Examen de Estadística I. 05-06-2006.

1. X, importe de la inscripción en euros en una carrera popular en la Comunidad Valenciana.
Y, número de inscritos.

a) $\overline{x} = \frac{1}{n}\sum_{i=1}^{4} x_i\, n_{i\cdot} = \frac{(0)(9)+(3)(4)+(5)(7)+(10)(3)}{23} = \underline{3.3478}$
$\overline{y} = \frac{1}{n}\sum_{j=1}^{3} y_j\, n_{\cdot j} = \frac{(100)(10)+(500)(8)+(1000)(5)}{23} = 434.7826$
$s_y^2 = \frac{1}{n}\sum_{j=1}^{3} y_j^2\, n_{\cdot j} - \overline{y}^2 = \frac{(100)^2(10)+(500)^2(8)+(1000)^2(5)}{23} - (434.7826)^2 = \underline{119659.7354}$

b) Por ejemplo, $\underline{n_{11} = 6 \neq \frac{(9)(10)}{23}} = \frac{n_{1\cdot}\, n_{\cdot 1}}{N}$, luego no se cumple la condición de independencia y las variables son dependientes.

c) $s_x^2 = \frac{1}{n}\sum_{i=1}^{4} x_i^2\, n_{i\cdot} - \overline{x}^2 = \frac{(0)^2(9)+(3)^2(4)+(5)^2(7)+(10)^2(3)}{23} - (3.3478)^2 = 11.0095$
$g_0(X) = \frac{s_x}{\overline{x}} = \frac{\sqrt{11.0095}}{3.3478} = \underline{0.9911 > 0.7956} = \frac{\sqrt{119659.7354}}{434.7826} = \frac{s_y}{\overline{y}} = g_0(Y)$, es decir, el número de inscritos presenta una mayor regularidad que el importe de la inscripción.

d) W, importe de la inscripción en euros en una carrera popular en las Islas Canarias.
$\overline{w} = 6, s_w = 3$
$x_0 = 14, w_0 = 15$
$\frac{x_0 - \overline{x}}{s_x} = \frac{14-3.3478}{\sqrt{11.0095}} = \underline{3.2104 > 3} = \frac{15-6}{3} = \frac{w_0 - \overline{w}}{s_w}$, es decir, en términos relativos la cuota de inscripción es mayor en la media maratón de Valencia.

2. Y_t, ingresos anuales en el período t en millones de euros.
$y_t^* = 400 + 20\,t$ (t, años; origen, 2000).

a) $y_1^* = 400 + (20)(1) = 420$
$\underline{y_t^* = 420 + 20\,t}$ (t, años; origen, 2001).

b) El IPC del año 2005 con base el año 2002 puede ser obtenido mediante una sencilla regla de tres, $IPC_{05}^{02} = \frac{IPC_{05}^{01} IPC_{02}^{02}}{IPC_{02}^{01}} = \frac{(119.4)(100)}{105.6} = 113.06\widehat{81}$, luego la estimación de los ingresos del año 2005 en euros constantes de 2002 es $\hat{y}_4 = \frac{y_4^*}{IPC_{05}^{02}/100} = \frac{420+(20)(4)}{1.1307} = \underline{442.2111}$.
$y_1^* = 420 + (20)(1) = 440$
$\frac{\hat{y}_4 - y_1^*}{y_1^*} = \frac{442.2111-440}{440} = \underline{0.005}$, es decir, los ingresos han aumentado en términos reales poco más del 0.5% en el período 2002-2005.

c) $IVE(1S) = 0.9$
$IVE(2S) = 2 - IVE(1S) = 2 - 0.9 = 1.1$
$y_t^* = \frac{420}{2} + \frac{20}{2^2}\,t = 210 + 5t$ (t, semestres; origen, semestre central 2001).
Del semestre central de 2001 al segundo de 2006 hay $(5)(2) + 0.5 = 10.5$ semestres.
$y_{10.5}^* = 210 + (5)(10.5) = 262.5$
$\hat{y}_{2S06} = y_{10.5}^*\, IVE(2S) = \underline{288.75}$

3. X, número de motocicletas vendidas semanalmente.
$X \sim Po(\lambda),\ \mu = \lambda = 2,\ f(x) = \exp(-\lambda)\frac{\lambda^x}{x!},\ x \in \mathbb{N}$

a) $p(X = 2) = \exp(-2)\frac{2^2}{2!} = \underline{0.2707}$

b) Y, número de semanas, de entre las 10, en las que no se vende ninguna motocicleta.
$Y \sim Bi(10, \theta),\ f(y) = \begin{pmatrix} 10 \\ y \end{pmatrix} \theta^y (1-\theta)^{10-y},\ y = 0, \ldots, 10$
$\theta = p(X = 0) = \exp(-2)\frac{2^0}{0!} = 0.1353$
$p(Y \geq 2) = 1 - p(Y < 2) = 1 - p(Y = 0) - p(Y = 1) = 1 - \begin{pmatrix} 10 \\ 0 \end{pmatrix} (0.1353)^0 (0.8647)^{10} - \begin{pmatrix} 10 \\ 1 \end{pmatrix} (0.1353)^1 (0.8647)^9 = \underline{0.4008}$

4. X, tiempo invertido en el trayecto en horas.
$F(x)$

a) La <u>función de distribución</u> ya que ésta determina completamente la distribución de probabilidad de una variable aleatoria. A partir de la función de distribución puede obtenerse cualquier característica de la variable aleatoria (incluidas media y varianza), así como cualquier probabilidad sobre la misma.

b) $p(a < X < b) = \underline{F(b) - F(a) - p(X = b)}$

c) $X \sim Un(2, 4)$
$p(2.5 < X < 3.5) = \int_{2.5}^{3.5} \frac{1}{4-2}\, dx = \frac{3.5-2.5}{4-2} = \underline{0.5}$
$X \sim N(3, 0.25)$
$p(2.5 < X < 3.5) = p(\frac{2.5-3}{\sqrt{0.25}} < Z < \frac{3.5-3}{\sqrt{0.25}}) = p(-1 < Z < 1) = F(1) - F(-1) = F(1) - [1 - F(1)] = 2F(1) - 1 = 2(0.8413) - 1 = \underline{0.6826}$, donde $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
Podría ser considerado más ajustado a la realidad el <u>modelo Normal</u> para así tener probabilidades no nulas para tiempos de trayecto inferiores a dos horas o superiores a cuatro.

5. X, distancia recorrida por un atleta en kilómetros.
Y, distancia recorrida por un ciclista en kilómetros.

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 2.5 \\ 15 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 49 \end{pmatrix}\right)$$

a) $p(2 \leq X \leq 4) = p\left(\frac{2-2.5}{\sqrt{1}} \leq \frac{X-\mu_x}{\sigma_x} \leq \frac{4-2.5}{\sqrt{1}}\right) = p(-0.5 \leq Z_x \leq 1.5) = p(Z_x \leq 1.5) - p(Z_x \leq -0.5) = p(Z_x \leq 1.5) - (1 - p(Z_x > 0.5)) = 0.9332 - 0.3085 = 0.6247$, donde $Z_x = \frac{X-\mu_x}{\sigma_x} \sim N(0, 1)$
$p(Y < 20) = p\left(\frac{Y-\mu_y}{\sigma_y} < \frac{20-15}{\sqrt{49}}\right) = p(Z_y < 0.7143) = 0.7626$, 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 consiguiente, $p(2 \leq X \leq 4,\ Y < 20) = p(2 \leq X \leq 4)\, p(Y < 20) = (0.6247)(0.7626) = \underline{0.4764}$

b) $W_1 = 1.2X + 1.1Y$, kilómetros recorridos por atleta y ciclista entrenados por el médico.
$W_2 = 1.15X + 1.15Y$, kilómetros recorridos por atleta y ciclista no entrenados por el médico.
$T = W_1 - W_2 = 0.05(X - Y)$, diferencia entre los kilómetros recorridos por atletas y ciclistas de ambos grupos .
$\underline{T \sim N}(\mu_t = 0.05(\mu_x - \mu_y) = (0.05)(2.5 - 15) = \underline{-0.625}, \sigma_t^2 = (0.05)^2(\sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}) = (0.05)^2(1 + 49 - 2(0)) = \underline{0.125})$

c) V_1, número de atletas, de entre los 10, pertenecientes a la categoría $A_1 = \{$menos de 3 horas en maratón$\}$, $p(A_1) = 0.2$.
V_2, número de atletas, de entre los 10, pertenecientes a la categoría $A_2 = \{$entre 3 horas y 3 horas y media en maratón$\}$, $p(A_2) = 0.4$.
V_3, número de atletas, de entre los 10, pertenecientes a la categoría $A_3 = \{$no completar el maratón en menos de 3 horas y media$\}$, $p(A_3) = 0.4$.
Admitiendo independencia en cuanto a la clasificación en las tres categorías de los 10 atletas, se tiene que el vector aleatorio cuyas componentes recogen el número de atletas en cada una de las posibles categorías sigue una distribución de probabilidad Multinomial.
$(V_1, V_2, V_3)' \sim Mn(n = 10, \boldsymbol{\theta} = (0.2, 0.4, 0.4)')$
$p(V_1 = 3, V_2 = 5, V_3 = 2) = \frac{10!}{3!\,5!\,2!}(0.2)^3(0.4)^5(0.4)^2 = \underline{0.033}$