

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

PRIMERA PARTE: CUESTIONES

1.- ¿Cuál es la interpretación del concepto “grados de libertad” a la hora de utilizar estimadores?

Respuesta.-

En una muestra aleatoria de tamaño n , las n variables X_1, X_2, \dots, X_n son independientes. Solemos decir que el conjunto de las n variables $\{X_1, X_2, \dots, X_n\}$ contiene n grados de libertad. Ahora bien es posible que un estimador cualquiera $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ mantenga los n grados de libertad o no. Por ejemplo, si en una población en la que desconocemos la media poblacional, utilizamos como estimador $\hat{\theta} = \sum_{i=1}^n (X_i - \bar{X})^2$ (por ejemplo para hacer estimaciones sobre la varianza) ocurre que hemos perdido un grado de libertad puesto que sabemos que $\sum_{i=1}^n X_i = n\bar{X}$ de donde, conociendo solamente $n-1$ valores de la muestra, podemos despejar el que queda. Así pues $\hat{\theta}$ posee $n-1$ grados de libertad.

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:

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

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 (Sesgo de $\hat{\theta} = 0$) el error cuadrático medio de $\hat{\theta}$ coincide con su varianza.

3.- ¿Porqué se considera más aconsejable la utilización de los estimadores máximo verosímiles frente a los estimadores obtenidos por el método de los momentos? ¿Existe alguna ventaja del método de los momentos frente al de máxima verosimilitud?

Respuesta.-

En general los estimadores máximo verosímiles son más eficientes que los obtenidos por el método de los momentos.

Una ventaja del método de los momentos frente al de máxima verosimilitud es su cálculo más simple.

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.- Un agricultor sabe que el precio de venta de un determinado producto sigue una distribución normal. Toma una muestra aleatoria simple de 6 clientes y tantea el precio de un kilo de dicho producto, obtenido que, según el cliente, puede venderlo a

1'21 €/kg, 1'32 €/kg, 1'14 €/kg, 1'62€/kg, 1'51 €/kg.

Obtenga un intervalo de confianza para el precio medio de venta al nivel de confianza de un 95%.

Solución.-

Como desconocemos la varianza de la población, consideraremos el estadístico $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ cuya distribución es t-Student con $n-1$ grados de libertad.

$$\text{El intervalo de confianza es } \left[\bar{x} - \frac{t_{\alpha/2} s}{\sqrt{n}}, \bar{x} + \frac{t_{\alpha/2} s}{\sqrt{n}} \right].$$

Para un nivel de confianza del 95%, obtenemos de las tablas de la t-Student con 4 grados de libertad que $t_{0,025} = 2,776$. Además $\bar{x} = \frac{1,21 + 1,32 + 1,14 + 1,62 + 1,51}{5} = 1,36$ y

$$s = \sqrt{\frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2} \cong 0,2016$$

Luego el intervalo es:

$$\left[1,36 - \frac{2,776 \cdot 0,2016}{\sqrt{5}}, 1,36 + \frac{2,776 \cdot 0,2016}{\sqrt{5}} \right] \cong [1,11 ; 1,61]$$

2.- Sea una población cuya distribución de probabilidad viene dada por

$$\begin{aligned}P(X=1) &= p^3 \\P(X=2) &= 3p^2q \\P(X=3) &= 3pq^2 \\P(X=4) &= q^3\end{aligned}$$

Siendo $0 \leq p \leq 1$ y $p+q=1$.

Obtener la estimación máximo verosímil del parámetro p utilizando la realización de una muestra aleatoria simple de tamaño 18, en la cual el valor 1 se presenta tres veces, el valor 2 se presenta cuatro veces, el valor 3 se presenta cinco veces y el valor 4 aparece seis veces.

Solución.-

La función de probabilidad: $P[X=x] = \binom{3}{x-1} p^{4-x} (1-p)^{x-1}$, $X \in \{1, 2, 3, 4\}$.

La función de verosimilitud: $L = \prod_{i=1}^{18} \binom{3}{x_i-1} p^{4-x_i} (1-p)^{x_i-1}$. Tomando logaritmos:

$$\ln L = \sum_{i=1}^{18} \left[\ln \binom{3}{x_i-1} + (4-x_i) \ln p + (x_i-1) \ln(1-p) \right]; \text{ derivando respecto de } p \text{ e igualando a}$$

$$\text{cero: } \frac{\partial \ln L}{\partial p} = \sum_{i=1}^{18} \left(\frac{4-x_i}{p} - \frac{x_i-1}{1-p} \right) = 0 \Leftrightarrow \frac{72 - \sum_{i=1}^{18} x_i}{p} = \frac{\sum_{i=1}^{18} x_i - 18}{1-p} \Leftrightarrow 54p = 72 - \sum_{i=1}^{18} x_i.$$

Dividiendo por 18 y despejando, tenemos el estimador: $\hat{p} = \frac{4 - \bar{X}}{3}$

Para la muestra obtenida, calculamos \bar{x} :

x_i	n_i	$x_i \cdot n_i$
1	3	3
2	4	8
3	5	15
4	6	24
	18	50

De donde $\bar{x} = \frac{50}{18}$. Así pues, la estimación sería $\hat{p} = \frac{4 - \frac{25}{9}}{3} = \frac{11}{27} \cong 0,4074$