Examen de Estadística I. 09-09-2002.

1. X_M, ingresos de los establecimientos de Madrid en euros.
X_B, ingresos de los establecimientos de Barcelona en euros.
X_V, ingresos de los establecimientos de Valencia en euros.
$\overline{x}_M = 12500,\ N_M = 5,\ \overline{x}_B = 10000,\ N_B = 4,\ \overline{x}_V = 5000,\ N_V = 2.$

a) La media de un conjunto de poblaciones es una media ponderada con pesos los distintos tamaños poblacionales. Si llamamos X a los ingresos de la cadena hotelera en euros, entonces

$$\overline{x} = \frac{\overline{x}_M N_M + \overline{x}_B N_B + \overline{x}_V N_V}{N_M + N_B + N_V} = \frac{12500 \cdot 5 + 10000 \cdot 4 + 5000 \cdot 2}{5+4+2} = 10227.\widehat{27}.$$

Como los costes globales son 8000 euros, el beneficio medio es $10227.\widehat{27} - 8000 = \underline{2227.\widehat{27}}$.

b) Para analizar la concentración de la distribución utilizamos el índice de Gini. Si denominamos Y al salario del trabajador en euros, se tiene que

y_i	n_i	N_i	$y_i n_i$	$\sum_{j=1}^{i} y_j n_j$	$p_i = \frac{N_i}{N}\ 100$	$q_i = \frac{\sum_{j=1}^{i} y_j n_j}{N\overline{y}}\ 100$
610	150	150	91500	91500	50	34.8571
900	90	240	81000	172500	80	65.7143
1500	60	300	90000	262500	100	100

$$I_G = \frac{\sum_{i=1}^{I}(p_i - q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{(50-34.8571)+(80-65.7143)+(100-100)}{50+80} = \underline{0.2264},$$

que indica que si bien los salarios no están equidistribuidos, la concentración que presentan es relativamente baja por lo que se encuentran bastante uniformemente repartidos.

2. Y_t, importaciones españolas en el período t en miles de millones de euros.

a) $y_t^* = a + bt$
$b = \frac{s_{ty}}{s_t^2} = \frac{\frac{1237}{5} - \frac{550}{5}\frac{10}{5}}{\frac{30}{5} - \left(\frac{10}{5}\right)^2} = \frac{27.4}{2} = 13.7,\ a = \overline{y} - b\overline{t} = \frac{550}{5} - (13.7)\frac{10}{5} = 82.6$
$y_t^* = 82.6 + 13.7\, t$ (t, años; origen, 1995).

$$R^2 = r_{ty}^2 = \frac{s_{ty}^2}{s_t^2\, s_y^2} = \frac{(27.4)^2}{2\left[\frac{62392}{5} - \left(\frac{550}{5}\right)^2\right]} = \underline{0.992},$$

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

b) $y_6^* = 82.6 + (13.7)(6) = 164.8$
$\hat{y}_6 = y_6^* \frac{100}{I_{2001}^{1999}} = (164.8)\frac{93}{102.9} = \underline{148.9446}$.

c) $y_t^* = \frac{82.6}{4} + \frac{13.7}{4^2}\, t = 20.65 + 0.8563\, t$ (t, trimestres; origen, trimestre central 1995). Del trimestre central de 1995 al tercero de 2001 hay $(6)(4) + 0.5 = 24.5$ semestres.

$y^*_{24.5} = 20.65 + (0.8563)(24.5) = 41.6294$
$IVE(3T) = 4 - IVE(1T) - IVE(2T) - IVE(4T) = 4 - 0.9 - 1.1 - 0.6 = 1.4$
$\hat{y}_{3T01} = y^*_{24.5}\ IVE(3T) = (41.6294)(1.4) = \underline{58.2812}$.

3. X, número de entradas de cine vendidas en ventanilla.
Y, número de entradas de cine vendidas por internet.
$\mu_x = 1200,\ \sigma_x = 200,\ \mu_y = 500,\ \sigma_y = 100$.

a) $X \sim N(\mu_x, \sigma_x^2),\ Y \sim N(\mu_y, \sigma_y^2)$.
$T = 6.01\,Y - 5.4\,X$, diferencia entre los ingresos por venta de entradas por internet y en ventanilla, en euros. Asumiendo independencia entre X e Y se tiene que cualquier combinación lineal de ellas sigue una distribución Normal. En particular, $D \sim N(\mu_d = (6.01)\mu_y - (5.4)\mu_x = (6.01)(500) - (5.4)(1200) = -3475, \sigma_d^2 = (6.01)^2\sigma_y^2 + (5.4)^2\sigma_x^2 = (6.01)^2(100)^2 + (5.4)^2(200)^2 = 1527601)$.
$p(D > 0) = p\left(Z_d > \frac{0-(-3475)}{\sqrt{1527601}}\right) = p(Z_d > 2.8116) = \underline{0.0025}$, donde $Z_d = \frac{D-\mu_d}{\sigma_d} \sim N(0,1)$.

b) El hecho de que las variables estén incorreladas no garantiza su independencia y, consecuentemente, que cualquier combinación lineal de ambas siga una distribución Normal. Por ello, no sería conocida la distribución conjunta de X e Y y no podríamos calcular la probabilidad requerida.

c) En el caso de que las variables sean conjuntamente Normales entonces cualquier combinación lineal de ellas sigue una distribución Normal, sin tener en consideración su posible independencia.
$T = X + Y$, número de entradas vendidas por internet y en ventanilla.
$T \sim N(\mu_t = \mu_x + \mu_y = 1200 + 500 = 1700, \sigma_t^2 = \sigma_x^2 + \sigma_y^2 = (200)^2 + (100)^2 = 50000)$.
$p(T > 1700) = p\left(Z_t > \frac{1700-1700}{\sqrt{50000}}\right) = p(Z_t > 0) = \underline{0.5}$, donde $Z_t = \frac{T-\mu_t}{\sigma_t} \sim N(0,1)$.

4. X, número de títulos conseguidos en una temporada.
$X \sim Po(1.2),\ \mu_y = 1.2 \rightarrow \lambda = 1.2$.

a) $p(X = 3) = \exp(-1.2)\,\frac{(1.2)^3}{3!} = \underline{0.0867}$.

b) $p(X \geq 1) = 1 - p(X = 0) = 1 - \exp(-1.2)\,\frac{(1.2)^0}{0!} = \underline{0.6988}$.
Considerando independencia entre el número de títulos conseguidos por el entrenador en una temporada u otra, la probabilidad requerida es la de una intersección de sucesos y por tanto el producto de las probabilidades: $(0.6988)^5 = \underline{0.1666}$.

c) Y, momento mensual de la destitución o conclusión en el cargo del entrenador.
Asumiendo que la destitución puede producirse en cualquier momento de la temporada, una vez concluido el período de confianza, sin existir ningún momento preferible a cualquier otro, resultaría adecuado modelizar la variable aleatoria Y mediante una distribución Uniforme, $Y \sim Un(2, 9)$.