Inferencia Estadística II

Junio de 2003

Problema 1 (3 puntos). En la tabla siguiente aparecen datos sobre el nivel sonoro X (expresado en decibelios) en diez fábricas mexicanas de cementos y la frecuencia media de casos de sordera traumática Y entre los trabajadores en las respectivas fabricas

X	85	65	62	50	68	52	40	43	47	58
Y	0.56	0.50	0.49	0.45	0.53	0.43	0.20	0.23	0.24	0.55

1. Representar los datos en un diagrama de dispersión.

2. Calcular el coeficiente de correlación lineal entre Y y X.

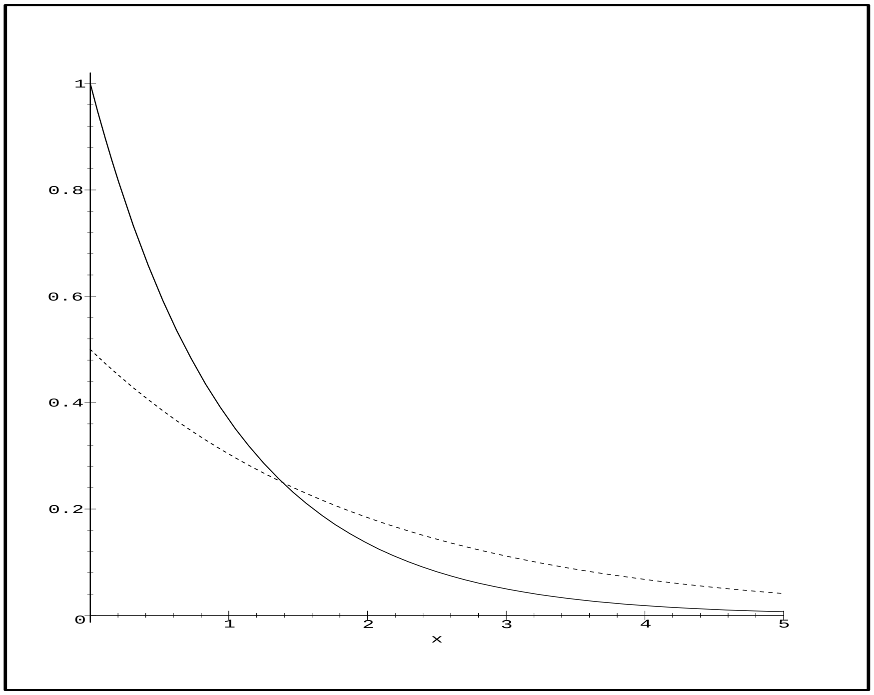
3. Contrastar con nivel de significación $\alpha = 5\%$ la asociación entre X e Y mediante el coeficiente de Spearman. Interpretar.

Problema II (3 puntos). Ciertos detectores de mentiras están basados en la presión sanguínea en el rostro del individuo. A partir de varios estudios se concluyó que la presión sanguínea X se distribuye según una distribución exponencial, con media 2 si el individuo miente y con media 1 si el individuo dice la verdad.

1. Hallar el test UMP basado en una observación de X para contrastar (con nivel $\alpha = 5\%$) las hipótesis H_0 :"El individuo dice la verdad" frente a H_1 :"El individuo miente".

2. Indicar en el gráfico que se muestra a continuación los riesgos α y β (errores de tipo I y II). Interpretar estos riesgos en el contexto anteriormente descrito.

3. Calcular la potencia del test ¿Qué podemos concluir?



Problema III (4 puntos). El Talgo Barcelona-Cáceres suele tener retraso. Este retraso, que denotamos θ, varía todos los días y sus variaciones dependen de factores que desconocemos. Sin embargo, un estudio empírico permitió poner en evidencia que θ (expresado en minutos) tiene una distribución (*a priori)* exponencial con parámetro $\lambda = 0.1$ (la media es entonces $1/\lambda = 10$). El objetivo es aquí predecir el retraso del Talgo y se escoge para medir el error de una predicción d de θ, la función cuadrática: $w(d, \theta) = (d - \theta)^2$.

1. Hallar la mejor predicción de θ mediante su distribución *a priori*. Dar el riesgo (o error) de esta predicción.

2. Con el fin de mejorar la predicción anterior, se pide a Renfe información acerca de las causas del retraso. La empresa de transporte público esgrime problemas relacionados con la adaptación de las vías al tren de alta velocidad. Esta adaptación requiere la presencia de obreros en varios puntos de la vía y obliga al Talgo a frenar en dichos puntos. De

hecho, se comprobó estadísticamente que el número X de puntos de obra en la vía sigue condicionalmente al retraso θ, una distribución de Poisson con media θ.

(a) Demostrar que la distribución de θ condicionada por X es la Gamma
$$G\left(X+1,\ \lambda+1\right).$$

(b) Deducir la mejor predicción de θ basada en la distribución *a priori* y la información muestral X. Expresar esta predicción como una combinación de X y de la media *a priori* de θ. Interpretar.

(c) Demostrar que el riesgo bayesiano medio de la predicción anterior es
$$W_Q\left(d_Q(X)\right) = \frac{1}{\lambda\left(\lambda+1\right)}$$
Comparar este riesgo con el obtenido en el apartado 1. Justificar.

Indicaciones:

- La función de densidad de una distribución gamma $G(a,b)$ se puede escribir como
$$f(x) = \frac{x^{a-1}\exp(-bx)}{C}$$
para $x > 0$ y 0 en caso contrario, donde C es una constante que no depende de x.
La media de esta distribución es $\frac{a}{b}$ y su varianza es $\frac{a}{b^2}$.

- La distribución exponencial con parámetro λ es una Gamma $G(1,\lambda)$, su función de densidad es $f(x) = \lambda\exp(-\lambda x)$ para $x > 0$.

- Una variable X sigue una distribución de Poisson con media θ si $P(X = x) = \frac{\theta^x}{x!}\exp(-\theta)$ para cada $x = 0, 1, 2, \ldots$

- La tabla del coeficiente de Spearman (para un tamaño muestral $n = 10$) indica que la problabilidad $P(|r_S| > 0.648) = 5\%$.

Duración del examen: 2h30min.

Publicación de las notas: Viernes 6 a partir de las 16h.
Revisión: Martes 10 a las 11h (despacho 10.0.09).