1. (2 p.) Sea (X, Y) una variable estadística bidimensional, cuyas rectas de regresión son

$$9x - 2y = 10 \qquad \text{y} \qquad 2x - y = 5$$

Sabiendo que

$$N = 10 \qquad \text{y} \qquad \sum_{i=1}^{10} x_i y_i = 40$$

a) Calcule el coeficiente de correlación lineal de Pearson.

b) Si se descubre que uno de los puntos considerados, el $(2,5)$, estaba equivocado y que el verdadero valor es el $(0,-5)$, determine la nueva recta de regresión $Y|X$, y estudie si el ajuste ha mejorado, o no.

Solución

a) Puesto que las ecuaciones explícitas de las rectas del enunciado son

$$y = \frac{9}{2}x - 5 \qquad , \qquad y = 2x - 5$$

la pendiente de cada una de ellas es $\frac{9}{2}$ y 2, respectivamente, y se cumple que $2 \leq \frac{9}{2}$, podemos afirmar que la ecuación de la recta de regresión de $Y|X$ es $y = 2x - 5$, mientras que la ecuación de la recta de regresión de $X|Y$ es $x = \frac{2}{9}y + \frac{10}{9}$.
Por otra parte, también sabemos que ambas rectas pasan por el punto $(\overline{x}, \overline{y})$. Así, puesto que $(0, -5)$ es la única solución del sistema lineal

$$\left.\begin{array}{rcl} 9x - 2y & = & 10 \\ 2x - y & = & 5 \end{array}\right\} \text{podemos afirmar que } \overline{x} = 0 \text{ y que } \overline{y} = -5$$

Además, teniendo en cuenta que $\sigma_{xy} = m_{11} - \overline{x}\,\overline{y}$, siendo $m_{11} = \frac{\sum_{i=1}^{N} x_i y_i}{N}$, y puesto que, como se afirma en el enunciado, $N = 10$ y $\sum_{i=1}^{10} x_i y_i = 40$, tenemos que $\sigma_{xy} = \frac{40}{10} - 0 \cdot (-5) = 4$
Por otro lado, la pendiente de la recta de regresión de $Y|X$ (pendiente que, en nuestro caso, vale 2) coincide con $\frac{\sigma_{xy}}{\sigma_x^2}$. Así, teniendo en cuenta

que, según hemos probado anteriormente, $\sigma_{xy} = 4$, tenemos que $2 = \dfrac{4}{\sigma_x^2}$, de donde podemos afirmar que $\sigma_x^2 = 2$ (y, por tanto, $\sigma_x = \sqrt{2}$)

De forma análoga, la pendiente de la recta de regresión de $X|Y$ (pendiente que, en nuestro caso, vale $\dfrac{9}{2}$) coincide con $\dfrac{\sigma_y^2}{\sigma_{xy}}$. Así, teniendo en cuenta que, según hemos probado anteriormente, $\sigma_{xy} = 4$, tenemos que $\dfrac{9}{2} = \dfrac{\sigma_y^2}{4}$, de donde podemos afirmar que $\sigma_y^2 = 18$ (y, por tanto, $\sigma_y = 3\sqrt{2}$)

Con ello, puesto que el coeficiente de correlación lineal de Pearson se define como $\rho = \dfrac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$, y teniendo en cuenta los resultados ya probados, podemos afirmar que $\rho = \dfrac{4}{\sqrt{2} \cdot 3\sqrt{2}} = \dfrac{2}{3}$

b) Para este segundo apartado, vamos a denotar por (X^*, Y^*) a la variable estadística bidimensional con el dato corregido, y vamos a mantener la notación (X, Y) para la variable estadística bidimensional original. Así, teniendo en cuenta los resultados probados en el apartado anterior, tenemos que

$$\overline{x} = 0 = \frac{\sum_{i=1}^{10} x_i}{10} \Rightarrow \boxed{\sum_{i=1}^{10} x_i = 0} \tag{1}$$

$$\sigma_x^2 = 2 = \frac{\sum_{i=1}^{10} x_i^2}{10} - (\overline{x})^2 = \frac{\sum_{i=1}^{10} x_i^2}{10} - 0^2 \Rightarrow \boxed{\sum_{i=1}^{10} x_i^2 = 20} \tag{2}$$

$$\overline{y} = -5 = \frac{\sum_{i=1}^{10} y_i}{10} \Rightarrow \boxed{\sum_{i=1}^{10} y_i = -50} \tag{3}$$

$$\sigma_y^2 = 18 = \frac{\sum_{i=1}^{10} y_i^2}{10} - (\overline{y})^2 = \frac{\sum_{i=1}^{10} y_i^2}{10} - (-5)^2 = \frac{\sum_{i=1}^{10} y_i^2}{10} - 25 \Rightarrow$$

$$\Rightarrow \boxed{\sum_{i=1}^{10} y_i^2 = 10(18 + 25) = 430} \tag{4}$$

Así, para las variables corregidas, tenemos que

$$\overline{x^*} = \frac{1}{10}\left(\sum_{i=1}^{10} x_i^*\right) = \frac{1}{10}\left[\left(\sum_{i=1}^{10} x_i\right) - 2 + 0\right] \overset{(1)}{=} \frac{0 - 2 + 0}{10} = \frac{-1}{5} \tag{5}$$

$$\sigma^2_{x^*} = \frac{\sum_{i=1}^{10}(x_i^*)^2}{10} - \left(\overline{x^*}\right)^2 \overset{(5)}{=} \frac{\left(\sum_{i=1}^{10} x_i^2\right) - 2^2 + 0^2}{10} - \left(\frac{-1}{5}\right)^2 \overset{(2)}{=}$$

$$\overset{(2)}{=} \frac{20-4}{10} - \frac{1}{25} = \frac{16}{10} - \frac{1}{25} = \frac{8}{5} - \frac{1}{25} = \frac{40-1}{25} = \frac{39}{25}$$

$$\overline{y^*} = \frac{1}{10}\left(\sum_{i=1}^{10} y_i^*\right) = \frac{1}{10}\left[\left(\sum_{i=1}^{10} y_i\right) - 5 + (-5)\right] \overset{(3)}{=} \frac{-50-5-5}{10} = -6 \tag{6}$$

$$\sigma^2_{y^*} = \frac{\sum_{i=1}^{10}(y_i^*)^2}{10} - \left(\overline{y^*}\right)^2 \overset{(6)}{=} \frac{\left(\sum_{i=1}^{10} y_i^2\right) - 5^2 + (-5)^2}{10} - (-6)^2 \overset{(4)}{=}$$

$$\overset{(4)}{=} \frac{430-25+25}{10} - 36 = 43 - 36 = 7$$

$$\sigma_{x^*y^*} = \frac{\sum_{i=1}^{10} x_i^* y_i^*}{10} - \overline{x^*}\cdot\overline{y^*} = \frac{\left(\sum_{i=1}^{10} x_i y_i\right) - 2\cdot 5 + 0\cdot(-5)}{10} - \frac{-1}{5}\cdot(-6) =$$

$$= \frac{40-10}{10} - \frac{6}{5} = 3 - \frac{6}{5} = \frac{15-6}{5} = \frac{9}{5}$$

Así, sabiendo que la ecuación recta de regresión de $Y^*|X^*$ es $y = a^* + b^*\cdot x$, siendo $a^* = \overline{y^*} - b^*\cdot\overline{x^*}$, $b^* = \dfrac{\sigma_{x^*y^*}}{\sigma^2_{x^*}}$, y teniendo en cuenta los resultados anteriores, tenemos que

$$b^* = \frac{\frac{9}{5}}{\frac{39}{25}} = \frac{45}{39}$$

$$a^* = -6 - \frac{45}{39}\cdot\frac{-1}{5} = -6 + \frac{9}{39} = \frac{-39\cdot 6 + 9}{39} = \frac{-234+9}{39} = \frac{-225}{39}$$

con lo que la ecuación de la recta de regresión de $Y^*|X^*$ es

$$y = \frac{-225}{39} + \frac{45}{39}\cdot x$$

Podemos determinar si el ajuste ha mejorado o no, calculando el coeficiente de correlación lineal (ρ^*) entre X^* e Y^*. Así, puesto que

$$\rho^* = \frac{\sigma_{x^*y^*}}{\sigma_{x^*}\cdot\sigma_{y^*}} = \frac{\frac{9}{5}}{\frac{\sqrt{39}}{5}\cdot\sqrt{7}} = \frac{9}{\sqrt{273}} \approx 0'545$$

es menor que ρ, el nuevo ajuste es peor que el inicial. ♣

2. (3 p.) Consideremos la siguiente tabla de frecuencias absolutas obtenida de los datos la variable estadística bidimensional (X, Y)

$Y \backslash X$	$[-10,-2]$	$(-2,2]$	$(2,8]$	$(8,10]$
-1	4	4	2	0
0	1	5	3	1
1	0	1	5	4

Se pide

a) Representar gráficamente la distribución de frecuencias de la variable X condicionada al valor 0 de la variable Y, y calcular su media, mediana, moda y varianza.

b) Utilizar el método de los mínimos cuadrados para ajustar, como modelo del tipo $X|Y$, uno de la forma $x = ay^3$

Solución

a) Si, por comodidad, denotamos por Z a la variable $X|Y=0$, ésta toma los valores indicados en la siguiente tabla. En la columna z_i se indican las correspondientes marcas de clase, en la columna N_i se indican las frecuencias absolutas acumuladas, en la columna a_i se indican las correspondientes amplitudes y, en la columna h_i se indican las alturas correspondientes que se han utilizado para la elaboración del histograma que se presenta en la figura 1, teniendo en cuenta que $h_i = \dfrac{n_i}{a_i}$

Z	z_i	n_i	N_i	a_i	h_i
$[-10,-2)$	-6	1	1	8	$1/8$
$[-2,\ 2)$	0	5	6	4	$5/4 = 10/8$
$[\ 2,\ 8)$	5	3	9	6	$3/6 = 4/8$
$[\ 8,\ 10]$	9	1	10	2	$1/2 = 4/8$

Así, puesto que $N = n_1+n_2+n_3+n_4 = 1+3+5+1 = 10$, y utilizando las marcas de clase para el cálculo de la media y de la varianza, obtenemos que

$$\mu_z = \frac{-6\cdot 1 + 0\cdot 5 + 5\cdot 3 + 9\cdot 1}{10} = \frac{-6+15+9}{10} = \frac{18}{10} = \frac{9}{5}$$

$$m_2(z) = \frac{(-6)^2\cdot 1 + 0^2\cdot 5 + 5^2\cdot 3 + 9^2\cdot 1}{10} = \frac{36+75+81}{10} = \frac{192}{10} = \frac{96}{5}$$

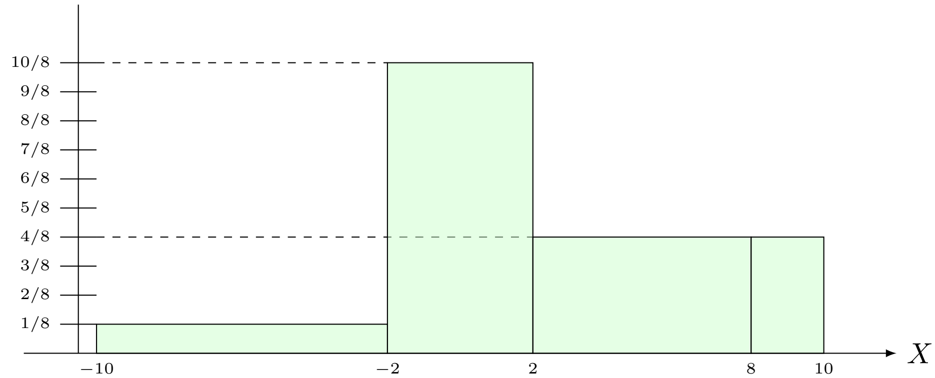

Figura 1: Histograma correspondiente a la variable $X|Y=0$

$$\sigma_z^2 = m_2(z) - \mu_z^2 = \frac{96}{5} - \left(\frac{9}{5}\right)^2 = \frac{96}{5} - \frac{81}{25} = \frac{96 \cdot 5 - 81}{25} = \frac{480 - 81}{25} = \frac{399}{25}$$

Para el cálculo de la moda, teniendo en cuenta que la variable Z es continua, podemos que aplicar la fórmula

$$\text{Mo} = L_{i-1} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot a_i \tag{7}$$

siendo $[-2, 2)$ el intervalo modal (es decir, $i = 2$, puesto que el intervalo $[-2, 2)$ es de mayor altura en el histograma ya que es el correspondiente al mayor valor de h_i). Así, $L_{i-1} = -2$, $\Delta_1 = \frac{10}{8} - \frac{1}{8} = \frac{9}{8}$, $\Delta_2 = \frac{10}{8} - \frac{4}{8} = \frac{6}{8}$ y $a_i = 4$, con lo que, sin más que sustituir en (7), obtenemos que

$$\text{Mo} = L_{i-1} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot a_i = -2 + \frac{\frac{9}{8}}{\frac{9}{8} + \frac{6}{8}} \cdot 4 =$$

$$= -2 + \frac{3}{5} \cdot 4 = -2 + \frac{12}{5} = \frac{-10 + 12}{5} = \frac{2}{5}$$

Para el cálculo de la mediana, podemos aplicar la fórmula

$$\text{Me} = L_{i-1} + \frac{N/2 - N_{i-1}}{n_i} \cdot a_i \tag{8}$$

siendo también $[-2, 2)$ el intervalo mediana (es decir, $i = 2$, puesto que $i = 2$ es el primer entero para el que se cumple que $N_i \geq N/2 = 5$). Así, $L_{i-1} = -2$ y puesto que $N_{i-1} = 1$, $n_2 = 5$ y $a_2 = 4$, sin más que sustituir en (8), obtenemos que

$$\text{Me} = L_{i-1} + \frac{N/2 - N_{i-1}}{n_i} \cdot a_i = -2 + \frac{5 - 1}{5} \cdot 4 =$$

$$= -2 + \frac{16}{5} = \frac{-10 + 16}{5} = \frac{6}{5}$$

b) Para ajustar una curva de la forma $x = a\,y^3$, por el método de mínimos cuadrados y como un modelo del tipo $X|Y$, a la nube de puntos descrita por N parejas de la forma (x_i, y_i) tenemos que determinar el valor de a que minimiza la función $F(a) = \sum_{i=1}^{N}(x_i - a\,y_i^3)^2$. Por tanto, el correspondiente valor de a tiene que anular a $\frac{\partial F}{\partial a}(a)$. Así, puesto que $\frac{\partial F}{\partial a}(a) = \frac{-2}{3}\sum_{i=1}^{N} y_i^3(x_i - a\,y_i^3)$, el parámetro a tiene que verificar que $\sum_{i=1}^{N} y_i^3(x_i - a\,y_i^3) = 0$, de donde podemos deducir que $a = \frac{\sum_{i=1}^{N} x_i\,y_i^3}{\sum_{i=1}^{N} y_i^6}$

Por otra parte, para simplificar los cálculos, vamos a calcular en primer lugar $\mu_{X|Y=-1}$, $\mu_{X|Y=0}$ y $\mu_{X|Y=1}$ y utilizar los resultados correspondientes para ajustar un modelo a una nube simplificada:

$$\mu_{X|Y=-1} = \frac{1}{10}\,(-6\cdot 4 + 0\cdot 4 + 5\cdot 2 + 9\cdot 0) = \frac{-24+0+10+0}{10} = \frac{-14}{10}$$

$$\mu_{X|Y=0} = \frac{1}{10}\,(-6\cdot 1 + 0\cdot 5 + 5\cdot 3 + 9\cdot 1) = \frac{-6+0+15+9}{10} = \frac{18}{10}$$

$$\mu_{X|Y=1} = \frac{1}{10}\,(-6\cdot 0 + 0\cdot 1 + 5\cdot 5 + 9\cdot 4) = \frac{0+0+25+36}{10} = \frac{61}{10}$$

Así, puesto que las frecuencias marginales de cada valor de Y son iguales, ajustar un modelo del tipo $x = a\,y^3$ a la nube de puntos inicial es equivalente a ajustarlo a los datos de la siguiente tabla

X	$\frac{-14}{10}$	$\frac{18}{10}$	$\frac{61}{10}$
Y	-1	0	1

En consecuencia, el correspondiente valor de a tiene que ser

$$a = \frac{\sum_{i=1}^{3} x_i\,y_i^3}{\sum_{i=1}^{3} y_i^6} = \frac{\frac{-14}{10}\cdot(-1)^3 + \frac{8}{10}\cdot 0^3 + \frac{61}{10}\cdot 1^3}{(-1)^6 + 0^6 + 1^6} = \frac{\frac{14}{10} + \frac{61}{10}}{2} = \frac{75}{20} = \frac{15}{4}$$

También podíamos haber llegado al mismo resultado utilizando la nube de puntos original. En tal caso, el correspondiente valor de a lo podríamos calcular como

$$a = \frac{\sum_{i=1}^{30} x_i\, y_i^3}{\sum_{i=1}^{30} y_i^6} = \frac{1}{(-1)^6 \cdot 10 + 0^6 \cdot 10 + 1^6 \cdot 10}\Big((-6)\cdot(-1)^3\cdot 4 + (-6)\cdot 0^3 \cdot 1 +$$

$$+(-6)\cdot 1^3\cdot 0 + 0\cdot(-1)^3\cdot 4 + 0\cdot 0^3\cdot 5 + 0\cdot 1^3\cdot 1 + 5\cdot(-1)^3\cdot 2 + 5\cdot 0^3\cdot 3 + 5\cdot 1^3\cdot 5 +$$

$$+9\cdot(-1)^3\cdot 0 + 9\cdot 0^3\cdot 1 + 9\cdot 1^3\cdot 4\Big) = \frac{24 - 10 + 25 + 36}{20} = \frac{75}{20} = \frac{15}{4}$$

De cualquier forma, la curva de la forma $x = a\,y^3$ que más se ajusta a la nube de puntos es $x = \dfrac{15}{4} y^3$ ♣

3. (3 p.) Consideremos la siguiente serie estadística correspondiente a los datos de precios (P) y cantidades (Q), por cuatrimestres (C_1, C_2 y C_3), de los productos A y B distribuidos por una empresa malagueña en los últimos tres años.

	2008			2009			2010		
	C_1	C_2	C_3	C_1	C_2	C_3	C_1	C_2	C_3
P_A	15	12	20	20	18	30	30	20	35
Q_A	10	5	20	10	30	15	40	20	50
P_B	30	24	40	40	36	60	60	40	70
Q_B	20	10	40	20	60	30	80	40	100

Se pide:

a) Calcular el índice ideal de Fisher para el primer cuatrimestre del año 2010 tomando como periodo base, el primer cuatrimestre del año 2009 e interprete el resultado.

b) Determinar los valores desestacionalizados de la serie de cantidades del producto B, aplicando el método de la media móvil en porcentajes, suponiendo la hipótesis multiplicativa.

c) Sin desestacionalizar la serie, utilizar el método de los mínimos cuadrados para predecir la tendencia de la serie temporal de valores totales del producto A, correspondiente al tercer cuatrimestre del año 2011, y calcular el coeficiente de determinación para evaluar la calidad de la predicción.

Solución

a) Sabiendo que el índice de Fisher del primer cuatrimestre del año 2010, tomando como base el primer cuatrimestre del año 2009, se define como la media geométrica entre los índices de Laspeyres y Paasche del primer cuatrimestre del año 2010, tomando como base el primer cuatrimestre del año 2009, y teniendo en cuenta que

$$L_{C_1(2010)|C_1(2009)} = \frac{30 \cdot 10 + 60 \cdot 20}{20 \cdot 10 + 40 \cdot 20} \cdot 100 = \frac{1500}{1000} \cdot 100 = 150$$

$$P_{C_1(2010)|C_1(2009)} = \frac{30 \cdot 40 + 60 \cdot 80}{20 \cdot 40 + 40 \cdot 80} \cdot 100 = \frac{6000}{4000} \cdot 100 = 150$$

podemos afirmar que

$$F_{C_1(2010)|C_1(2009)} = \sqrt{L_{C_1(2010)|C_1(2009)} \cdot P_{C_1(2010)|C_1(2009)}} =$$

$$= \sqrt{150 \cdot 150} = 150$$

entendiéndose que, respecto al primer cuatrimestre del año 2009, el precio de una cesta compuesta por los productos A y B, se ha encarecido, en el primer cuatrimestre del año 2010, en un 50 %

b) Antes de empezar el segundo apartado, indiquemos que los valores correspondientes de la serie que queremos desestacionalizar son los que se indican en la siguiente tabla:

Y	2008	2009	2010
C_1	20	20	80
C_2	10	60	40
C_3	40	30	100

Así, si suponemos, bajo la hipótesis multiplicativa, que $Y = T \cdot E \cdot C \cdot A$, podemos eliminar $E \cdot A$ calculando las medias móviles de orden 3 (puesto que 3 son los cuatrimestres de cada año). Por tanto, teniendo en cuenta que no tendríamos que centrar puesto que el orden es impar, obtenemos que

$T \cdot C$	2008	2009	2010
C_1	–	120/3	150/3
C_2	70/3	110/3	220/3
C_3	70/3	170/3	–

$T \cdot C$	2008	2009	2010
C_1	–	40	50
C_2	$23'\hat{3}$	$36'\hat{6}$	$73'\hat{3}$
C_3	$23'\hat{3}$	$56'\hat{6}$	–

Si dividimos los datos de la serie, entre los de $T \cdot C$, obtenemos $E \cdot A$, siendo los valores correspondientes los indicados en la siguiente tabla:

$E \cdot A$	2008	2009	2010
C_1	–	60/120	240/150
C_2	30/70	180/110	120/220
C_3	120/70	90/170	–

$E \cdot A$	2008	2009	2010
C_1	–	$0'5$	$1'6$
C_2	$0'429$	$1'636$	$0'545$
C_3	$1'714$	$0'529$	–

Calculando las medias por filas (sumando los valores de cada fila y dividiendo la suma entre 2, 3 y 2 respectivamente) de los elementos de $E \cdot A$, obtenemos los valores indicados en la primera columna de la siguiente tabla, cuyos índices respecto a la media $\left(\dfrac{3'042}{3}\right)$ se calculan dividiendo el correspondiente valor entre la media y multiplicando el resultado por

100 y son los que indican en la segunda columna de la siguiente tabla:

$\dfrac{\frac{60}{120}+\frac{240}{150}}{2}=1'05$	$103'551\,\%$
$\dfrac{\frac{60}{120}+\frac{240}{150}+\frac{120}{220}}{3}=0'870$	$85'812\,\%$
$\dfrac{\frac{120}{70}+\frac{90}{170}}{2}=1'122$	$110'637\,\%$

Así, puesto que no es necesario normalizar ya que la suma de los valores de la última columna de la tabla recién obtenida es 300, los valores estimados de los índices de estacionalidad son

E	2008	2009	2010
C_1	$103'551\,\%$	$103'551\,\%$	$103'551\,\%$
C_2	$85'812\,\%$	$85'812\,\%$	$85'812\,\%$
C_3	$110'637\,\%$	$110'637\,\%$	$110'637\,\%$

con lo que los valores de la serie desestacionalizada los podemos obtener dividiendo los datos de la serie Y entre los de la tabla recién obtenida y son los que se indican en la tabla siguiente:

$Y/E = TCA$	2008	2009	2010
C_1	$19'314$	$19'314$	$77'257$
C_2	$11'653$	$69'920$	$46'613$
C_3	$36'154$	$27'116$	$90'386$

c) En primer lugar, indiquemos claramente cuáles son los valores de la serie de la que, sin desestacionalizar los datos y utilizando el método de los mínimos cuadrados, queremos estimar la tendencia. Para ello, si denotamos por -4, -3, -2, -1, 0, 1, 2, 3 y 4 a los instantes de tiempo correspondientes a cada uno de los cuatrimestres y años indicados en la tabla inicial (es decir, el instante $t = -4$ se corresponde con el primer cuatrimestre del año 2008; el instante $t = -3$, con el segundo cuatrimestre del año 2008... y así sucesivamente) y tenemos en cuenta que los valores totales del producto A se definen como el producto de los precios del producto A (P_A) por las cantidades del producto A (Q_A), los valores de la serie temporal son los indicados en la siguiente tabla:

t	-4	-3	-2	-1	0	1	2	3	4
w	150	60	400	200	540	450	1200	400	1750

para los que se tiene que

$$\bar{t} = \frac{-4-3-2-1+0+1+2+3+4}{9} = 0$$

$$m_2(t) = \frac{(-4)^2+(-3)^2+(-2)^2+(-1)^2+0^2+1^2+2^2+3^2+4^4}{9} =$$

$$= \frac{16+9+4+1+0+1+4+9+16}{9} = \frac{60}{9} = \frac{20}{3}$$

$$\sigma_t^2 = m_2(t) - \left(\bar{t}\right)^2 = \frac{20}{3} - 0^2 = \frac{20}{3}$$

$$\overline{w} = \frac{150+60+400+200+540+450+1200+400+1750}{9} = \frac{5150}{9}$$

$$m_2(w) = \frac{150^2+60^2+400^2+200^2+540^2+450^2+1200^2+400^2+1750^2}{9} =$$

$$= \frac{100}{9}\left(15^2+6^2+40^2+20^2+54^2+45^2+120^2+40^2+175^2\right) =$$

$$= \frac{100}{9}\left(225+36+1600+400+2916+2025+14400+1600+30625\right) =$$

$$= \frac{5382700}{9}$$

$$\sigma_w^2 = m_2(w) - (\,\overline{w}\,)^2 = \frac{5382700}{9} - \left(\frac{5150}{9}\right)^2 =$$

$$= \frac{5382700\cdot 9 - 5150\cdot 5150}{81} = \frac{48444300-26522500}{81} = \frac{21921800}{81}$$

$$m_{11} = \frac{1}{9}\Big((-4)\cdot 150 + (-3)\cdot 60 + (-2)\cdot 400 + (-1)\cdot 200 + 0\cdot 540 +$$

$$+1\cdot 450 + 2\cdot 1200 + 3\cdot 400 + 4\cdot 1750\Big) =$$

$$= \frac{-600-180-800-200+450+0+2400+1200+7000}{9} = \frac{9270}{9} = 1030$$

$$\sigma_{tw} = m_{11} - \bar{t}\,\overline{w} = 1030 - 0\cdot\frac{5150}{9} = 1030$$

Así, puesto que los parámetros de un ajuste rectilíneo del tipo $W|T$ (es decir, $w = a + b\,t$) a la nube de puntos descrita por la tabla anteriormente indicada, vienen determinados por las igualdades

$$b = \frac{\sigma_{tw}}{\sigma_t^2} \qquad , \qquad a = \overline{w} - b\,\bar{t}$$

sin más que llevar a esas igualdades los valores obtenidos tenemos que

$$b = \frac{\sigma_{tw}}{\sigma_t^2} = \frac{1030}{\frac{20}{3}} = \frac{3090}{20} = \frac{309}{2} = 154'5$$

$$a = \overline{w} - b\,\bar{t} = \frac{5150}{9} - \frac{309}{2}\,0 = \frac{5150}{9} = 572'\hat{2}$$

con lo que el modelo de mínimos cuadrados resultante es

$$w = a + b\,t = \frac{5150}{9} + \frac{309}{2}\,t = \frac{10300 + 2781\,t}{18}$$

La predicción para el tercer cuatrimestre del año 2011 es la correspondiente al instante de tiempo $t = 7$, y así

$$w(7) = \frac{10300 + 2781 \cdot 7}{18} = \frac{10300 + 19467}{18} = \frac{29767}{18} = 1653'7\hat{2}$$

Puesto que el ajuste es rectilíneo, el coeficiente de determinación R^2 cumple que $R^2 = \rho^2$ siendo ρ el coeficiente de correlación lineal de Pearson. Por tanto

$$R^2 = \rho^2 = \frac{\sigma_{tw}^2}{\sigma_t^2 \cdot \sigma_w^2} = \frac{1030^2}{\frac{20}{3} \cdot \frac{21921800}{81}} = \frac{103^2 \cdot 81 \cdot 3}{20 \cdot 219218} = \frac{2577987}{4384360} \approx 0'588$$

con lo que la calidad de la predicción es “bastante pobre”. ♣

4. (2 p.) Un juego consiste en lanzar simultáneamente un dado y dos monedas, y obtenemos un premio de un euro por cada punto obtenido en el dado, más un euro por cada cara obtenida en las monedas. Por ejemplo, si tenemos la suerte de sacar un 5 en el dado, una cara en una moneda y una cruz en la otra moneda, entonces recibiría un premio total de 5+1+0=6 euros.

    a) Determine el espacio muestral y la función de probabilidad correspondiente a los posibles premios totales que se pueden obtener con este juego.

    b) Si jugamos 2 veces a este juego, ¿qué probabilidad hay de ganar 7 euros en cada partida, sabiendo que el premio total es 14 euros?

Solución

a) Si indicamos, en primer lugar, el resultado obtenido en el dado y, en segundo lugar, el obtenido en las monedas, los posibles resultados de una partida son:

$$\left\{\begin{array}{c}
\{1 \text{ y dos cruces}\}, \{1, \text{ una cara y una cruz}\}, \{1 \text{ y dos caras}\}, \\
\{2 \text{ y dos cruces}\}, \{2, \text{ una cara y una cruz}\}, \{2 \text{ y dos caras}\}, \\
\{3 \text{ y dos cruces}\}, \{3, \text{ una cara y una cruz}\}, \{3 \text{ y dos caras}\}, \\
\{4 \text{ y dos cruces}\}, \{4, \text{ una cara y una cruz}\}, \{4 \text{ y dos caras}\}, \\
\{5 \text{ y dos cruces}\}, \{5, \text{ una cara y una cruz}\}, \{5 \text{ y dos caras}\}, \\
\{6 \text{ y dos cruces}\}, \{6, \text{ una cara y una cruz}\}, \{6 \text{ y dos caras}\}
\end{array}\right\}$$

y la probabilidad de obtener cada uno de ellos viene indicada, respectivamente, en la siguiente tabla

$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2=\frac{2}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$
$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2=\frac{2}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$
$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2=\frac{2}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$
$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2=\frac{2}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$
$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2=\frac{2}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$
$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2=\frac{2}{24}$	$\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{24}$

pudiendo comprobarse que la suma de todas las probabilidades vale 1.

Por otra parte los posibles premios totales son, respectivamente,

1	2	3
2	3	4
3	4	5
4	5	6
5	6	7
6	7	8

y, denotando por X al premio total que puede consistir, exclusivamente, en uno, dos, tres, cuatro, cinco, seis, siete u ocho euros, se tiene que

$$P[X=1]=\frac{1}{24}$$

$$P[X=2]=\frac{2}{24}+\frac{1}{24}=\frac{3}{24}$$

$$P[X=3]=\frac{1}{24}+\frac{2}{24}+\frac{1}{24}=\frac{4}{24}$$

$$P[X=4]=\frac{1}{24}+\frac{2}{24}+\frac{1}{24}=\frac{4}{24}$$

$$P[X=5]=\frac{1}{24}+\frac{2}{24}+\frac{1}{24}=\frac{4}{24}$$

$$P[X=6]=\frac{1}{24}\frac{2}{24}+\frac{1}{24}=\frac{4}{24}$$

$$P[X=7]=\frac{2}{24}+\frac{1}{24}=\frac{3}{24}$$

$$P[X=8]=\frac{1}{24}$$

Obsérvese que $\sum_{k=1}^{8} P[X=k]=\dfrac{1+3+4+4+4+4+3+1}{24}=1$

b) Si denotamos por A al suceso "ganar 7 euros en cada una de las dos partidas" y por B al suceso "ganar, entre las dos partidas, 14 euros", la probabilidad que nos piden es $P(A|B)$ que podemos calcular como $\dfrac{P(A\cap B)}{P(B)}$. En la siguiente tabla, denotando por X_1 al premio que se puede obtener en la primera partida, y por X_2, al premio que se puede obtener en la segunda partida, se muestran los posibles premios totales obtenidos entre las dos partidas:

$X_2 \backslash X_1$	1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8	9
2	3	4	5	6	7	8	9	10
3	4	5	6	7	8	9	10	11
4	5	6	7	8	9	10	11	12
5	6	7	8	9	10	11	12	13
6	7	8	9	10	11	12	13	14
7	8	9	10	11	12	13	14	15
8	9	10	11	12	13	14	15	16

Así, en términos de X_1 y de X_2 se tiene que

$$P(A \cap B) = P(A) = P\Big[\big(X_1 = 7\big) \cap \big(X_2 = 7\big)\Big] = \frac{3}{24} \cdot \frac{3}{24} = \frac{9}{576}$$

$$\begin{gathered} P(B) = P\Big(\Big[\big(X_1 = 6\big) \cap \big(X_2 = 8\big)\Big] \cup \\ \cup \Big[\big(X_1 = 7\big) \cap \big(X_2 = 7\big)\Big] \cup \Big[\big(X_1 = 8\big) \cap \big(X_2 = 6\big)\Big]\Big) = \\ = P\Big[\big(X_1 = 6\big) \cap \big(X_2 = 8\big)\Big] + \\ + P\Big[\big(X_1 = 7\big) \cap \big(X_2 = 7\big)\Big] + P\Big[\big(X_1 = 8\big) \cap \big(X_2 = 6\big)\Big] = \\ = \frac{4}{24} \cdot \frac{1}{24} + \frac{3}{24} \cdot \frac{3}{24} + \frac{1}{24} \cdot \frac{4}{24} \\ = \frac{4 + 9 + 4}{576} = \frac{17}{576} \end{gathered}$$

y, por tanto, $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{9}{256}}{\frac{17}{256}} = \dfrac{9}{17}$ ♣