Curso de Estadística

Examen final (Solución) 3-7-2013

☞ Completa esto antes de empezar

➤ **Nº de matrícula:** ☐☐☐☐☐

➤ **Apellidos:** ....

➤ **Nombre:**..

❶ – Ejercicio

a) En una distribución bidimensional la varianza para la variable X es $3'75$ y la varianza para la variable Y es $7'40$. ¿Puede ser el coeficiente de regresión de Y sobre X $b = 1'44$? (1 punto)

b) Las medias en una distribución bidimensional son $M_x = 16$ $M_y = 12$, y sus varianzas verifican $5\sigma_x = 4\sigma_y$, siendo el coeficiente de correlación $r = 0'75$. ¿Qué valor se puede esperar para la variable x cuando $y = 4$? (1 punto)

Solución:

a) Es conocido por teoría

$$b = \frac{\sigma_{xy}}{\sigma_x^2} \Longrightarrow \sigma_{xy} = b\sigma_x^2 = 1'44 \cdot 3'75 = 5'4$$

de esta forma, el coeficiente de correlación será

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{5'40}{\sqrt{3'75}\sqrt{7'40}} = \frac{5'4}{5'27} = 1'025 \qquad \textbf{absurdo}.$$

b) Se ha visto en teoría:

$$r_{xy} \equiv x - M_x = \frac{\sigma_{xy}}{\sigma_y^2}(y - M_y)$$

Puesto que

$$r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} = \frac{3}{4} \Longrightarrow \sigma_{xy} = \frac{3}{4}\sigma_x\sigma_y \quad \text{además} \quad \sigma_x = \frac{4}{5}\sigma_y$$

reulta

$$\sigma_{xy} = \frac{3}{4}\sigma_y^2$$

Así pues,

$$b' = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{\frac{3}{5}\sigma_y^2}{\sigma_y^2} = \frac{3}{5}$$

De esta forma

$$r_{xy} \equiv x - 16 = \frac{3}{5}(y - 12)$$

Para $y = 4$

$$x = 16 + \frac{3}{5}(4 - 12) = 11'2$$

⊠

❷ – **Ejercicio**

Una máquina envasa azúcar en bolsas de peso teórico de 1 kilogramo. El peso real en kilos de las bolsas es una v.a. X que sigue una distribución $N(1, \sigma = 0'025)$. Estas bolsas se distribuyen en cajas de 30 unidades cada una. Una bolsa se considera aceptable si su peso está entre 950 y 1050 gramos.

a) ¿Cuál es la probabilidad de qué una bolsa elegida al azar sea aceptable? (1 punto)

b) ¿Cuál es el porcentaje de cajas que tienen más de cuatro bolsas no aceptables? (2 puntos)

c) A cada caja se le asigna un valor Y que depende del número de bolsas no aceptables que contenga, de modo que

$$Y = \begin{cases} 0 & \text{si el n}^{\text{o}} \text{ de bolsas no aceptables es mayor que 4} \\ 20 & \text{si el n}^{\text{o}} \text{ de bolsas no aceptables es menor que 2} \\ 10 & \text{resto} \end{cases}$$

Se pide la función de probabilidad, función de distribución y media de esta v.a. (3 puntos)

Solución:

a) Sea la v.a X =peso de una bolsa con $X \equiv N(1, \sigma = 0'025)$
$P(\text{bolsa aceptable}) = P(0'95 < X < 1'05) = P\left(\dfrac{0'95-1}{0'025} < \dfrac{X-1}{0'025} < \dfrac{1'05-1}{0'025}\right) =$
$= P(-2 < N(0,1) < 2) = 0'9545$

b) $P(\text{bolsa no aceptable}) = 1 - 0'9545 \approx 0'04$.
Sea la v.a T =número de bolsas no aceptables en una caja con $T \equiv B(30, 0'04)$
$P(\text{caja con más de 4 bolsas no aceptables}) = P(T > 4) \approx P(\mathcal{P}(1'2) > 4)$
pues $T \approx \mathcal{P}(30 \times 0'04)$
$P(\mathcal{P}(1'2) > 4) = 1 - 0'992 = 0'008$.
Por tanto, el $0'8\,\%$ de las cajas tienen más de 4 bolsas no aceptables.

c) $Y = \begin{cases} 0 & T > 4 \\ 10 & 2 \leq T \leq 4 \\ 20 & T < 2 \end{cases}$
$P(Y = 20) = P(T < 2) \approx P(\mathcal{P}(1'2) < 2) = P(\mathcal{P}(1'2) \leq 1) = 0'663$
$P(Y = 0) = P(T > 4) = 0'008$
$P(Y = 10) = 1 - (P(Y = 20) + P(Y = 10)) = 0'329$.
La función de distribución de la v.a. Y será

$$F(y) = \begin{cases} 0 & y < 0 \\ 0'008 & 0 \leq y < 10 \\ 0'337 & 10 \leq y < 20 \\ 1 & 20 \leq y \end{cases}$$

$E(Y) = \sum kP(Y = k) = 10P(Y = 10) + 20P(Y = 20) = 16'55$

⊠

❸ – **Ejercicio**

Se ha estudiado el beneficio anual (pérdida en el caso de valores negativos) de las empresas de una determinada localidad, y se ha caracterizado por una distribución poblacional Normal, con dos millones de euros de desviación típica. Si se elige una *m.a.s.* de 25 empresas con $\overline{x} = 0'5$ millones, determina el intervalo de confianza del $90\,\%$ y del $95\,\%$ para el beneficio medio anual de las empresas de la localidad. ¿Cuál de los dos intervalos es mayor? ¿por qué? Justifica la respuesta. (2 puntos)

Solución:

A la vista del enunciado del problema se define la v.a. y se establece su distribución poblacional

$$X \equiv \textit{Beneficio o pérdida anual de las empresas} \rightsquigarrow N(\mu, 2)$$

Hay que estimar el valor de μ mediante intervalos de confianza. Se parte de una *m.a.s.* de $n = 25$ empresas. Puesto que las distribuciones poblacionales de la *m.a.s.* son Normales se considera como *cantidad pivotal*

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

lo que nos lleva a considerar un intervalo de confianza para μ de la forma

$$\left(\overline{x} - \frac{\sigma}{\sqrt{n}}\, z_{\alpha/2},\ \overline{x} + \frac{\sigma}{\sqrt{n}}\, z_{\alpha/2}\right)$$

a) Para un nivel del confianza del 90 % resulta

$$P(Z > z_{\alpha/2}) = \alpha/2 \Longrightarrow P(Z > z_{0'05}) = 0'05 \iff P(Z \leqslant z_{0'05}) = 0'95 \Longrightarrow z_{0'05} = 1'64$$

con lo cual el intervalo de confianza queda

$$I_{90} = \left(0'5 - 1'64 \cdot \frac{2}{\sqrt{25}}, 0'5 + 1'64 \cdot \frac{2}{\sqrt{25}}\right) = (-0'158, 1'158)$$

b) Para un nivel del confianza del 95 % resulta

$$P(Z > z_{\alpha/2}) = \alpha/2 \Longrightarrow P(Z > z_{0'025}) = 0'025 \iff P(Z \leqslant z_{0'025}) = 0'975 \Longrightarrow z_{0'025} = 1'96$$

con lo cual el intervalo de confianza queda

$$I_{95} = \left(0'5 - 1'96 \cdot \frac{2}{\sqrt{25}}, 0'5 + 1'96 \cdot \frac{2}{\sqrt{25}}\right) = (-0'284, 1'284)$$

Es inmediato darse cuenta que

$$I_{90} \subset I_{95}$$

Por tanto al aumentar el nivel de confianza aumenta el intervalo de confianza y disminuye la precisión. Esto es así, como consecuencia de la expresión

$$P\left(|\overline{X} - \mu| \leqslant z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

En cierta manera, la precisión se puede relacionar con la cantidad $|\overline{X} - \mu|$. Si se quiere que la precisión sea elevada entonces $|\overline{X} - \mu|$ debe ser pequeña. Si la confianza se impone grande, entonces el valor $z_{\alpha/2}$ es grande y la cota $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ es grande. Como consecuencia, el intervalo de confianza es grande y disminuye la precisión.

$\boxtimes$