

TEMA 13 : MUESTREO POR LISTA  
 por: Enrique Wabo

**1 INTRODUCCIÓN**

En ocasiones se deben hacer estimaciones a partir de unidades de distinto tamaño. En estos casos, un muestreo simple no tomaría en cuenta la importancia del tamaño de las unidades, ya que le daría el mismo peso a las unidades grandes y pequeñas. Hay métodos de muestreo que permiten asignarle a las unidades de la población un peso diferencial al momento de seleccionar la muestra. Uno de estos métodos es el Muestreo Por Lista, también llamado Muestreo con Probabilidad Proporcional al Tamaño (PPS, en inglés: *Probability Proportional to Size*).

En este diseño, el peso diferencial se asigna a través de una variable auxiliar X. En casos como el ejemplo recién mencionado, la variable X, que es la superficie de cada rodal, surge por sí sola; en otros casos vamos a tener libertad para elegir esa variable. Las características principales del Muestreo por Lista son:

- a) es un método de muestreo con reemplazo;
- b) utiliza una variable auxiliar X, que debe estar alta y positivamente correlacionada con Y;
- c) las unidades de muestreo son seleccionadas con probabilidad proporcional al valor asociado de la variable auxiliar X; y
- d) para seleccionar la muestra se requiere una lista de todas las unidades de muestreo que conforman la población, cada una con el correspondiente valor de X.

Como ejemplo supongamos una empresa que posee un alto número de rodales de distinto tamaño (X) y que está interesada en conocer el volumen de madera (Y) contenido en ellos. Un diseño simple llevaría al riesgo de una estimación sesgada, ya que los rodales tienen distinto tamaño y distinto peso relativo. En este caso, sería deseable que las unidades de muestreo fuesen seleccionadas con probabilidad proporcional a su volumen; el volumen sería la variable de interés Y, en tanto que la superficie de cada rodal sería la variable auxiliar X.

**2 SELECCIÓN DE LA MUESTRA**

El mecanismo de muestreo se inicia con la confección de un listado de las unidades de muestreo de la población, cada una con su correspondiente valor X. Los valores de X se suman en forma acumulativa para cada unidad; los valores acumulados van desde el 1 hasta el total acumulado (suma de todos los X).

Si n es el tamaño de la muestra, se seleccionan n números aleatorios desde 1 hasta el total; una unidad de muestreo es seleccionada cuando el número aleatorio seleccionado es menor o igual al acumulado de X.

La Tabla 1 muestra un ejemplo de organización del muestreo. La columna 4 muestra el intervalo que le corresponde a cada unidad de muestreo. Las columnas 2 y 5 muestran idénticos valores, indicando que los rangos de la columna 4 representan los tamaños de X.

TABLA 1

Unidad N°	Valor X Xi	Valores Acumulados	Subseries asociadas	N° de términos
1	2	3	4	5
1	5	5	1 - 5	5
2	10	15	6 - 15	10
3	7	22	16 - 22	7
4	12	34	23 - 34	12
5	6	40	35 - 40	6

Una vez completada la tabla se seleccionan tantos números aleatorios como unidades se pretende en la muestra: n. Para el ejemplo, desde el 1 hasta el 40. Supongamos que el primer número aleatorio fue el 27, eso quiere decir que la unidad 4 ha sido seleccionada en la muestra. Si una misma unidad es seleccionada más de una vez, se debe repetir el muestreo en esa unidad tantas veces como veces haya sido seleccionada.

**3 ESTIMACIONES**

**3.1 ESTIMADOR DE MEDIA DE RAZONES**

Como fuera mencionado, los valores de la variable Y varían de una unidad a otra en parte debido a que el “tamaño” de la unidad también lo hace. Por eso, la estimación recurre a la media de las razones entre los valores de la variable Y de interés y la variable X observados en cada unidad de la muestra. Definimos a la razón individual entre Y y X, como:

$$r_i = \frac{y_i}{x_i} \tag{1}$$

El promedio de r en la muestra es:

$$\bar{r} = \frac{1}{n} \sum_1^n r \tag{2}$$

con varianza estimada:

$$S_{\bar{r}}^2 = \frac{\sum r^2 - \frac{(\sum r)^2}{n}}{n(n-1)} \tag{3}$$

El total estimado de Y es:

$$\hat{Y} = \bar{r} X = \bar{r} \sum_1^N x \tag{4}$$

con error estándar:  $s_{\hat{Y}} = X^2 s_{\bar{r}}$

**3.2 MUESTREO BIETÁPICO**

La primera etapa del muestreo se hace siguiendo un muestreo por lista; en la segunda etapa se hace un muestreo simple. Curiosamente, las fórmulas son las mismas que se emplean en el muestreo simple.

Dentro de cada unidad primaria se obtiene el promedio por unidad secundaria. Posteriormente se promedian esas medias y se obtiene su error estándar:

$\bar{y}_i$  = media por US (parcela) en la i-ésima UP (rodal).

$$\bar{y}_2 = \text{media por US en la población}; \bar{y}_2 = \frac{\sum_{i=1}^n \bar{y}_i}{n}$$

$$S_{\bar{y}_2}^2 = \frac{S_{\bar{y}}^2}{n} = \frac{\sum \bar{y}_i^2 - \frac{(\sum \bar{y}_i)^2}{n}}{n(n-1)}$$

$$\hat{Y} = M \bar{y}_2 \quad s_{\hat{Y}} = M s_{\bar{y}_2}$$

**4 APLICACIONES ESPECIALES**

Hay situaciones en donde el muestreo por lista ocurre en forma automática, sin requerir el armado de la lista inicial. Por ejemplo, si disponemos de un mosaico aerofotográfico que muestra los rodales forestales de una región, podemos seleccionar una "lista muestra" sin necesidad de usar una lista. Una muestra puede ser seleccionada de la siguiente manera. Una grilla de coordenadas es superpuesta sobre el mosaico, y a partir de una tabla de números aleatorios n conjuntos de coordenadas (x , y) son seleccionados. Cada conjunto define un punto aleatorio sobre el mosaico. Obviamente, la probabilidad de que tal punto caiga sobre un rodal, es proporcional al área (foto aérea) de la carta. Si el rodal i recibe  $a_i$  puntos, participa  $a_i$  veces en la muestra. Por medio de una grilla de puntos, las áreas individuales  $x_i$  de los rodales seleccionados son estimadas, junto con el área forestal total  $\sum x_i$ . Así, las  $p_i$  de los rodales seleccionados son conocidas y sus  $Y_i$ 's son medidas en el campo.

**5 EJEMPLO**

Una empresa posee 10 rodales ocupados por árboles de una misma especie y de la misma clase de edad. Es su interés estimar el volumen de madera existente en los 10 rodales (N). Se aplica un muestreo por lista, donde la variable auxiliar es la superficie boscosa de cada rodal. Se seleccionan 4 rodales (n) y a cada uno se le mide el volumen al 100 por ciento. En la tabla que se adjunta se indican los valores de interés para el muestreo.

Unidad	X	ΣX
1	5	5
2	2	7
3	30	37
4	120	157
5	15	172
6	10	182
7	13	195
8	25	220
9	140	360
10	160	520

Se seleccionaron 4 números aleatorios: 108, 176, 296 y 470; de manera que los rodales seleccionados fueron el 4, el 6, el 9 y el 10. En el siguiente cuadro se indican los valores obtenidos en la muestra:

Unidad	X	Volumen Total (m³)	$n_i$ (m³/ha)
4	120	52.120	52.120/120 = 434,333
6	10	5.680	5.680/ 10 = 568,000
9	140	65.400	65.400/140 = 467,143
10	160	70.220	70.220/160 = 438,875

$\Sigma r = 1.908,351 \quad \Sigma r^2 = 922.103,003$

$\bar{r} = 1.908,351/4 = 477,088 \text{ m}^3/\text{ha}$

$s_r^2 = 11.652,118/(4-1) = 3.884,039$

$s_{\bar{r}} = (3.884,039/4)^{0.5} = 31,16 \text{ m}^3/\text{ha}$

Total estimado:

$(477,088 \text{ m}^3/\text{ha}) (520 \text{ ha}) = 248.086 \text{ m}^3$

Error estándar del total:

$(31,16 \text{ m}^3/\text{ha}) (520 \text{ ha}) = 16.203 \text{ m}^3$

**BIBLIOGRAFÍA**

Fresse F. 1962. Muestreo Forestal Elemental.

## EJERCICIO

1. Una propiedad contiene 600 ha plantadas con una misma especie, divididas en 15 rodales. Usted se dispone a hacer un inventario de esas plantaciones y ha optado por un muestreo por lista con una muestra de tamaño 5; a cada rodal seleccionado se le mide el volumen en su totalidad. La siguiente tabla muestra la superficie de cada rodal (X). Usted está aplicando muestreo por lista en una propiedad de 720 ha de extensión. Estime el volumen total (Y) y su intervalo de confianza para un 95% (use  $t = 2$ ).

Rodal	X (ha)	Volumen (m3)
1	5	
2	2	
3	30	
4	120	
5	15	
6	10	
7	13	
8	25	
9	131	
10	160	
11	15	
12	25	
13	13	
14	24	
15	12	

2. En una propiedad con 20 rodales y que ocupan una superficie de 800 ha se aplicó un muestreo bietápico, en el cual se aplicó muestreo por lista para la primera etapa del muestreo. Los datos se indican en el cuadro adjunto. Estime el volumen total y su error estándar.

U. Primaria	U. Secundaria	Vol. (m3/ha)
1	1	230
	2	181
	3	240
2	1	120
3	1	110
	2	130
	3	150
	4	180
4	1	240
	2	220
	3	225
	4	205
5	1	300

## RESPUESTAS

## Ejercicio 1:

Mediante números aleatorios se seleccionó la muestra de tamaño 5; las unidades seleccionadas fueron: 4, 5, 9, 10 y 15. Los valores de área y volumen seleccionados fueron:

X (ha)	Y (m3)	$r_i$
120	42.000	350,00
15	5.100	340,00
131	50.200	383,21
160	46.600	291,25
12	3.600	300,00

Suma de y = 1664,46      Suma de  $y^2$  = 559.773,48

Media de r =  $1.664,46/5 = 332,89$  m<sup>3</sup>/ha

Error Estándar de r =  $37,72/\sqrt{5} = 16,86$  m<sup>3</sup>/ha

Total estimado =  $332,89$  m<sup>3</sup>/ha x 600 ha = 199.734 m<sup>3</sup>

E.E. del total est. =  $16,86$  m<sup>3</sup>/ha x 600 ha = 10.116 m<sup>3</sup>

IC(95%) =  $\pm 20.232$  m<sup>3</sup>    Li = 179.502 m<sup>3</sup>    Ls = 219.966 m<sup>3</sup>

## Ejercicio 2:

Promedios por US:

U.P.	Media/US
1	217,00
2	120,00
3	142,50
4	222,50
5	300,00

$$\bar{y} = \frac{\sum_{i=1}^5 \bar{y}_i}{5} = \frac{1.002}{5} = 200,4 \text{ m}^3/\text{ha}$$

$$S_{\bar{y}} = \frac{71,59}{\sqrt{5}} = 32,02 \text{ m}^3$$

Total Estimado = 800 ha x 200,4 m<sup>3</sup>/ha = 160.320 m<sup>3</sup>

EE Total Estimado = 800 ha x 32,02 = 25.616 m<sup>3</sup>