Muestreo estratificado

Daniel Czarnievicz

2017

El muestreo estratificado consiste en particionar la población $U = \{1; \dots; k; \dots; N\}$ en H estratos distintos $U_1; \dots; U_h; \dots; U_H$, y tomar una muestra aleatoria de forma independiente en cada uno de ellos. Para construir los estratos se debe utilizar algún tipo de información auxiliar. Los estratos deben ser tales que:

$$\bigcup_{h=1}^{H} U_h = U \qquad \text{y} \qquad U_i \cap U_j = \emptyset \;\; \forall i \neq j$$

Llamaremos N_h al tamaño del estrato h, es decir, $\# U_h = N_h$. Luego entonces $N = \sum_{h=1}^{H} N_h$. Por su parte, el total poblacional puede escribirse de varias formas:

$$\star \;\; t_y = \sum\nolimits_U y_k = \sum_{h=1}^{H} \sum\nolimits_{U_h} y_k = \sum_{h=1}^{H} t_{y_h} = \sum_{h=1}^{H} N_h \, \bar{y}_{U_h}$$

Por otro lado, llamando $w_h = \dfrac{N_h}{N}$ al peso relativo de cada estrato, tenemos que:

$$\star \;\; \bar{y}_U = \frac{1}{N} \sum\nolimits_U y_k = \frac{1}{N} \sum_{h=1}^{H} N_h \, \bar{y}_{U_h} = \sum_{h=1}^{H} w_h \, \bar{y}_{U_h}$$

Estrategia de selección

Para cada estrato U_h se selecciona una muestra s_h de tamaño n_{S_h} bajo el diseño $p_h(.)$. La selección dentro de cada estrato se realiza de forma independiente. La muestra queda conformada por $s = \bigcup_{h=1}^{H} s_h$. El tamaño muestral es $n_s = \sum_{h=1}^{H} n_{S_h}$. Dada la independencia, el diseño viene dado por:

$$p(s) = \prod_{h=1}^{H} p_h(s_h)$$

Probabilidades de inclusión

Las probabilidades de inclusión inducidas por el diseño serán:

$$\star \;\; \pi_k = P(k \in s) = P(k \in s_h) \;\; \forall k \in U; \;\; h = 1; \dots; H$$

$$\star \;\; \pi_{kl} = P(k; l \in s) = \begin{cases} \pi_{kl} & \text{si} \quad k; l \in U_h, \;\; h = 1; \dots; H \\ \pi_k \, \pi_l & \text{si} \quad k \in U_i \text{ y } l \in U_j, \;\; \text{con } i \neq j \end{cases}$$

$$\star \;\; \Delta_{kl} = \begin{cases} \pi_{kl} - \pi_k \, \pi_l & \text{si} \quad k; l \in U_h, \;\; h = 1; \dots; H \\ 0 & \text{si} \quad k \in U_i \text{ y } l \in U_j, \;\; \text{con } i \neq j \end{cases}$$

Por lo tanto, los π_{kl} son inducidos por $p_h(.)$ o por $p_i(.)$ y $p_j(.)$

El estimador $\hat{t}_\pi$

El estimador π de un total será entonces la suma de los estimadores π para los totales de cada estrato:

$$\star\ \hat{t}_\pi = \sum\nolimits_s y_k^{\checkmark} = \sum_{h=1}^{H} \sum\nolimits_{s_h} y_k^{\checkmark} = \sum_{h=1}^{H} \hat{t}_{\pi_h}$$

$$\star\ E_{ST}(\hat{t}_\pi) = E_{ST}\left(\sum_{h=1}^{H} \sum\nolimits_{s_h} y_k^{\checkmark}\right) = \sum_{h=1}^{H} E_{ST}\left(\sum\nolimits_{s_h} y_k^{\checkmark}\right) = \sum_{h=1}^{H} \sum\nolimits_{U_h} E_{ST}(I_k)\, y_k^{\checkmark} =$$

$$= \sum_{h=1}^{H} \sum\nolimits_{U_h} y_k = \sum_{h=1}^{H} t_{y_h} = t_y$$

$$\star\ V_{ST}(\hat{t}_\pi) = V_{ST}\left(\sum_{h=1}^{H} \hat{t}_{\pi_h}\right) = \sum_{h=1}^{H} V_{p_h(s_h)}(\hat{t}_{\pi_h}) = \sum_{h=1}^{H} \left[\sum \sum\nolimits_{U_h} \Delta_{kl}\, y_k^{\checkmark}\ y_l^{\checkmark}\right]$$

$$\star\ \hat{V}_{ST}(\hat{t}_\pi) = \sum_{h=1}^{H} \hat{V}_{p_h(s_h)}(\hat{t}_{\pi_h}) = \sum_{h=1}^{H} \left[\sum \sum\nolimits_{s_h} \Delta_{kl}^{\checkmark}\, y_k^{\checkmark}\ y_l^{\checkmark}\right]$$

$$\star\ E_{ST}\left(\hat{V}_{ST}(\hat{t}_\pi)\right) = E_{ST}\left(\sum_{h=1}^{H} \left[\sum \sum\nolimits_{s_h} \Delta_{kl}^{\checkmark}\, y_k^{\checkmark}\ y_l^{\checkmark}\right]\right) = \sum_{h=1}^{H} \left[E_{ST}\left(\sum \sum\nolimits_{s_h} \Delta_{kl}^{\checkmark}\, y_k^{\checkmark}\ y_l^{\checkmark}\right)\right] =$$

$$= \sum_{h=1}^{H} \left[\sum \sum\nolimits_{U_h} E_{ST}(I_k; I_l)\, \Delta_{kl}^{\checkmark}\, y_k^{\checkmark}\ y_l^{\checkmark}\right] = \sum_{h=1}^{H} \left[\sum \sum\nolimits_{U_h} \Delta_{kl}\, y_k^{\checkmark}\ y_l^{\checkmark}\right] = V_{ST}(\hat{t}_\pi)$$