INFERENCIA ESTADÍSTICA

Definición, características, usos

DEFINICIÓN

Se le llama estimador a una función de la muestra utilizada para estimar un parámetro desconocido de la población.

El problema matemático de la estimación surge de generar resultados fiables de muestras obtenidas a partir de una población.

Ejemplos de estos son la media y la varianza.

PROPIEDADES DE UN ESTIMADOR

Todos los estimadores son estadísticos, o sea una función real medible de la muestra de una variable aleatorio.

Para que un estadístico sea un estimador debe de cumplir varias condiciones.

Puede haber mas de un estimador para un mismo parámetro.

SUFICIENCIA

Un estimador es suficiente porque trabaja con todos los datos de la muestra.

La muestra ya es un subconjunto de la población, un estimador tiene que tener en cuenta al 100% de la información de la muestra porque los supuestos de errores se obtienen en función al tamaño muestral.

SESGO

El sesgo es la diferencia entre la esperanza y el verdadero valor del parámetro estimado.

Un estimador debe de ser insesgado (o centrado), esto quiere decir que su esperanza es igual al parámetro.

Por ejemplo:

$$E(\bar{X}) = \mu$$

CONSISTENCIA

Es una propiedad que dicta que cuando crece el tamaño muestral, el valor del estimador se acerca al valor del parámetro.

Hay muchas definiciones matemáticas de esta propiedad, la mas utilizada, llamada consistencia en media cuadrática dicta que la esperanza del estimador tiende al parámetro, cuando n tiende a infinito; y la varianza del estimador tiende a 0, cuando n tiende a infinito.

ROBUSTEZ

Un estimador robusto arroja resultados similares a los reales por mas que las hipótesis de partida sean incorrectas.

EFICIENCIA

Tiene en cuenta la varianza del estimador en relación o otros que estimen lo mismo. El mas eficiente será el que tenga menor varianza.

La eficiencia de los estimadores está limitada por las características de la distribución de probabilidad de la muestra de la que proceden.

ESTIMACIÓN PUNTUAL

Utilizando la fórmula de los estimadores podemos llegar a un único valor numérico, estimado del parámetro.

Son estimaciones puntuales los cálculos de media muestral o varianza muestral por ejemplo.

ESTIMACIÓN POR INTERVALOS

Si queremos saber e indicar la precisión de una estimación, será necesario presentar la estimación como un par de valores entre los cuales sabemos con cierta seguridad que se encuentra el parámetro.

ERRORES INTRODUCIDOS EN EL MUESTREO

Es necesario recordar que al tomar una muestra se introducen dos tipos de errores:

- *Errores no aleatorios, o sistemáticos*: Son los que se introducen por los sesgos de la selección de la muestra. Esto se controla con un buen diseño de estudio y muestreo.
- *Errores aleatorios, o experimentales*: Es el problema que surge del proceso mismo de muestreo, ya que no observamos a la población total y la muestra es aleatoria. Es importante recordar que este error disminuye cuando se aumenta el tamaño de la muestra y puede ser medido.

Es importante evitar el primero y reducir en la medida de lo posible el segundo.

DISTRIBUCIÓN MUESTRAL

Cuando se toma una muestra aleatoria de tamaño n y calculamos su media, se sabe que debido al error de muestreo ésta será diferente de la verdadera media.

Es de interés conocer cuan diferentes son la estimación del parámetro al valor real del parámetro.

Suponiendo que se toman muchas muestras todas del mismo tamaño n y de la misma población, se pueden obtener sus respectivas medias, si estas son similares entre si se puede decir que el error de muestreo es pequeño.

DISTRIBUCIÓN MUESTRAL

La magnitud de este error se sabe que depende fundamentalmente de dos aspectos:

- Cuando mayor sea el tamaño de la muestra menor error (en un censo completo no existe error)
- La variabilidad (dispersión) de los valores en la población estudiada. Una población heterogénea tendrá un error de muestreo mayor que el de una población homogénea. Esto ocurre porque a menor dispersión, los valores seleccionados para la muestra tenderán a estar mas cercanos a la media poblacional.

DISTRIBUCIÓN MUESTRAL

Si se obtuvieran todas las muestras posibles para una población (manteniendo el n) y se observase la distribución de sus medias se podría ver que:

- Esta nueva distribución tiende a ser normal sin importar la característica de la variable original.
- La media de la distribución de estas medias es igual a la media de la población.
- La desviación estándar de esta distribución es:

$$EE\bar{x} = \frac{\sigma}{\sqrt{n}}$$

- A este valor se le llama error estándar.

DISTRIBUCIÓN MUESTRAL

Como esta distribución de las muestras es normal, podemos aplicar propiedades de esta y decir por ejemplo que:

$\mu \pm 1{,}96\ EE\bar{x}$ incluye al 95% de las medias muestrales

$\mu \pm 2{,}58\ EE\bar{x}$ incluye al 99% de las medias muestrales

INTERVALOS DE CONFIANZA

Como lo anterior surge del supuesto de tomar todas las posibles muestras para una población, que utilidad tiene esto cuando tomamos solo una muestra?

Anteriormente se vio que en una distribución de medias muestrales el 95% estarán dentro del intervalo:

$$\mu \pm 1{,}96\ EE\bar{x}$$

Dicho de otra manera, ahora sabemos que cualquier muestra única que obtengamos tiene un 95% de tener una media dentro de ese intervalo.

Podemos construir un intervalo dentro del cual caiga el 99% de las medias, o el 99.9%, o cualquier otro, pero nunca podremos definir un intervalo tal en que la certeza de nuestras afirmaciones sea absoluta (recuérdese que la distribución normal es asintótica)

Todo este mismo razonamiento se puede aplicar para estimar una proporción de una variable binomial

INTERVALOS DE CONFIANZA

Para el caso real de una población finita existe un factor de corrección al error típico que se calcula como N-n/N-1.

Si n/N es menor a 0,1 el factor de corrección tiende a 1 y no es necesario utilizarlo.

INTERVALOS DE CONFIANZA

Idealmente, el intervalo de confianza va a contener el valor real del parámetro y cuanto mayor precisión tengamos en la estimación, mas estrecho será el intervalo.

A la probabilidad de que el intervalo contenga al valor del parámetro se le llama coeficiente de confianza, y si lo expresamos en porcentaje le llamamos nivel de confianza.

Al complementario del coeficiente de confianza se le llama alfa.