Técnicas de Muestreo I

Patricia Isabel Romero Mares

Departamento de Probabilidad y Estadística
IIMAS UNAM

noviembre 2015

Muestreo con probabilidades de selección arbitrarias y sin reemplazo

Probabilidad de inclusión de primer orden:

$$\pi_i = P\left(U_i \text{ esté en muestra}\right)$$

Probabilidad de inclusión de segundo orden:

$$\pi_{ij} = P\left(U_i \text{ y } U_j \text{ estén en muestra}\right)$$

El estimador Horvitz-Thompson del total Y de los valores en la población, basado en una muestra de tamaño n es:

$$\hat{Y}_\pi = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{n} w_i y_i$$

donde $w_i = \frac{1}{\pi_i}$ es el factor de expansión, es decir, cada dato se expande a los que representa en la población. $\hat{Y}_\pi$ es un estimador insesgado de Y.

Si $\pi_i = \pi \, \forall i$ se llama diseño autoponderado, entonces, $\hat{Y}_\pi = \frac{1}{\pi} \sum_{i=1}^{n} y_i$.

La varianza del estimador anterior es:

$$V\left(\hat{Y}_\pi\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\pi_{ij} - \pi_i \pi_j\right) \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j}$$

Si $\pi_{ij} > 0$ para toda i y j en la población, un estimador insesgado de la varianza es:

$$\hat{V}\left(\hat{Y}_\pi\right) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\left(\pi_i \pi_j - \pi_{ij}\right)}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

Ver Sukhatme(1984), Särndal et al.(1991)

Los valores de π_i y π_{ij} para algunos diseños de muestra son:

1. m.a.s.

$$\pi_i = \frac{n}{N} \ \ \forall i$$

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)} \ \ \forall i,j$$

2. estratificado y m.a.s. en cada estrato

$$\pi_{hi} = \frac{n_h}{N_h} \ \forall i \text{ en el estrato } h$$

$$\pi_{hij} = \frac{n_h(n_h-1)}{N_h(N_h-1)} \ \ \text{si } i,j \text{ en el mismo estrato}$$

$$\pi_{hih'j} = \pi_{hi}\pi_{h'j} \ \ \text{si } i,j \text{ en diferentes estratos}$$

3. Sistemático con intervalo de selección $k = \frac{N}{n}$

$$\pi_i = \frac{1}{k} = \frac{n}{N} \; \forall i$$

$$\pi_{ij} = \begin{cases} \frac{n}{N} & \text{si } i \text{ y } j \text{ están en el mismo conglomerado} \\ 0 & \text{si } i \text{ y } j \text{ no están en el mismo conglomerado} \end{cases}$$

4. ppt sin reemplazo se complica mucho el cálculo de π_{ij}
Por ejemplo, para $n = 2$

$$\begin{aligned} \pi_i &= P(U_i \text{ en 1a.}) + P(U_i \text{ en 2a.}) \\ &= \frac{X_i}{X} + \sum_{k=1 \neq i}^{N} \frac{X_k}{X} \left(\frac{X_i}{X - X_k} \right) \end{aligned}$$

$$
\begin{aligned}
\pi_{ij} &= P(U_i \text{ en 1a.})P(U_j \text{ en 2a.}|U_i \text{ en 1a.}) \\
&\quad + P(U_j \text{ en 1a.})P(U_i \text{ en 2a.}|U_j \text{ en 1a.}) \\
&= \frac{X_i}{X}\frac{X_j}{X - X_i} + \frac{X_j}{X}\frac{X_i}{X - X_j}
\end{aligned}
$$

Para $n > 2$ resulta muy complicado el cálculo de π_{ij}.

Factores de expansión

$$
\begin{aligned}
n &= \sum_{i=1}^{N} \pi_i \\
\hat{N} &= \sum_{i=1}^{n} \frac{1}{\pi_i} = \sum_{i=1}^{n} w_i
\end{aligned}
$$

Por ejemplo, en m.a.s. $\pi_i = \frac{n}{N}$

$$\hat{N} = \sum_{i=1}^{n} \frac{N}{n} = N$$

$$n = \sum_{i=1}^{N} \pi_i = \sum_{i=1}^{N} \frac{n}{N} = n$$