

ESTADÍSTICA TEÓRICA II. Tercer curso de Economía
CURSO 2007/2008. CONVOCATORIA DE SEPTIEMBRE
Código de la Carrera 43 Código de la Asignatura 305.

PRIMERA PARTE: CUESTIONES

1.- Definición de estadístico, estimador y estimación. Ponga un ejemplo de cada uno de estos conceptos en los que se aprecien las diferencias entre ellos.

Respuesta.-

Estadístico es cualquier función de la variable muestral (X_1, X_2, \dots, X_n) :

$$\hat{\theta} = g(X_1, X_2, \dots, X_n)$$

Estimador es un estadístico que usamos para hallar el valor de un parámetro poblacional desconocido.

Estimación es un valor concreto del estimador, para una muestra concreta elegida.

Por ejemplo, estadísticos serían el máximo de (X_1, X_2, \dots, X_n) , la media muestral

$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, la varianza muestral $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$; cualquier momento muestral

respecto del origen $\frac{1}{n} \sum_{i=1}^n X_i^n$, etc

Un estimador de la media poblacional sería por ejemplo la media muestral $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Si como estimador de la media poblacional elegimos la media muestral, para muestras de tamaño 5 y, elegida una muestra, se obtienen los valores $\{2,5; 3; 2; 3,1; 2,7\}$, entonces la estimación correspondiente sería $\frac{2,5 + 3 + 2 + 3,1 + 2,7}{5} = \frac{13,3}{5} = 2,66$

2.- Error cuadrático medio de un estimador: concepto. ¿Para qué se utiliza?.

Respuesta.-

Sea $\hat{\theta}$ el estimador de un parámetro poblacional θ . Se define el error cuadrático medio como el valor de $E[(\hat{\theta} - \theta)^2]$.

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

Si desarrollamos dicho valor:

$$\begin{aligned} E[(\hat{\theta} - \theta)^2] &= (restamos y sumamos $E(\hat{\theta})$) = $E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}))^2] +$
 $+ E[(E(\hat{\theta}) - \theta)^2] + 2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)] = \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 + 2(E(\hat{\theta}) - \theta)E[(\hat{\theta} - E(\hat{\theta}))] =$
 $= (\text{teniendo en cuenta que } E[(\hat{\theta} - E(\hat{\theta}))] = 0) = \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2.$$$

El valor de $E(\hat{\theta}) - \theta$ recibe el nombre de Sesgo de $\hat{\theta}$, luego podemos escribir:

$$ECM(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{Sesgo de } \hat{\theta})^2$$

En el problema de la estimación puntual interesa que el error cuadrático medio sea lo menor posible, lo cual se consigue cuanto menores sean la varianza de $\hat{\theta}$ y el valor absoluto del sesgo. Si el estimador es insesgado ($\text{Sesgo de } \hat{\theta} = 0$) el error cuadrático medio de $\hat{\theta}$ coincide con su varianza.

3.- Condiciones que debe cumplir el pivote en el método pivotal.

Respuesta.-

Si θ es el parámetro desconocido, entonces el pivote es una función $T(X_1, X_2, \dots, X_n, \theta)$ que debe cumplir las condiciones:

- 1) El valor que toma $T(X_1, X_2, \dots, X_n, \theta)$ para cada muestra depende exclusivamente de θ .
- 2) La distribución muestral del estadístico $T(X_1, X_2, \dots, X_n, \theta)$ no depende de θ .

4.- Concepto de nivel de significación y potencia de un contraste. Relación entre ambos.

Respuesta.-

Nivel de significación α de un contraste es la probabilidad de cometer error de tipo I, es decir, es la probabilidad de rechazar la hipótesis nula, siendo ésta cierta. También se denomina tamaño de la región crítica (o de rechazo) ya que la probabilidad de que el estimador pertenezca a la región crítica es precisamente α .

La potencia del contraste es la probabilidad de rechazar la hipótesis nula. Si representamos por β la probabilidad de error de tipo II, es decir, la probabilidad de aceptar la hipótesis nula, siendo falsa, entonces:

$$\text{Potencia del contraste} = \begin{cases} \alpha, & \text{si } H_0 \text{ cierta} \\ 1 - \beta, & \text{si } H_0 \text{ falsa} \end{cases}$$

Para un tamaño muestral n fijo, si α aumenta, entonces β disminuye y, por lo tanto, $1 - \beta$ aumenta, y viceversa.

PROBLEMAS

1.- Dada una población binomial $B(1, p)$ se extrae una muestra aleatoria simple de tamaño n y se

utiliza como estimador del parámetro p la suma de los valores muestrales: $T = \sum_{i=1}^n x_i$

¿Es suficiente el estimador propuesto?.

Solución.-

Calculamos la función de probabilidad de la muestra (X_1, X_2, \dots, X_n) condicionada por el estadístico T :

$$P(X_1=x_1, X_2=x_2, \dots, X_n=x_n / T=t) = \begin{cases} \frac{p^t (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}}, & \text{si } \sum_{i=1}^n x_i = t \\ 0, & \text{si } \sum_{i=1}^n x_i \neq t \end{cases} = \begin{cases} \frac{1}{\binom{n}{t}}, & \text{si } \sum_{i=1}^n x_i = t \\ 0, & \text{si } \sum_{i=1}^n x_i \neq t \end{cases}$$

Dado que esta probabilidad no depende del parámetro a estimar p concluimos que T contiene toda la información que la muestra puede proporcionar sobre el valor de p , es decir, T es suficiente.

2.- En una población $N(\mu, 2)$ se desconoce el valor de la media, estableciéndose dos hipótesis: bien es igual a 4 o bien es igual a 6. Se decide realizar un contraste de hipótesis con nivel de significación 0'05, para lo cual se selecciona una muestra aleatoria simple de tamaño $n = 4$, cuyos elementos son: 1'8; 6'6; 2'4 y 3'5.

- a) Utilizando el lema de Neymann-Pearson, determine la mejor región crítica y explique la decisión que se tomará en consecuencia.
- b) Halle la potencia del contraste.

Solución.-

a) Consideramos $H_0: \mu = 4$ y $H_1: \mu = 6$. Aplicamos el lema de Neymann-Pearson:

$$\frac{L(X_1, \dots, X_4 / H_0)}{L(X_1, \dots, X_4 / H_1)} = \frac{\left(\frac{1}{2\sqrt{2\pi}}\right)^4 e^{-\sum_{i=1}^4 \frac{(x_i-4)^2}{8}}}{\left(\frac{1}{2\sqrt{2\pi}}\right)^4 e^{-\sum_{i=1}^4 \frac{(x_i-6)^2}{8}}} = (\text{simplificando}) = e^{\sum_{i=1}^4 \frac{20-4x_i}{8}} = e^{\sum_{i=1}^4 \frac{5-x_i}{2}} \leq k.$$

Tomando logaritmos: $\sum_{i=1}^4 \frac{5-x_i}{2} \leq \ln k \Leftrightarrow -\sum_{i=1}^4 x_i \leq 2 \ln k - 20 \Leftrightarrow \bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i \geq \frac{10 - \ln k}{2} = \bar{x}_c.$

Bajo la hipótesis nula, la media muestral es normal $N(4, 1)$ y para el nivel de significación requerido:

$$0,05 = P[\bar{X} \geq \bar{x}_c] = P[Z \geq \bar{x}_c - 4] = (\text{tablas}) = P[Z \geq 1,64] \Leftrightarrow \bar{x}_c = 5,64$$

Puesto que $\bar{x} = \frac{1,8 + 6,6 + 2,4 + 3,5}{4} = \frac{14,3}{4} = 3,575$, aceptaremos la hipótesis nula $\mu = 4$.

b) La probabilidad de rechazar H_0 siendo cierta H_1 sería:

$$P[\bar{x} \geq 5,64 / \mu = 6] = P[Z \geq 5,64 - 6] = P[Z \geq -0,36] = (\text{tablas}) = 0,6406$$

Luego la potencia del contraste:

$$\text{Probabilidad de rechazar } H_0 = \begin{cases} 0,05, & \text{si } \mu = 4 \\ 0,6406, & \text{si } \mu = 6 \end{cases}$$