Ejercicios T5- REGRESIÓN

1.-El gasto(Y) (en miles de euros) de una empresa depende linealmente y positivamente del número del número de unidades producidas (X) . De información de una muestral de tamaño 100 conocemos que la suma de estos cien gastos fue 300 , la media de la producción 2 , el coeficiente de determinación entre gasto y producción que fue 0,81 y los $a_{2,0} = 16$ **y** $a_{0,2} = 16$. **Con esta información predecir el valor del gasto si conocemos que la producción es de 4 unidades**

Y= a+bX gasto = a + b·ventas (en miles de euros)

$$\bar{x} = 2 \qquad a_{2,0} = 16 \qquad s_x^2 = a_{2,0} - \bar{x}^2 = 16 - 4 = 12$$

$$\bar{y} = 3$$

$$s_y^2 = a_{0,2} - \bar{x}^2 = 16 - 3 = 13$$

$r_{x,y} = \sqrt{R^2} = \sqrt{0,81} = 0,9$ positivo pues relación positiva

$$r_{x,y} = 0,9 = \frac{s_{x,y}}{s_x \cdot s_y} = \frac{s_{x,y}}{\sqrt{12} \cdot \sqrt{13}} \rightarrow s_{xy} = 0,9 \cdot \sqrt{12} \cdot \sqrt{13} = 11,24$$

$$b = \frac{s_{xy}}{s_x^2} = \frac{11,24}{12} = 0,9366 \qquad \bar{y} = a + 0,9366\bar{x}$$

3 = a + 0,9366·2 a = 1,1266 luego la recta de regresión Y = 1,1266 +0,9 X así para X = 4 Y = 4,726 miles de euros

2.-El gasto(Y) (en miles de euros) de una empresa depende linealmente y positivamente del número del número de unidades producidas (X) . De información de una muestral de tamaño 100 conocemos que la suma de estos cien gastos fue 300 , la media de la producción 2 , el coeficiente de determinación entre gasto y producción que fue 0,81 y los $a_{1,1} = 17,24$ **y** $a_{2,0} = 16$. **Con esta información predecir el valor de la producción (x) si conocemos que el gasto será de 5000 euros (y=5)**

Y= a+bX gasto = a + b·ventas (en miles de euros)

$\bar{x} = 2$ $\quad a_{2,0} = 16 \quad s_x^2 = a_{2,0} - \bar{x}^2 = 16 - 4 = 12$

$\bar{y} = 3$

$s_{xy} = a_{1,1} - \bar{x}\bar{y} = 17{,}24 - 2 \cdot 3 = 11{,}24$

$r_{x,y} = \sqrt{R^2} = \sqrt{0{,}81} = 0{,}9$ positivo pues relación positiva

$r_{x,y} = 0{,}9 = \frac{S_{x,y}}{S_x \cdot S_y} = \frac{11{,}24}{S_y \sqrt{12}} \rightarrow S_y = \frac{11{,}24}{0{,}9 \cdot \sqrt{12}} = 3{,}6$ luego $s_y^2 = 13$

$b = \frac{S_{xy}}{S_y^2} = \frac{11{,}24}{13} = 0{,}8646 \qquad \bar{x} = a + 0{,}8645\bar{y}$

2 = a + 0,8645·3 a = -0,5935 luego la recta de regresión x = -0,5935 +0,8645 y

así para y = 5 x = 3,729 unidades

3.- Sean dos rectas de regresión 2x+3y =10 ; x + y = 4

sabiendo además que $a_{0,2} = \Sigma y_i^2 / N = 16$.

a) Averiguar cuál es la regresión X/Y e Y/X. .(Recuérdese que $r^2 = bb'$)

b) Obtener el coeficiente de correlación r . Comentando el resultado.

c) Calcular el centro de gravedad .

d) Obtener la varianza residual S^2_r de la regresión Y/X 3 puntos

opción A 2x+3y =10 es Y= a+ b X por lo que y = 10/3 – 2/3x

x+y =4 es X = a´+b´ Y por lo que x = 4 –y

en este caso r^2 = bb´ = (-1) ·(- 2/3)= 2/3 posible

opción B 2x+3y =10 es X = a´+b´ por lo que x = 10/2 – 3/2y

x+y =4 es Y= a+ b X por lo que y = 4 –x

en este caso r^2 = bb´ = (-3/2) ·(- 1)= 3/2 mayor que uno luego imposible

a) luego 2x+3y = 10 es la recta Y/X y = 10/3 – 2/3x

siendo x + y = 4 la X/Y x = 4 –y

b) el coeficiente de correlación r será : si r^2 = bb´ = (-1) ·(- 2/3)= 2/3

$r_{x,y} = \sqrt{2/3} = 0{,}8164$ con signo negativo pues ambos b son negativos

c) ambas rectas de regresión pasan por el centro de gravedad , luego

$$\left.\begin{array}{l}\bar{y}=\frac{10}{3}-\frac{2}{3}\bar{x}\\ \bar{x}=4-\bar{y}\end{array}\right\}\bar{y}=\frac{10}{3}-\frac{2}{3}(4-\bar{y})=\frac{10}{3}-\frac{8}{3}+\frac{2}{3}\bar{y}=$$

$$\bar{y}=\frac{2}{3}+\frac{2}{3}\bar{y}\quad\rightarrow\frac{1}{3}\bar{y}=\frac{2}{3}\ \rightarrow\bar{y}=2$$

$$luego\ \bar{x}=4-\bar{y}\quad\bar{x}=4-2\quad\rightarrow\bar{x}=2$$

luego centro de gravedad es 2,2

d) varianza residual de regresión Y/X y = 10/3 – 2/3x

sabemos que $S_Y^2 = S_{Y^*}^2 + S_r^2$ conocemos que el porcentaje de la varianza explicada por la regresión es r = 0,8164 luego el no explicado es del 1-0,8164 = 0,1836
por lo que la Varianza residual será el 18,36% de la varianza de la Y

dado que $$\begin{array}{l} s_y^2 = a_{2,0} - \bar{y}^2 = 16 - 2^2 = 12 \\ por\ lo\ que\ s_{r(y/x)}^2 = 0{,}1836 \cdot 12 = 2{,}2 \end{array}$$

4.- A partir de los datos de las siguientes cinco empresas de una determinada industria sobre el número de horas trabajadas en determinado proceso y el número de unidades de producto producidas en ese proceso:

Horas trabajadas	Unidades producidas
4	**20**
6	**25**
7	**35**
3	**20**
5	**25**

Determinar:

a) La recta de regresión que nos permita explicar el número de unidades producidas en función de las horas trabajadas.
b) Analizar la calidad del ajuste, obteniendo la varianza residual y el coeficiente de determinación lineal.Comentar los resultados.
c) Dar una predicción para la producción de una empresa que dedique 8 horas de trabajo a ese proceso.
d) Comparar la variabilidad relativa de las dos variables analizadas.
e) Estudiar la simetría de la primera variable.

Indicadores	Y=unidades	X=horas	
Media	25	5	
Varianzas y covarianza	30	2	7
Desv.Típica	5.477	1.414	

REGRESIÓN	
C.Correlación	0.904
C.Determinación	0.817
Varianza Explicada	24.51
Varianza Residual	5.49
Coeficiente a	7.5
Coeficiente b	3.5
RECTA	Y*= 7.5 + 3.5X

Luego la predicción pedida para X=8 será 35,5. Calidad (bondad del ajuste) 0,817 , el 81,7 de la variabilidad de la Y viene explicada por el modelo.

Indicadores

valores	X =horas	Y =unidades
Media	5	25
Varianza	2	30
Desv.Típica	1.41	5.48
C.V.Pearson	0.28	0.22
C.Asimetría	0:simetrica	0.911:asim.por derecha
C.Curtósis	-1.28:platicúrtica	-0.505:platicúrtica
Cuasi Varianza	2.5	37.5
Cuasi.Desv.Típica	1.58	6.12
C.Asimetría Insesgado	0:simetrica	1.363:asim.por derecha
C.Curtósis Insesgado	-1.18:platicúrtica	2.024:leptocúrtica
número de casos	5	5

Variabilidad relativa viene dada por el coeficiente de variación de Pearson en el caso de X es ligeramente mayor , luego mayor variabilidad relativa , sin embargo la variabilidad absoluta es mayor en la variable Y, lógico si percibimos que las unidades de medida son mayores y a la varianza le afectan las unidades en las que está medida la variable.

La variable X es totalmente simetrica. Miestras que la Y es asimétrica por la derecha ,

5.- Dadas dos variables X e Y con : $\vec{m} = \begin{bmatrix} 50 \\ 10 \end{bmatrix}$ $\vec{V} = \begin{bmatrix} 144 & 100 \\ 100 & 196 \end{bmatrix}$

a) Calcular la recta de regresión X/Y
b) Establecer la bondad del ajuste.
c) Predecir el valor de X conociendo que Y es igual a 12.

Recta X/Y X=a+bY $b=\frac{S_{xy}}{S_y^2}=\frac{100}{196}=0{,}5102$

Dado que $\bar{x}=a+b\bar{y}$ así 50 = a + 0,5102 · 10 donde a =44,898

Luego la recta quedará X= 44,898 + 0,5102

Bondad del ajuste si $r_{xy}=\frac{S_{xy}}{S_x \cdot S_y}=\frac{100}{12\cdot 14}=\frac{100}{168}=0{,}595$ luego $R^2=0{,}3543$

Predicción para Y =12 X=44,898+0,5102·12=51

6.-Dadas dos variables X e Y con de la que se tiene informacion conjunta referente a 50 individuos conocemos :

media de X = 50 , $\sum_{j=1}^{50} y_j=500$, $a_{2,0}=2644$, $S_y^2=196$, $a_{1,1}=600$

a) Calcular la recta de regresión Y/X
b) Establecer la bondad del ajuste.
c) Predecir el valor de Y conociendo que X es igual a 12.

Según datos media x = 50 , media de y = 500/50=10, varianza de x = $a_{2,0}-\bar{x}^2$ = 2644-2500=144

Varianza de y = 196 , covarianza = $a_{1,1}-\bar{x}\cdot\bar{y}=600-10\cdot 50=100$

Recta Y/X Y=a+bX $b=\frac{S_{xy}}{S_x^2}=\frac{100}{144}=0{,}694$

Dado que $\bar{Y}=a+b\bar{X}$ así 10 = a + 0,694 · 50 donde a = -24,72

Luego la recta quedará Y= -24,72 + 0,694 X

Bondad del ajuste si $r_{xy}=\frac{S_{xy}}{S_x \cdot S_y}=\frac{100}{12\cdot 14}=\frac{100}{168}=0{,}595$ luego $R^2=0{,}3543$

Predicción para X =12 Y=-24,898+0,694·12 = -16,57