Muestreo Estadístico

Dr.
Jesús Alberto
Mellado

I Introducción

Aplicaciones de muestreo

- Levantamiento de encuestas
- Muestreo para la investigación
- Muestreo para el control de calidad

Elemento

Objeto del que se puede obtener cierta información, ya sea por sus características o por su comportamiento.

Población

Conjunto bien definido de elementos de los cuales se desea hacer alguna inferencia.

Parámetro

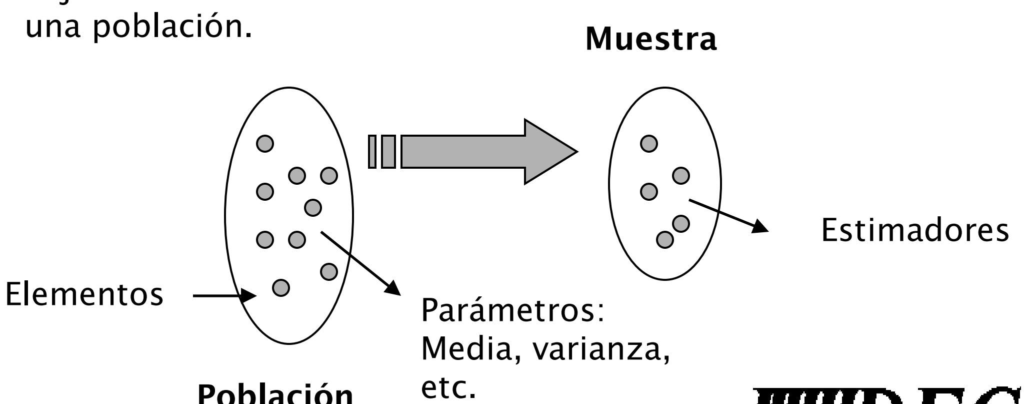
Un parámetro es un valor que describe en términos estadísticos las características de una población (media, varianza, etc.)

Muestra

Es un conjunto de elementos tomados bajo cierto criterio de una población.

Estimador

Es una función de variables aleatoria de datos obserbables para inferir sobre los parámetro de acuerdo a una muestra.



Departamento de
Estadística y Cálculo

Uso de muestras

Cuando las poblaciones son pequeñas lo recomendable es realizar un censo en lugar de una muestra, ya que se tienen resultados confiables, pero cuando la población es demasiado grande o estadísticamente infinita, es necesario la recolección de una muestra.

Objetivo del Muestreo

El objetivo del muestreo es obtener inferencia sobre una población de interés, de la forma más eficiente y confiable.

II Encuestas

Objetivo

El objetivo de la elaboración y la aplicación de encuestas es la recopilación de datos suficientes para inferir sobre una población. Normalmente las encuestas se usan para investigar sbre más de una variable, ya que procedimiento no es sencillo en la mayoría de los casos.

Métodos de recolección de la información

Entrevista Personal: Es el procedimiento donde una persona acude al sujeto seleccionado para obtener la información.

Estrevista por Teléfono: Si el objetivo son personas cuyo nivel económico les permite tener teléfono, entonces es válida la entrevista por teléfono

Cuestionarios Autoaplicados: Se pueden entrevar o enviar por correo cuestionarios que las personas respondes por algún interés especial.

Observación directa: Si las variables a evaluar lo permiten, es posible realizar la observación directa:

Diseño del cuestionario

Explicación: Es conveniente que el encuestado esté enterado del procedimiento de cómo se llevará a cabo la entrevista. Ejemplo: Esta encuesta consiste en 10 preguntas y usted debe escoger una de las opciones que le voy a dar.

Orden de las preguntas: La preguntas deben estar agrupadas por temas y de lo general a lo particular. Se ha demostrado que la organización de las preguntas afecta las respuestas. Ejemplo: Qué opina de la pena de muerte en el mundo ?. Qué opina de la pena de muerte en México ?.

Preguntas abiertas y cerradas: Dependiendo del objetivo de la pregunta, se debe escoger si es abierta o cerrada. Se usa más la pregunta cerrada por su fácil manejo. Ejemplo: Qué opina del desempeño de Fox. Qué opina del desempeño de Fox a) Muy Bueno b) Bueno …

Redacción de las preguntas: Las preguntas deben ser equilibradas y no inducir la respuesta. Ejemplo: Verdad que la pena de mueste es mala ?. Apoya usted la pena de muerte ?. Apoya usted la pena de muerte o la rechaza ?.

Planeación de una encuesta

Establecimiento de objetivos.

Definir población objetivo.

Definir el marco de la muestra. El método qué define qué personas se van a entrevistar.

Seleccionar el diseño de muestreo. Diseño aleatorio irrestricto, conglomerados, etc.

Seleccionar el método de medición. Entrevista personal, telefónica, etc.

Diseñar el instrumento de medición

Selección y adiestramiento de investigadores de campo

Preferentemente realizar una prueba piloto

Organización del trabajo de campo

Organización del manejo de datos

Análisis de los datos.

Dr. Jesús Mellado Bosque

III Muestreo simple aleatorio

Definición

Cuando se realiza un muestreo, donde se seleccionan los elementos al azar y todos los elementos tienen la misma probabilidad de ser seleccionados, a este muestreo se le llama Simple Aleatorio o Muestreo Irrestricto Aleatorio.

Selección de elementos

A cada elemento se le asigna un número consecutivo. Si no es posible, como el muestreo de la población de pinos en una sierra, se puede usar una fotografía aérea y trazar la cuadrícula más pequeña posible y numerar los cuadrantes. En caso de líneas de producción donde los productos salen en contenedores con charolas, se puede numerar cada contenedor y luego marcar cada charola y el número de producto.

Se obtienen los números de los elementos seleccionados aleatoriamente. Se puede usar una tabla de números aleatorios o se puede usar una computadora con la función “random” o “aleatorio”.

Normalmente los números aleatorios van de 0 a 1 con cinco o mas decimales (0.3423, 0.5678, etc). El número aleatorio se multiplica por N y se obtiene el número del elemento

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:

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

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.

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

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

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

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 aleatoriamente. 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:

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

El límite del error es:

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

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 tamaña 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:

a) Investigar pruebas anteriores donde se haya calculado la varianza

b) Si se conoce el valor mínimo y el máximo, el rango dividido entre cuatro puede ser un estimador.

c) 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 su estimador

Dr. Jesús Mellado Bosque

Ejemplo

En una productora de chorizos se desea conocer el contenido de grasa promedio de la producción diaria. Es por eso que es necesario encontrar el tamaño de la muestra. Al día se producen 200 productos y se sabe por experimentos previos que la varianza poblacional es de 21 gr. Y se desea un error de 2 gr. para la media.

El valor B = 2

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

El valor de n es: $n = \frac{200(21)}{(200-1)1 + (21)} = 20$

Nótese que el resultado se redondea al límite superior.

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:

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

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 raiz cuadrada (se convierte a desviación estándar) y se multiplica por 2, se le resta y se le suma a ala media y se conforma el intervalo con el 95% de seguridad.

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

Ejemplo

En una productora de chorizos se desea conocer el contenido total de grasa de la producción diaria. Al día se producen 200 productos y se seleccionaron 15 aleatoriamente. Indique el total estimado 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

El total es: 18.67(200) = 3733 gr

La varianza muestral es: 21.24

La varianza de del total es:

$$\hat{V}(\hat{\tau}) = 200^2 \frac{21.24}{15}\left(\frac{200-15}{200}\right) = 52387$$

El límite del error es:

$$2\sqrt{\hat{V}(\hat{\tau})} = 2\sqrt{52387} = 457.8$$

Sumando y restando 457.8 del total se obtiene el intervalo de confianza

$$3276 < \bar{\tau} < 4191$$

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

7

Ejemplo

En una productora de chorizos se desea conocer el contenido de grasa total de la producción diaria. Es por eso que es necesario encontrar el tamaño de la muestra. Al día se producen 200 productos y se sabe por experimentos previos que la varianza poblacional es de 21 gr. Y se desea un error de 300 gr. para el total.

El valor B = 4

El valor D es: $D = \frac{300^2}{4(200)^2} = 0.56$

El valor de n es: $n = \frac{200(21)}{(200-1)0.56 + (21)} = 32$

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.

IV Muestreo cualitativo

Dr. Jesús Mellado Bosque

Las variables se dividen en variables cuantitativas y cualitativas, las cuantitativas tratan características que se pueden medir (peso, altura, etc.). Las variables cualitativas tratan características que no se pueden medir (color, textura, opinión, etc.).

Para manejar las variables cualitativas se puede optar por el muestreo de proporciones. 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.

IV Muestreo cualitativo

Las variables se dividen en variables cuantitativas y cualitativas, las cuantitativas tratan características que se pueden medir (peso, altura, etc.). Las variables cualitativas tratan características que no se pueden medir (color, textura, opinión, etc.).

Para manejar las variables cualitativas se puede optar por el muestreo de proporciones. 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, estonces se puede estimar mudiante una muestra. El esimador 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 es:

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

Ejemplo

En una majada se tienen 500 chivas de las cuales se tiene la sospecha de que algunas están enfermas, para calcular la proporción de chivas enfermas se tomó una muestra aleatoriamente de 20 chivas y los resultados son los siguientes, donde las infectadas se tomaron como 1.

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

La proporción es 7/20 = 0.35

La varianza muestral es: 0.35(1 – 0.35) = 0.227

La varianza de la proporción es:

$$\hat{V}(\hat{p}) = \frac{0.227}{20-1}\left(\frac{500-20}{500}\right) = 0.011$$

El límite del error es:

$$2\sqrt{\hat{V}(\hat{\tau})} = 2\sqrt{0.11} = 0.214$$

Sumando y restando 0.214 de la proporción se obtiene el intervalo de confianza

0.135 < p < 0.564

Tamaño de la muestra para estimar la proporción

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

Hay que recordar que las proporciones van de 0 a 1, el valor de de debe ser arpoximado a 0.2 o menor

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

Ejemplo

En una majada se tienen 500 chivas, se desea realizar una muestra para conocer la proporción de chivas enfermas. Encontrar el tamaño de la muestra si se supone que la proporción debe estar cerca del 30% y se requiere um error mázimo de 0.15

El valor B = 0.15

El valor D es: $D = \frac{0.15^2}{4} = 0.0056$

El valor de n es: $n = \frac{500(0.3)(0.7)}{(500-1)0.0056 + (0.3)(0.7)} = 35$

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