Examen de Estadística I. 01-09-2008.

1. X, tiempo en media maratón en minutos.
Y, tiempo en maratón en minutos.

a) $y^* = a + bx$
$b = \frac{s_{xy}}{s_x^2} = \frac{\frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})(y_i-\overline{y})}{\frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2} = \frac{\frac{1}{n}\sum_{i=1}^n x_i y_i - \overline{x}\,\overline{y}}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2} = \frac{16241-(87)(183.8)}{7685-(87)^2} = \frac{250.4}{116} = 2.1586$, $a = \overline{y} - b\overline{x} = 183.8 - (2.1586)(87) = -4$
$\underline{y^* = -4 + 2.1586\,x}$

b) $x_0 = 92$
$y_0^* = a + b\,x_0 = -4 + (2.1586)(92) = \underline{194.5912}$
$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{(250.4)^2}{(116)(542.56)} = \underline{0.9962}$

2. X, importe de las ventas en unidades monetarias.

Calculamos el índice de Gini de X para obtener un indicador de su grado de concentración.

$[L_{i-1}, L_i)$	p_i	q_i
bajas	40	10
intermedias	90	70
altas	100	100

$I_G = \frac{\sum_{i=1}^{I}(p_i - q_i)}{\sum_{i=1}^{I-1} p_i} = \frac{(40-10)+(90-70)+(100-100)}{40+90} = \underline{0.3846}$, que como resulta relativamente cercano a cero, indica que las ventas en estos grandes almacenes <u>están poco concentradas</u>.

3. Y_t, ventas de un comercio en el período t en unidades monetarias.
$y_t^* = 12 + 2\,t$ (t, años; origen, 2002).

a) $IVE(1T) = 1.2,\ IVE(2T) = 1.3,\ IVE(3T) = 0.6$
$y_t = \frac{12}{4} + \frac{2}{4^2}\,t = 3 + 0.125\,t$ (t, trimestres; origen, trimestre central 2002).
Del trimestre central de 2002 al cuarto de 2008 hay $(6)(4) + 1.5 = 25.5$ trimestres.
$y_{25.5} = 3 + (0.125)(25.5) = 6.1875$
$IVE(4T) = 4 - IVE(1T) - IVE(2T) - IVE(3T) = 0.9$, luego $\hat{y}_{4T08} = y_{25.5}\,IVE(4T) = (6.1875)(0.9) = \underline{5.5688}$

b) $y_0 = 12 + (2)(0) = 12$, $y_5 = 12 + (2)(5) = 22$
$\hat{y}_0 = \frac{12}{1.035} = 11.5942$, $\hat{y}_5 = \frac{22}{1.209} = 18.1969$
$\frac{\hat{y}_7 - \hat{y}_0}{\hat{y}_0} = \frac{18.1969-11.5942}{11.5942} = \underline{0.5695}$, es decir, los ingresos <u>han aumentado en términos reales casi un 57% en el período 2002-2007</u>.

4. X, cotización de ciertas acciones en unidades monetarias.
$X \sim N(\mu = 500, \sigma^2 = 25^2)$

a) Y, número de días, de entre 10, en los que el inversor decide comprar acciones.
$Y \sim Bi(10, \theta),\ f(y) = \begin{pmatrix} 10 \\ y \end{pmatrix} \theta^y (1-\theta)^{10-y},\ y = 0, \ldots, 10$
$\theta = p(525 < X < 575) = p(\frac{525-500}{25} < Z_x < \frac{575-500}{25}) = p(1 < Z_x < 3) = p(Z_x < 3) - p(z_x < 1) = 0.9987 - 0.8413 = 0.1574$, donde $Z_x = \frac{X-\mu_x}{\sigma_x} \sim N(0,1)$.
$p(Y \leq 3) = \begin{pmatrix} 10 \\ 3 \end{pmatrix} (0.1574)^3 (0.8426)^7 + \begin{pmatrix} 10 \\ 2 \end{pmatrix} (0.1574)^2 (0.8426)^8 + \begin{pmatrix} 10 \\ 1 \end{pmatrix} (0.1574)^1 (0.8426)^9 + \begin{pmatrix} 10 \\ 0 \end{pmatrix} (0.1574)^0 (0.8426)^{10} = \underline{0.9417}$

b) La desigualdad de Chebychev proporciona una cota para ciertas probabilidades conociendo tan sólo la media y la varianza de la variable aleatoria, $p(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$, así para una cota de probabilidad $1 - \frac{1}{k^2} = 0.75$ se tiene que $k = 2$ y consecuentemente los dos valores entre los cuales se situaría el índice con una probabilidad de al menos el 75% son $\mu - k\sigma = 500 - (2)(25) = \underline{450}$ y $\mu + k\sigma = 500 + (2)(25) = \underline{550}$.

5. X, ingresos diarios comercio 1 en euros.
Y, ingresos diarios comercio 2 en euros.
$X \sim N(\mu = 2000, \sigma^2 = 300^2)$
$Y \sim N(\mu = 1000, \sigma^2 = 360^2)$

a) $B = \sum_{i=1}^{6}(X_i + Y_i)$, ingresos totales semanales en euros.
Admitiendo <u>independencia entre los ingresos de ambos comercios y también independencia entre los ingresos de los diferentes días de la semana para cada uno de los comercios</u>, se tiene que $B \sim N(\mu_b = 6(\mu_x + \mu_y) = 6(2000 + 1000) = 18000, \sigma_b^2 = (6)(\sigma_x^2 + \sigma_y^2) = 6(300^2 + 360^2) = 1317600)$.
$p(B > 18000) = p(Z_b > \frac{18000-18000}{\sigma_b}) = p(Z_b > 0) = \underline{0.5}$, donde $Z_b = \frac{B-\mu_b}{\sigma_b} \sim N(0,1)$

b) No. Si no se asume la independencia entre las variables no se tiene la normalidad para la variable aleatoria B.