CAPÍTULO 9: MUESTREO POR CONGLOMERADOS DE UNA Y DOS ETAPAS

por: Enrique Wabo

1 MUESTREO POR CONGLOMERADOS DE UNA ETAPA

1.1 INTRODUCCIÓN

El Muestreo por Conglomerados de Una Etapa es comúnmente llamado Muestreo por Conglomerados, simplemente. En el muestreo por conglomerados, cada unidad de muestreo está conformada por un conjunto o conglomerado de elementos. Los elementos, a su vez, pueden considerarse como si fueran unidades secundarias, con lo cual, el conglomerado sería equivalente a una unidad primaria. Podemos decir, entonces, que en el muestreo por conglomerados, una vez que una unidad primaria o conglomerado es seleccionada, son medidas todas las unidades secundarias o elementos dentro de la unidad primaria.

Un ejemplo de este tipo de muestreo es el muestreo mediante parcelas de muestreo de tamaño fijo: mientras la parcela es la unidad primaria, cada árbol es una unidad secundaria. Sin embargo, deben hacerse las siguientes aclaraciones sobre el uso de las parcelas de tamaño fijo y su relación con el muestreo por conglomerados:

a) cuando el propósito de las mediciones es estimar alguna característica a expresarse por unidad de superficie, como volumen por hectárea o área basal por hectárea, se considera como un muestreo aleatorio (sistemático) simple y corresponde aplicar entonces las fórmulas de este diseño, ya vistas; y

b) cuando el propósito es estimar una proporción de alguna clase de árboles, como proporción de árboles enfermos, se considera una aplicación del estimador de razón y corresponde aplicar las fórmulas de este mecanismo.

En consecuencia, el muestreo por conglomerados de elementos queda reservado a aquellos casos en los que la característica a estimar corresponde a los árboles individuales, como diámetro medio por árbol o altura media por árbol o volumen medio por árbol. Aquí, la verdadera unidad de muestreo es el árbol y los árboles son seleccionados por conglomerados.

1.2 SÍMBOLOS A USAR

N = número total de conglomerados en la población; surge como N = área total/área de la parcela

M = número total de elementos en la población

n = número de conglomerados seleccionados en la muestra

m = número total de elementos en la muestra:

$$m = \sum_{1}^{n} m_i$$

m_i = número de elementos en el i-ésimo conglomerado

$\overline{M}$ = número medio de elementos por conglomerado; que surge como M/N.

$\overline{m}$ = número medio estimado de elementos por conglomerado; es:

$$\overline{m} = \frac{\sum_{1}^{n} m_i}{n}$$

y es el estimador de $\overline{M}$.

y_{ij} = valor de Y observado en el j-ésimo elemento (con j= 1- m_i) del i-ésimo conglomerado (con i = 1- n)

y_i = total de Y en el i-ésimo conglomerado;

$$y_i = \sum_{1}^{m_i} y_{ij}$$

$\overline{y}_E$ = media de Y por elemento en la muestra.

$\overline{y}_C$ = media de Y por conglomerado en la muestra.

1.3 ESTIMACIONES

1.3.1 Media por elemento y su error estándar

- Media estimada por elemento

$$\overline{y}_E = \frac{\sum_{1}^{n} y_i}{\sum_{1}^{n} m_i} = \tag{1}$$

- Error estándar estimado de la media

La varianza de la media se estima con:

$$s_{\overline{y}_E}^{\ 2} = \frac{1}{\overline{M}^2}\frac{s_E^{\ 2}}{n}\frac{(N-n)}{N} \tag{2}$$

pero como $\overline{M}$ rara vez es conocido, en lugar de (2) se puede usar:

$$s_{\overline{y}_E}^{\ 2} = \frac{1}{\overline{m}^2}\frac{s_E^{\ 2}}{n}\frac{(N-n)}{N} \tag{3}$$

siendo:

$$s_E^{\ 2} = \frac{\sum_{1}^{n} y_i^2 + \overline{y}_E^{\ 2}\sum_{1}^{n} m_i^2 - 2\,\overline{y}_E\sum_{1}^{n} y_i\, m_i}{n-1} \tag{4}$$

El total estimado de Y para la población puede estimarse a partir de $\overline{y}_E$ mediante:

$$\hat{Y} = \overline{y}_E\, M \tag{5}$$

Sin embargo, el término M suele ser desconocido, por lo que se usa otra expresión, la que se verá en el siguiente punto.

1.3.2 Estimación del total y su error estándar

Una alternativa para estimar el total de la población es la siguiente fórmula:

$$\hat{Y} = M\,\bar{y}_C \qquad (5)$$

Puede observarse que esta expresión no es otra que la correspondiente al muestreo aleatorio (sistemático) con parcelas de tamaño fijo y el objetivo es conocer el volumen medio por parcela para después por expansión determina el volumen total. Así, (5) se convierte en:

$$\hat{Y} = N\,\bar{y}_C = N\,\bar{y} \qquad (6)$$

con error estándar: $S_{\hat{Y}} = N\,s_{\bar{y}}$ (7)

En otras palabras y como ya fuera aclarado en la introducción, cuando lo que se pretende saber es el total de Y para la población aplicamos las fórmulas de estimación del muestreo aleatorio (sistemático) simple.

1.3.3 Ejemplo

En un rodal de 12 ha se instalan 5 parcelas de 500 m2 cada una y se pretende conocer el volumen medio por árbol. La tabla siguiente muestra los valores observados (NOTA: la muestra es chica porque es a los fines de mostrar un ejemplo).

Nº	Volumen (dm3)	yi	mi
1	231 234 262	727	3
2	210 268	478	2
3	421 368	789	2
4	189	189	1
5	246 438	684	2

$\Sigma y = 2.867$ $\Sigma m = 10$ $\Sigma ym = 6.278$

$\Sigma y^2 = 1.883.111$ $\Sigma m^2 = 22$ $n = 5$

$N = 12/0{,}05 = 240$ $\bar{m} = 10/5 = 2$ árb/cong.

- Media/conglomerado = 2.867/10 = 286,7 = $\bar{y}_E$

$$- S_E^2 = \frac{1.883.111 + (286{,}7)^2 (22) - (2)\,(286{,}7)\,(6278)}{4}$$

$$= 22.909{,}345$$

$$S_{\bar{y}_E}^{\;2} = \frac{1}{(2^2)}\,\frac{22.909{,}345}{5}\,\frac{(240-5)}{240} = 1.121{,}60$$

$$S_{\bar{y}_E} = \sqrt{1.121{,}60} = 33{,}49 \text{ dm3/árbol}$$

Final:

- Media : 286,7 dm³/árbol
- E. estándar: 33,5 dm³/árbol
- IC(95%) : 286,7 ± (2) (33,5)

Lím. Inferior : 219,7 dm³

Lím. Superior: 353,7 dm³

Tenemos una confianza del 95 % de que el volumen medio por árbol está entre 220 y 354 dm cúbicos.

2 MUESTREO POR CONGLOMERADOS EN DOS ETAPAS O MUESTREO BIETÁPICO

2.1 INTRODUCCIÓN

El muestreo por conglomerados en dos etapas en comúnmente llamado Muestreo en Dos Etapas o Muestreo Bietápico, ya que el muestreo se realiza en dos etapas

En la primera etapa, una muestra de unidades, llamadas unidades primarias, es seleccionada de la población. Las unidades primarias pueden ser rodales, lotes, estratos, parcelas de muestreo, etc. Cada unidad primaria está a su vez dividida en unidades más pequeñas llamadas unidades secundarias.

En la segunda etapa, sólo una parte de las unidades secundarias situadas dentro de cada unidad primaria es observada. Por eso este diseño recibe también el nombre de muestreo con submuestreo.

Las razones para aplicar este diseño son variadas, pero puede ser especialmente apto cuando no hay tiempo o fondos suficientes para observar todos los rodales, lotes o estratos.

Por cierto, se puede incorporar una tercera etpa de muestreo, una cuarta, etc.; pero el más usado es el bietápico.

2.2 NOMENCLATURA

Se selecciona una muestra de Unidades Primarias (UP) y dentro de cada unidad primaria se selecciona una muestra de Unidades Secundarias (US).

N: número de UP en la población.

M: número total de US dentro de cada UP.

n: número de UP seleccionadas en la muestra;
i = 1, 2, ..., n.

m: número US seleccionadas dentro de cada UP;
j = 1, 2, ..., m

$\bar{\bar{y}}$: media por US.

S_b^2 : varianza entre UP's.

S_W^2 : varianza dentro de las UP's; es equivalente al término s_y^2 del muestreo simple y se calcula con las mismas fórmulas.

El subíndice b representa a la palabra *between* (*entre*, en inglés); el subíndice w representa la palabra *within* (dentro, en inglés).

2.3 ESTIMADORES

Los componentes que se utilizan para las estimaciones más importantes son las tres siguientes:

a) la sumatoria de todos los valores (STV) observados en las n x m unidades secundarias, expresada como:

$$STV = \sum_{i}^{n} \sum_{j}^{m} y_{ij}$$

b) la sumatoria de los cuadrados (SCV) de todos los valores observados en las n x m unidades secundarias, expresada como:

$$SCV = \sum_{i}^{n} \sum_{j}^{m} y_{ij}^2$$

c) la suma de los cuadrados de los totales por unidad primaria (T^2), correspondiente a las m unidades secundarias:

$$ST^2 = \sum_{i}^{n} T_i^2$$

A continuación se indican los principales estimadores:

- media por unidad secundaria:

$$\bar{\bar{y}} = \frac{STV}{n \times m} \qquad (8)$$

-

- error estándar de la media:

$$s_{\bar{\bar{y}}} = \sqrt{\frac{1}{m\,n}\left[s_b^2 \left(1 - \frac{n}{N}\right) + \frac{n}{N} \times s_w^2 \left(1 - \frac{m}{M}\right)\right]} \qquad (9)$$

siendo:

$$s_b^2 = \frac{\dfrac{\sum_{1}^{n} T_i^2}{m} - \dfrac{STV^2}{m\,n}}{(n-1)} \qquad (10)$$

$$s_w^2 = \frac{SCV - \dfrac{\sum_{1}^{n} T_i^2}{m}}{n\,(m-1)} \qquad (11)$$

2.4 EJEMPLO DE BIETÁPICO

Ejemplo extraído de F. Freese. Se va a estimar el volumen total de madera de una superficie total de 30.000 ha. La superficie fue dividida en 1500 UP de 20 ha cada una; cada UP contiene 160 US de 0,1250 m2 de superficie. Se aplica un muestreo bietápico con unidades primarias de igual tamaño. La muestra está conformada por 4 UP y 3 US por UP. En la Tabla 1 se indican los volúmenes por hectárea registrados en cada US.

Tabla 1

UP	Vol. m^3/ha	
1	147 180 206	Σy = 533 Σy^2 = 96.445
2	312 265 300	Σy = 877 Σy^2 = 257.569
3	220 280 210	Σy = 710 Σy^2 = 170.900
4	250 232 185	Σy = 667 Σy^2 = 150.549
	2.787	

N = 1500 n = 4
M = 160 m = 3

STV = 147 + 180 + ··· + 232 + 185 = 2.787
SCV = $147^2 + 180^2 + \cdots + 185^2$ = 675.463
ST2 = $533^2 + 877^2 + \cdots + 667^2$ = 2.002.207

Cálculo de componentes:

$$s_b^2 = \frac{\dfrac{2.002.207}{3} - \dfrac{(2.787)^2}{3 \times 4}}{(4-1)}$$

$$s_b^2 = \frac{667.402{,}33 - 647.28075}{3} = 6.707{,}19$$

$$s_w^2 = \frac{675.463 - 667.402{,}33}{4\,(3-1)}$$

$$s_w^2 = \frac{675.463 - 667.402{,}33}{8} = 1.007{,}58$$

Estimaciones:

$$\boxed{\bar{\bar{y}} = \frac{2.787}{12} = 232{,}25 \text{ m}^3/\text{ha}}$$

$$s_{\bar{\bar{y}}} = \sqrt{\frac{1}{4 \times 3}\left[6.707{,}19 \left(1 - \frac{4}{1500}\right) + \frac{4}{1500}\ 1007{,}58 \left(1 - \frac{3}{160}\right)\right]}$$

$$s_{\bar{\bar{y}}} = \sqrt{\frac{1}{12}\left[6.707{,}19\ (0{,}9973) + 2{,}687\ (0{,}98125)\right]}$$

$$\boxed{s_{\bar{\bar{y}}} = \sqrt{\frac{6.691{,}72}{12}} = 23{,}61 \text{ m}^3/\text{ha}}$$

El total estimado y su error estándar son:

$$\boxed{\hat{Y} = \bar{\bar{y}}\ M\ N = 232{,}25 \times 160 \times 1500 = 55.740.000 \text{ m}^3}$$

$$\boxed{s_{\hat{Y}} = s_{\bar{\bar{y}}}\ M\ N = 23{,}61 \times 160 \times 1500 = 5.666.400\ m^3}$$

El intervalo de confianza del total para una probabilidad del 95 % es:

IC(95%) = 55.740.000 ± 11.332.800

Li = 44.407.200 m^3

Ls = 67.072.800 m^3

BIBLIOGRAFÍA

Sampling Techniques for Forest Resource Inventory. Barry D. Shiver and Bruce E. Borders. 1996. John Wiley & Sons INC.

Sampling Theory for Forest Inventory. 1986. Pieter G. de Vries. Springer-Verlag.

F. Fresse. Muestreo Forestal Elemental. 1962 .

Octubre 2002