CAPÍTULO 10: MUESTREO DOBLE

por: Enrique Wabo

1 INTRODUCCIÓN

El Muestreo Doble o Muestreo en Dos Fases se desarrolló para poder utilizar los estimadores de Regresión y de Razón cuando se desconocen la media y el total poblacional de X pero se puede lograr una buena estimación de ellos. De esta forma se pueden aprovechar las ventajas de aquellos estimadores.

El mecanismo recibe el nombre genérico de Muestreo Doble, o Muestreo en Dos Fases, porque el muestreo se lleva a cabo, precisamente, en dos fases. En la primera fase (Fase 1), una muestra grande de tamaño n' se selecciona de una población, y en ella sólo se observan valores de X. En la segunda fase (Fase 2), se selecciona una muestra pequeña de tamaño n de la misma población, y en ella se observan pares de valores de X e Y. La notación a emplearse será la siguiente:

n' = tamaño de la muestra en la Fase 1 (m. grande).

x' = valor de X observado en la Fase 1.

$\bar{x}'$ = media de X estimada en la Fase 1.

n = tamaño de la muestra en la Fase 2 o muestra chica, en la que ambas variables son medidas.

x = valor de X observado en la Fase 2 (m. chica).

$\bar{x}$ = media de X estimada en la Fase 2.

$\bar{y}$ = media de la variable Y estimada en la Fase 2.

En resumen:

Fase 1 (m. grande)	Fase 2 (m. chica)
n' x' $\bar{x}'$	n x $\bar{x}$ y $\bar{y}$

Algunos autores representan a los valores de la Fase 1 con el subíndice 1 y a los valores de la fase 2 con el subíndice 2; otros autores representan a los valores de la fase 1 con la letra z. Las equivalencias serían:

Fase 1 (m. grande)	Fase 2 (m. chica)
$n_1 = n'$	$n_2 = n$
$x_1 = x' = z$	$x_2 = x$
$\bar{x}_1 = \bar{x}' = \bar{z}$	$\bar{x}_2 = \bar{x}$
	$\bar{y}_2 = \bar{y}$

Mientras la Fase 1 aporta información sobre las características de la variable auxiliar X, la Fase 2 aporta los pares de valores que permiten estimar la relación lineal entre X e Y, bajo la forma de una razón o de una regresión.

La Fase 2 puede ser dependiente o independiente de la fase 1, dando lugar a dos variantes del Muestreo Doble.

Una de las variantes es el Muestreo Doble con Submuestreo o Muestreo Doble con Fases Dependientes; en este caso, de la muestra grande de tamaño n' de la Fase 1 se selecciona una submuestra de tamaño n en forma aleatoria, en la cual sólo la variable Y es observada. Los valores de X asociados a cada valor Y deben estar identificados y corresponden a los valores observados en la fase 1.

La otra variante es el Muestreo Doble con Fases Independientes. En este caso, la muestra chica (Fase 2) no es una submuestra de la muestra grande (Fase 1).

La importancia de estas dos variantes radica en que las fórmulas para determinar el error estándar de la media y del total varía levemente de una situación a otra. Las medias y totales estimados, en cambio, usan las mismas fórmulas para ambos casos.

A continuación veremos ambas variantes. En primer lugar, veremos cómo se calculan las medias y totales de Y para cada estimador, y luego veremos cómo se determinan los errores estándar según las fases sean dependientes o independientes.

2 ESTIMACIÓN DE MEDIAS Y TOTALES

2.1 ESTIMADOR DE RAZÓN

El estimador de razón se determina con los valores de la Fase 2, usando la fórmula tradicional:

$$\hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\sum_{1}^{n} y}{\sum_{1}^{n} x} \qquad (1)$$

La media estimada de Y por unidad de muestreo y el total estimado de Y se obtienen con las siguientes fórmulas:

$$\bar{y}_R = \hat{R}\,\bar{x}' \qquad (2)$$

$$\hat{Y}_R = N\,\bar{y}_R \qquad (3)$$

donde N es el total de unidades de muestreo en la población, que surge de dividir la superficie total por el área de las parcelas.

2.2 ESTIMADOR DE REGRESIÓN

La estimación de la media de Y por unidad para la población y el total estimado, toman las siguientes formas:

$$\hat{\mu}_Y = \bar{y}_{reg} = \bar{y} + b\,(\bar{x}' - \bar{x}) \qquad (4)$$

$$\hat{Y}_{reg} = N\,\bar{y}_{reg} \qquad (5)$$

3 ESTIMACIÓN DE ERRORES ESTÁNDAR

3.1 ESTIMADOR DE RAZÓN

3.1.1 FASES DEPENDIENTES

Definimos los siguientes términos:

$$A = \frac{s_y^2 + \hat{R}^2 s_{x'}^2 - 2\hat{R}\,s_{xy}}{n} \qquad (6)$$

$$B = \frac{2\hat{R}\,s_{xy} - \hat{R}^2 s_{x'}^2}{n'} \qquad (7)$$

$$C = -\left(\frac{s_y^2}{N}\right) \qquad (8)$$

De Vries (1986) propone la siguiente fórmula para el error estándar de la media estimada:

$$S_{\bar{y}_R} = \sqrt{A + B + C} \qquad (9)$$

En las fórmulas donde interviene la varianza de X, se dará prioridad al empleo del valor correspondiente a la muestra grande (Fase 1) por considerarse más confiable; de no ser posible su uso se utilizará la correspondiente a la Fase 2.

3.1.2 FASES INDEPENDIENTES

Cuando las fases son independientes, el error estándar de la media toma la siguiente forma:

$$S_{\bar{y}_R} = \sqrt{(A + B)\left(\frac{N - n}{N}\right)} \qquad (10)$$

siendo:

$$A = \frac{s_y^2 + \hat{R}^2 s_{xp}^2 - 2\hat{R}\,s_{xy}}{n} \qquad (11)$$

$$B = \frac{\hat{R}^2\ s_{xp}^2}{n'} \qquad (12)$$

$$s_{xp}^2 = \frac{SCDX' + SCDX}{n' + n - 2} \qquad (13)$$

3.2 ESTIMADOR DE REGRESIÓN

3.2.1 FASES DEPENDIENTES

El error estándar de la media estimada, cuando las fases son dependientes, es la raíz cuadrada de la siguiente expresión:

$$s_{\bar{y}}^2 = \frac{s_{reg}^2}{n}\left(1 + \frac{1}{n} + \frac{1}{n'}\right) + b^2\,\frac{s_{x'}^2}{n'} \qquad (14)$$

siendo b la pendiente estimada de la recta de regresión

3.2.2 FASES INDEPENDIENTES

El error estándar de la media estimada, cuando las fases son independientes, es la raíz cuadrada de la siguiente expresión:

$$s_{\bar{y}}^2 = \frac{s_{reg}^2}{n}\left(1 + \frac{1}{n} - \frac{1}{n'}\right) + b^2\,\frac{s_{x'}^2}{n'} \qquad (15)$$

La diferencia con la fórmula (14) está en el término (1/n’), que en este caso se resta.

3.2.3 FASES INDEPENDIENTES Y DEPENDIENTES

Cuando el tamaño de la muestra es grande, una aproximación al error estándar de la media, aplicable tanto si las fases son dependientes o independientes, es la raíz cuadrada de la siguiente expresión:

$$s_{\bar{y}}^2 = \frac{s_{reg}^2}{n} + b^2\,\frac{s_{x'}^2}{n'} \qquad (16)$$

También para muestras grandes, De Vries recomienda usar la siguiente ecuación, tanto para fases dependientes como independientes:

$$s_{\bar{y}}^2 = \frac{s_y^2}{n}\left[1 - \left(\frac{n' - n}{n'}\right) r^2\right] \qquad (17)$$

siendo r la estimación del coeficiente de correlación lineal entre X e Y, que se estima con:

$$r = \frac{SPXY}{\sqrt{SCDX\ SCDY}} = \frac{Sxy}{Sx\ Sy} \qquad (18)$$

De ser posible la varianza de X se estima a partir de los datos de la muestra grande.

4 EJEMPLO

El siguiente ejemplo pertenece a F. Freese; las unidades originales han sido modificadas para expresar valores en el sistema decimal.

En el año 1992 se seleccionó una muestra de 200 parcelas de 250 m^2 de un monte de 80 ha. El resultado fue un volumen medio de 372 m^3/ha.

En el año 2002 se selecciona una submuestra de 40 parcelas, escogidas al azar, entre las 200 parcelas iniciales, con el objeto de estimar el volumen medio por hectárea en el año 2002. En la siguiente tabla se muestra un esquema de los valores X e Y.

Nº	Vol./ha en 1992 X	Vol./ha en 2002 Y
1	280	370
2	240	290
3	410	520
...	...	...
...	...	...
39	480	580
40	420	540

Datos generales:

Superficie de bosques = 80 ha

Superficie de la parcela = 250 m2 c/u

N = 3.200 parcelas

Fase 1:

$n' = 200$ $\bar{x}' = 372\ m^3/ha$

Fase 2:

$n = 40$

$\bar{x} = 369{,}75\ m^3/ha$ $\bar{y} = 470{,}5$

SCDX = 192.697,5 $S_x^2 = 4.940{,}9615$

SCDY = 302.590 $S_y^2 = 7.758{,}7179$

SPXY = 227.605,0 $S_{xy} = 5.836{,}0256$

r = 0,942

Se utilizará el estimador de regresión en muestreo doble con fases dependientes. La pendiente estimada de la recta es:

b = 227.605,0/192.697,5 = 1,181152

La varianza de la regresión (desvíos respecto a la regresión) estimada, es:

$S_{reg}^2 = 888{,}2616$

La media estimada de Y para el año 2002 es:

$\bar{y}_{reg} = 470{,}50 + 1{,}181152\ (372 - 369{,}75)$

$= 473{,}16$ m3/ha

El error estándar de la media es:

$$S_{\bar{y}_{reg}} = \sqrt{\frac{(888{,}26)\ (1{,}030)}{40} + \frac{6.893{,}23}{200}}$$

$= 7{,}57$

Resumiendo:

- Media estimada para el 2002 = 473,16 m3/ha
- E. estándar de la media est.. = 7,57 m3/ha

Si se hubiese aplicado muestreo simple, los valores que se habrían obtenido son:

- Media estimada para el 2002 = 470,50 m3/ha
- E. estándar de la media est.. = 12,46 m3/ha

Puede observarse el aumento de precisión obtenido por el estimador de regresión; el error estándar estimado con el estimador de regresión (7,57) es el 60% del error estándar estimado con el muestreo aleatorio simple (12,46).

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.
- Muestreo Forestal Elemental. 1962. F. Freese. Forest Service.

Septiembre 2002