Muestreo Estratificado
Ejercicio 2 *

García Prado, Sergio
sergio@garciparedes.me

12 de noviembre de 2017

1. Introducción

[TODO]

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 del método *m.a.s.* sobre cada estrato, caracterizado porque el tamaño de la muestra se fija *a-priori* y se seleccionan las observaciones *sin reemplazamiento*, es decir, una vez seleccionada una observación, esta desaparece del conjunto de candidatos 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 $\mathcal{T} = \sum_U y_k$.

Denotaremos por P a la proporción poblacional de una determinada variable de interés Y de carácter binario denotando como $y_k \quad \forall k \in U$ al *k-ésimo* valor de Y. Es fácil entender por tanto que la proporción poblacional se define como $P = \frac{\sum_U y_k}{N}$.

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 $\widehat{\mathcal{T}} = \sum_{s_h} \frac{y_k}{W_h}$. Para el caso del estimador de la proporción poblacional, un buen estimador es $\widehat{P} = \frac{\sum_{s_h} \frac{y_k}{W_h}}{N}$

La varianza del estimador del total total se define a continuación:

$$Var(\widehat{\mathcal{T}}) = \tag{1}$$

$$= \sum_{h=1}^{L} \frac{N_h^2 (1 - f_h) \sigma_h^{*2}}{n_h} \tag{2}$$

$$= \sum_{h=1}^{L} \frac{N_h^2 (1 - \frac{n_h}{N_h}) \sigma_h^{*2}}{n_h} \tag{3}$$

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

$$= \sum_{h=1}^{L} \frac{(N_h^2 - N_h n_h)\sigma_h^{*2}}{n_h} \tag{4}$$

$$= \sum_{h=1}^{L} \left(\frac{N_h^2}{n_h} - N_h \right) \sigma_h^{*2} \tag{5}$$

La varianza del estimador de la proporción total se define a continuación:

$$Var(\widehat{P}) = \tag{6}$$

$$= \sum_{h=1}^{L} W_h^2 \frac{1 - f_h}{n_h} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{7}$$

$$= \sum_{h=1}^{L} W_h^2 \frac{1 - \frac{n_h}{N_h}}{n_h} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{8}$$

$$= \sum_{h=1}^{L} \frac{W_h^2 - \frac{W_h^2 n_h}{N_h}}{n_h} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{9}$$

$$= \sum_{h=1}^{L} \frac{W_h^2 - \frac{n_h * N_h}{N^2}}{n_h} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{10}$$

$$= \sum_{h=1}^{L} \left(\frac{W_h^2}{n_h} - \frac{N_h}{N^2} \right) \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{11}$$

$$\tag{12}$$

$$w_h = \frac{n_h}{n} \tag{13}$$

$$\frac{N_h}{N_h - 1} \simeq 1 \tag{14}$$

$$\frac{B^2}{k^2} = Var(\widehat{\theta}) \tag{15}$$

$$w_h = W_h \qquad \text{(afijación proporcional)} \tag{16}$$

$$w_h = \frac{N_h \sigma_h^*}{\sum_{i=1}^{L} N_i \sigma_i^*} \qquad \text{(afijación mínima varianza)} \tag{17}$$

Una vez desarrollada la ecuación que a partir de la cual determinar el tamaño de muestra global n, fijado un erro de estimación b a una confianza de nivel k, el siguiente paso es particularizar esto para estimadores concretos. Esto consiste simplemente en la substitución del valor de la varianza por el del estimador en cuestión. En la sección 2 se realiza dicha demostración para el caso del total poblacional $\widehat{\mathcal{T}_h}$ y en la sección 3 se realiza para el caso del estimador de proporción poblacional $\widehat{P_h}$.

2. Demostración para estimador del total poblacional $\widehat{\mathcal{T}}$

[TODO]

$$\frac{B^2}{k^2} = Var(\widehat{\mathcal{T}}) \tag{18}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L} \left(\frac{N_h^2}{n_h} - N_h \right) \sigma_h^{*2} \tag{19}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L} \left(\frac{N_h^2}{n * w_h} - N_h \right) \sigma_h^{*2} \tag{20}$$

$$\frac{B^2}{k^2} + \sum_{h=1}^{L} N_h \sigma_h^{*2} = \sum_{h=1}^{L} \frac{N_h^2}{n * w_h} \sigma_h^{*2} \tag{21}$$

$$n = \frac{\sum_{h=1}^{L} \frac{N_h^2}{w_h} \sigma_h^{*2}}{\frac{B^2}{k^2} + \sum_{h=1}^{L} N_h \sigma_h^{*2}} \tag{22}$$

$$\tag{23}$$

3. Demostración para estimador de proporción poblacional $\widehat{P}$

[TODO]

$$\frac{B^2}{k^2} = Var(\widehat{P}) \tag{24}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L} \frac{W_h^2}{n_h} - \frac{N_h}{N^2} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{25}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L} \frac{W_h^2}{n * w_h} - \frac{N_h}{N^2} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{26}$$

$$\frac{B^2}{k^2} + \sum_{h=1}^{L} \frac{N_h}{N^2} \frac{N_h}{N_h - 1} P_h(1 - P_h) = \sum_{h=1}^{L} \frac{W_h^2}{n * w_h} \frac{N_h}{N_h - 1} P_h(1 - P_h) \tag{27}$$

$$n = \frac{\sum_{h=1}^{L} \frac{W_h^2}{w_h} \frac{N_h}{N_h - 1} P_h(1 - P_h)}{\frac{B^2}{k^2} + \sum_{h=1}^{L} \frac{N_h}{N^2} \frac{N_h}{N_h - 1} P_h(1 - P_h)} \tag{28}$$

$$n = \frac{\sum_{h=1}^{L} \frac{W_h^2}{w_h} \frac{N_h}{N_h - 1} P_h(1 - P_h)}{\frac{B^2}{k^2} + \sum_{h=1}^{L} \frac{W_h}{N} \frac{N_h}{N_h - 1} P_h(1 - P_h)} \tag{29}$$

$$\tag{30}$$

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.