Estimación

Definición, características, usos

Instituto CBQ

Educamos diferente

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.

Medias UN GRUPO → Tamaño de muestra: $n \geq (z\sigma / e)^2$

¿La variable es normalmente distribuida o bien n>30?

No → **Cambio de Procedimiento**

Si

¿La varianza poblacional es conocida?

Si

Intervalo de Confianza

$$\mu \in \left(\bar{x} \pm z\sigma \big/ \sqrt{n} \right)$$

Prueba de Hipótesis Ho:μ = k

$$Z_c = (\bar{x} - \mu) \big/ \left(\sigma / \sqrt{n} \right)$$

No

Intervalo de Confianza

$$\mu \in \left(\bar{x} \pm t\hat{s} \big/ \sqrt{n} \right)$$

Prueba de Hipótesis Ho: μ = k

$$t_c = (\bar{x} - \mu) \big/ \left(\hat{s} \big/ \sqrt{n} \right)$$

Si n es grande

Estimación de medias con varianza poblacional conocida

- Supuestos:
  - La variable sigue una distribución Normal o bien
  - Tamaño muestral mayor a 30
- El estadístico para el intervalo de confianza es:

$$\mu \in \left(\bar{x} \pm z \sigma \middle/ \sqrt{n} \right)$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- El procedimiento consiste en:
  - Calcular el z crítico.
  - Transformar el z en x.

Ejercicio 1

Un investigador necesita conocer el nivel medio de una enzima en una población. Los datos disponibles corresponden a determinaciones hechas en un conjunto de 400 individuos y se obtuvo una media de 91 unidades, se sabe que el CV poblacional es de 15%.

Estimar el intervalo de la media poblacional con un nivel de confianza del 99%.

Calcular el menor tamaño muestral para cometer un error no mayor que 5 unidades, a un nivel de confianza del 95%.

Ejercicio 2

Se quiere estimar el tiempo de espera medio en el triage de una emergencia para el cual se conoce la desviación típica (0,5 minutos) y se sabe que sigue una distribución normal.

El tiempo medio de espera en una muestra aleatoria de 25 usuarios se midió en 5,2 minutos.

Calcular el intervalo de confianza al 95% para el tiempo medio.

Calcular el tamaño muestral mas pequeño que me permita estimar la media a un nivel de confianza del 95% con un error no mayor de 0,5 minutos.

Estimación de medias con varianza poblacional desconocida

- Supuestos:
  - La variable sigue una distribución Normal o bien
  - Tamaño muestral mayor a 30
- El estadístico para el intervalo de confianza es:

$$\mu \in \left(\bar{x} \pm t\hat{s} / \sqrt{n} \right)$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \qquad gl = n-1$$

- El procedimiento consiste en:
  - Calcular el t crítico.
  - Transformar el t en x.

Ejercicio 1

La puntuación media de una muestra de 20 estudiantes de la EUTM para una misma prueba presenta una media de 9,8525 en 15 y una desviación típica muestral de 0,0965. Cual es el intervalo de confianza a un 95% para el medio de la puntuación.

Estimación de proporciones

- Supuestos:
  - La variable se puede aproximar a una distribución normal porque la muestra es grande (se aplica la teoría del límite central)
- El estadístico para el intervalo de confianza es:

$$p \in \left(p \pm z\sqrt{pq/n} \right)$$

$$z = \frac{X - np}{\sqrt{npq}} = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

- El procedimiento consiste en:
  - Calcular el z crítico.
  - Transformar el z en valores límites de p.

Ejercicio 1

Para predecir el tiempo de recambio de un receptor de imagen digital de un equipo de radiología se quiere saber cual es la proporción de repetición en el total de estudios realizados en 1 año. Se tomó una muestra aleatoria de 100 estudios y se vio que 10 de ellos requirieron una repetición.

Determinar el IC al 95% para la proporción poblacional de repeticiones de estudios.

Cual sería el mínimo tamaño muestral que me permite calcular el anterior IC con un error menor a un 2% de la población.

Ejercicio 2

En una muestra de 100 pacientes que recibieron un nuevo tratamiento se encontró que 80 se curaron.

Cual es la proporción poblacional de curación (IC 95%)

Que tamaño mínimo de muestra es necesario para estimar la proporción con un 99% de confianza y un error no mayor al 3%