

VI Muestreo Sistemático



Dr.
Jesús
Mellado
Bosque

Definición

El muestreo sistemático, es aquel que se realiza en un lugar donde pasan las unidades muestrales y se selecciona una unidad cada k elementos que pasan, hasta cumplir los n necesarios.

Ejemplos:

Se muestrean latas de atún de una línea de producción cada 50 unidades.

A la salida de un supermercado se encuesta una persona después de de que han salido 10.

En una línea de empaque se muestrean bolsas de carne de pollo tomando uno de cada 35 que pasan.

En una oficina se muestrean los registro contables, tomando uno de cada 40 que se ingresan al archivero.

Ventajas

-Es más económico, ya que se evita el proceso de la selección aleatoria y evita la movilidad del encuestador.

- Es más preciso cuando se tiene una movilidad constante de elementos muestrales, ya que muestrea uniformemente.

Procedimiento

Si se conoce N , el tamaño de la población, y si por experiencias anteriores, se puede estimar n , entonces

$$k = N/n$$

El muestreo se realizará cada que pasen k elementos, es decir 1 en k .

Antes de iniciar el muestreo, se selecciona aleatoriamente un número entre 1 y k , que ese será la primera muestra, después se sigue la regla de 1 en k .

En caso de desconocer N , se escoge n como un valor exploratorio y luego se determina el valor de la muestra

Estimación de la media

El estimador de la media poblacional es la media muestral.

$$y = \frac{\sum y_i}{n}$$

Intervalo de confianza de la media

Para encontrar el intervalo de confianza para la media primero se encuentra la varianza de la muestra:

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

Luego se obtiene la varianza de la media, que es menor que la varianza muestral, así que se obtiene el estimador de la varianza de la media. Donde N es el tamaño de la población y n es el tamaño de la muestra. Cuando N es muy grande, el factor del paréntesis se puede eliminar.

$$\hat{V}(\bar{y}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

La varianza de la media se le obtiene la raíz cuadrada (se convierte a desviación estándar) y se multiplica por 2, se le resta y se le suma a la media y se conforma el intervalo con el 95% de seguridad.

$$\bar{y} - 2\sqrt{\hat{V}(\bar{y})} < \bar{y} < \bar{y} + 2\sqrt{\hat{V}(\bar{y})}$$

Ejemplo

En una productora de chorizos se desea conocer el contenido de grasa promedio de la producción diaria. Al día se producen 200 productos y se seleccionaron 15 sistemáticamente. Indique la media estimada y su intervalo de confianza al 95% de seguridad. Los datos son en gramos (datos ficticios):

21 14 13 12 14 13 16 20
23 22 20 19 25 25 23

La media es 18.67

La varianza muestral es:
21.24

La varianza de la media es:

$$\hat{V}(\bar{y}) = \frac{21.24}{15} \left(\frac{200-15}{200} \right) = 1.31$$

El límite del error es:

$$2\sqrt{\hat{V}(\bar{y})} = 2\sqrt{1.31} = 2.28$$

Sumando y restando 2.28 de la media se obtiene el intervalo de confianza

$$16.38 < \bar{y} < 20.96$$



Tamaño de la muestra para estimar la media

Entre mas grande sea la muestra, el intervalo de confianza para la media será más pequeño, entonces, para definir el tamaño de la muestra primero se define el tamaño del intervalo esperado. Se define **B** como la distancia que se desea entre la media y el límite superior de la media, según las unidades utilizadas.

Se calcula el valor de D $D = \frac{B^2}{4}$

Se calcula n: $n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$

Desafortunadamente, para encontrar el tamaño de la muestra, se requiere conocer de antemano la varianza poblacional, la cual se puede hacer de la siguiente manera:

- Investigar pruebas anteriores donde se haya calculado la varianza
- Si se conoce el valor mínimo y el máximo, el rango dividido entre cuatro puede ser un estimador.
- Obtener una pequeña muestra, luego calcular el tamaño de la muestra correcta utilizando los datos ya recopilados.

Donde σ^2 es la varianza poblacional o s^2 , que es su estimador

Ejemplo

En una productora de chorizos se desea conocer el contenido de grasa promedio de la producción diaria. Si al día se producen 200 productos y por experimentos previos se sabe que la varianza es de 21.24 grs. Encontrar el tamaño de la muestra necesaria si se permite un error máximo de 1 grs.

El valor B = 1

El valor D es: $D = \frac{1^2}{4} = 0.25$

El valor de n es: $n = \frac{200(21.24)}{(200-1)0.25 + (21.24)} = 59.83$

El resultado se redondea a 60. La cantidad es muy alta porque el límite de error se seleccionó muy pequeño.

En algunas condiciones es difícil conocer el valor de N, por ejemplo, el número de plantas pequeñas en una parcela, el número de árboles de cierta región o el número de fauna silvestre de un área. En tal caso el tamaño de la muestra se calcula con la siguiente ecuación

$$n = \left(\frac{1.96s}{B} \right)^2$$

Donde s es la desviación estándar muestral y B es el error máximo permitido. Esta ecuación es con el 95% de seguridad

Estimación del total de una variable

El estimador del total es: $\hat{\tau} = Ny$

Intervalo de confianza del total

Para encontrar el intervalo de confianza del total primero se encuentra la varianza de la muestra:

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

Luego se obtiene la varianza del estimador del total, se le agrega el término N porque se multiplica por el total de elementos t está al cuadrado porque está dentro de la varianza.

$$\hat{V}(\hat{\tau}) = N^2 \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

La varianza de la media se le obtiene la raíz cuadrada (se convierte a desviación estándar) y se multiplica por 2, se le resta y se le suma a la media y se conforma el intervalo con el 95% de seguridad.

$$\hat{\tau} - 2\sqrt{\hat{V}(\hat{\tau})} < \hat{\tau} < \hat{\tau} + 2\sqrt{\hat{V}(\hat{\tau})}$$

Ejemplo

En una huerta de 600 manzanos de primer año de producción se desea estimar el total de la producción, la muestra se realizó sistemáticamente muestreando 1 de cada 30 árboles, además se desea conocer el intervalo de confianza del total al 95% de seguridad. Los datos son en kilos (datos ficticios):

25 24 18 26 32 26 22 23
24 26 24 24 32 28 26 21
20 26 20 22

La media es 24.45

El total es: $24.45(600) = 14,670$ kg.

La varianza muestral es: 13

La varianza de del total es:

$$\hat{V}(\hat{\tau}) = 600^2 \frac{13}{20} \left(\frac{600-20}{600} \right) = 226,154$$

El límite del error es:

$$2\sqrt{\hat{V}(\hat{\tau})} = 2\sqrt{226,154} = 951.11$$

Sumando y restando 951.11 del total se obtiene el intervalo de confianza

$$13,719 < \tau < 15,629$$

Tamaño de la muestra para estimar el total

Se define B como la distancia que se desea entre el total y el límite superior, según las unidades utilizadas.

Se calcula el valor de D
$$D = \frac{B^2}{4N^2}$$

Se calcula n :
$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

Donde σ^2 es la varianza poblacional o su estimador s^2

Ejemplo

En una huerta de 600 manzanos de primer año de producción se desea estimar el total de la producción, pero primero se desea saber el tamaño de la muestra. En una muestra previa se encontró una varianza de 13 kilos. El error máximo permitido es de 1000 kg.

El valor $B = 1000$

El valor D es:
$$D = \frac{1000^2}{4(600)^2} = 0.694$$

El valor de n es:
$$n = \frac{600(13)}{(600-1)0.694 + (13)} = 18.8$$

Nótese que el resultado se redondea al límite superior.

Para calcular el total se debe conocer N , es por eso que no existe ecuación simplificada para el tamaño de la muestra.

Estimación de una proporción

Los objetos a observar se dividen entre los que tienen una cualidad y los que no. A los objetos que tienen la cualidad se les asigna el número 1, y a los demás el número 0.

Si p es la proporción de elementos que cumplen cierta cualidad, entonces se puede estimar mediante una muestra. El estimador es:
$$\hat{p} = \frac{\sum y_i}{n}$$

Una vez obtenida la muestra se puede obtener el valor q ($q=1-p$) y la varianza muestral como:
$$s^2 = \hat{p}\hat{q}$$

La varianza del estimador es:
$$\hat{V}(\hat{p}) = \frac{\hat{p}\hat{q}}{n-1} \left(\frac{N-n}{N} \right)$$

El intervalo de confianza para la proporción es:

$$\hat{p} - 2\sqrt{\hat{V}(\hat{p})} < \hat{p} < \hat{p} + 2\sqrt{\hat{V}(\hat{p})}$$

Ejemplo

En la sierra de Arteaga se desea saber cuántos pinos han sido atacados por una plaga de un escarabajo de la corteza, se estima que existen 1050 pinos en la zona a muestrear. La muestra se realiza de manera sistemática caminando en una línea tratando de recorrer todos los pinos y muestrear uno en cada 50. Se muestrearon 21 pinos con los siguientes resultados:

0 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0

La proporción es $4/21 = 0.190$

La varianza muestral es: $0.190(1 - 0.190) = 0.154$

La varianza de la proporción es: $\hat{V}(\hat{p}) = \frac{0.154}{21-1} \left(\frac{1050-21}{1050} \right) = 0.0078$

El límite del error es: $2\sqrt{\hat{V}(\hat{p})} = 2\sqrt{0.0078} = 0.177$

Sumando y restando 0.177 de la proporción se obtiene el intervalo de confianza $0.131 < p < 0.367$

Tamaño de la muestra para estimar la proporción

Se calcula el valor de D $D = \frac{B^2}{4}$

Se calcula n: $n = \frac{N\hat{p}\hat{q}}{(N-1)D + \hat{p}\hat{q}}$

Ejemplo

En una sierra se desea muestrear de 1050 pinos aquellos que estén enfermos. La muestra se desea proporción sistemática y se sabe que la varianza es 0.154; el límite máximo del error es 0.08

El valor B = 0.08

El valor D es: $D = \frac{0.08^2}{4} = 0.0016$

El valor de n es: $n = \frac{1050(0.154)}{(1050-1)0.0016 + (0.154)} = 88.34$

En algunas condiciones es difícil conocer el valor de N, por ejemplo, el número de plantas pequeñas en una parcela, el número de árboles de cierta región o el número de fauna silvestre de un área. En tal caso el tamaño de la muestra se calcula con la siguiente ecuación

$$n = \left(\frac{1.96}{B} \right)^2 \sqrt{pq}$$

Donde pq es la varianza muestral y B es el error máximo permitido. Esta ecuación es con el 95% de seguridad



Dr. Jesús Mellado



Departamento de
Estadística y
Cálculo