



**ESTADÍSTICA TEÓRICA II. Tercer curso de Economía CURSO 2004/2005.
FEBRERO 2ª semana. Código de la Carrera 43 Código de la Asignatura 305.**

PRIMERA PARTE: CUESTIONES TEÓRICO-CONCEPTUALES

1. Explique la diferencia entre estimador y estimación. Ponga un ejemplo.

Respuesta.-

En una población consideramos una muestra de tamaño n y sean X_1, X_2, \dots, X_n las n variables muestrales. Sea θ un parámetro muestral. Se llama **estimador** de θ a una cierta función de la muestra:

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

elegida de acuerdo con ciertas propiedades de idoneidad.

Efectuada una realización muestral (x_1, x_2, \dots, x_n) , se llama **estimación**, al valor del estimador para esa realización:

$$\hat{\theta} = g(x_1, x_2, \dots, x_n)$$

Ejemplo: Supongamos una cierta población, una variable aleatoria X y que desconocemos la media poblacional μ , que deseamos estimar. Elegimos un tamaño muestral, por ejemplo, $n = 5$ y utilizamos como estimador la media muestral:

$$\hat{\mu} = \frac{1}{5} \sum_{i=1}^5 X_i$$

Sea por ejemplo $\{3, 5, 8, 3, 7\}$ una realización muestral. Entonces $\frac{3+5+8+3+7}{5} = 5,2$ es una estimación de μ .

2. Explique conceptualmente que mide la potencia de un contraste.

La potencia de un contraste es la probabilidad de rechazar la hipótesis nula H_0 . Si esta es cierta, la potencia del contraste coincide con el error de tipo I, y si H_0 es falsa, la potencia del contraste sería $1 - \beta$, donde β es el error de tipo II, a saber, la probabilidad de aceptar H_0 siendo falsa.

3. ¿Cuándo se utiliza la desigualdad de Chebychev para obtener intervalos de confianza? Razone la respuesta.

Respuesta.-

Cuando se desea un intervalo de confianza para la media y se desconoce la distribución de la población pero se conoce la varianza.

En efecto, para una muestra de tamaño n de una población con media μ desconocida y varianza σ^2 conocida, usando como estimador de μ la media muestral \bar{X} , sabemos que

$$E[\bar{X}] = \mu \text{ y } \text{Var}[\bar{X}] = \frac{\sigma^2}{n}$$

Entonces sustituyendo en la desigualdad de Chebychev $P[|\bar{X} - E(\bar{X})| \leq k] \geq 1 - \frac{\text{Var}(\bar{X})}{k^2}$, queda:

$$P[|\bar{X} - \mu| \leq k] \geq 1 - \frac{\sigma^2}{nk^2} \text{ que equivale a } P[\bar{X} - k \leq \mu \leq \bar{X} + k] \geq 1 - \frac{\sigma^2}{nk^2}.$$

La expresión $1 - \frac{\sigma^2}{nk^2}$ es el coeficiente de confianza $1 - \alpha$, lo que nos permite despejar k , obteniéndose $k = \frac{\sigma}{\sqrt{n\alpha}}$ de forma que el intervalo buscado sería:

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n\alpha}}, \bar{X} + \frac{\sigma}{\sqrt{n\alpha}} \right]$$

4. Explique razonadamente la diferencia entre un contraste de independencia y uno de homogeneidad.

Respuesta.-

En un contraste de independencia se establece la hipótesis nula

H_0 : X e Y son independientes.

donde X e Y son dos características de la población.

Para realizar el contraste se utiliza el estadístico $\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, donde O_{ij} son

las frecuencias observadas de la muestra y E_{ij} son las frecuencias esperadas bajo el supuesto de independencia.

En un contraste de homogeneidad se parte de r muestras independientes clasificadas en s características. La hipótesis nula:

H_0 : $p_{1j} = p_{2j} = p_{3j} = \dots = p_{rj}$, $j = 1, 2, 3, \dots, s$

es decir, la probabilidad de cada categoría es la misma en todas las muestras.

Para el contraste se utiliza el mismo estadístico que en el de independencia.

PROBLEMAS

1.- Un Patronato de Turismo selecciona una muestra aleatoria de los beneficios (pérdidas si son negativos) generados por diferentes establecimientos hoteleros de una zona turística determinada. La información suministrada por la muestra utiliza como datos base los beneficios o pérdidas del último ejercicio cerrado, pendientes de aplicación:

	Establecimiento Hotelero							
	1º	2º	3º	4º	5º	6º	7º	8º
Pérdidas y Ganancias								
31/12/2003	112	82	-62	92	120	87	101	-72

(Datos en miles de euros)

Con el fin de conocer la evolución que el sector turístico ha experimentado en dicha zona y suponiendo que los beneficios (pérdidas) siguen una distribución normal.

- ¿Se puede admitir con un 95% de confianza que la desviación típica de los beneficios (pérdidas) es igual a 40.000 euros?
- ¿Qué sucedería si aumentamos el nivel de confianza a un 99%?. Explicar el resultado obtenido.

Solución.-

a) La variable $\frac{(n-1)S^2}{\sigma^2}$ se distribuye χ^2_{n-1} . En nuestro caso:

$$n = 8$$

La media muestral resulta ser 57,5

La varianza muestral $S^2 = \frac{1}{7} \sum_{i=1}^8 (x_i - 57,5)^2 = 6068,57$ de donde la desviación típica

muestral $S = 77,90$.

De las tablas de la χ^2 con 7 grados de libertad se obtiene:

$$P\left(1,69 \leq \frac{(n-1)S^2}{\sigma^2} \leq 16,01\right) = 0,95. \text{ Sustituyendo tenemos el intervalo } 1,69 \leq \frac{7 \cdot 6068,57}{\sigma^2} \leq 16,01$$

Simplificando se obtiene que $2652,88 \leq \sigma^2 \leq 25138,04 \Leftrightarrow 51,51 \leq \sigma \leq 158,55$ en miles de euros. Por lo tanto no podemos admitir con un 95% de confianza, que la desviación típica sea igual a 40000 euros.

b) Al aumentar el nivel de confianza, aumenta la amplitud del intervalo. En este caso de las tablas se obtiene: $P\left(0,99 \leq \frac{(n-1)S^2}{\sigma^2} \leq 20,28\right) = 0,99$. Sustituyendo tenemos el intervalo

$$0,99 \leq \frac{7 \cdot 6068,57}{\sigma^2} \leq 20,28 \text{ Simplificando se obtiene que } 2094,91 \leq \sigma^2 \leq 42941,38 \Leftrightarrow$$

$\Leftrightarrow 45,77 \leq \sigma \leq 207,22$ en miles de euros. Tampoco podemos admitir con un 99% de confianza, que la desviación típica sea igual a 40000 euros.

2.- Después de las últimas subidas del petróleo, se ha realizado un estudio sobre su influencia en el precio de la gasolina en España. De este estudio se deduce que la mediana del precio del “gasóleo A” está en los 0’90 euros/litro. Se toma una muestra aleatoria de 18 gasolineras de la Comunidad de Madrid, en las que el precio del “gasóleo A” en euros/litro es:

0’89	0’91	0’88	1	0’77	0’85	0’99	0’79	0’98
0’87	1’07	0’74	1’02	0’94	0’83	1’05	0’96	1’04

Se tiene la idea de que el precio del “gasóleo A” en la Comunidad de Madrid está por encima de la mediana del precio, ¿soportan los datos muestrales esta idea al nivel de significación del 5%?

Solución.-

Supuesta la distribución de precios simétrica, efectuaremos un contraste de rangos-signos de Wilcoxon. Sea pues $D_i = X_i - 0,9$ y T^+ la suma de los rangos de las D_i positivas. Como $n = 18$ es suficientemente grande ($n > 15$), podemos considerar que T^+ se distribuye asintóticamente normal con media $E(T^+) = \frac{n(n+1)}{4} = \frac{18 \cdot 19}{4} = 85,5$ y desviación típica

$$DT(T^+) = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{18 \cdot 19 \cdot 37}{24}} \cong 22,96, \text{ por lo que } Z = \frac{T^+ - 85,5}{22,96} \text{ es normal } N(0,1).$$



Siendo Me la mediana del precio del gasóleo en la comunidad de Madrid, el contraste será:

$$H_0: Me > 0,90$$

$$H_1: Me \leq 0,90$$

y la región crítica, para un nivel de significación del 5%,

$$Z \leq -1,645$$

para obtener Z_{exp} , construimos la tabla:

X_i = Precio del gasóleo	$D_i = X_i - 0,9$	Signo	$ D_i $	Rango	Rangos de los $D_i > 0$
0,89	-0,01	-	0,01	1,5	
0,91	0,01	+	0,01	1,5	1,5
0,88	-0,02	-	0,02	3	
0,87	-0,03	-	0,03	4	
0,94	0,04	+	0,04	5	5
0,85	-0,05	-	0,05	6	
0,96	0,06	+	0,06	7	7
0,83	-0,07	-	0,07	8	
0,98	0,08	+	0,08	9	9
0,99	0,09	+	0,09	10	10
1	0,1	+	0,1	11	11
0,79	-0,11	-	0,11	12	
1,02	0,12	+	0,12	13	13
0,77	-0,13	-	0,13	14	
1,04	0,14	+	0,14	15	15
1,05	0,15	+	0,15	16	16
0,74	-0,16	-	0,16	17	
1,07	0,17	+	0,17	18	18
				Suma:	105,5

de donde el valor de $Z_{exp} = \frac{105,5 - 85,5}{22,96} \cong 0,8710 > -1,645$. Luego se acepta la hipótesis nula H_0 .