Examen de Estadística I. 04-09-2006.

1. X, duración de la llamada en segundos del primer hijo.
Y, duración de la llamada en segundos del segundo hijo.

a) $\overline{x} = 120, s_x = 50, \overline{y} = 210, s_y = 40$
$g_0(X) = \frac{s_x}{\overline{x}} = \frac{50}{120} = \underline{0.41\widehat{6} > 0.1905} = \frac{40}{210} = \frac{s_y}{\overline{y}} = g_0(Y)$, es decir, $\underline{\overline{y}}$ representa mejor a la variable ya que presenta una mayor regularidad.

b) C, coste de la llamada en euros.
$\overline{c}_x = 1.2, n_x = 60, \overline{c}_y = 2.4, n_y = 30$
$\overline{c} = \frac{1}{n_x + n_y}(\overline{c}_x n_x + \overline{c}_y n_y) = \underline{1.6}$

c) D, duración media de las llamadas en segundos del segundo hijo.

d_i	n_i	N_i	$d_i n_i$	$\sum_{j=1}^{i} d_j n_j$	$p_i = \frac{N_i}{N}100$	$q_i = \frac{\sum_{j=1}^{i} d_j n_j}{N\overline{d}}100$
130	27	27	3510	3510	45	27.6814
210	19	46	3990	7500	$76.\widehat{6}$	59.1483
370	14	60	5180	12680	100	100

El índice de Gini es $I_G = \frac{\sum_{i=1}^{I}(p_i - q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{(45-27.6814)+(76.\widehat{6}-59.1483)+(100-100)}{45+76.\widehat{6}} = 0.2863$, que como resulta menor que 0.44, el de la primera factura, indica que la distribución del tiempo medio en las llamadas de diferente tipo está más concentrada en el primer hijo.

2. Y, ingresos anuales de la empresa en millones de euros.
$\overline{y} = 60,\ \overline{t} = 0.5,\ s_y^2 = 2000,\ s_t^2 = 6,\ s_{ty} = 96$

a) $y^* = a + bt$
$b = \frac{s_{ty}}{s_t^2} = \frac{96}{6} = 16,\ a = \overline{y} - b\overline{t} = 60 - (16)(0.5) = 52$
$\underline{y^* = 52 + 16\,t}$ (t, años; origen, 2001)
$R^2 = r_{ty}^2 = \frac{s_{ty}^2}{s_t^2\, s_y^2} = \frac{(96)^2}{(6)(2000)} = \underline{0.768}$, coeficiente que por su proximidad a uno indica que el ajuste realizado es razonablemente bueno.

b) $IVE(1T) = 1.1,\ IVE(2T) = 1$ e $IVE(4T) = 1.1$
$IVE(3T) = 4 - IVE(1T) - IVE(2T) - IVE(4T) = 4 - 1.1 - 1 - 1.1 = 0.8$
$y^* = \frac{52}{4} + \frac{16}{4^2}\,t = 13 + t$ (t, trimestres; origen, trimestre central 2001).
Del trimestre central de 2001 al tercero de 2006 hay $(5)(4) + 0.5 = 20.5$ trimestres.
$y^*_{20.5} = 13 + 20.5 = 33.5$
$\hat{y}_{3T06} = y^*_{33.5}\, IVE(3T) = (33.5)(0.8) = \underline{26.8}$

c) $\hat{y}_3 = \frac{y_3^*}{P_{04}^{01}/100} = \frac{52+(16)(3)}{1.1} = 90.\widehat{90}$
$\hat{y}_4 = \frac{y_4^*}{P_{05}^{01}/100} = \frac{52+(16)(4)}{1.14} = 101.7544$

$\frac{\hat{y}_4-\hat{y}_3}{\hat{y}_3} = \frac{101.7544-90.\widehat{90}}{90.\widehat{90}} = \underline{0.1193}$, es decir, los ingresos han aumentado en términos reales el 11.93% en el período 2004-2005.

3. X, ventas al contado de un agente en u.m.
Y, ventas a crédito de un agente en u.m.
$(x_1, \ldots, x_n)$
$(y_1, \ldots, y_n)$

No sería correcto realizar la comparación de x_i con y_i directamente, dado que es más oportuno relativizar con respecto a las ventas realizadas por el resto de agentes. Así, lo correcto sería comparar los valores tipificados o estandarizados $z_x = \frac{x_i-\overline{x}}{s_x}$ y $z_y = \frac{y_i-\overline{y}}{s_y}$.

4. X, tiempo de realización del examen en minutos.
$X \sim Un(30, 90),\ f(x) = \frac{1}{90-30} = \frac{1}{60},\ x \in [30, 90]$

a) $p(X > 50) = \int_{50}^{90} f(x)\, dx = \int_{50}^{90} \frac{1}{60}\, dx = \frac{90-50}{60} = \underline{\frac{2}{3}}$

b) Y, número de alumnos, de entre 10, que tardan menos de 50 minutos en realizar el examen.
$Y \sim Bi(10, \theta),\ f(y) = \begin{pmatrix} 10 \\ y \end{pmatrix} \theta^y (1-\theta)^{10-y},\ y = 0, \ldots, 10$
$\theta = 1 - p(X > 50) = \frac{1}{3}$
$p(Y \geq 4) = 1 - p(Y < 4) = 1 - p(Y = 0) - p(Y = 1) - p(Y = 2) - p(Y = 3) = 1 - \begin{pmatrix} 10 \\ 0 \end{pmatrix} (0.\widehat{3})^0 (0.\widehat{6})^{10} - \begin{pmatrix} 10 \\ 1 \end{pmatrix} (0.\widehat{3})^1 (0.\widehat{6})^9 - \begin{pmatrix} 10 \\ 2 \end{pmatrix} (0.\widehat{3})^2 (0.\widehat{6})^8 - \begin{pmatrix} 10 \\ 3 \end{pmatrix} (0.\widehat{3})^3 (0.\widehat{6})^7 = 1 - 0.5593 = \underline{0.4407}$

5. X, ventas diarias de leche en litros.
Y, ventas diarias de paquetes de yogures en unidades.
$B = 0.8X + 1.05Y - 430$, beneficio diario en euros.
$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 900 \\ 200 \end{pmatrix}, \begin{pmatrix} 50^2 & \sigma_{xy} \\ \sigma_{xy} & 15^2 \end{pmatrix}\right)$

a) $\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$, luego $\sigma_{xy} = \rho_{xy}\, \sigma_x\, \sigma_y = (-0.4)(50)(15) = -300$
$\mu_b = 0.8\mu_x + 1.05\mu_y - 430 = (0.8)(900) + (1.05)(200) - 430 = 500,\ \sigma_b^2 = (0.8)^2(50)^2 + (1.05)^2(15)^2 + (2)(0.8)(1.05)(-300) = 1344.0625$
$p(B < 510) = p\left(\frac{B-\mu_b}{\sigma_b} < \frac{510-500}{\sqrt{1344.0625}}\right) = p(Z_b \leq 0.2728) = \underline{0.6074}$, donde $Z_b = \frac{B-\mu_b}{\sigma_b} \sim N(0, 1)$

b) $p(X < k) = 0.8 \Leftrightarrow p\left(\frac{X-\mu_x}{\sigma_x} < \frac{k-900}{50}\right) = 0.8 \Leftrightarrow p\left(Z_x < \frac{k-900}{50}\right) = 0.8 \Leftrightarrow \frac{k-900}{50} = 0.84 \Leftrightarrow k = \underline{942}$, donde $Z_x = \frac{X-\mu_x}{\sigma_x} \sim N(0, 1)$

c) W_1, número de familias, de entre las 10, pertenecientes a la categoría $A_1 = \{$consumir preferentemente leche$\}$, $p(A_1) = 0.5$.
W_2, número de familias, de entre las 10, pertenecientes a la categoría $A_2 = \{$consumir preferentemente yogures$\}$, $p(A_2) = 0.35$.
W_3, número de familias, de entre las 10, pertenecientes a la categoría $A_3 = \{$consumir preferentemente otro tipo de productos$\}$, $p(A_3) = 0.15$.
Admitiendo independencia en cuanto a la clasificación en las tres categorías de las 10 familias,

se tiene que el vector aleatorio cuyas componentes recogen el número de familias en cada una de las posibles categorías sigue una distribución de probabilidad Multinomial.

$(W_1, W_2, W_3)' \sim Mn(n = 10, \boldsymbol{\theta} = (0.5, 0.35, 0.15)')$

$p(W_1 = 6, W_2 = 3, W_3 = 1) = \frac{10!}{6!\,3!\,1!}(0.5)^6(0.35)^3(0.15)^1 = \underline{0.0844}$