Muestreo Aleatorio Estratificado 06|Marzo|2021

Una muestra aleatoria estratificada es aquella que se obtiene cuando se divide a la población de interés en grupos que no presentan traslapes (subconjuntos distintos), de tal manera que la unión de todos estos grupos es la población completa, en donde en cada uno de estos subconjuntos se aplica un muestreo aleatorio simple.

Notación y Propiedades

Sean L = Número de estratos y E_i = Estrato i, $i = 1, \dots, L$. Tales que P = Población donde:

$$P = \bigcup_{i=1}^{L} E_i$$

tal $E_i \cap E_j \neq \emptyset$ y $i \neq j$. Además anteriormente definimos N = tamaño de la población y n = tamaño de la muestra. Variables que pueden ser reescritas bajo esta nueva notación, de la siguiente manera:

$$N = \sum_{i=1}^{L} N_i$$

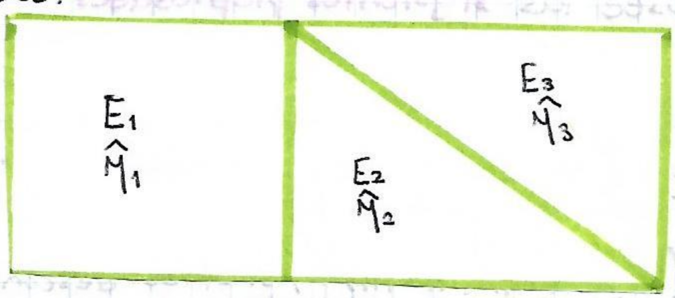
donde N_i = tamaño del estrato i. De manera similar:

$$n = \sum_{i=1}^{L} n_i$$

donde n_i = tamaño de la muestra del estrato i.

Estimación de una media poblacional μ apartir de un Muestreo Aleatorio Estratificado

Definir un estimador que se encuentre en función de las medias estimadas en cada estrato y el tamaño de estas, evitando sesgos en los resultados.



Propongamos como estimador de la media poblacional

$$\hat{\mu} = \frac{\sum_{i=1}^{L} N_i \hat{\mu}_i}{N}$$

el cual es un estimador insesgado de la media poblacional.

Muestreo Aleatorio Estratificado. 06/Marzo/2021

Demostración

$$E[\hat{\mu}] = E[\bar{y}] = E\left[\frac{\sum N_i \bar{y}_i}{N}\right] = \frac{1}{N} E\left[\sum_{i=1}^{L} N_i \bar{y}_i\right] = \frac{1}{N}\sum_{i=1}^{L} E[N_i \bar{y}_i]$$

$$= \frac{1}{N}\sum_{i=1}^{L} N_i E(\bar{y}_i) = \frac{1}{N}\sum_{i=1}^{L} N_i \mu = \frac{1}{N} N\mu = \mu$$

Ahora para conocer finalmente la distribución de este nuevo valor, calculemos:

$$Var(\hat{\mu}) = \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)\right]$$

Demostración

$$Var(\hat{\mu}) = Var(\bar{y}) = Var\left(\frac{\sum N_i \bar{y}_i}{N}\right) = \frac{1}{N^2} Var\left(\sum_{i=1}^{L} N_i \bar{y}_i\right)$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 Var(\bar{y}_i) + 2\sum_{i}\sum_{j} N_i N_j cov(\bar{y}_i, \bar{y}_j)\right]$$

como son pares distintos, $cov(\bar{y}_i, \bar{y}_j) = 0$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 Var(\bar{y}_i)\right] = \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)\right]$$

En resumen, posee las siguientes propiedades:

1. $E[\hat{\mu}] = \mu$

2. $$Var(\hat{\mu}) = \begin{cases} \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)\right], & \text{si } \sigma_i^2 \text{ es conocida} \\ \frac{1}{N^2}\left[\sum_{i=1}^{L} \frac{S_i^2}{n_i} N_i (N_i - n_i)\right], & \text{si } \sigma_i^2 \text{ es desconocida} \end{cases}$$

Con una cota en el error de estimación dada por:

$$B = 2\sqrt{Var(\hat{\mu})}$$

y un intervalo de confianza con un nivel de confianza al 95%, dado por los límites de error de estimación que pueden obtener-

se mediante:
- Límite Inferior: $\hat{\mu} - B$
- Límite Superior: $\hat{\mu} + B$

El margen de error puede estimarse como:

$$M.E. = \frac{B}{\hat{\mu}} * 100\%$$

Estimación de un total poblacional t a partir de un Muestreo Aleatorio Estratificado

Propongamos como estimador del total poblacional:

$$\hat{t} = N\hat{\mu} = N\frac{\sum N_i\hat{\mu}_i}{N} = \sum_{i=1}^{L} N_i\hat{\mu}_i = \sum_{i=1}^{L}\hat{t}_i$$

El cual posee las siguientes propiedades:

1. $E(\hat{t}) = t$

2. $Var(\hat{t}) = \begin{cases} \sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right), & \text{si } \sigma_i^2 \text{ es conocida} \\ \sum_{i=1}^{L} N_i \frac{S_i^2}{n_i}(N_i - n_i), & \text{si } \sigma_i^2 \text{ es desconocida} \end{cases}$

Demostración

$$E[\hat{t}] = \sum_{i=1}^{L} N_i E(\hat{\mu}_i) = \sum_{i=1}^{L} N_i \mu_i = t$$

$$Var[\hat{t}] = Var(N\hat{\mu}) = N^2 Var(\hat{\mu}) = N^2 \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)\right] = \sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)$$

Con una cota en el error de estimación dada por:

$$B = 2\sqrt{Var(\hat{t})}$$

Un intervalo de confianza dado por:
- Límite inferior: $\hat{t} - B$
- Límite superior: $\hat{t} + B$

Y un margen de error que se puede calcular como:

Muestreo Aleatorio Estratificado

$$M.E = \frac{B}{\hat{t}} \times 100\%$$

Nota

σ_i^2 es conocida si se conoce para toda i, si no se conoce para alguna i, entonces σ_i^2 es desconocida.

Estimación de una proporción poblacional p a partir de un Muestreo Aleatorio Estratificado

Propongamos como estimador de total poblacional:

$$\hat{p} = \frac{\sum^{L} N_i \hat{p_i}}{N}$$

El cual posee las siguientes propiedades:

1. $E[\hat{p}] = p$

2. $$Var(\hat{p}) = \begin{cases} \frac{1}{N^2} \sum_{i=1}^{L} N_i^2 \frac{p_i(1-p_i)}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right), & \text{si } \sigma_i^2 \text{ es conocida} \\ \frac{1}{N^2} \sum_{i=1}^{L} N_i \frac{(1-\hat{p}_i)\hat{p_i}}{n_i - 1}(N_i - n_i), & \text{si } \sigma_i^2 \text{ es desconocida} \end{cases}$$

Demostración

$$E[\hat{p}] = \frac{1}{N}\sum_{i=1}^{L} N_i E(\hat{p}_i) = \frac{1}{N}\sum_{i=1}^{L} N_i p_i = \frac{1}{N} Np = p$$

$$Var(\hat{p}) = Var\left(\frac{\sum N_i \hat{p_i}}{N}\right) = \frac{1}{N^2} Var\left(\sum_{i=1}^{L} N_i \hat{p_i}\right) = \frac{1}{N^2}\left[\sum_{i=1}^{L} N_i^2 Var(\hat{p}_i) + 2\sum_i\sum_j N_i N_j cov(\hat{p}_i, \hat{p}_j)\right]$$

$$= \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 Var(\hat{p}_i) = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \frac{p_i(1-p_i)}{n_i}\left(\frac{N_i - n_i}{N_i - 1}\right)$$

Con una cota en el error de estimación dado por:

$$B = 2\sqrt{Var(\hat{p})}$$

Un intervalo de confianza al 95% dado por:

- Límite inferior: $\hat{p} - \hat{B}$
- Límite superior: $\hat{p} + \hat{B}$

y un margen de error que se obtiene mediante:

$$M.E = \frac{B}{\hat{p}} * 100\%$$

Nota

σ_i^2 es conocida si se conoce para toda i, si no se conoce para alguna i entonces σ_i^2 es desconocida.

Selección de un tamaño de muestra

Para seleccionar un tamaño de muestra adecuado se emplea lo siguiente:

Para estimar una media poblacional

$$n = \frac{\sum_{i}^{L}\left[\frac{N_i^2\sigma_i^2}{w_i}\right]}{DN^2 + \sum_{i=1}[N_i\sigma_i^2]}, \text{ donde } D = \frac{B^2}{4}$$

Para estimar un total poblacional

$$n = \frac{\sum\left[\frac{N_i^2\sigma_i^2}{w_i}\right]}{DN^2 + \sum_{i=1}[N_i\sigma_i^2]}, \text{ donde } D = \frac{B^2}{4N^2}$$

Para estimar una proporción poblacional

$$n = \frac{\sum_{i}\left[\frac{N_i^2\sigma_i^2}{w_i}\right]}{DN^2 + \sum_{i=1}[N_i\sigma_i^2]}, \text{ donde } D = \frac{B^2}{4} \text{ y } \sigma_i^2 = p_i(1-p_i)$$

Selección de asignación de un estrato

Criterios o factores que hay que tomar en cuenta para realizar la asignación de una muestra en distintos estratos.

1. El número de elementos que hay en cada estrato E_i (directamente proporcional)
2. La variabilidad de datos en cada estrato (inversamente proporcional)
3. El costo por observación dentro de cada estrato (inversamente proporcional), entre ellos contemplamos:
   a) Mapas
   b) Sistemas, ambientes y plataformas
   c) Diseño de la muestra
   d) Trabajo de los encuestadores
   e) El total de encuestas a realizar

f) Alimentos
g) Renta de vehículos
h) Trabajo de supervisores
i) Papelería
j) Capturistas

Existe una asignación, a la cual se le conoce como asignación óptima:

$$w_i = \frac{\frac{N_i \sigma_i}{\sqrt{C_i}}}{\sum_{i=1}^{L} \frac{N_i \sigma_i}{\sqrt{C_i}}}$$

la cual nos asegura la varianza mínima del estimador poblacional. Además de esta asignación, existen algunas otras que son consideradas a partir de la asignación óptima:

Asignación con estratos iguales

$$w_i = \frac{\frac{\sigma_i}{\sqrt{C_i}}}{\sum_{i=1}^{L} \frac{\sigma_i}{\sqrt{C_i}}}$$

Asignación con costos iguales

$$w_i = \frac{N_i \sigma_i}{\sum N_i \sigma_i}$$

Asignación de misma volatilidad

$$w_i = \frac{\frac{N_i}{\sqrt{C_i}}}{\sum_{i=1}^{L} \frac{N_i}{\sqrt{C_i}}}$$

Asignación proporcional

$$w_i = \frac{N_i}{N}$$

Asignación equitativa

$$w_i = \frac{1}{L}$$