Técnicas de Muestreo I

Patricia Isabel Romero Mares

Departamento de Probabilidad y Estadística
IIMAS UNAM

septiembre 2015

Estimación en Dominios

Dominios

Las técnicas del muestreo probabilístico no solo se utilizan para obtener información de poblaciones enteras (estimaciones de parámetros), sino que, en muchas situaciones, la población se subdivide en un cierto número de clases, donde el interés recae sobre cada una de estas clases o subpoblaciones, o bien, solo sobre una específica.

Un **Dominio de Estudio** es una subdivisión de la población que posee ciertas caracterísicas, (geográficas, socio-demográficas, etc.) para la cual se requieren estimaciones por separado.

La estimación en dominios se aplica por ejemplo:

- En una encuesta familiar se desea hacer estimaciones por separado para las familias con 0,1,... niños, o para propietarios de casa, o bien para familias en grupos de ocupación diferentes.
- Estimar el tiempo de vida promedio de las hembras de una población de cierta especie animal.
- El gasto total de las entidades públicas por diferentes conceptos: salarios, mobiliario, papelería, etc.

El objetivo de la estimación en dominios es estimar características generales de una o varias subpoblaciones en estudio, como pueden ser el promedio, el total o el porcentaje.

Frecuentemente no se cuenta con un marco que liste los elementos específicos del dominio en estudio, sino que el marco con que se cuenta es el de la población entera, el cual contiene a las unidades del dominio de interés. Por lo que no sabemos cuáles son las unidades que pertenecen al dominio hasta obtener la muestra.

Así, el número de unidades de la muestra que pertenecen al dominio es una variable aleatoria, con valor desconocido al momento de diseñar la encuesta.

A nivel poblacional: Se tiene una población de interés

$$P = \{U_1, U_2, \ldots, U_N\}$$

sobre la cual se establece un dominio de estudio d. Sea

$$P_d = \{U_i : i \in d\}$$

donde $P_d \subseteq P$

N	Número de elementos en la población
D	Número de dominios en la población
N_d	No. de elementos en el dominio d;$d = 1, 2, \ldots, D$
Y_d	Total del dominio d
$\bar{Y}_d = \frac{Y_d}{N_d}$	Media del dominio d

A nivel muestral: Se seleciona una muestra aleatoria simple de tamaño n de la población.
Sea n_d el número de elementos en la muestra que caen en el dominio d.

$$n_d \leq n.$$

Al seleccionar una muestra aleatoria simple de una población, y establecer un dominio d, las unidades dentro de la muestra que pertenecen al dominio se identifican con:

$$x_i = \begin{cases} 1 & i \in d \\ 0 & i \notin d \end{cases}$$

donde x_i es el elemento i de la muestra, $i = 1, 2, \ldots, n$.
De esta manera el tamaño de la muestra para el dominio d se obtiene:

$$n_d = \sum_{i=1}^{n} x_i$$

Por lo tanto, aunque n es fijo, n_d variará de una muestra a otra.

Notación

La variable de interés se define como:

$$u_i = \begin{cases} y_i & i \in d \\ 0 & i \notin d \end{cases}$$

Suponga que se selecciona una m.a.s. de tamaño n, y se define un dominio de interés d.

Para estimar la Media del dominio d, se utiliza un estimador de la razón

$$\widehat{\bar{Y}}_d = \bar{y}_d = \frac{1}{n_d}\sum_{i=1}^{n_d} y_i = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} x_i}$$

Con varianza

$$\begin{aligned}
\widehat{V}\left(\bar{y}_d\right) &= \left(1-\frac{n}{N}\right)\frac{1}{n\bar{x}^2}\frac{\sum_{i=1}^{n}\left(u_i-\bar{y}_d x_i\right)^2}{n-1};\ \bar{x}=\frac{n_d}{n} \\
&= \left(1-\frac{n}{N}\right)\frac{1}{n\bar{x}^2}\frac{\sum_{i=1}^{n_d}\left(y_i-\bar{y}_d\right)^2}{n-1} \\
&= \left(1-\frac{n}{N}\right)\frac{n^2}{nn_d^2}\frac{\sum_{i=1}^{n_d}\left(y_i-\bar{y}_d\right)^2}{n-1}
\end{aligned}$$

$$
\begin{aligned}
\widehat{V}\left(\bar{y}_d\right) &\approx \left(1-\frac{n}{N}\right)\frac{1}{n}\left(\frac{N}{N_d}\right)^2\frac{\left(n_d-1\right)\widehat{S}^2_{yd}}{n-1}\\
&\approx \left(1-\frac{n}{N}\right)\frac{1}{n}\frac{N^2}{N_d^2}\frac{n_d}{n}\widehat{S}^2_{yd}\\
&\approx \left(1-\frac{n}{N}\right)\frac{1}{n}\frac{N^2}{N_d^2}\frac{N_d}{N}\widehat{S}^2_{yd}\\
&= \left(1-\frac{n}{N}\right)\frac{N_d}{N}\frac{1}{n_d}\frac{N_d}{N}\frac{N^2}{N_d^2}\widehat{S}^2_{yd}\\
&= \left(1-\frac{n}{N}\right)\frac{\widehat{S}^2_{yd}}{n_d}
\end{aligned}
$$

La aproximación depende de un tamaño de muestra grande en el dominio d; si la muestra es lo bastante grande entonces esperamos que

$$n_d/n \approx N_d/N \Rightarrow n \approx n_d \frac{N}{N_d}$$

y

$$(n_d - 1)/(n-1) \approx n_d/n$$

Es decir, en una muestra grande, la varianza de $\bar{y}_d$ es aproximadamente igual a la obtenida al usar la expresión de la varianza de $\bar{y}_d$ sin considerar que es un estimador de razón.

Estimador del total del dominio

En el caso de estimar un total, se tienen dos situaciones:

1. Si se conoce N_d:

$$\widehat{Y}_d = N_d \bar{y}_d$$

$$\widehat{V}\left(\widehat{Y}_d\right) = N_d^2 \widehat{V}\left(\bar{y}_d\right)$$

2. Si no se conoce N_d:

$$\widehat{N}_d = N \frac{n_d}{n}$$

$$\widehat{Y}_d = N \frac{n_d}{n} \frac{\sum_{i=1}^{n} u_i}{n_d} = N\bar{u}$$

$$\widehat{V}\left(\widehat{Y}_d\right) = N^2 \widehat{V}\left(\bar{u}\right)$$