Técnicas de Muestreo I

Patricia Isabel Romero Mares

Departamento de Probabilidad y Estadística
IIMAS UNAM

septiembre 2015

Ejemplos muestreo estratificado

La información que aparece a continuación representa la estratificación de todas las propiedades agrícolas en un estado, clasificadas por tamaño.

Para una muestra de 100 ranchos, calcule los tamaños de muestra en cada estrato bajo

a) distribución proporcional

b) distribución óptima y

c) compare las precisiones de estos métodos con la del m.a.s.

Tamaño de la propiedad (has.)	Número de propiedades N_h	Promedio de has. de maíz $\bar{Y}_h$	Desviación estándar S_h
0-40	394	5.4	8.3
41-80	461	16.3	13.3
81-120	391	24.3	15.1
121-160	334	34.5	19.8
161-200	169	42.1	24.5
201-240	113	50.1	26.0
241 -	148	63.8	35.2

a) Distribución proporcional

$$n_h = n\,\frac{N_h}{N}$$

$n_1 = 20$, $n_2 = 23$, $n_3 = 19$, $n_4 = 17$, $n_5 = 8$, $n_6 = 6$, $n_7 = 7$.

b) Distribución óptima

$$n_h = n\,\frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}$$

$n_1 = 10$, $n_2 = 18$, $n_3 = 17$, $n_4 = 19$, $n_5 = 12$, $n_6 = 9$, $n_7 = 15$.

c) Comparación de las precisiones de estos dos métodos con la del m.a.s.

Primero, se demostrará que la varianza total se puede escribir como la varianza dentro de estratos más la varianza entre estratos.

$$
\begin{aligned}
(N-1)S^2 &= \sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^2=\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi}-\bar{Y}\right)^2 \\
&= \sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi}-\bar{Y}_h+\bar{Y}_h-\bar{Y}\right)^2 \\
&= \sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi}-\bar{Y}_h\right)^2+\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(\bar{Y}_h-\bar{Y}\right)^2 \\
&\quad +2\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi}-\bar{Y}_h\right)\left(\bar{Y}_h-\bar{Y}\right).
\end{aligned}
$$

Pero,

$$
\begin{aligned}
&\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi}-\bar{Y}_h\right)\left(\bar{Y}_h-\bar{Y}\right)= \\
&= \sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(Y_{hi}\bar{Y}_h-Y_{hi}\bar{Y}-\bar{Y}_h^2+\bar{Y}_h\bar{Y}\right) \\
&= \sum_h \bar{Y}_h\sum_i Y_{hi}-\bar{Y}\sum_h\sum_i Y_{hi}-\sum_h N_h\bar{Y}_h^2+\bar{Y}\sum_h N_h\bar{Y}_h \\
&= \sum_h N_h\bar{Y}_h^2-N\bar{Y}^2-\sum_h N_h\bar{Y}_h^2+N\bar{Y}^2 \\
&= 0.
\end{aligned}
$$

Por lo tanto,

$$
\begin{aligned}
(N-1)S^2 &= \sum_h \left[\sum_i \left(Y_{hi}-\bar{Y}_h\right)^2\right] + \sum_h N_h \left(\bar{Y}_h-\bar{Y}\right)^2 \\
&= \sum_h \left(N_h-1\right) S_h^2 + \sum_h N_h \left(\bar{Y}_h-\bar{Y}\right)^2 \\
S^2 &= \sum_h \frac{\left(N_h-1\right) S_h^2}{N-1} + \sum_h \frac{N_h \left(\bar{Y}_h-\bar{Y}\right)^2}{N-1}.
\end{aligned}
$$

Si N_h y, por lo tanto, N son grandes, entonces $N_h - 1 \approx N_h$ y $N - 1 \approx N$, entonces

$$S^2 \approx \underbrace{\sum_h W_h S_h^2}_{\text{dentro}} + \underbrace{\sum_h W_h \left(\bar{Y}_h - \bar{Y}\right)^2}_{\text{entre}}$$

Regresando al ejercicio:

$$S^2 = \sum_h W_h S_h^2 + \sum_h W_h \left(\bar{Y}_h - \bar{Y}\right)^2 = 343.28 + 332.76 = 676.04$$

La varianza del estimador del promedio con m.a.s. es:

$$V_{m.a.s.}\left(\hat{\bar{Y}}\right) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} = \left(1 - \frac{100}{2010}\right)\frac{676.04}{100} = 6.424$$

La varianza del estimador del promedio en muestreo estratificado y m.a.s. en cada estrato es:

$$V_{est}\left(\hat{\bar{Y}}\right) = \sum_{h=1}^{L} W_h^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h} \tag{1}$$

Si la distribución de la muestra a los estratos es la óptima, es decir,

$$n_h = n\,\frac{N_h S_h}{\sum_h N_h S_h}$$

y sustituimos n_h en la expresión (1):

$$V_{opt}\left(\hat{\bar{Y}}\right) = \frac{\left(\sum_h W_h S_h\right)^2}{n} - \frac{\sum_h W_h S_h^2}{N}$$

Sustituyendo valores:

$$V_{opt}\left(\hat{\bar{Y}}\right) = \frac{289.625}{100} - \frac{343.279}{2010} = 2.725$$

$$\frac{V_{opt}\left(\hat{\bar{Y}}\right)}{V_{m.a.s.}\left(\hat{\bar{Y}}\right)} = \frac{2.725}{6.424} = 0.4241$$

$1 - 0.4241 = 0.576$, es decir, $57.6\,\%$ de reducción de varianza.

Si la distribución de la muestra a los estratos es proporcional, es decir,

$$n_h = n \frac{N_h}{N}$$

si sustituimos la expresión de n_h en (1), la varianza del estimador de la media con distribución proporcional es:

$$V_{prop}\left(\hat{\bar{Y}}\right) = \left(1 - \frac{n}{N}\right) \frac{\sum_h W_h S_h^2}{n}$$

Sustituyendo valores:

$$V_{prop}\left(\hat{\bar{Y}}\right) = \left(1 - \frac{100}{2010}\right) \frac{1}{100} \sum_h W_h S_h^2 = 3.262$$

$$\frac{V_{prop}\left(\hat{\bar{Y}}\right)}{V_{m.a.s.}\left(\hat{\bar{Y}}\right)} = \frac{3.262}{6.424} = 0.5077$$

$1 - 0.5077 = 0.492$, es decir, $49.2\,\%$ de reducción de varianza.