Examen de Estadística I. 18-06-2009.

1. X, precio de un coche de segunda mano en euros.
Y, kilometraje de un coche de segunda mano.
$\overline{x} = 1200$, $g_0(x) = 0.17$, $\overline{y} = 60000$, $g_0(x) = 0.5$, $r_{xy} = -0.9$

a) $U = \frac{X}{1000}$, precio de un coche de segunda mano en miles de euros.
$V = \frac{Y}{1000}$, miles de kilómetros de un coche de segunda mano.
$u^* = a + b\,v$
$\overline{u} = \frac{\overline{x}}{1000} = 1.2$, $g_0(u) = g_0(x) = 0.17 \Rightarrow \frac{s_u}{\overline{u}} = 0.17 \Rightarrow s_u = (0.17)(1.2) = 0.204$
$\overline{v} = \frac{\overline{y}}{1000} = 60$, $g_0(v) = g_0(y) = 0.5 \Rightarrow \frac{s_v}{\overline{v}} = 0.5 \Rightarrow s_v = (0.5)(60) = 30$
$r_{uv} = r_{xy} = -0.9 \Rightarrow \frac{s_{uv}}{s_u s_v} = -0.9 \Rightarrow s_{uv} = (-0.9)(0.204)(30) = -5.508$
$b = \frac{s_{uv}}{s_v^2} = \frac{-5.508}{30^2} = -0.0061$, $a = \overline{u} - b\overline{v} = 1.2 - (-0.0061)(60) = 1.5672$, luego
$\underline{u^* = 1.5672 - 0.0061\,v}$

b) $v_0 = 80$
$u_0^* = a + b\,v_0 = 1.5672 - (0.0061)(80) = \underline{1.0776}$, es decir, debería pedirse 1077.6 euros.

c) $R^2 = r_{uv}^2 = (-0.9)^2 = \underline{0.81}$, que por su cercanía a uno indica que el ajuste realizado es razonablemente bueno.

2. X, inversión en publicidad en miles de euros
Y, importe de las ventas en miles de euros.

a) $x_0 = 456$, $\hat{x}_1 = \frac{x_1}{IPC_{08}/100} = (456)(1.05) = 478.8$, $IPC_{08} = (102.8)(1.02) = 104.856$, luego $x_1 = (478.8)(1.0486) = \underline{502.0505}$

b) $IVE(1T) = 1$, $IVE(3T) = 0.8$, $IVE(4T) = 0.4$
$y^* = \frac{1770}{4} + \frac{190}{4^2}\,t = 442.5 + 11.875\,t$ (t, trimestres; origen, trimestre central 1997).
Del trimestre central de 1997 al segundo de 2010 hay $(13)(4) - 0.5 = 51.5$ trimestres.
$y_{51.5}^* = 442.5 + (11.875)(51.5) = 1054.0625$
$IVE(2T) = 4 - IVE(1T) - IVE(3T) - IVE(4T) = 1.8$, luego $\hat{y}_{2T10} = y_{51.5}^*\,IVE(2T) = (1054.0625)(1.8) = \underline{1897.3125}$

3. X, ingresos en euros.

a) Calculamos el índice de Gini de X.

$[L_{i-1}, L_i)$	x_i	n_i	N_i	$x_i n_i$	$\sum_{j=1}^{i} x_j n_j$	$p_i = \frac{N_i}{N}\,100$	$q_i = \frac{\sum_{j=1}^{i} x_j n_j}{N\overline{x}}\,100$
$[0, 8000)$	4000	10	10	40000	40000	20	5.2459
$[8000, 15000)$	11500	15	25	172500	212500	50	27.8689
$[15000, 25000)$	20000	20	45	400000	612500	90	80.3279
$[25000, 35000)$	30000	5	50	150000	762500	100	100

$I_G = \frac{\sum_{i=1}^{I-1}(p_i - q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{(20-5.2459)+(50-27.8689)+(90-80.3279)}{20+50+90} = \underline{0.291}$, que como resulta relativamente cercano a cero, indica que los ingresos de estos ciudadanos están similarmente distribuidos entre las distintas categorías salariales.

b) El orden del cuantil 20000, que pertenece al tercer intervalo, es
$20000 = Q_r = L_{i-1} + a_i \, \frac{Nr - N_{i-1}}{n_i} = 15000 + 10000 \, \frac{50r-25}{20} \Rightarrow r = 0.7$ y por lo tanto el porcentaje de ciudadanos que supera los 20000 euros es el $\underline{30\%}$.

4. X, número aleatorio entre 0 y 1.
$X \sim U(0,1),\ f(x) = 1,\ x \in [0,1]$

a) $F(x) = p(X \leq x) = \int_{-\infty}^{x} f(x)\, dx$, por lo tanto

$$F(x) \quad = \quad \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \in [0,1] \\ 1 & \text{si } x > 1 \end{cases}$$

b) $p(X < 0.1) = \int_0^{0.1} 1\, dx = \underline{0.1}$

c) W, números aleatorios, de entre 5, que pueden ser considerados bajos.
$W \sim Bi(n = 5, \theta = 0.1),\ f(w) = \left(\begin{array}{c} 5 \\ w \end{array} \right) \theta^w (1-\theta)^{5-w},\ w = 0, \ldots, 5$
$p(W = 0) = (0.9)^5 = \underline{0.5905}$

d) Y, número aleatorio entre 0 y 1.
$p(X = Y) = \underline{0}$, ya que las probabilidades puntuales en variables continuas son nulas.

5. X, número de telespectadores por minuto de la final de la Copa del Rey, en millones.
Y, número de telespectadores por minuto de la final de la Champions, en millones.
$X \sim N(\mu_x = 10, \sigma_x^2 = (0.4)^2)$, $Y \sim N(\mu_y = 11.5, \sigma_x^2 = 1^2)$, independientes.

a) $T = X + Y$, número total de telespectadores que verán el logo publicitario, en millones.
Cualquier combinación lineal de variables Normales independientes es Normal, de modo que $T \sim N(\mu_t = \mu_x + \mu_y = 10 + 11.5 = 21.5, \sigma_t^2 = \sigma_x^2 + \sigma_y^2 = (0.4)^2 + 1^2 = 1.16)$.
$p(T > 22) = p(Z_t > \frac{22-21.5}{\sqrt{1.16}}) = p(Z_t > 0.4642) = \underline{0.3211}$, donde $Z_t = \frac{T-\mu_t}{\sigma_t} \sim N(0,1)$.

b) No podría ser resuelto. Si no se asume la independencia entre las variables no se tiene la normalidad para la variable aleatoria T.

c) $p(Y > 14.577) = p(Z_y > \frac{14.577-11.5}{1}) = p(Z_y > 3.077) = \underline{0.001}$, donde $Z_y = \frac{Y-\mu_y}{\sigma_y} \sim N(0,1)$.