.4
Promedio	35.4								

Número	Datos muestreo estratificado comunidad Ayutla								
	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro	Número	Ahorro
1	45.0	16	40.5	31	85.0	46	53.4	61	78.0
2	40.5	17	56.9	32	102.3	47	58.9	62	70.0
3	47.5	18	36.9	33	44.6	48	53.6	63	102.6
4	62.0	19	69.8	34	49.5	49	80.6	64	69.1
5	50.0	20	52.7	35	58.9	50	45.0	65	79.0
6	45.3	21	44.2	36	90.4	51	57.7	66	134.6
7	36.9	22	45.4	37	49.5	52	48.7	67	54.0
8	88.0	23	48.7	38	45.4	53	65.7	68	36.8
9	44.9	24	77.3	39	53.6	54	60.5	69	53.6
10	61.8	25	65.8	40	65.0	55	46.1	70	44.6
11	90.4	26	65.8	41	36.8	56	65.0	71	86.1
12	45.4	27	50.0	42	81.0	57	53.6	72	69.0
13	78.0	28	80.5	43	65.8	58	69.0	73	47.0
14	44.2	29	45.4	44	61.8	59	110.0	74	61.0
15	49.0	30	90.3	45	86.3	60	77.3	75	52.0
Promedio	62.1								

Bibliografía

Anderson, D. R., D. J. Sweeney y T. A. Williams. (2008). *Estadística para la administración y la economía.* (10ª ed). México: CENGAGE Learning. 260-262.

Levine, D. M., T. C. Krehbiel y M. L. Berenson. (2006). *Estadística para la administración.* (4ta ed). México: Pearson Prentice Hall. 221.

Lind, D. A., W. G. Marchal, y S. A. Wathen. (2008). *Estadística aplicada a los negocios y a la economía.* (13ª ed). México: McGraw-Hill. 262, 265, 266

Muestreo estratificado con asignación proporcional. La muestra se reparte entre los estratos proporcionalmente a los tamaños de éstos. Este tipo de asignación se utiliza cuando los costos y las varianzas de los estratos no son muy diferentes.