Examen de Estadística I. 14-06-2004.

1. X, superficie en metros cuadrados.
Y, valor catastral en miles de euros.

	x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
	100	24	10000	576	2400
	150	36	22500	1296	5400
	120	30	14400	900	3600
	80	20	6400	400	1600
	90	20	8100	400	1800
$\sum_{i=1}^{5}$	540	130	61400	3572	14800
$\frac{1}{5}\sum_{i=1}^{5}$	108	26	12280	714.4	2960

a) $y^* = a + bx$
$b = \frac{s_{xy}}{s_x^2} = \frac{2960-(108)(26)}{12280-(108)^2} = \frac{152}{616} = 0.2468$, $a = \overline{y} - b\overline{x} = 26 - \frac{152}{616}(108) = -0.6494$
$\underline{y^* = -0.6494 + 0.2468\,x}$

b) $y_0 = 25$
$x^* = a' + b'y$
$b' = \frac{s_{xy}}{s_y^2} = \frac{2960-(108)(26)}{714.4-(26)^2} = \frac{152}{38.4} = 3.958\widehat{3}$, $a' = \overline{x} - b'\overline{y} = 108 - (3.958\widehat{3})(26) = 5.08\widehat{3}$
$x^* = 5.08\widehat{3} + 3.958\widehat{3}\,y$
$x_0^* = a' + b'y_0 = 5.08\widehat{3} + (3.958\widehat{3})(25) = \underline{104.041\widehat{6}}$

c) Las dos rectas de regresión sólo pueden ser coincidentes en el caso de que la relación lineal sea perfecta, esto es, sólo cuando $R^2 = 1$.

d) $x_0 = 200, y_0 = 45$
$\frac{x_0-\overline{x}}{s_x} = \frac{200-108}{\sqrt{616}} = \underline{3.7068 > 3.0661} = \frac{45-26}{\sqrt{38.4}} = \frac{y_0-\overline{y}}{s_y}$, es decir, <u>en términos relativos es mayor la superficie que el valor catastral</u>.

2. Y_t, ventas anuales en el período t en millones de euros.
$y_t^* = 20 + 16\,t$ (t, años; origen, 2000).
$s_y^2 = 520,\ s_{ty} = 32$

a) $y_4^* = 20 + (16)(4) = \underline{84}$ millones de euros.
$R^2 = r_{ty}^2 = \frac{s_{ty}^2}{s_t^2\,s_y^2} = \frac{b\,s_{ty}}{s_y^2} = \frac{(16)(32)}{(520)} = \underline{0.9846}$, coeficiente que por su gran proximidad a uno indica que la predicción realizada utilizando la recta de regresión lineal mínimo cuadrática presenta una <u>muy elevada fiabilidad</u>.

b) $y_1^* = 20 + (16)(1) = 36$, $y_3^* = 20 + (16)(3) = 68$
$\hat{y}_1 = y_1^*\,\frac{100}{I_{2001}} = \frac{36}{1.05} = 34.2857$, $\hat{y}_3 = y_3^*\,\frac{100}{I_{2003}} = \frac{68}{1.3} = 52.3077$
$\frac{\hat{y}_3-\hat{y}_1}{\hat{y}_1}100 = \underline{52.5641\%}$

c) $IVE(1T) = 0.9$, $IVE(3T) = 0.9$ e $IVE(4T) = 1$

$IVE(2T) = 4 - IVE(1T) - IVE(3T) - IVE(4T) = 4 - 0.9 - 0.9 - 1 = 1.2$
$y_t^* = \frac{20}{4} + \frac{16}{4^2}\, t = 5 + t$ (t, trimestres; origen, trimestre central 2000).
Del trimestre central de 2000 al segundo de 2005 hay $(5)(4) - 0.5 = 19.5$ trimestres.
$y_{19.5}^* = 5 + 19.5 = 24.5$
$\hat{y}_{2T05} = y_{19.5}^* \, IVE(2T) = \underline{29.4}$

3. X, número de corredores atendidos por los servicios médicos.
$X \sim Po(\lambda),\ f(x) = \exp(-\lambda)\frac{\lambda^x}{x!},\ x \in \mathbb{N}$

a) $\mu = \lambda = 5$
$p(X > 2) = 1 - p(X \leq 2) = 1 - \exp(-5)\left[\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!}\right] = 1 - \exp(-5)\left[1 + 5 + \frac{25}{2}\right] = \underline{0.8753}$

b) Y, tiempo en media maratón en minutos.
$Y \sim Un(90, 120),\ f(y) = \frac{1}{120-90} = \frac{1}{30},\ y \in [90, 120]$
$p(110 \leq Y \leq 115) = \int_{110}^{115} \frac{1}{30}\, dy = \underline{0.1\widehat{6}}$

c) Z, número de corredores, de entre los 4, con tiempo en carrera entre 1h. 35 y 1h. 40.
$Z \sim Bi(4, \theta),\ f(z) = \left(\begin{array}{c} 4 \\ z \end{array}\right) \theta^z (1-\theta)^{4-z},\ z = 0, 1, 2, 3, 4$
$\theta = p(95 \leq Y \leq 100) = p(110 \leq Y \leq 115) = 0.1\widehat{6}$
$p(Z \geq 3) = p(Z = 3) + p(Z = 4) = \left(\begin{array}{c} 4 \\ 3 \end{array}\right)\left(0.1\widehat{6}\right)^3\left(0.8\widehat{3}\right)^1 + \left(\begin{array}{c} 4 \\ 4 \end{array}\right)\left(0.1\widehat{6}\right)^4\left(0.8\widehat{3}\right)^0 =$
$(4)(4.63 \times 10^{-3})\left(0.8\widehat{3}\right) + (1)(7.72 \times 10^{-4})(1) = \underline{0.0162}$

4. X, número mensual de consultas de saldo.
Y, número mensual de transferencias realizadas.
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 2100 \\ 2000 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & 3600 \end{pmatrix}\right)$$

a) $p(X > 1900) = 0.99$
$p\left(Z > \frac{1900-2100}{\sigma_x}\right) = 0.99$, donde $Z = \frac{X-\mu_x}{\sigma_x} \sim N(0,1)$, entonces $\frac{1900-2100}{\sigma_x} = -2.33 \Rightarrow \sigma_x = 85.8369$
$\gamma_0(X) = \frac{\sigma_x}{\mu_x} = \frac{85.8369}{2100} = \underline{0.0409 > 0.03} = \frac{60}{2000} = \frac{\sigma_y}{\mu_y} = \gamma_0(Y)$, es decir, el servicio de transferencias presenta una mayor regularidad que el de consultas de saldo.

b) $T = X - Y \sim N(\mu_t = \mu_x - \mu_y = 2100 - 2000 = 100, \sigma_t^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} = 85.8369^2 + 3600 - (2)(0) = 10967.9751)$
$p(X > Y) = p(T > 0) = p(Z > frac0 - 100\sqrt{10967.9751}) = p(Z > -0.9549) = p(Z < 0.9549) = \underline{0.8302}$, donde $Z = \frac{T-\mu_t}{\sigma_t} \sim N(0,1)$

c) No, porque U_1, U_2, U_3 y U_4 no son categorías mutuamente excluyentes. Un mismo individuo puede pertenecer a más de una categoría, por ejemplo, un mismo cliente puede realizar mensualmente consultas de saldo, transferencias bancarias, gestiones de operativa bursátil e incluso utilizar otros servicios de la web de la Caja de Ahorros.