

CURSO 2007/2008.	JUNIO. 2ª semana
Código de la Carrera 42	Código de la Asignatura 209.

PRIMERA PARTE: CUESTIONES TEÓRICO-CONCEPTUALES

1. ¿Qué diferencia existe entre la covarianza de dos variables aleatorias dependiendo de si son independientes o no?

Respuesta.-

En general dada una variable aleatoria bidimensional (X, Y) se tiene que:

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] = E[XY - E(X)Y - XE(Y) + E(X)E(Y)] = \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y) \end{aligned}$$

Si X e Y son independientes, entonces $E(XY) = E(X) \cdot E(Y)$, luego $\text{Cov}(X, Y) = 0$.

Si X e Y no son independientes, entonces la covarianza ya no da necesariamente cero, aunque también podría anularse.

2. ¿Qué relación existe entre la función de distribución empírica de una muestra y la función de distribución de la variable aleatoria de donde se ha extraído dicha muestra?

Respuesta.-

La función de distribución empírica de una muestra observada $\{x_1, x_2, \dots, x_n\}$ de una variable aleatoria X , es la función $F_n(x) = \frac{N(x)}{n}$, siendo $N(x)$ el número de valores de la muestra que son menores o iguales que x , mientras que la función de distribución $F(x) = P[X \leq x]$.

El teorema de Glivenko-Canteli establece que la función de distribución empírica converge en probabilidad a la función de distribución $F(x)$, cuando $n \rightarrow \infty$, es decir, dado $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P \left[\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| < \varepsilon \right] = 0.$$

3. ¿Qué es y para qué se usa el error cuadrático medio de un estimador?

Respuesta.-

Se denomina error cuadrático medio de un estimador $\hat{\theta}$ de un parámetro θ al valor de $E[(\hat{\theta} - \theta)^2]$. Un estimador será tanto mejor cuanto menor sea su error cuadrático medio.

Si lo desarrollamos se obtiene que $E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}))^2] + (E(\hat{\theta}) - \theta)^2 = \text{Var}(\hat{\theta}) + \text{Sesgo}^2(\hat{\theta})$. Por tanto, si el sesgo del estimador es nulo (el estimador es insesgado), entonces su error cuadrático medio coincide con su varianza.

4. ¿De qué depende el valor de la potencia de un contraste?

Respuesta.-

Si $\beta = P[\text{aceptar } H_0/H_0 \text{ falsa}]$ es la probabilidad del error de tipo II, entonces la potencia del contraste es $1 - \beta$. Si α es el tamaño de la región crítica, entonces la potencia del contraste depende del valor de α y del tamaño n de la muestra:

- para n fijo, si α disminuye, $1 - \beta$ disminuye y viceversa, si $1 - \beta$ aumenta, α aumenta
- si n aumenta, entonces es posible que α disminuya y $1 - \beta$ aumente.

PROBLEMAS

1.- El número de veces al mes que, por diferentes motivos, suena la alarma de una fábrica, sigue una distribución de Poisson. Se ha observado que la alarma no se activa el 13,53% de los meses.

a) ¿Cuál es el número medio de veces al mes en que suena la alarma de la mencionada fábrica?

b) ¿Cuál será la probabilidad de que la alarma suene 5 veces en tres meses? ¿y de que suene más de 18 veces al año?

c) Si el mantenimiento de la alarma tiene un coste fijo de 30 euros al mes y cada vez que se activa y acude el servicio de emergencia supone un pago de 180 euros, ¿cambiaría la fábrica su sistema de seguridad por el de otra empresa que le pide únicamente un coste fijo mensual de 300 euros?

Solución.-

Sea X la variable “nº de veces que la alarma suena al mes”. Que no se activa el 13,53% de las veces quiere decir que, para un mes cualquiera, $P[X = 0] = 0,1353$. Ahora bien, como

$P[X = x] = \frac{\lambda^x}{x!} e^{-\lambda} \rightarrow P[X = 0] = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-\lambda} = 0,1353$. Tomando logaritmos neperianos queda que $-\lambda = -2 \leftrightarrow \lambda = 2$. Luego:

a) El número medio de veces al mes en que suena la alarma es 2.

b) La variable aleatoria Y = “número de veces que la alarma sonará en tres meses” será una variable aleatoria de Poisson, de parámetro $3\lambda = 6$. Luego $P[Y = 5] = (\text{tablas}) = 0,1606$

La variable aleatoria A = “número de veces que la alarma sonará en un año” será una variable aleatoria de Poisson, de parámetro $12\lambda = 24$. Para calcular $P[A > 18]$ aproximaremos por la normal $N(24, \sqrt{24})$. Luego $P(A > 18) = (\text{corrección por continuidad}) = P(A \geq 18,5) =$
 $= (\text{tipificando}) = P\left(Z \geq \frac{18,5 - 24}{\sqrt{24}}\right) = P(Z \geq -1,12) = (\text{tablas}) = 1 - 0,1314 = 0,8686$

c) Puesto que el número medio de veces que se dispara la alarma al mes es 2, el coste medio mensual sería de $30 + 2 \cdot 180 = 390$ €. Por tanto debería cambiarse de empresa de seguridad.

2.- Los niveles de audiencia (en miles de personas) de un programa de televisión, medidos en 10 emisiones elegidas aleatoriamente, han sido los siguientes:

682, 553, 555, 666, 657, 649, 522, 568, 700, 552

Suponiendo que los niveles de audiencia siguen una distribución normal:

a) ¿Se podría afirmar, con un 95% de confianza, que la audiencia media del programa es de 600.000 espectadores?

b) La compañía productora del programa televisivo afirmó, durante las negociaciones para la venta del programa, que éste acapararía una audiencia fiel y que la desviación típica del número de espectadores sería de 15.000. ¿Se puede dar como cierta dicha afirmación con un 95% de confianza?

Solución.-

Para la muestra obtenida, la media muestral es $\bar{x} = 610,4$ y la varianza muestral

$$s^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 \cong 4366,04 \text{ de donde } s \cong 66,08. \text{ Luego:}$$

a) La variable $\frac{\bar{X} - \mu}{S} \sqrt{n}$ se distribuye t-Student con n-1 grados de libertad. Entonces, para 9 grados de libertad obtenemos de las tablas el intervalo del 0,95 de probabilidad:

$$[-2,262 ; 2,262]$$

luego, con un 95% de confianza podemos afirmar que

$$-2,262 < \frac{610,4 - \mu}{66,08} \sqrt{10} < 2,262 \quad \Leftrightarrow \quad 610,4 - \frac{2,262 \cdot 66,08}{\sqrt{10}} < \mu < 610,4 + \frac{2,262 \cdot 66,08}{\sqrt{10}} \quad \Leftrightarrow$$

$$\Leftrightarrow 563,13 < \mu < 657,67$$

Así pues, 600000 sería un valor admisible para la audiencia media del programa, con un 95% de confianza.

b) La variable $\frac{(n-1)S^2}{\sigma^2}$ se distribuye χ^2 con n-1 grados de libertad. Entonces, para 9 grados de libertad obtenemos de las tablas el intervalo del 0,95 de probabilidad:

$$[2,7 ; 19,02]$$

luego, con un 95% de confianza podemos afirmar que:

$$2,7 < \frac{9 \cdot 4366,04}{\sigma^2} < 19,02 \quad \Leftrightarrow \quad \frac{9 \cdot 4366,04}{19,02} < \sigma^2 < \frac{9 \cdot 4366,04}{2,7} \quad \Leftrightarrow \quad 2065,95 < \sigma^2 < 14553,48 \quad \Leftrightarrow$$

$\Leftrightarrow 45,45 < \sigma < 120,64$. es decir la desviación típica, con un 95% de confianza, estaría entre 45450 y 120640 espectadores. Luego no se puede dar por cierta la afirmación del problema.