CU-10 MUESTREO PROBABILÍSTICO UTILIZANDO EXCEL

Juan Luis Arias Vargas
Magíster en la enseñanza de la Matemática
Especialista en Administración de la Informática Educativa
Ingeniero Industrial
Docente Asistente Universidad Católica Popular del Risaralda
Docente Catedrático Auxiliar Universidad Tecnológica de Pereira
Grupo de Investigación GEMA
Juan.arias@ucpr.edu.co

RESUMEN

El presente trabajo muestra conceptos básicos sobre muestreo probabilístico sus bondades y cómo haciendo uso de herramientas sencillas de Excel, se puede calcular el tamaño de la muestra y seleccionarla aleatoriamente.

Palabras Clave: muestreo, muestras aleatorias, Excel.

ABSTRACT

This text presents basic concepts about probabilistic sampling, its strengths, and the way how using simple tools for Excel, the size of the sample can be calculated and selected in an aleatory way.

Key Words: sampling, aleatory samples, Excel.

Introducción

El muestreo estadístico o muestreo probabilístico son técnicas que permiten calcular el tamaño de la muestra apropiado, para estimar los parámetros, en una investigación de tipo cuantitativo, lo cual permite hacer inferencias y tomar decisiones acerca de la población; sin embargo en ocasiones se usan muestreos no probabilísticos, con el fin de tener una idea del valor de los parámetros, no obstante a partir de estos resultados no se puede inferir sobre la población.

Tabla No. 1 Muestreo probabilístico Vs muestreo no probabilístico

Muestreo probabilístico	Muestreo no probabilístico
El tamaño de la muestra se calcula por medio de una regla matemática	El tamaño de la muestra es decisión del investigador
La selección de la muestra es bajo un procedimiento aleatorio	La selección de muestra es de criterio del investigador o del entrevistador
Se puede calcular el error muestral	No se puede calcular el error muestral
Se pueden hacer inferencias a cerca de la población de donde se extrajo la muestra	Las conclusiones son solo para los elementos de la muestra

Razones por la cuales usar el muestreo

Ahorro de dinero: El número de instrumento para la recolección de información es menor, el tiempo para la aplicación de los mismo, la digitación y tabulación de ellos es menor, lo que redunda en ahorro de recursos económicos.

Ahorro de tiempo: Se necesita menos horas/hombre para la aplicación de instrumentos, la digitación y tabulación de la información recolectada, lo que implica un tiempo menor para conocer los resultados de la investigación, permitiendo tomar decisiones a tiempo.

Por exactitud, existen errores que se cometen bien sea en el cálculo de los parámetros o en la estimación de los mismos, en este sentido, si se realiza un censo, para calcular los parámetros, se cometen errores no muestrales que son mucho mayores, por los volúmenes de información a manejar, que la suma de los errores muestrales y no muestrales que se cometen cuando se aplica una técnica de muestreo.

Cuando el elemento es destruido o contaminado, es lógico que se deba hacer la observación solo a unos cuentos de ellos.

Algunas definiciones importantes

Población objetivo: Es un conjunto de elementos que tienen unas características comunes y están delimitados en tiempo y en espacio.

Muestra: Es un subconjunto de elementos que pertenecen a una población, se dice que la muestra es representativa de la población si tiene un tamaño óptimo y su selección se realiza de manera aleatoria.

Elemento: Es la unidad de la cual se solicita la información.

Unidad de muestro: Son los elementos disponibles para su selección en alguna etapa del muestreo, en ocasiones la unidad de muestreo corresponde al mismo elemento.

Marco muestral: Es el conjunto de elementos del cual se saca la muestra.

Población de estudio: Es aquella de donde realmente se sacó la muestra.

Muestreo probabilístico

Tamaño de la muestra

Una parte importante en una investigación es la cantidad de información que se usa para tomar decisiones acerca de los parámetros de una población a partir de la estimación de los parámetros o las pruebas de hipótesis, ya que muy poca información, es decir una muestra de tamaño pequeño, podría ser insuficiente y mucha información, es decir una muestra muy grande sería muy costosa y se desperdiciarían recursos, por lo tanto se debe de usar un tamaño de muestra que sea adecuado al estudio que se esté realizando, a la población de estudio y a su tamaño; los métodos utilizados se basan en el error de estimación que es admitido por el investigador y en el nivel de confianza que se desee para la investigación y dependen del tipo de muestreo a utilizar.

El error de estimación admitido por el investigador se denota por la letra β y debe estar en las mismas unidades de la variable de estudio.

En forma general si se está estimando el parámetro ϕ, entonces el estimador por intervalo es $\hat{\phi} \pm a\sigma_{\hat{\phi}} \Rightarrow \hat{\phi} \pm B \therefore a\sigma_{\hat{\phi}} = B$

Tipos de muestreo probabilístico más utilizados

Muestreo aleatorio simple (MAS)

Este tipo de muestreo consiste en seleccionar una muestra aleatoria de tamaño n de una población de tamaño N, para este caso cada elemento de la población tiene la misma probabilidad de ser seleccionado o lo que es lo mismo cada muestra de tamaño n tiene la misma probabilidad.

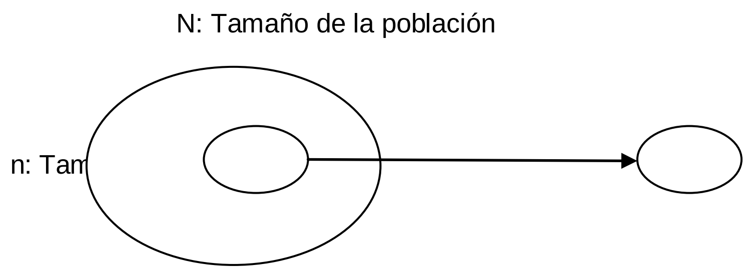


Figura No. 1 Muestreo Aleatorio Simple

Tamaño de la muestra para estimar a μ

El intervalo de confianza para μ al $(1-\alpha)100\%$ cuando N es infinita está dado por:

$\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ por lo tanto $B = z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ de donde $n = \dfrac{z_{\alpha/2}^2\sigma^2}{B^2}$

El intervalo de confianza para μ al $(1-\alpha)100\%$ cuando N es finita está dado por:

$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ por lo tanto $B = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ de donde $n = \frac{N z_{\alpha/2}^2 \sigma^2}{(N-1)B^2 + z_{\alpha/2}^2 \sigma^2}$

Nótese que el valor de la varianza involucrado en el cálculo del tamaño de la muestra es poblacional, es muy común que este valor no se conozca, para solucionar este problema se puede optar por cualquiera de las opciones siguientes:

Se puede tomar la varianza de una investigación anterior pero que todavía tanga validez.
Se puede tomar una premuestra aleatoria, es decir realizar una muestra piloto y a partir de estos datos estimar la varianza de la población, en este caso los elementos de la premuestra se pueden usar en el estudio final.

Si la población sigue una distribución normal y se conoce el rango (R) de los datos en la población, entonces $\sigma = \frac{R}{4}$

Tamaño de la muestra para estimar a p

El intervalo de confianza para p al $(1-\alpha)100\%$ cuando N es infinita está dado por:

$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{pq}{n}}$ por lo tanto, $B = z_{\alpha/2} \sqrt{\frac{pq}{n}}$ de donde $n = \frac{z_{\alpha/2}^2 pq}{B^2}$

El intervalo de confianza para p al $(1-\alpha)100\%$ cuando N es finita está dado por:

$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$ por lo tanto, $B = z_{\alpha/2} \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$ de donde $n = \frac{N z_{\alpha/2}^2 pq}{(N-1)B^2 + z_{\alpha/2}^2 pq}$

Nótese que el valor de la proporción involucrado en el cálculo del tamaño de la muestra es poblacional, es muy común que este valor no se conozca, para solucionar este problema se puede optar por cualquiera de las opciones siguientes:

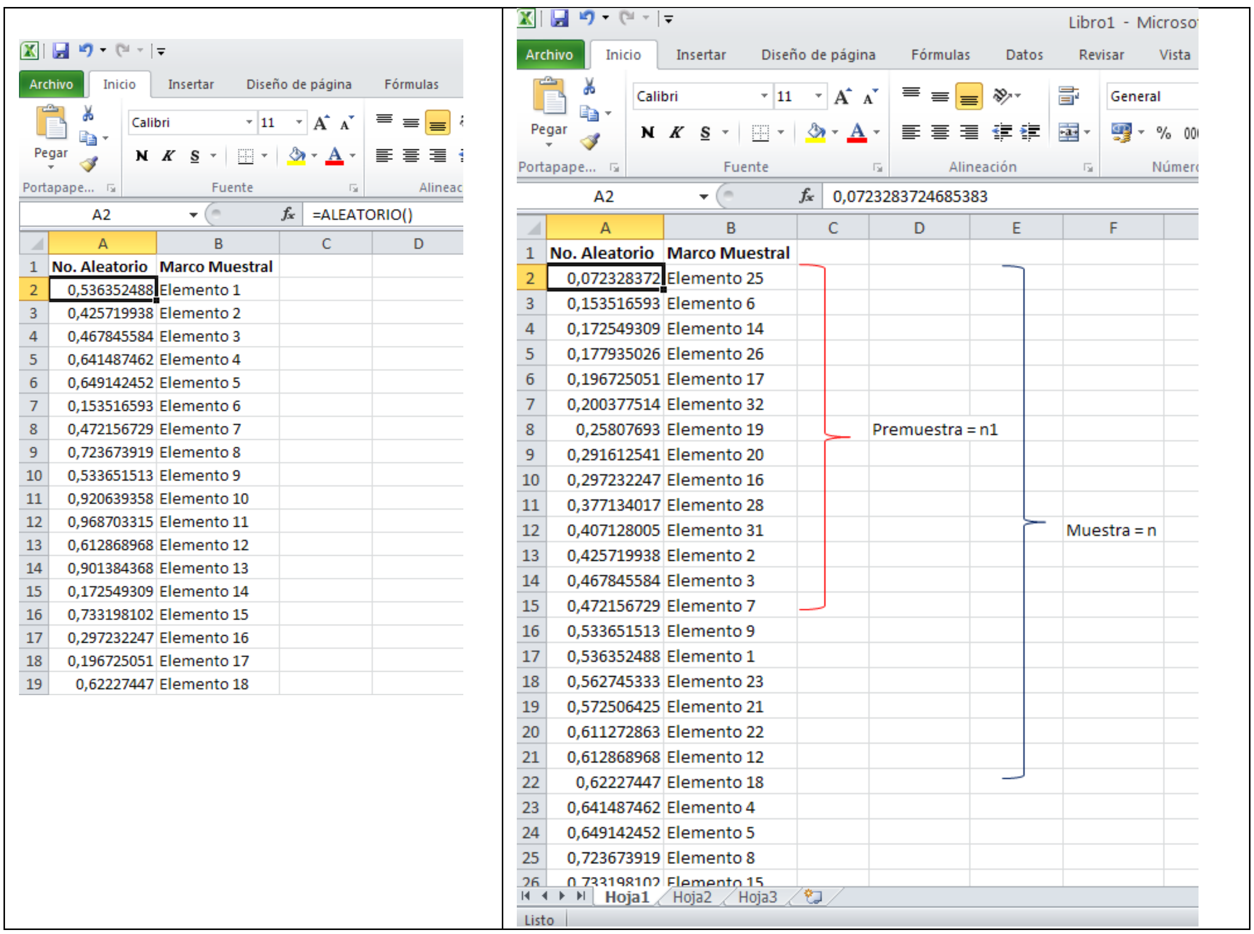
Se puede tomar la proporción de una investigación anterior pero que todavía tanga validez.

Se puede tomar una premuestra aleatoria, es decir realizar una muestra piloto y a partir de estos datos estimar la proporción de la población, en este caso los elementos de la premuestra se pueden usar en el estudio final.
Lo más común es dar a la proporción el valor de 0.5 en cuyo caso el tamaño de la muestra será el más grande posible con las otras condiciones fijas.

Nota: Aunque se podría tener una regla para calcular el tamaño de la muestra para estimar el total población, se puede omitir ya que el tamaño de la muestra para este estimador es el mismo que para estimar la media o la proporción según sea la necesidad para estimar el total poblacional.

Selección de la muestra

Para la selección de la muestra aleatoria se puede usar un método muy sencillo utilizando una hoja de cálculo de Excel, en la hoja de Excel se tiene una lista de números aleatorios en la primera columna y en la segunda columna el listado del marco muestral, con pegado especial se pegan los números aleatorios como valor y luego se ordenan las dos columnas tomando como fila principal la de los números aleatorios, así que queda la segunda fila ordenada de manera aleatoria y de allí se pueden elegir los primeros elementos para realizar la premuestra (n1) si es del caso y después de calcular el tamaño de n, este se completa con los elementos siguientes.



A2 =ALEATORIO()

	A	B
1	No. Aleatorio	Marco Muestral
2	0,536352488	Elemento 1
3	0,425719938	Elemento 2
4	0,467845584	Elemento 3
5	0,641487462	Elemento 4
6	0,649142452	Elemento 5
7	0,153516593	Elemento 6
8	0,472156729	Elemento 7
9	0,723673919	Elemento 8
10	0,533651513	Elemento 9
11	0,920639358	Elemento 10
12	0,968703315	Elemento 11
13	0,612868968	Elemento 12
14	0,901384368	Elemento 13
15	0,172549309	Elemento 14
16	0,733198102	Elemento 15
17	0,297232247	Elemento 16
18	0,196725051	Elemento 17
19	0,62227447	Elemento 18

A2 0,0723283724685383

	A	B
1	No. Aleatorio	Marco Muestral
2	0,072328372	Elemento 25
3	0,153516593	Elemento 6
4	0,172549309	Elemento 14
5	0,177935026	Elemento 26
6	0,196725051	Elemento 17
7	0,200377514	Elemento 32
8	0,25807693	Elemento 19
9	0,291612541	Elemento 20
10	0,297232247	Elemento 16
11	0,377134017	Elemento 28
12	0,407128005	Elemento 31
13	0,425719938	Elemento 2
14	0,467845584	Elemento 3
15	0,472156729	Elemento 7
16	0,533651513	Elemento 9
17	0,536352488	Elemento 1
18	0,562745333	Elemento 23
19	0,572506425	Elemento 21
20	0,611272863	Elemento 22
21	0,612868968	Elemento 12
22	0,62227447	Elemento 18
23	0,641487462	Elemento 4
24	0,649142452	Elemento 5
25	0,723673919	Elemento 8
26	0,733198102	Elemento 15

Figura No. 2 Pasos para la selección de la muestra

Muestreo aleatorio sistemático lineal (MSL)

N: Tamaño de la población

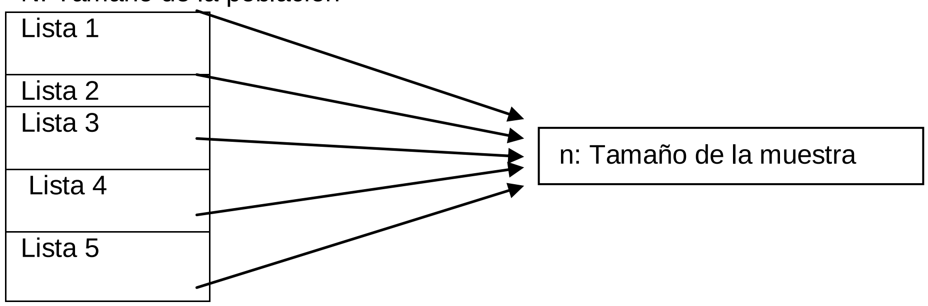


Figura No. 3 Muestreo Aleatorio Sistemático Lineal

El muestreo Sistemático Lineal, es muy usado por la facilidad en la selección de la muestra, para la aplicación de un MSL es necesario tener un marco muestral y los elementos del marco muestral deben ocupar cada posición de manera aleatoria con el fin de que la precisión del estimador sea igual a la de un MAS, sin embargo también se puede aplicar en situaciones donde no se conoce el marco muestral pero se quiere estudiar una situación en particular como por ejemplo el tiempo promedio que se demora un cajero en atender un cliente, en este caso se decide tomar el tiempo que se demora con un cliente que llega cada X tiempo o cada Y clientes; conseguir la proporción de clientes con determinada característica en un archivo, en este caso se puede sacar una muestra cada Z longitud del archivo, la longitud Z resulta de la división de la longitud total del archivo por el tamaño de muestra n; conocer la opinión de los clientes que frecuentan un establecimiento público respecto al servicio que reciben, en este caso se puede decidir en entrevistar al cliente que llega cada X tiempo o cada Y clientes.

Selección de una muestra aleatoria si se conoce N y se tiene el marco muestral:

La población esta numerada de 1 a N, se divide N por n teniendo como resultado k que se denomina tamaño del intervalo muestral, suponiendo que k sea un número entero, se elige un número (r) aleatorio entre 1 y k (utilizando esta opción en Excel) , siendo r el primer elemento seleccionado, seguidamente se suma k al este número aleatorio r, para obtener el segundo elemento seleccionado y así hasta completar el tamaño de la muestra, es decir que los elementos seleccionados son r, r + k, r + 2k, … , r + (n-1)k; de acuerdo con lo anterior solo se pueden tener k posibles muestras sistemáticas.

Si k no es un número entero, se puede proceder de dos formas, la primera, tomar la parte entera de k, multiplicarla por n y realizar la resta N – nk, y este resultado es el número de elemento que deben ser eliminados aleatoriamente de la población, por ejemplo se tiene N=1345, n=160, k=1345/160=8.406, la parte entera es 8 de tal forma que 160*8=1280, por lo tanto se deben eliminar aleatoriamente 65 elementos de la población y finalmente sacar la muestra aleatoria de los 1280 elementos restantes; la segunda es elegir aleatoriamente un punto de partida entre 1 y k , siendo este el primer elemento de la muestra y a este se le suma k hasta completar la muestra de tamaño n.

Para los casos en que se desconoce N este se puede estimar de acuerdo a cada caso específico, por ejemplo si en una longitud Z hay w elemento entonces N=Z*w, si se toma un elemento cada determinado tiempo se deberá tener en cuenta todos los elementos que hay entre cada intervalo de tiempo y sumarlos, estos más los n elementos muestreados conforman la población, si se desea muestrear uno de cada 10 personas que llegan a un determinado lugar, primero se elige un número aleatorio entre 1 y 10, suponga que salió el 8 y que el ultimo muestreado fue el 608 y después de este llegaron 6 personas más entonces N = 608 + 6 = 614.

Bajo el supuesto de que el ordenamiento de la variable de estudio es aleatorio el tamaño de la muestra es el mismo que para un MAS.

Muestreo aleatorio estratificado (MAE)

Este tipo de muestreo se caracteriza porque la población se organiza en estratos, cada uno de ellos es homogéneo al interior y heterogéneo entre ellos, en este sentido lo que se busca es un mejor estimador del parámetro y un ahorro de recursos, ya que la variabilidad por estrato es menor, por lo tanto lo que se hace es sacar una muestra aleatoria en cada estrato, lo más común es que cada tamaño de la muestra sea proporcional al tamaño de cada estrato, que generalmente son de diferente tamaño.

N: Tamaño de la población

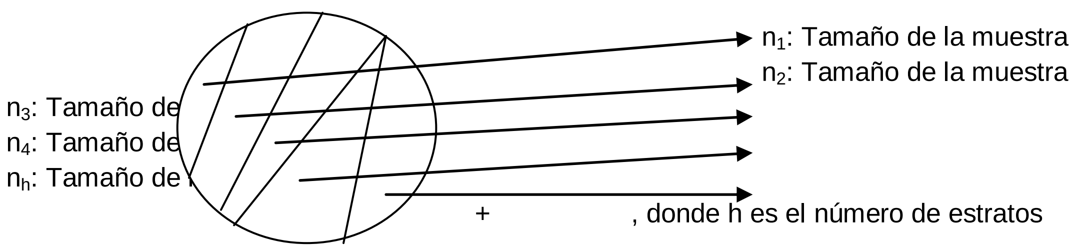


Figura No. 4 Muestreo Aleatorio Estratificado

Afijación de la muestra:

Si todos los estratos tienen aproximada mente el mismo tamaño y no se conoce la variabilidad dentro de cada estrato se puede asignar un tamaño de muestra igual para cada estrato dado por:

$$n_h = \frac{n}{H}$$

Si los estratos son de diferentes tamaños se puede usar la afijación proporcional, con el fin de darle a todas las unidades poblacionales la misma probabilidad:

$$n_h = nW_h$$

Sin embargo las afijaciones anteriores no cumple con el fin último de la estratificación, ya que no optimizan el estimador, debido a que no tienen en cuenta la homogeneidad de cada estrato, por lo que se sugieren las siguientes formas de afijar el tamaño de la muestra encada en cada estrato:

Si no se tienen en cuenta los costos por cada elemento muestreado:

$$n_h = n\frac{W_h\sigma_h}{\sum_{h=1}^{H} W_h\sigma_h}, \text{ donde } \sigma_h \approx s_h$$

Si se tiene en cuenta el costo por cada elemento muestreado:

$$n_h = n\frac{\frac{W_h\sigma_h}{\sqrt{C_h}}}{\sum_{h=1}^{H}\frac{W_h\sigma_h}{\sqrt{C_h}}}, \text{ donde } \sigma_h \approx s_h,\ C = C_0 + \sum_{h=1}^{H} C_h n_h$$

C es el presupuesto asignado a la recolección de información, C_0 son los costos fijos que no dependen del número de elementos a seleccionar y C_h es el costo por cada elemento muestreado.

Tamaño de la muestra para la estimación de la media (μ) y del total (T)

El tamaño de la muestra que minimiza la varianza de la media total dado un presupuesto fijo es:

$$n = \frac{(C - C_0)\sum_{h=1}^{H}\left(\mathrm{W_h}\ \sigma_\mathrm{h} / \sqrt{\mathrm{C_h}}\right)}{\sum_{\mathrm{h}=1}^{H}\left(\mathrm{W_h}\ \sigma_\mathrm{h}\sqrt{C_h}\right)}, \text{ donde } \sigma_h \approx s_h$$

El tamaño de la muestra que minimiza el costo dado un error admitido y un nivel de confianza es:

$$n = \frac{n_0}{1 + \frac{z_{\alpha/2}^2}{N\beta^2}\sum_{h=1}^{N} W_h\sigma_h^2}, \text{ donde } n_0 = \frac{\sum_{\mathrm{h}=1}^{H}\left(\mathrm{W_h}\ \sigma_\mathrm{h}\sqrt{C_h}\right)\left(\sum_{\mathrm{h}=1}^{H}(\mathrm{W_h}\ \sigma_\mathrm{h}) / \left(\sqrt{C_h}\right)\right)}{\beta^2 / \mathrm{z}_{\alpha/2}^2}, \text{ Donde } \sigma_h \approx s_h$$

Si los costos son desconocidos el tamaño de la muestra que optimiza el estimador es:

$$n = \frac{\mathrm{z}_{\alpha/2}^2\left(\sum_{h=1}^{H} W_h\sigma_h\right)^2}{\beta^2 + \frac{\mathrm{z}_{\alpha/2}^2}{N}\sum_{h=1}^{H} W_h\sigma_h^2}, \text{ donde } \sigma_h \approx s_h$$

Si se desea hacer una afijación proporcional el tamaño de la muestra se puede calcular así:

$$n = \frac{n_0}{1 + {n_0}/{N}}, \text{ Donde } n_0 = \frac{\sum_{h=1}^{H} W_h \sigma_h^2}{\beta^2 / z_{\alpha/2}^2}, \text{ donde } \sigma_h \approx s_h$$

Tamaño de la muestra para la estimación de la proporción (p) y del total (T)

El tamaño de la muestra que minimiza la varianza de la media total dado un presupuesto fijo es:

$$n = \frac{(C - C_0)\sum_{h=1}^{H}\left(W_h \sqrt{{p_h q_h}/{C_h}}\right)}{\sum_{h=1}^{H}\left(W_h \sqrt{p_h q_h C_h}\right)}, \text{ donde } p_h \approx \hat{p}_h$$

El tamaño de la muestra que minimiza el costo dado un error admitido y un nivel de confianza es:

$$n = \frac{n_0}{1 + \frac{z_{\alpha/2}^2}{N\beta^2}\sum_{h=1}^{N} W_h p_h q_h}, \text{ donde } n_0 = \frac{\sum_{h=1}^{H}\left(W_h \sqrt{p_h q_h C_h}\right)\left(\sum_{h=1}^{H} W_h \sqrt{{p_h q_h}/{C_h}}\right)}{\beta^2 / z_{\alpha/2}^2}, \text{ donde } p_h \approx \hat{p}_h$$

Para los dos casos anteriores $n_h = n \dfrac{N_h \sqrt{{p_h q_h}/{C_h}}}{\sum_{h=1}^{H} N_h \sqrt{{p_h q_h}/{C_h}}}$, donde $p_h \approx \hat{p}_h$

Si los costos son desconocidos el tamaño de la muestra que optimiza el estimador es:

$$n = \frac{n_0}{1 + \frac{z_{\alpha/2}^2}{N\beta^2}\sum_{h=1}^{H} W_h p_h q_h}, \text{ donde } n_0 = \frac{\left(\sum_{h=1}^{H} W_h \sqrt{p_h q_h}\right)^2}{\beta^2 / z_{\alpha/2}^2}, \text{ Donde } p_h \approx \hat{p}_h$$

La afijación de la muestra es:

$$n_h = n\frac{W_h\sqrt{p_h q_h}}{\sum_{h=1}^{H} W_h\sqrt{p_h q_h}}$$, donde $p_h \approx \hat{p}_h$

Si se desea hacer una afijación proporcional el tamaño de la muestra se puede calcular así:

$$n = \frac{n_0}{1+ n_0/N}$$, Donde $$n_0 = \frac{\sum_{h=1}^{H} W_h p_h q_h}{\beta^2 / z_{\alpha/2}^2}$$, Donde $p_h \cong \hat{p}_h$

Selección de la muestra

Para seleccionar la muestra se procede como se indicó en el numeral 4.2.1.3, pero realizando el mismo procedimiento en cada estrato.

Muestreo aleatorio por conglomerados (MCON)

N: Número de conglomerados en que se divide la población

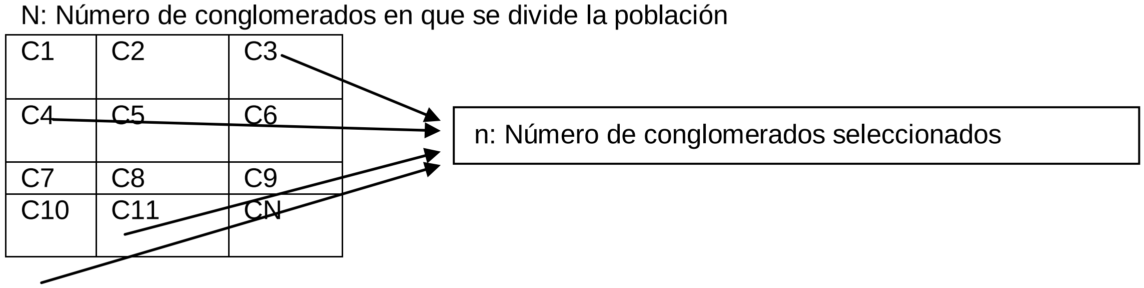


Figura No. 5 Muestreo Aleatorio por Conglomerado

En un muestreo por conglomerados, la población se divide en subgrupos de tal forma que al interior de cada conglomerado los elementos sean heterogéneos y los conglomerados entre sí sean homogéneos, la técnica de muestreo consiste en elegir aleatoriamente n conglomerados de los N conglomerados y realizar el estudio con todos los elementos que componen cada conglomerado, ejemplos de conglomerados pueden ser las manzanas de una ciudad, los edificios, los cajones de un archivador, las canastas de gaseosa en una línea de producción, una repartición geográfica delimitada por limites claramente definidos, en la práctica los conglomerados pueden ser de igual o de diferente tamaño, pudiéndose definir de diferente forma dentro de la población por ejemplo manzanas, edificios y conjuntos residenciales.

Al aplicar un MCON generalmente se ahorra tiempo y dinero, pero se pierde precisión en la estimación debido a que generalmente la varianza es mayor que si el muestreo es MAS o MEA, cuando se va realizar un MCON es necesario tener en cuenta los siguientes factores:

1. Los conglomerados deben ser mutuamente excluyentes.
2. Debe existir una estimación del número de elementos en cada conglomerado.

3. Los conglomerados deben ser lo suficientemente pequeños para que se posibilite el ahorro en los costos.

En este caso se usaran conglomerados de igual tamaño

La selección se realiza utilizando un muestreo aleatorio simple y así seleccionar n conglomerados, (la selección se realiza de igual forma que en la explicada en el numeral 4.2.1.3) la población está compuesta por N conglomerados cada uno con M unidades, por lo tanto el número de elementos en la población es

Estimación de la media poblacional:

$\overline{x} = \sum_{i=1}^{n} \sum_{j=1}^{M} \frac{x_{ij}}{n}$ Media muestral por conglomerado

$\overline{\overline{x}} = \frac{\overline{x}}{M}$ Media muestral por unidad

Con error estándar $ee\left[\overline{\overline{x}}\right] = \sqrt{\frac{\frac{N(N-n)}{n} \sum_{i=1}^{n} \frac{(X_i - \overline{x})^2}{n-1}}{N^2 M^2}}$, X_i: Es el total por conglomerado

Intervalo de confianza al $(1-\alpha)100\%$ para μ

$\overline{\overline{x}} \pm t_{\left(n-1, \alpha/2\right)} ee\left[\overline{\overline{x}}\right]$

4.2.4.3 Estimación para el total poblacional

$\hat{T}_{con} = N\overline{x} = M_o \overline{\overline{x}}$

Con error estándar $ee\left[\hat{T}_{con}\right] = \sqrt{\frac{N(N-n)}{n} \sum_{i=1}^{n} \frac{(X_i - \overline{x})^2}{n-1}}$, X_i: Es el total por conglomerado

Intervalo de confianza al $(1-\alpha)100\%$ para T

$\hat{T}_{con} \pm t_{\left(n-1, \alpha/2\right)} ee\left[\hat{T}_{con}\right]$

Tamaño de la muestra

$$n = \frac{\dfrac{z^2_{\alpha/2} s^2_{con}}{\beta^2 M^2}}{1 + \dfrac{1}{N}\left(\dfrac{z^2_{\alpha/2} s^2_{con}}{\beta^2 M^2}\right)}$$, donde $s^2_{con} = \sum_{i=1}^{n} \frac{(X_i - \bar{x})^2}{n_1 - 1}$, n_1 : Premuestra.

Estimación de la proporción poblacional

$\hat{p}_i = \frac{x_i}{M}$, proporción estimada en el i-ésimo conglomerado

$$\hat{p}_{con} = \sum_{i=1}^{n} \frac{p_i}{n}$$

Con error estándar $ee[\hat{p}_{con}] = \sqrt{\frac{N-n}{Nn} \sum_{i=1}^{n} \frac{(\hat{p}_i - \hat{p}_{con})^2}{n-1}}$

Intervalo de confianza al $(1-\alpha)100\%$ para p

$$\hat{p}_{con} \pm t_{\left(n-1, \alpha/2\right)} ee[\hat{p}_{con}]$$

Estimación para el total poblacional

$$\hat{T}_{con} = N\hat{p}_{con}$$

Con error estándar $ee[\hat{T}_{con}] = Nee[\hat{p}_{con}]$, X_i: Es el total por conglomerado

Intervalo de confianza al $(1-\alpha)100\%$ para T

$$\hat{T}_{con} \pm t_{\left(n-1, \alpha/2\right)} ee[\hat{T}_{con}]$$

Tamaño de la muestra

$$n = \frac{\dfrac{z^2_{\alpha/2} s^2_{con}}{\beta^2 M^2}}{1 + \dfrac{1}{N}\left(\dfrac{z^2_{\alpha/2} s^2_{con}}{\beta^2 M^2}\right)}$$, donde $s^2_{con} = \sum_{i=1}^{n} \frac{(\hat{p}_i - \hat{p}_{con})^2}{n_1 - 1}$, n_1 : Premuestra.

Calculo del tamaño de la muestra

Para calcular el tamaño de la muestra en cada uno de los casos anteriores, se pueden utilizar las funciones de Excel, con el fin de poder variar el error y el nivel de confianza y así poder optimizar los recursos asignados a cada investigación.

Bibliografía

Cochran, W. G. (2000). Técnicas de Muestreo. (1 ed.). México: CECSA.

KINNEAR, T.C. & Taylor, J.R. (1999) Investigación de mercados (5 ed.). México: Mc Graw Hill.

Lohr, S.L. (2000) Muestreo: Diseño y Análisis. (1 ed.). México: Thomson.

Mendenhall, W. (1990). Elementos de Muestreo. (1 ed.). México: Iberoamercana

Ospina, D.B. (2001). Introducción al Muestreo (1 ed.). Colombia: UNIBIBLOS