Examen de Estadística I. 06-09-2004.

1. X, cuota íntegra en miles de euros.
Y, edad en años.

	x_i	y_i	x_i^2	y_i^2	$x_i\, y_i$
	28	24	784	576	672
	32	32	1024	1024	1024
	58	45	3364	2025	2610
	30	40	900	1600	1200
	43	54	1849	2916	2322
$\sum_{i=1}^{5}$	191	195	7921	8141	7828
$\frac{1}{5}\sum_{i=1}^{5}$	38.2	39	1584.2	1628.2	1565.6

a) $x^* = a' + b'y$
$b' = \frac{s_{xy}}{s_y^2} = \frac{1565.6-(38.2)(39)}{1628.2-(39)^2} = \frac{75.8}{107.2} = 0.7071,\ a' = \overline{x} - b'\overline{y} = 38.2 - \frac{75.8}{107.2}(39) = 10.6235$
$\underline{x^* = 10.6235 + 0.7071\, y}$
$y_0 = 35$
$x_0^* = a' + b'y_0 = 10.6235 + 0.7071(35) = \underline{35.3716}$

b) $g_0(X) = \frac{s_x}{\overline{x}} = \frac{\sqrt{1584.2-(38.2)^2}}{38.2} = \frac{11.1786}{38.2} = \underline{0.2926 > 0.2655} = \frac{10.3537}{39} = \frac{\sqrt{107.2}}{39} = \frac{s_y}{\overline{y}} = g_0(Y)$, es decir, la edad presenta una mayor regularidad que la cuota íntegra.

c) El índice de Gini es un indicador numérico del grado de concentración de una distribución.

$[L_{i-1}, L_i[$	p_i	q_i
$[0, 20[$	60	50
$[20, 60[$	90	90
$[60, +\infty[$	100	100

$$I_G = \frac{\sum_{i=1}^{I}(p_i - q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{(60-50)+(90-90)+(100-100)}{60+90} = \underline{0.0\widehat{6}}.$$

2. El IPC general del año 1999 con base el año 2001 puede ser obtenido mediante una sencilla regla de tres, $IPC_{99}^{01} = \frac{IPC_{99}^{96} IPC_{01}^{01}}{IPC_{01}^{96}} = \frac{(107)(100)}{112} = 95.5357$, luego el salario real del año 1999 en euros constantes de 2001 fue $\frac{1500}{IPC_{99}^{01}/100} = \frac{(1500)(107)}{112} = 1570.0935$.

Por su parte, el salario real del año 2003 en euros constantes de 2001 fue $\frac{1800}{IPC_{03}^{01}/100} = \frac{1800}{1.07} = 1682.243$.

El incremento en términos reales del salario en el período $1999-2003$ ha sido de un $\frac{1682.243-1570.0935}{1570.0935}100 = 7.1429\%$, bastante inferior al 16% de aumento del alquiler, por lo que el inquilino no debería estar conforme con la subida del mismo.

3. Y_t, ventas anuales en el período t en número de unidades.
$Y_t^* = 10 + 15\,t$ (t, años; origen, 2000).

a) $s_y^2 = 472,\ s_t^2 = 2$

$R^2 = r_{ty}^2 = \frac{s_{ty}^2}{s_t^2\, s_y^2} = \frac{b^2\, s_t^2}{s_y^2} = \frac{(15)^2(2)}{472} = \underline{0.9534}$, coeficiente que por su gran proximidad a uno indica que la recta de regresión lineal mínimo cuadrática ajusta <u>bastante bien</u> los datos.

b) $IVE(1C) = 0.2$ e $IVE(2C) = 1.8$
$IVE(3C) = 3 - IVE(1C) - IVE(2C) = 3 - 0.2 - 1.8 = 1$
$y_t^* = \frac{10}{3} + \frac{15}{3^2}\, t = 3.\widehat{3} + 1.\widehat{6}\, t$ (t, cuatrimestres; origen, segundo cuatrimestre 2000).
Del segundo cuatrimestre de 2000 al tercero de 2005 hay $(5)(3) + 1 = 16$ cuatrimestres.
$y_{16}^* = 3.\widehat{3} + \left(1.\widehat{6}\right)(16) = 30$
$\hat{y}_{3C05} = y_{16}^*\, IVE(3C) = (30)(1) = \underline{30}$, esto es, la predicción corregida por estacionalidad y sin corregir coinciden dado que <u>no ha habido variación estacional</u> (índice igual a uno) en el período de referencia.

4. X, tiempo de espera en horas.
$X \sim Ex(\lambda = 0.5),\ f(x) = (0.5)\exp(-0.5\, x),\ x > 0$

a) $\mu = \frac{1}{\lambda} = \frac{1}{0.5} = \underline{2}$ y $\sigma^2 = \frac{1}{\lambda^2} = \frac{1}{(0.5)^2} = \underline{4}$

b) $p(X > 2.25) = \int_{2.25}^{+\infty} f(x)\, dx = \int_{2.25}^{+\infty} (0.5)\exp(-0.5\, x)\, dx = -\int_{2.25}^{+\infty} (-0.5)\exp(-0.5\, x)\, dx = [-\exp(-0.5\, x)]_{2.25}^{+\infty} = \exp((-0.5)(2.25)) - \lim_{x\to+\infty} \exp(-0.5\, x) = \exp(-1.125) - 0 = \underline{0.3247}$

c) El rango de variación de la distribución Normal es el conjunto de todos los números reales, mientras que el de la distribución Exponencial es sólo el de los reales positivos. Una segunda diferencia es el carácter simétrico de la Normal, que la distribución Exponencial no tiene.

5. X, ingresos diarios de los teléfonos móviles de tarjeta en miles de euros.
Y, ingresos diarios de los teléfonos móviles de contrato en miles de euros.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 5 \\ 8 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 6 \end{pmatrix}\right)$$

a) $p(3 \le X \le 6) = p\left(\frac{3-5}{\sqrt{3}} \le \frac{X-\mu_x}{\sigma_x} \le \frac{6-5}{\sqrt{3}}\right) = p(-1.1547 \le Z_x \le 0.5774) = p(Z_x \le 0.5774) - p(Z_x \le -1.1547) = p(Z_x \le 0.5774) - (1 - p(Z_x > 1.1547)) = 0.719 - 0.1251 = 0.5939$, donde $Z_x = \frac{X-\mu_x}{\sigma_x} \sim N(0,1)$
$p(7 \le Y \le 10) = p\left(\frac{7-8}{\sqrt{6}} \le \frac{Y-\mu_y}{\sigma_y} \le \frac{10-8}{\sqrt{6}}\right) = p(-0.4082 \le Z_y \le 0.8165) = p(Z_y \le 0.8165) - p(Z_y \le -0.4082) = p(Z_y \le 0.8165) - (1 - p(Z_y > 0.4082)) = 0.7925 - 0.3409 = 0.4516$, donde $Z_y = \frac{Y-\mu_y}{\sigma_y} \sim N(0,1)$
Como se tiene la incorrelación entre X e Y y se admite para ellas una distribución conjunta binormal entonces X e Y son independientes. Por lo tanto, $p(3 \le X \le 6,\ 7 \le Y \le 10) = p(3 \le X \le 6)\, p(7 \le Y \le 10) = (0.5939)(0.4516) = \underline{0.2682}$

b) $W_1 = 1.02X + 1.01Y$, ingresos diarios con la primera oferta en miles de euros.
$W_2 = 1.015X + 1.015Y$, ingresos diarios con la segunda oferta en miles de euros.
$T = W_2 - W_1 = 0.005(Y - X)$, diferencia entre los ingresos diarios con ambas ofertas en miles de euros.
$T \sim N(\mu_t = 0.005(\mu_y - \mu_x) = (0.005)(8-5) = 0.015, \sigma_t^2 = (0.005)^2(\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}) = (0.005)^2(3 + 6 - 2(0)) = 2.25 \times 10^{-4})$