Muestreo Estratificado
Ejercicio 4 *

García Prado, Sergio
sergio@garciparedes.me

13 de noviembre de 2017

Resumen

En este trabajo se desarrollan las expresiones del tamaño de muestra n global para estimar la media poblacional $\widehat{\mu}$ fijado un error de estimación global B y una confianza $(1-\alpha)$ para *m.a.s.* con y sin reemplazamiento

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 $s = s_1 \cup ... \cup s_h \cup ... \cup s_L$.

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

Denotaremos por μ a la media poblacional de una determinada variable de interés Y siendo $y_k \quad \forall k \in U$ la *k-ésima* observación de Y. Es fácil entender por tanto que la media poblacional se define como $\mu = \frac{\sum_U y_k}{N} = \frac{\mathcal{T}}{N}$.

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:

$$n_h = n * w_h \tag{1}$$

Donde w_h depende del tipo de afijación sobre el cual se esté trabajando. En las ecuaciones (2) y (3) se muestra para los casos de *afijación proporcional* y *mínima varianza* respectivamente.

$$w_h = W_h \qquad \text{(afijación proporcional)} \tag{2}$$

$$w_h = \frac{N_h \sigma_h^*}{\sum_{i=1}^{L} N_i \sigma_i^*} \qquad \text{(afijación mínima varianza)} \tag{3}$$

Para tamaños de estrato N_h grandes es fácil comprobar que se cumple la propiedad $\frac{N_h}{N_h - 1} \simeq 1$, lo cual permite simplificar la ecuación del tamaño de la muestra.

Puesto que lo que se pretende demostrar en este trabajo es la ecuación para estimar el tamaño n de la muestra global, fijando un error de estimación B a un nivel de confianza k, esto es equivalente a despejar el valor n en la ecuación (4).

$$B^2 = k^2 Var(\widehat{\theta}) \tag{4}$$

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

2. Demostración para *m.a.s.* sin reemplazamiento.

Bajo la hipótesis de *muestreo aleatorio simple (m.a.s.)*, a partir del cual se pretende obtener una aproximación lo más precisa posible de la media poblacional μ, un buen estimador del total poblacional es $\widehat{\mathcal{T}}_{m.a.s.} = \sum_{h=1}^{n} \sum_{k \in s_h} \frac{y_k}{n_h} N_h$. Para el caso del estimador de la proporción poblacional, un buen estimador es $\widehat{\mu}_{m.a.s.} = \frac{\widehat{\mathcal{T}}_{m.a.s.}}{N}$.

En las ecuaciones (5) y (6) se definen respectivamente las varianza del estimador del total poblacional $\widehat{\mathcal{T}}$ así como la proporción poblacional $\widehat{\mu}$.

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

$$
\begin{aligned}
Var(\widehat{\mu}_{m.a.s.}) &= \\
&= Var\left(\frac{\widehat{\mathcal{T}}_{m.a.s.}}{N}\right) \\
&= \frac{Var\left(\widehat{\mathcal{T}}_{m.a.s.}\right)}{N^2} \\
&= \frac{\sum_{h=1}^{L}\left(\frac{N_h^2}{n_h} - N_h\right)\sigma_h^{*2}}{N^2} \\
&= \sum_{h=1}^{L}\left(\frac{W_h^2}{n_h} - \frac{W_h^2}{N_h}\right)\sigma_h^{*2}
\end{aligned} \tag{6}
$$

Para llegar a la expresión que relacione el tamaño de la muestra global n con el error de estimación B a un nivel de confianza $(1-\alpha)$, definiremos $k = Z_{1-\alpha/2}$, siendo $Z_{1-\alpha/2}$ el valor crítico $(1-\alpha/2)$ de la *distribución normal estándar* $N(0,1)$. Además, nos apoyaremos en la ecuación (4) y la estimación de la varianza de la ecuación (6).

$$B^2 = k^2 Var(\widehat{\mu}_{m.a.s.}) \tag{7}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L}\left(\frac{W_h^2}{n_h} - \frac{W_h^2}{N_h}\right)\sigma_h^{*2} \tag{8}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L}\left(\frac{W_h^2}{n * w_h} - \frac{W_h^2}{N_h}\right)\sigma_h^{*2} \tag{9}$$

$$\frac{B^2}{k^2} + \sum_{h=1}^{L}\frac{W_h^2}{N_h}\sigma_h^{*2} = \sum_{h=1}^{L}\frac{W_h^2}{n * w_h}\sigma_h^{*2} \tag{10}$$

$$n = \frac{\sum_{h=1}^{L}\frac{W_h^2}{w_h}\sigma_h^{*2}}{\frac{B^2}{k^2} + \sum_{h=1}^{L}\frac{W_h^2}{N_h}\sigma_h^{*2}} \tag{11}$$

$$n = \frac{\sum\limits_{h=1}^{L} \frac{W_h^2}{w_h} \sigma_h^{*2}}{\frac{B^2}{Z^2_{1-\frac{\alpha}{2}}} + \sum\limits_{h=1}^{L} \frac{W_h^2}{N_h} \sigma_h^{*2}} \tag{12}$$

3. Demostración para *m.a.s.* con reemplazamiento

Bajo la hipótesis de *muestreo aleatorio simple con reeemplazamiento (m.a.s.con)*, a partir del cual se pretende obtener una aproximación lo más precisa posible de la media poblacional μ, un buen estimador del total poblacional es $\widehat{\mathcal{T}}_{m.a.s.con} = \sum\limits_{h=1}^{n} \sum\limits_{k \in s_h} \frac{y_k}{n_h} N_h$. Para el caso del estimador de la proporción poblacional, un buen estimador es $\widehat{\mu}_{m.a.s.con} = \frac{\widehat{\mathcal{T}}_{m.a.s.con}}{N}$.

En las ecuaciones (13) y (14) se definen respectivamente las varianza del estimador del total poblacional $\widehat{\mathcal{T}}$ así como la proporción poblacional $\widehat{\mu}$.

$$Var(\widehat{\mathcal{T}}_{m.a.s.con}) = \sum_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{n_h} \tag{13}$$

$$\begin{aligned} Var(\widehat{\mu}_{m.a.s.con}) &= \\ &= Var\left(\frac{\widehat{\mathcal{T}}_{m.a.s.con}}{N}\right) \\ &= \frac{Var\left(\widehat{\mathcal{T}}_{m.a.s.con}\right)}{N^2} \\ &= \frac{\sum\limits_{h=1}^{L} \frac{N_h^2 \sigma_h^2}{n_h}}{N^2} \\ &= \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h} \end{aligned} \tag{14}$$

Para llegar a la expresión que relacione el tamaño de la muestra global n con el error de estimación B a un nivel de confianza $(1-\alpha)$, definiremos $k = Z_{1-\alpha/2}$, siendo $Z_{1-\alpha/2}$ el valor crítico $(1-\alpha/2)$ de la *distribución normal estándar* $N(0,1)$. Además, nos apoyaremos en la ecuación (4) y la estimación de la varianza de la ecuación (6).

$$B^2 = k^2 Var(\widehat{\mu}_{m.a.s.con}) \tag{15}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h} \tag{16}$$

$$\frac{B^2}{k^2} = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n * w_h} \tag{17}$$

$$n = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{B^2 * w_h} k^2 \tag{18}$$

$$n = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{B^2 * w_h} \left(Z_{1-\frac{\alpha}{2}}\right)^2 \tag{19}$$

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.