Examen de Estadística I. 11-06-2002.

1. X, ventas de automóviles en la factoría asiática.
Y, ventas de automóviles en la factoría española.
$\mu_x = 100,\ \sigma_x = 10,\ \mu_y = 75,\ \sigma_y = 8.$

a) Utilizando los correspondientes coeficientes de variación,

$$g_o(X) = \frac{\sigma_x}{\mu_x} = \frac{10}{100} = \underline{0.1 < 0.1067} = \frac{8}{75} = \frac{\sigma_y}{\mu_y} = g_o(Y)\,,$$

obtenemos que las ventas medias de la factoría asiática son más representativas.

b) $\frac{x_0-\mu_x}{\sigma_x} = \frac{105-100}{10} = \underline{0.5 < 0.625} = \frac{80-75}{8} = \frac{y_0-\mu_y}{\sigma_y}$,

es decir, en términos relativos son mayores las ventas mensuales de la factoría española.

c) Para analizar la concentración de la distribución utilizamos el índice de Gini, que puede ser obtenido a partir de los datos.

$p_i = \frac{N_i}{N}\,100$	$q_i = \frac{\sum_{j=1}^{i} x_j n_j}{N\overline{x}}\,100$
70	60
90	85
100	100

$$I_G = \frac{\sum_{i=1}^{I}(p_i-q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{(70-60)+(90-85)+(100-100)}{70+90} = \underline{0.0938},$$

que indica que los salarios están bastante uniformemente repartidos ya que la distribución presenta una concentración muy baja.

2. Y_t, ingresos por venta de entradas de cine en el período t en millones de euros.

	t	y_t	t^2	y_t^2	$t\,y_t$
	0	531	0	281961	0
	1	608	1	369664	608
	2	643	4	413449	1286
	3	694	9	481636	2082
$\sum_{i=1}^{4}$	6	2476	14	1546710	3976
$\frac{1}{4}\sum_{i=1}^{4}$	1.5	619	3.5	386677.5	994

a) $y_t^* = a + bt$
$b = \frac{s_{ty}}{s_t^2} = \frac{994-(1.5)(619)}{3.5-(1.5)^2} = \frac{65.5}{1.25} = 52.4,\ a = \overline{y} - b\overline{t} = 619 - (52.4)(1.5) = 540.4$
$\underline{y_t^* = 540.4 + 52.4\,t \text{ (t, años; origen, 1998)}}$.

b) $y_5^* = 540.4 + (52.4)(5) = \underline{802.4}$
$R^2 = r_{ty}^2 = \frac{s_{ty}^2}{s_t^2\, s_y^2} = \frac{(65.5)^2}{(1.25)(386677.5-619^2)} = \underline{0.976},$

coeficiente que por su gran proximidad a uno indica que las predicciones realizadas utilizando la recta de regresión lineal mínimo cuadrática presentan una muy elevada fiabilidad.

c) $y_t^* = \frac{540.4}{2} + \frac{52.4}{2^2}\, t = 270.2 + 13.1\, t$ (t, semestres; origen, semestre central 1998).

d) Del semestre central de 1998 al segundo de 2002 hay $(2)(4) + 0.5 = 8.5$ semestres.
$y_{8.5}^* = 270.2 + (13.1)(8.5) = 381.55$
$IVE(2S) = 2 - IVE(1S) = 2 - (1.05) = 0.95$
$P_{2002}^{2000}(p) = \frac{(152.25)(100)}{142.69} = 106.6998$
$\hat{y}_{2S02} = y_{8.5}^* \frac{IVE(2S)}{P_{2002}^{2000}(p)/100} = (381.55)\frac{0.95}{1.067} = \underline{339.7124}$.

3. X, duración en años de una pieza del ordenador.
$X \sim Ex(\mu_x = 5),\ f(x) = \frac{1}{5}\exp\left(-\frac{1}{5}x\right),\ x > 0.$

a) $p(X < 6) = \int_0^6 \frac{1}{5}\exp\left(-\frac{1}{5}x\right)\, dx = -\left[\exp\left(-\frac{1}{5}x\right)\right]_0^6 = \underline{0.6988}$.

b) Z, número de reclamaciones mensuales.
$\mu_z = 2$.
Se trata del conteo del número, no acotado, de ocurrencias de un determinado suceso con una baja probabilidad, en una unidad temporal. El número de reclamaciones mensuales puede tomar cualquier valor natural, por lo que $\underline{Z \sim Po(\lambda)}$. Además, como el parámetro de la distribución Poisson coincide con la media, $\underline{\lambda = 2}$.

4. X, tiempo invertido en la carrera a pie en minutos.
Y, tiempo invertido en la carrera en bicicleta en minutos.
$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 45 \\ 35 \end{pmatrix}, \begin{pmatrix} 100 & 80 \\ 80 & 100 \end{pmatrix}\right).$

a) $T = X + Y$, tiempo invertido en las dos pruebas en minutos.
$T \sim N(\mu_t = \mu_x + \mu_y = 45 + 35 = 80, \sigma_t^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy} = 100 + 100 + (2)(80) = 360)$
$p(T < 90) = p(Z_t < \frac{90-80}{\sqrt{360}}) = p(Z_t < 0.527) = \underline{0.7019}$, donde $Z_t = \frac{T-\mu_t}{\sigma_t} \sim N(0,1)$.

b) V, número de participantes, de entre los 5 amigos, que acaban la prueba ciclista en menos de media hora.

Si definimos el éxito de una prueba Bernouilli como acabar la prueba ciclista en menos de media hora, el parámetro de la distribución es $p(Y < 30) = p(Z_y < \frac{30-35}{\sqrt{100}}) = p(Z < -0.5) = 1 - p(Z > 0.5) = 0.3085$, donde $Z_y = \frac{Y-\mu_y}{\sigma_y} \sim N(0,1)$.

Admitiendo independencia entre los tiempos de la prueba ciclista de los cinco amigos, se tiene que $V \sim Bi(n = 5, \theta = 0.3085),\ f(x) = \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x},\ x = 0, 1, \ldots, 5.$

$p(V \geq 3) = p(V = 3) + p(V = 4) + p(V = 5) = \begin{pmatrix} 5 \\ 3 \end{pmatrix}(0.3085)^3(1 - 0.3085)^2 + \begin{pmatrix} 5 \\ 4 \end{pmatrix}(0.3085)^4(1 - 0.3085)^1 + \begin{pmatrix} 5 \\ 5 \end{pmatrix}(0.3085)^5(1 - 0.3085)^0 = \underline{0.1745}$.

c) W_1, número de participantes, de entre los 50 inscritos, pertenecientes a la categoría $A_1 =$ *senior masculino*, $p(A_1) = 0.65$.

W_2, número de participantes, de entre los 50 inscritos, pertenecientes a la categoría $A_2 =$ *senior femenino*, $p(A_2) = 0.15$.

W_3, número de participantes, de entre los 50 inscritos, pertenecientes a la categoría $A_3 =$ *juveniles*, $p(A_3) = 0.2$.

Admitiendo independencia en cuanto a la clasificación en las tres categorías de los 50 atletas inscritos, se tiene que el vector aleatorio cuyas componentes recogen el número de participantes en cada una de las posibles categorías sigue una distribución de probabilidad Multinomial, $\underline{(W_1, W_2, W_3)' \sim Mn(n = 50, \boldsymbol{\theta} = (0.65, 0.15, 0.2))}$.