Examen de Estadística I. 22-06-2005.

1. X, valor anual del activo en millones de euros.
Y, importe anual de la deuda total en millones de euros.
$\overline{x} = 11.38$, $\overline{y} = 7.93$, $s_x^2 = 34.63$, $s_y^2 = 22.06$, $s_{xy} = 24.69$

a) $y^* = a + bx$
$b = \frac{s_{xy}}{s_x^2} = \frac{24.69}{34.63} = 0.713$, $a = \overline{y} - b\overline{x} = 7.93 - \frac{24.69}{34.63}(11.38) = -0.1835$
$\underline{y^* = -0.1835 + 0.713\,x}$
$R^2 = r_{xy}^2 = \frac{s_{xy}^2}{s_x^2\, s_y^2} = \frac{(24.69)^2}{(34.63)(22.06)} = \underline{0.798}$, coeficiente que por su proximidad a uno indica que el ajuste lineal realizado es razonable.

b) Z_1, deuda bancaria en millones de euros.
Z_2, deuda no bancaria en millones de euros.
$Y = Z_1 + Z_2$, deuda total en millones de euros.
$\overline{z}_1 = 1.7$, $s_{z_1}^2 = 1.63$, $s_{z_1 z_2} = -3.09$
$\overline{z}_2 = \overline{y} - \overline{z}_1 = 7.93 - 1.7 = \underline{6.23}$, $s_{z_2}^2 = s_y^2 - s_{z_1}^2 - 2s_{z_1 z_2} = 22.06 - 1.63 - (2)(-3.09) = \underline{26.61}$

c) No sería adecuado realizar un ajuste lineal de la deuda total en función del inmovilizado material dado que no existe relación lineal entre ambas variables al ser $R^2 = 0$. No podemos asegurar que las dos variables sean estadísticamente independientes, tan sólo que están incorreladas, que es una condición más débil.

2. Y_t, ingresos anuales en el período t en miles de euros.
$y_t^* = 60 + 12\,t$ (t, años; origen, 2002).

a) El índice de precios de Paasche del año 2005 con base el año 2002 puede ser obtenido mediante una sencilla regla de tres, $P_{05}^{02} = \frac{P_{05}^{?} P_{02}^{02}}{P_{02}^{?}} = \frac{(140)(100)}{110} = 127.\widehat{27}$, luego la previsión de los ingresos del año 2005 en euros constantes de 2002 es $\hat{y}_3 = \frac{y_3^*}{P_{02}^{05}/100} = \frac{60+(12)(3)}{1.\widehat{27}} = \underline{75.4286}$.

b) $IVE(1T) = 0.9$, $IVE(2T) = 1.1$ e $IVE(4T) = 0.8$
$IVE(3T) = 4 - IVE(1T) - IVE(2T) - IVE(4T) = 4 - 0.9 - 1.1 - 0.8 = 1.2$
$y_t^* = \frac{60}{4} + \frac{12}{4^2}\,t = 15 + 0.75t$ (t, trimestres; origen, trimestre central 2002).
Del trimestre central de 2002 al tercero de 2006 hay $(4)(4) + 0.5 = 16.5$ trimestres.
$y_{16.5}^* = 15 + (0.75)(16.5) = 27.375$
$\hat{y}_{3T06} = y_{16.5}^*\, IVE(3T) = \underline{32.85}$

3. X, horas perdidas por los trabajadores de A que estuvieron de baja.
Y, horas perdidas por los trabajadores de B que estuvieron de baja.
$\overline{x} = 130$, $\overline{y} = 90$, $s_x = 60$, $s_y = 45$

a) $g_0(X) = \frac{s_x}{\overline{x}} = \frac{60}{130} = \underline{0.4615 < 0.5} = \frac{45}{90} = \frac{s_y}{\overline{y}} = g_0(Y)$, es decir, en B hubo una menor regularidad que en A en el número de horas perdidas por enfermedad y, por consiguiente, la media en A es más representativa que en B.

b) $x_0 = 80, y_0 = 70$

$\frac{x_0-\overline{x}}{s_x} = \frac{80-130}{60} = \underline{-0.8\hat{3} < -0.\hat{4}} = \frac{70-90}{45} = \frac{y_0-\overline{y}}{s_y}$, es decir, <u>en términos relativos perdió más horas el trabajador de la empresa B</u>.

4. X, tiempo de trayecto Valencia-Madrid, en horas.

$X \sim Un(a = 2.5, b = 4)$

a) $f(x) = \frac{1}{b-a} = \frac{1}{4-2.5} = \underline{\frac{2}{3},\ x \in [2.5, 4]}$

b) $p(X < 3) = \int_{2.5}^{3} f(x)\,dx = \int_{2.5}^{3} \frac{2}{3}\,dx = \underline{\frac{1}{3}}$

c) $Y = 6X + 12$, importe del trayecto, en euros.

$p(Y > 30) = p(6X + 12 > 30) = p(X > 3) = 1 - p(X \leq 3) = 1 - \frac{1}{3} = \underline{\frac{2}{3}}$

d) Z, número de trayectos semanales, de entre 6, con tiempo mayor que la media.

$Z \sim Bi(6, \theta),\ f(z) = \begin{pmatrix} 6 \\ z \end{pmatrix} \theta^z (1-\theta)^{6-z},\ z = 0, 1, 2, 3, 4, 5, 6$

$\theta = p(X > \mu_x) = 0.5$, pues la distribución Uniforme es simétrica respecto a la media.

$p(Z = 2) = \begin{pmatrix} 6 \\ 2 \end{pmatrix} (0.5)^2 (0.5)^4 = \underline{0.2344}$

5. X, ventas mensuales del comercial A en miles de euros.

Y, ventas mensuales del comercial B en miles de euros.

$\mu_x = 300,\ \sigma_x = 20,\ \mu_y = 310,\ \sigma_y = 50$

a) $X \sim N(\mu_x, \sigma_x^2),\ Y \sim N(\mu_y, \sigma_y^2)$

Considerando <u>independencia entre las variables X e Y</u>, esto es, que las ventas del comercial A nada tienen que ver con las de B, entonces toda combinación lineal de ambas es Normal, en particular, $W = X - 1.4Y \sim N(\mu_w = \mu_x - 1.4\,\mu_y = 300 - (1.4)(310) = -130, \sigma_w^2 = \sigma_x^2 + (1.4)^2\sigma_y^2 = (20)^2 + (1.4)^2(50)^2 = 5300)$.

$p(X > 1.4Y) = p(X - 1.4Y > 0) = p\left(Z > \frac{0-(-130)}{\sqrt{5300}}\right) = p(Z > 1.7857) = \underline{0.0371}$, donde $Z = \frac{W-\mu_w}{\sigma_w} \sim N(0, 1)$

b) <u>No, dado que no se puede asegurar que W sea Normal</u>. Si las variables están correladas entonces ya no es de aplicación el resultado anterior y no necesariamente la combinación lineal de variables Normales sigue una distribución Normal.

c)

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 300 \\ 310 \end{pmatrix}, \begin{pmatrix} 20^2 & 400 \\ 400 & 50^2 \end{pmatrix}\right)$$

$T = 0.0125(X + Y)$, comisión conjunta en miles de euros.

$T \sim N(\mu_t = 0.0125(\mu_x + \mu_y) = 0.0125(300 + 310) = 7.625, \sigma_t^2 = (0.0125)^2(\sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}) = (0.0125)^2(20^2 + 50^2 + (2)(400)) = 0.5781)$

$p(T > k) = 0.75 \Leftrightarrow p\left(\frac{T-\mu_t}{\sigma_t} > \frac{k-7.625}{\sqrt{0.5781}}\right) = 0.75 \Leftrightarrow p\left(Z > \frac{k-7.625}{\sqrt{0.5781}}\right) = 0.75 \Leftrightarrow \frac{k-7.625}{\sqrt{0.5781}} = -0.675 \Leftrightarrow k = \underline{7.1118}$, donde $Z = \frac{T-\mu_t}{\sigma_t} \sim N(0, 1)$