Instituto Cántabro de ESTADÍSTICA

Instituto Cántabro de ESTADÍSTICA

Santander, 28 de marzo de 2018

GOBIERNO de CANTABRIA

Liébana Cantabria Año Jubilar

Instituto Cántabro de ESTADÍSTICA

Muestreo Aleatorio Simple

Muestreo Estratificado

GOBIERNO de CANTABRIA

Liébana Cantabria Año Jubilar

Muestreo Aleatorio Simple

- Consiste en **elegir en forma aleatoria "*n*" unidades muestrales del universo "N"**. El proceso debe otorgar la misma oportunidad de selección a todas las unidades muestrales en una sola ocasión.
- Para poder realizar este tipo de muestreo, todos los individuos de la población deben estar numerados en un listado.
- Normalmente, se hace con un programa estadístico ó a partir de un listado de números aleatorios.
- Una alternativa si no se dispone de números aleatorios es la selección sistemática:
  - Se elige un numero aleatorio entre 1 y N: a
  - Se divide N entre el tamaño de muestra "n": b=N/n
  - Se selecciona como primer elemento el que ocupe el lugar a, el segundo elemento a encuestar sería el que ocupe el lugar a+b, el tercero el a+2b, y así sucesivamente hasta seleccionar las n muestras.

MAS: Inferencia

- Es el conjunto de métodos y técnicas que permiten inducir, a partir de la información empírica proporcionada por una muestra, cual es el comportamiento de una determinada población con un riesgo de error medible en términos de probabilidad.
- El estimador de una media poblacional $\bar{Y}$ en un MAS es :

$$\mu = \bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}$$

- La varianza del estimador $\bar{Y}$ es:

$\hat{\sigma}_{\bar{Y}}^2 = \left(\frac{S^2}{n}\right)\left(\frac{N-n}{N}\right)$, siendo $S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$

MAS: Inferencia

- El estimador de un total poblacional τ en un MAS es :

$$\hat{\tau} = N\bar{Y} = \frac{N\sum_{i=1}^{n} Y_i}{n} = \sum_{i=1}^{n}\left(\frac{N}{n}\right)Y_i$$

Factor de elevación: $\frac{N}{n}$

- La varianza del estimador es $\hat{\sigma}_{\hat{\tau}}^2 = N^2\left(\frac{S^2}{n}\right)\left(\frac{N-n}{N}\right)$
- El estimador de una proporción poblacional en un MAS es:

$\hat{P} = \frac{y}{n}$

donde y son los casos de la muestra que reúnen la característica con la que se elabora la proporción (los que responden afirmativamente a un ítem).

- La varianza del estimador es $\hat{\sigma}_{\hat{p}}^2 = \left(\frac{\hat{p}\hat{q}}{n-1}\right)\left(\frac{N-n}{N}\right)$, siendo $\hat{p} + \hat{q} = 1$

¿Cuándo se debe usar el muestreo aleatorio simple?:

- Cuando se sabe que la variable de mayor interés se distribuye aleatoriamente en el universo.
- Para universos pequeños.
- Para universos de poca dispersión geográfica.
- Cuando no se conoce el patrón de distribución para la variable de interés.

Muestreo Estratificado

- Este tipo de muestreo se utiliza cuando el **universo original**, de tamaño N, es **fragmentado en estratos relativamente homogéneos.**
- Ejemplos de estratos: tipo de municipio (ciudad, intermedio, rural), sexo, grupos de edades.
- A cada uno de estos estratos se les trata independientemente como un universo, en cuanto al método de selección de las unidades muestrales y de estimación de parámetros.
- Dentro de cada estrato, las unidades muestrales se pueden seleccionar en forma aleatoria, por conglomerados o sistemáticamente.
- No es aconsejable formar muchos estratos

Muestreo estratificado: Afijación

- Cómo distribuir la muestra total entre los estratos
- La afijación uniforme asigna el mismo número de muestras en cada estrato . Ejemplo: si el tamaño de la muestra es de 400, se elegirán aleatoriamente 200 mujeres y 200 hombres.
- La afijación proporcional distribuye la muestra total en proporción al número de unidades de cada estrato. Ejemplo: se sabe que en la población la distribución es del 55% de mujeres y 45% de hombres. Si el tamaño de la muestra es de 400, se elegirán aleatoriamente 220 mujeres y 180 hombres.

Muestreo estratificado: Afijación

- La afijación de mínima varianza o afijación de Neyman consiste en determinar los valores de n_h (número de unidades que se extraen del estrato h-esimo para la muestra) de forma que para un tamaño de muestra fijo igual a n la varianza de los estimadores sea mínima.

$$n_h = n\frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}, h = 1,2,\ldots,L$$

donde S_h es la cuasivarianza poblacional del estrato h-esimo y $N_h$es el tamaño poblacional del estrato h-esimo y L el número de estratos

- La afijación óptima consiste en determinar los valores de n_h (número de unidades que se extraen del estrato h-esimo para la muestra) de forma que para un coste fijo C, la varianza de los estimadores sea mínima. El coste fijo es la suma de los costes derivados de la selección de unidades en cada estrato, $C = \sum_{h=1}^{l} n_h C_h$, siendo $C_h$el coste de realizar la encuesta en el estrato h:

$$n_h = n\frac{N_h S_h / \sqrt{C_h}}{\sum_{h=1}^{L} N_h S_h / \sqrt{C_h}}, h = 1,2,\ldots,L$$

Determinación del tamaño de la muestra en el muestreo estratificado

En el caso más general de muestreo estratificado aleatorio, sin especificar el tipo de afijación empleado, el tamaño de la muestra se determina:

$$n = \left(\sum \frac{W_h^2}{w_h} S_h^2\right) \Bigg/ \left(\frac{d^2}{Z_\infty^2} + \frac{\sum W_h S_h^2}{N}\right)$$

Donde:

$W_h = \frac{N_h}{N}$, y N_h es el tamaño poblacional del estrato h-esimo

S_h es la cuasivarianza poblacional del estrato h-esimo.

$w_h = \frac{n_h}{n}$ el peso correspondiente a cada estrato en la muestra

d es la precisión o el error máximo admisible

Z_∞ el grado de seguridad o confianza

En el caso de muestreo con afijación proporcional el tamaño de la muestra para un total se determina:

$$n = \frac{\sum_{h=1}^{k} N_h s_h^2}{\frac{N d^2}{Z_\alpha^2} + \frac{\sum_{h=1}^{k} N_h s_h^2}{N}}$$

En el caso de muestreo con afijación proporcional el tamaño de la muestra para un total se determina:

$$n = \frac{\sum_{h=1}^{k} N_h s_h^2}{\frac{d^2}{Z^2 N} + \frac{\sum_{h=1}^{k} N_h s_h^2}{N}}$$

Inferencia en Muestreo Estratificado

- Se estima un valor medio para cada estrato con una muestra aleatoria estratificada con:

$$\bar{Y}_i = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$$

- La varianza del estimador $\bar{Y}_i$ es: $S_i^2 = \frac{\sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2}{n_i - 1}$
- Para estimar la muestra poblacional $\bar{Y}$ de una muestra estratificada:

$$\bar{Y}_{est} = \frac{1}{N} \sum_{i=1}^{L} N_i \bar{Y}_i$$

- Y la varianza del estimador es: $\hat{\sigma}^2_{\bar{Y}_{est}} = \frac{1}{N^2} \sum_{i=1}^{L} \left(\frac{S_i^2}{n_i}\right) \left(\frac{N_i - n_i}{N_i}\right)$

Inferencia en Muestreo Estratificado

- Para estimar un total poblacional en una muestra estratificada: $\hat{\tau} = N\bar{Y}_{est}$, siendo su varianza $\hat{\sigma}^2_{\hat{\tau}} = N\hat{\sigma}^2_{\bar{Y}_{est}}$.
- Si estimamos una proporción con una muestra aleatoria estratificada: $\hat{p}_{est} = \frac{1}{N}\sum_{i=1}^{L} N_i\hat{p}_i$
- La varianza estimada del estimador de una proporción:

$$\hat{\sigma}^2_{\hat{p}_{est}} = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\left(\frac{\hat{p}_i\hat{q}_i}{n_i - 1}\right)\left(\frac{N_i - n_i}{N_i}\right)$$