Muestreo Estratificado
Ejercicio 5 *

García Prado, Sergio
sergio@garciparedes.me

13 de noviembre de 2017

Resumen

En este trabajo se ha desarrollado la demostración de la expresión de los pesos $w_h \quad h \in \{1, ..., L\}$ siendo L el número de estratos, fijando el tamaño n de muestra global *a-priori*. Esto se ha llevado a cabo bajo la el método de *muestreo aleatorio simple con reeemplazamiento (m.a.s.con)* y siguiendo la estrategia de *afijación de mínima varianza*.

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, nos referiremos a la muestra global por $s = s_1 \cup ... \cup s_h \cup ... \cup s_L$

Tal y como se ha indicado anteriormente se va a presuponer la utilización del método *m.a.s.con* sobre cada estrato, caracterizado porque el tamaño de la muestra se fija *a-priori* y se seleccionan las observaciones *con reemplazamiento*, es decir, cada observación puede ser seleccionada más de una vez. Por tanto, puede haber 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 $\mathcal{T} = \sum\limits_{U} y_k$.

Denotaremos por μ a la media 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 la media poblacional se define como $\mu = \frac{\sum_U y_k}{N} = \frac{\mathcal{T}}{N}$.

Las expresiones de los estimadores insesgados $\widehat{\mathcal{T}}_{m.a.s.con}$ y $\widehat{\mu}_{m.a.s.con}$ se muestran en las ecuaciones (1) y (2) respectivamente.

$$\widehat{\mathcal{T}}_{m.a.s.con} = \sum_{h=1}^{L} \sum_{k \in s_h} \frac{y_k N_h}{n_h} \tag{1}$$

$$\widehat{\mu}_{m.a.s.con} = \frac{\widehat{\mathcal{T}}_{m.a.s.con}}{N} \tag{2}$$

$$\tag{3}$$

Las expresiones de la varianza de los estimadores $\widehat{\mathcal{T}}_{m.a.s.con}$ y $\widehat{\mu}_{m.a.s.con}$ se muestran en las ecuaciones (4) y (5) respectivamente.

$$Var(\widehat{\mathcal{T}}_{m.a.s.con}) = \sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{n_h} \tag{4}$$

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

$$
\begin{aligned}
&Var(\widehat{\mu}_{m.a.s.con}) = \\
&\qquad = Var\left(\frac{\widehat{\mathcal{T}}_{m.a.s.con}}{N}\right) \\
&\qquad = \frac{Var\left(\widehat{\mathcal{T}}_{m.a.s.con}\right)}{N^2} \\
&\qquad = \frac{\sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{n_h}}{N^2} \\
&\qquad = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h}
\end{aligned} \tag{5}
$$

Sea N el tamaño de la población, N_h el del estrato U_h y n_h el de la muestra s_h. El tamaño relativo del estrato se define como $W_h = \frac{N_h}{N}$. Denotaremos por $f_h = \frac{n_h}{N_h}$ el tamaño relativo de la muestra.

Definiremos n_h como el tamaño de la muestra *h-ésima*. Esto también puede entenderse como el tamaño de la muestra global ponderado por un determinado peso w_h dependiente de la estrategia de afijación escogida. Esto se puede definir matemáticamente como:

$$
n_h = n * w_h \tag{6}
$$

Por tanto, para la obteción del la expressión de w_h, primeramente el problema se presenta como la obtención del tamaño óptimo $n_h \quad h \in \{1, ..., L\}$ que minimice la varianza global (**afijación de mínima varianza**) de la estimación, suponiendo como valor conocido N para la obtención del mejor estimador de un determinado estadístico. Definiremos la función $\phi(n_1, ..., n_L)$ como:

$$
\phi(n_1, ..., n_L) = Var(\widehat{\theta}) + \lambda \left(\sum_{h=1}^{L} n_h - n \right) \tag{7}
$$

Entonces el objetivo es minimizar dicha función, de tal manera que se minimiza la varianza global de la estimación $\widehat{\theta}$. Esto es equivalente a buscar los valores que hagan mínima dicha función, es decir:

$$
min_{n_h} \{\phi(n_1, ..., n_L)\} \tag{8}
$$

2. Demostración

El primer paso es derivar la función $\phi(n_1, ..., n_L)$ respecto del valor que se pretende minimizar:

$$
\begin{aligned}
&\frac{\partial \phi(n_1, ..., n_L)}{\partial n_h} = \\
&\qquad = \frac{\partial \left(Var(\widehat{\mu}_{m.a.s.con}) + \lambda \left(\sum_{h=1}^{L} n_h - n \right) \right)}{\partial n_h} \\
&\qquad = \frac{\partial \left(\left(\sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h} \right) + \lambda \left(\sum_{h=1}^{L} n_h - n \right) \right)}{\partial n_h} \\
&\qquad = \frac{-W_h^2 \sigma_h^{2*}}{n_h^2} + \lambda \\
&\qquad = 0
\end{aligned} \tag{9}
$$

Puesto que $n_1 + ... + n_L = n$ se obtiene:

$$
\sqrt{\lambda} = \frac{\sum_{h=1}^{L} W_h \sigma_h}{n} \tag{10}
$$

Lo siguiente es despejar el valor n_h, es decir, el tamaño de cada estrato:

$$\begin{aligned}
n_h &= \\
&= \frac{W_h \sigma_h}{\sqrt{\lambda}} \\
&= \frac{W_h \sigma_h}{\frac{\sum_{h=1}^{L} W_h \sigma_h}{n}} \\
&= n \frac{\frac{N_h}{N} \sigma_h}{\sum\limits_{h=1}^{L} \frac{N_h}{N} \sigma_h} \\
&= n \frac{N_h \sigma_h}{\sum\limits_{h=1}^{L} N_h \sigma_h}
\end{aligned} \tag{11}$$

Por último, para despejar el valor w_h, basta con aplicar la propiedad $w_h = \frac{n_h}{n}$:

$$n_h = n \frac{N_h \sigma_h}{\sum\limits_{h=1}^{L} N_h \sigma_h} \tag{12}$$

$$w = \frac{n_h}{n} = \frac{N_h \sigma_h}{\sum\limits_{h=1}^{L} N_h \sigma_h} \tag{13}$$

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.