Muestreo Estratificado
Ejercicio 3 *

García Prado, Sergio
sergio@garciparedes.me

13 de noviembre de 2017

Resumen

En este trabajo se ha realizado la demostración de que en la situación particular de diseño *m.a.s.* y la estimación parámetro total poblacional $\widehat{\mathcal{T}}$, para cometer un error de estimación global B basta con fijar el mismo error de estimación b en cada uno de los estratos, siendo b igual al cociente entre el error global B y la raíz cuadrada del número de estratos L.

1. Introducción

Denotaremos por $U = U_1 \cup ... \cup U_h \cup ... \cup U_L = \{1, ..., k, ..., N\}$ a la población, para la cual tenemos una división en L estratos denotando por U_h al estrato $h \in \{1, ..., L\}$. Sea I_h el conjunto de índices de las observaciones seleccionadas en el estrato U_h y s_h la muestra extraida de dicho estrato. Por tanto, podemos denotar a la muestra global por $s = s_1 \cup ... \cup s_h \cup ... \cup s_L$

En este caso, tal y como se ha indicado anteriormente, se va a presuponer la utilización de *muestreo aleatorio simple (m.a.s.)* sobre cada estrato. Esta método de muestreo se caracteriza por la fijación *a-priori* del tamaño de muestras y la selección de observaciones *sin reemplazamiento*. Esto es equivalente a decir que una vez seleccionada una observación esta desaparece del conjunto de candidatas a aparecer en la muestra. Por tanto, no hay observaciones repetidas en la muestra.

Denotaremos por $\mathcal{T}$ al total poblacional de una determinada variable de interés Y denotando como $y_k \quad \forall k \in U$ al *k-ésimo* valor de Y. Es fácil entender por tanto que el total poblacional se define como $\mathcal{T} = \sum_U y_k$.

Bajo la hipótesis de *muestreo aleatorio simple (m.a.s.)*, a partir del cual se pretende obtener una aproximación lo más precisa posible tanto del total poblacional $\mathcal{T}$ como de la proporción poblacional P. Para ello, es necesario apoyarse en los valores del tamaño de la población N, el del estrato U_h como N_h y el de la muestra s_h denotado por n_h. El tamaño relativo del estrato se define como $W_h = \frac{N_h}{N}$. También se define el tamaño relativo de la muestra como $f_h = \frac{n_h}{N_h}$. Entonces, en este caso un buen estimador del total poblacional es el π-estimador $\widehat{\mathcal{T}} = \sum_{s_h} \frac{y_k}{W_h}$.

En las ecuaciones (1) y (2) se definen las varianzas de los estimadores del total poblacional en el estrato h denotado $\widehat{\mathcal{T}_h}$ y total poblacional denotado por $\widehat{\mathcal{T}}$. Nótese la propiedad de adictividad de la varianza sobre los estratos para el estimador del total poblacional, ya que será de gran utilidad en la demostración.

$$Var(\widehat{\mathcal{T}_h}) = \left(\frac{N_h^2}{n_h} - N_h\right) \sigma_h^{*2} \tag{1}$$

$$Var(\widehat{\mathcal{T}}) = \sum_{h=1}^{L} Var(\widehat{\mathcal{T}_h}) = \sum_{h=1}^{L} \left(\frac{N_h^2}{n_h} - N_h\right) \sigma_h^{*2} \tag{2}$$

Para la demostración nos apoyaremos en la ecuación (3), que relaciona el error de estimación B a un nivel de confianza k, con la varianza del estimador $\widehat{\theta}$.

$$B^2 = k^2 Var(\widehat{\theta}) \tag{3}$$

*URL: `https://github.com/garciparedes/statistical-sampling-stratified`

2. Demostración

Para demostrar la propiedad de que para conseguir un error de estimación B con un nivel de confianza k en *m.a.s.* para el estimador del total poblacional, es necesario fijar en cada estrato el error de estimación en el valor b se ha utilizado la propiedad de adictividad de la varianza que se da en este caso concreto.

$$b_h = b = \frac{B}{\sqrt{L}} \qquad \forall h \in \{1, ..., L\} \tag{4}$$

$$B^2 = k^2 Var(\widehat{\mathcal{T}}) \qquad b_h^2 = \frac{B^2}{L} = k^2 Var(\widehat{\mathcal{T}_h}) \tag{5}$$

Debido a la propiedad de adictividad de la varianza para el caso del total poblacional se cumple la siguiente igualdad:

$$Var(\widehat{\mathcal{T}}) = \sum_{h=1}^{L} Var(\widehat{\mathcal{T}_h}) \tag{6}$$

$$k^2 Var(\widehat{\mathcal{T}}) = \sum_{h=1}^{L} k^2 Var(\widehat{\mathcal{T}_h}) \tag{7}$$

$$B^2 = \sum_{h=1}^{L} b_h^2 \tag{8}$$

$$B^2 = B^2 \sum_{h=1}^{L} \frac{1}{L} \tag{9}$$

$$B = B \tag{10}$$

Referencias

[1] Särndal, C.-E., Swensson, B., and Wretman, J. *Model assisted survey sampling.* Springer Science & Business Media, 2003.

[2] Tapia García, J. A. Muestreo Estadístico 1, 2017/18. Facultad de Ciencias: Departamento de Estadística.