FONEMATO 1

Se toma una muestra de 50 empresas, observando el número de trabajadores "X" y la producción "Y".

X	5	15	20	30	70
Y	4	10	15	20	40
n_i	15	10	10	5	10

1) Determine la media de "X" mediante un cambio de origen y de escala.
2) Determine la recta de regresión de "Y" sobre "X", el coeficiente de correlación lineal y la varianza residual.

SOLUCIÓN

1) Hacemos $\dfrac{X-20}{5} = Z$:

X	5	15	20	30	70
$Z = \dfrac{X-20}{5}$	-3	-1	0	2	10
n_i	15	10	10	5	10

$\Rightarrow$

$$\Rightarrow \bar{z} = \frac{(-3).15 + (-1).10 + 0.10 + 2.5 + 10.10}{50} = \frac{55}{50} = 1'1 \Rightarrow$$

$$\Rightarrow \bar{x} = 5.\bar{z} + 20 = 25'5$$

2) La recta de regresión de "Y" sobre "X", es

$$y^* = a.x + b = 0'548.x + 2'226$$

$$\bullet\ a = \frac{S_{XY}}{S_X^2} = \frac{302'9}{552'25} = 0'548$$

$$*\ S_X^2 = \frac{5^2.15 + 15^2.10 + 20^2.10 + 30^2.5 + 70^2.10}{50} - 25'5^2 = 552'25$$

$$*\ S_{XY} = a_{11} - \bar{x}.\bar{y} = 716 - 25'5.16'2 = 302'9$$

$$a_{11} = \frac{4.5.15 + 15.10.10 + 20.15.10 + 30.20.5 + 70.40.10}{50} = 716$$

$$\bar{y} = \frac{4.15 + 10.10 + 15.10 + 20.5 + 40.10}{50} = \frac{810}{50} = 16'2$$

$$\bullet\ b = \bar{y} - a.\bar{x} = 16'2 - 0'548.25'5 = 2'226$$

Varianza total : $S_Y^2 = \dfrac{4^2.15 + 10^2.10 + 15^2.10 + 20^2.5 + 40^2.10}{50} - 16'2^2 = 447'36$

Coeficiente de correlación lineal: $r = \dfrac{S_{XY}}{S_X.S_Y} = \dfrac{302'9}{\sqrt{552'25}.\sqrt{447'36}} = 0'608$

Varianza residual: $S_e^2 = (1 - r^2).S_Y^2 = 281'44.$

FONEMATO 2

Determine la recta de regresión de "Y" sobre "X", el coeficiente de correlación lineal y la varianza residual.

X	5	7	9	10
Y	2	3	5	5
n_i	1	2	2	3

SOLUCIÓN

$$y^* = a.x + b = 0'654.x - 1'36$$

$$\bullet\ a = \frac{S_{XY}}{S_X^2} = \frac{1'953}{2'98} = 0'654$$

$$*\ \bar{x} = \frac{5.1 + 7.2 + 9.2 + 10.3}{8} = \frac{67}{8}$$

$$*\ S_X^2 = \frac{5^2.1 + 7^2.2 + 9^2.2 + 10^2.3}{8} - \left(\frac{67}{8}\right)^2 = 2'98$$

$$*\ S_{XY} = a_{11} - \bar{x}.\bar{y} = \frac{292}{8} - \frac{67}{8}.\frac{33}{8} = 1'953$$

$$a_{11} = \frac{5.2.1 + 7.3.2 + 9.5.2 + 10.5.3}{8} = \frac{292}{8}$$

$$\bar{y} = \frac{2.1 + 3.2 + 5.2 + 5.3}{8} = \frac{33}{8}$$

$$\bullet\ b = \bar{y} - a.\bar{x} = \frac{33}{8} - 0'654.\frac{67}{8} = -1'36$$

Varianza total :

$$S_Y^2 = \frac{2^2.1 + 3^2.2 + 5^2.2 + 5^2.3}{8} - \left(\frac{33}{8}\right)^2 = 1'35$$

Coeficiente de correlación lineal:

$$r = \frac{S_{XY}}{S_X.S_Y} = \frac{1'953}{\sqrt{2'98}.\sqrt{1'35}} = 0'97$$

Varianza residual:

$$S_e^2 = (1 - r^2).S_Y^2 = 0'079$$

FONEMATO 3

La evolución temporal de la masa salarial de una empresa se recoge en el siguiente cuadro:

Año	0	1	2	3	4	5
Masa salarial	2'5	3'5	4	4'2	4'3	4'6

1) Determine la recta que explica el salario en función del tiempo, calculando el coeficiente de correlación lineal y la varianza residual.
2) Estime la masa salarial del sexto año.

SOLUCIÓN

1) Siendo "Y" la masa salarial y "X" el tiempo:

$$y^* = a.x + b = 0'37.x + 2'915$$

$$\bullet\ a = \frac{S_{XY}}{S_X^2} = \frac{1'091}{2'91} = 0'37$$

$$*\ \bar{x} = \frac{0+1+2+3+4+5}{6} = 2'5$$

$$*\ S_X^2 = \frac{0^2+1^2+2^2+3^2+4^2+5^2}{6} - 2'5^2 = 2'91$$

$$*\ S_{XY} = a_{11} - \bar{x}.\bar{y} = 10'71 - 2'5.3'85 = 1'091$$

$$a_{11} = \frac{0.2'5+1.3'5+2.4+3.4'2+4.4'3+5.4'6}{6} = 10'71$$

$$\bar{y} = \frac{2'5+3'5+4+4'2+4'3+4'6}{6} = 3'85$$

$$\bullet\ b = \bar{y} - a.\bar{x} = 3'85 - 0'37.2'5 = 2'915$$

Varianza total :

$$S_Y^2 = \frac{2'5^2+3'5^2+4^2+4'2^2+4'3^2+4'6^2}{6} - 3'85^2 = 2'855$$

Coeficiente de correlación lineal:

$$r = \frac{S_{XY}}{S_X.S_Y} = \frac{1'091}{\sqrt{2'91}.\sqrt{2'855}} = 0'337$$

Varianza residual:

$$S_e^2 = (1-r^2).S_Y^2 = 2'448$$

2) Si $x = 6 \Rightarrow y^* = 0'37.6 + 2'915 = 5'135$, que es poco fiable, pues el coeficiente de correlación lineal $r = 0'337$ es próximo a cero.

FONEMATO 4

De un sector productivo formado por 7 empresas se recogen los siguientes datos:

Empresa	A	B	C	D	E	F	G
Producción	15	20	30	50	80	100	150
Empleados (×100)	3	3'2	3'5	4	5	6	8

1) Determine la recta que explica la producción en función del número de empleados, calculando el coeficiente de determinación y la varianza residual.
2) Calcule la productividad marginal del sector por persona empleada.
3) Estime la producción de una empresa con 1000 empleados.

SOLUCIÓN

1) Siendo "Y" la producción y "X" el número de empleados:

$$y^* = a.x + b = 27'15.x + 63'24$$

$$\bullet\; a = \frac{S_{XY}}{S_X^2} = \frac{76'54}{2'819} = 27'15$$

$$* \; \bar{x} = \frac{3 + 3'2 + 3'5 + 4 + 5 + 6 + 8}{7} = 4'67$$

$$* \; S_X^2 = \frac{3^2 + 3'2^2 + 3'5^2 + 4^2 + 5^2 + 6^2 + 8^2}{7} - 4'67^2 = 2'819$$

$$* \; S_{XY} = a_{11} - \bar{x}.\bar{y} = 373'42 - 4'67.63'57 = 76'54$$

$$a_{11} = \frac{15.3 + 20.3'2 + 30.3'5 + 50.4 + 80.5 + 100.6 + 150.8}{7} = 373'42$$

$$\bar{y} = \frac{15 + 20 + 30 + 50 + 80 + 100 + 150}{7} = 63'57$$

$$\bullet\; b = \bar{y} - a.\bar{x} = 63'57 - 27'15.4'67 = 63'24$$

- **Varianza total:** $S_Y^2 = \dfrac{15^2 + 20^2 + 30^2 + \ldots + 150^2}{7} - 63'57^2 = 2090'8$
- **Varianza explicada por la regresión:**

  Siendo $y^* = a.x + b$, es $S_{Y^*}^2 = a^2.S_X^2 = 27'15^2.2'819 = 2077'2$.
- **Varianza residual:** $S_e^2 = S_Y^2 - S_{Y^*}^2 = 2090'8 - 2077'2 = 13'58$.
- **Coeficiente de determinación:** $R^2 = \dfrac{S_{Y^*}^2}{S_Y^2} = \dfrac{2077'2}{2090'8} = 0'993$.

2) Si $y^* = 27'15.x + 63'24$, productividad marginal del sector por cada 100 personas empleada es la pendiente $a = 27'15$.

3) Con 1000 empleados $\Rightarrow x = 10 \Rightarrow y^* = 27'15.10 + 63'24 = 334'74$, que es fiable, pues el coeficiente de determinación lineal $R^2 = 0'993$ es próximo a 1.

FONEMATO 5

Se recogen los siguientes datos relativos a 5 familias:

Renta	50	120	200	400	600
Gasto alimentario	30	40	60	100	120

1) Estime el gasto alimentario mediante una función lineal de la renta, calculando el coeficiente de determinación y la varianza residual.
2) Estime el gasto alimentario para una renta de 250.
3) Para una renta de 250, calcule la elasticidad del gasto respecto a la renta.
4) Calcule la propensión marginal al gasto para una renta de 700.

SOLUCIÓN

1) Siendo "Y" el gasto alimentario y "X" la renta:

$$y^* = a.x + b = 0'17.x + 23'22$$

$$\bullet\ a = \frac{S_{XY}}{S_X^2} = \frac{6880}{40304} = 0'17$$

$$*\ \bar{x} = \frac{50 + 120 + 200 + 400 + 600}{5} = 274$$

$$*\ S_X^2 = \frac{50^2 + 120^2 + 200^2 + 400^2 + 600^2}{5} - 274^2 = 40304$$

$$*\ S_{XY} = a_{11} - \bar{x}.\bar{y} = 26060 - 274.70 = 6880$$

$$a_{11} = \frac{50.30 + 120.40 + 200.60 + 400.100 + 600.120}{5} = 26060$$

$$\bar{y} = \frac{30 + 40 + 60 + 100 + 120}{5} = 70$$

$$\bullet\ b = \bar{y} - a.\bar{x} = 70 - 0'17.274 = 23'22$$

- **Varianza total:** $S_Y^2 = \dfrac{30^2 + 40^2 + 60^2 + 100^2 + 120^2}{5} - 70^2 = 1200$
- **Coeficiente de determinación:** $R^2 = \dfrac{S_{XY}^2}{S_X^2.S_Y^2} = \dfrac{6880^2}{40304.1200} = 0'978$.
- **Varianza residual:** $S_e^2 = (1 - r^2).S_Y^2 = (1 - R^2).S_Y^2 = 25'56$.

2) Si $x = 250 \Rightarrow y^* = 0'17.250 + 23'22 = 65'72$, que es fiable, pues el coeficiente de determinación lineal $R^2 = 0'978$ es próximo a 1.

3) Elasticidad del gasto respecto a la renta si $x = 250$:

$$\left(\frac{dy^*}{dx}.\frac{x}{y^*}\right)_{x=250} = 0'17.\frac{250}{65'72} = 0'647$$

4) Propensión marginal al gasto para una renta de 700: $\left(\dfrac{dy^*}{dx}\right)_{x=700} = 0'17$

FONEMATO 6

Datos sobre la renta "X" de 100 contribuyentes y los impuestos "Y" que pagan:

X / Y	0 – 2	2 – 4	4 – 6
6 – 10	20	0	0
10 – 20	5	30	5
20 – 40	0	20	20

1) Si el modelo impositivo es $Y = C + t.X$, determine el impuesto fijo "C" y el tipo impositivo "t". ¿Es bueno el modelo propuesto?
2) Determine la varianza explicada por la regresión y la varianza residual.
3) Si las rentas aumentan 0'1, ¿cuál es el aumento previsto en la cantidad pagada?

SOLUCIÓN

1) $$y^* = t.x + C = 0'1119.x + 0'806$$

• $$t = \frac{S_{XY}}{S_X^2} = \frac{8'8}{78'64} = 0'1119$$

* $$\bar{x} = \frac{8.20 + 15.40 + 30.40}{100} = 19'6$$

Marca de clase x_i	8	15	30
Frecuencia abs.	20	40	40

* $$S_X^2 = \frac{8^2.20 + 15^2.40 + 30^2.40}{100} - 19'6^2 = 78'64$$

* $$S_{XY} = a_{11} - \bar{x}.\bar{y} = 67'6 - 19'6.3 = 8'8$$

$$a_{11} = \frac{8.1.20 + 15.1.5 + 15.3.30 + 15.5.5 + 30.3.20 + 30.5.20}{100} = 67'6$$

x_i / y_i	1	3	5
8	20	0	0
15	5	30	5
30	0	20	20

$$\bar{y} = \frac{1.25 + 3.50 + 5.25}{100} = 3$$

y_i	1	3	5
Frecuencia abs.	25	50	25

• $$C = \bar{y} - t.\bar{x} = 3 - 0'1119.19'6 = 0'806$$

• El modelo es aceptable, pues el coeficiente de determinación es próximo a 1:

$$R^2 = S_{XY}^2 / S_X^2.S_Y^2 = 8'8^2/(78'64.2) = 0'701$$

$$\boxed{S_Y^2 = \left((1^2.25 + 3^2.50 + 5^2.25)/100\right) - 3^2 = 2}$$

2) Siendo $y^* = t.x + C$, es $S_{Y^*}^2 = t^2.S_X^2 = 0'1119^2.78'64$ ... y $S_e^2 = S_Y^2 - S_{Y^*}^2$.

3) Si la renta pasa de "x" a $x + 0'1$, la variación prevista de impuestos es $0'1.t$.

FONEMATO 7

Datos sobre antigüedad "X" y salario "Y" de los trabajadores de una empresa:

Antigüedad "X"	Salario "Y"	Trabajadores
a	$L_0 - 6$	b
c	$6 - 8$	2
2	$8 - 14$	4
1	$14 - 18$	4

1) Halle L_0, a, b y c sabiendo que $S_{XY} = -7$, que las rectas de regresión se cortan en el punto (3;9'5) y que la distribución de frecuencias relativas de "Y" es

Salario	$L_0 - 6$	$6 - 8$	$8 - 14$	$14 - 18$
f_i	d	0'125	0'25	0'25

2) Si se prescinde del 15% de los empleados con salarios más bajos y del 10% con salarios más altos, ¿entre qué valores están los salarios del 75% restante?

3) Halle la recta de regresión de "Y" sobre "X" y el coeficiente de determinación.

SOLUCIÓN

1) Si la frecuencia relativa de los trabajadores con antigüedad en $6 - 8$ es 0'125 y hay 2 trabajadores en esa situación, el número "N" de trabajadores es 16:

$$\frac{2}{N} = 0'125 \Rightarrow N = 16 \Rightarrow b = N - (2 + 2 + 4) = 16 - 8 = 8.$$

- La suma de las frecuencias relativas de los salarios ha de ser la unidad:

$$d + 0'125 + 0'25 + 0'25 = 1 \Rightarrow d = 0'375$$

- El valor de L_0 lo obtenemos al exigir que el salario medio sea $\overline{y} = 9'5$:

$$\overline{y} = 9'5 \Rightarrow \frac{L_0 + 6}{2}.0'375 + 7.0'125 + 11.0'25 + 16.0'25 = 9'5 \Rightarrow L_0 = 4$$

Salario	$L_0 - 6$	$6 - 8$	$8 - 14$	$14 - 18$
f_i	0'375	0'125	0'25	0'25

- Exijamos que la antigüedad media sea $\overline{x} = 3$:

$$\overline{x} = 3 \Rightarrow \frac{a.6 + c.2 + 2.4 + 1.4}{16} = 3 \Rightarrow 6.a + 2.c = 36 \quad \text{(I)}$$

Antigüedad	a	c	2	1
Trabajadores	6	2	4	4

- Exijamos que $S_{XY} = -7$:

$$-7 = \frac{a.5.6 + c.7.2 + 2.11.4 + 1.16.4}{16} - 9'5.3 \Rightarrow 30.a + 14.c = 192 \quad \text{(II)}$$

Antigüedad "X"	Salario "Y"	Trabajadores
a	$4 - 6$	6
c	$6 - 8$	2
2	$8 - 14$	4
1	$14 - 18$	4

- La solución de (I) y (II) es $a = 5$ y $c = 3$.

2) El percentil 15 de "Y" está en el intervalo $4-6$, que es el primero con frecuencia absoluta acumulada mayor o igual que $15.N/100 = 15.16/100 = 2'4$.

$$P_{15} = L_{i-1} + \frac{2'4 - N_{i-1}}{n_i}.(L_i - L_{i-1}) = 4 + \frac{2'4-0}{6}.(6-4) = 4'8$$

Salario	$4-6$	$6-8$	$8-14$	$14-18$
n_i	6	2	4	4
N_i	6	8	12	16

El percentil 90 de "Y" está en el intervalo $14-18$, que es el primero con frecuencia absoluta acumulada mayor o igual que $90.N/100 = 90.16/100 = 14'4$.

$$P_{90} = L_{i-1} + \frac{14'4 - N_{i-1}}{n_i}.(L_i - L_{i-1}) = 14 + \frac{14'4-12}{6}.(18-14) = 16'4$$

3) Regresión de "Y" (salario) sobre "X" (antigüedad):

$$y^* = a.x + b = -2'54.x + 17'13$$

- $a = \dfrac{S_{XY}}{S_X^2} = \dfrac{-7}{2'75} = -2'54$

  - $* \; S_{XY} = -7$
  - $* \; \overline{x} = 3$
  - $* \; S_X^2 = \dfrac{5^2.6 + 3^2.2 + 2^2.4 + 1^2.4}{16} - 3^2 = 2'75$

Antigüedad	5	3	2	1
Trabajadores	6	2	4	4

- $b = \overline{y} - a.\overline{x} = 9'5 - (-2'54).3 = 17'13$

- **Coeficiente de determinación:**

$$R^2 = \frac{S_{XY}^2}{S_X^2.S_Y^2} = \frac{(-7)^2}{2'75.19'5} = 0'914$$

$$S_Y^2 = \left(5^2.0'375 + 7^2.0'125 + 11^2.025 + 16^2.0'25\right) - 9'5^2 = 19'5$$

Salario	$4-6$	$6-8$	$8-14$	$14-18$
f_i	0'375	0'125	0'25	0'25

FONEMATO 8

La recta de regresión $x^* = 0'15.y + 3$ expresa la relación estadística entre en número "X" de unidades vendidas diariamente de un bien y el gasto mensual "Y" en hacerle publicidad. Se sabe que la covarianza es 22'5 y que la distribución marginal de "X" es la siguiente:

x_i	5	7	10	12
Frecuencia abs.	2	5	3	2

1) Determine las respectivas medias de "X" e "Y", y la varianza de "Y".
2) Determine la recta de regresión de "Y" sobre "X", su coeficiente de determinación y la varianza residual.

SOLUCIÓN

1) Media de "X": $\bar{x} = \dfrac{5.2 + 7.5 + 10.3 + 12.2}{2 + 5 + 3 + 2} = \dfrac{99}{12} = 8'25$

Como la recta de regresión $x^* = 0'15.y + 3$ pasa por el punto $(\bar{x}; \bar{y}) \equiv (8'25; \bar{y})$, es $8'25 = 0'15.\bar{y} + 3$, por lo que $\bar{y} = 35$.

- Siendo $x^* = 0'15.y + 3$, es $\dfrac{S_{XY}}{S_Y^2} = 0'15$, y como $S_{XY} = 22'5$, resulta $S_Y^2 = 150$.

2) Recta de regresión de "Y" sobre "X".

$$y^* = a.x + b = 4'07.x + 1'37$$

- $a = \dfrac{S_{XY}}{S_X^2} = \dfrac{22'5}{5'52} = 4'07$

$* \ S_{XY} = 22'5$

$* \ \bar{x} = 8'5$

$$* \ S_X^2 = \frac{5^2.2 + 7^2.5 + 10^2.3 + 12^2.2}{2 + 5 + 3 + 2} - 8'25^2 = 5'52$$

- $b = \bar{y} - a.\bar{x} = 35 - 4'07.8'5 = 1'37$

- El **coeficiente de determinación** es el producto de los coeficientes de regresión:

$$R^2 = 0'15.4'07 = 0'61$$

- **Varianza residual:**

$$S_e^2 = (1 - R^2).S_Y^2 = 58'425$$

FONEMATO 9

De una distribución $(X;Y)$ se conoce la distribución marginal de "X"

x_i	3	5	8	9
Frecuencia abs.	5	1	2	1

Si $\sum y_j^2.n_{\bullet j} = 3240$ y la recta de regresión de "Y" sobre "X" es $y^* = 5.x - 20$, determine la recta de regresión de "X" sobre "Y", su coeficiente de determinación y la varianza residual.

SOLUCIÓN

Es:

$$\bar{x} = \frac{3.5 + 5.1 + 8.2 + 9.1}{5 + 1 + 2 + 1} = 5$$

$$S_X^2 = \frac{3^2.5 + 5^2.1 + 8^2.2 + 9^2.1}{5 + 1 + 2 + 1} - 5^2 = 6$$

Si $y^* = a.x + b \equiv 5.x - 20$ es la recta de regresión de "Y" sobre "X", entonces:

$$a = \frac{S_{XY}}{S_X^2} = 5 \Rightarrow S_{XY} = 5.S_X^2 = 30$$

$$b = \bar{y} - a.\bar{x} \Rightarrow -20 = \bar{y} - 5.5 \Rightarrow \bar{y} = 5$$

La recta de regresión de "X" sobre "Y" es $x^* = a'.y + b'$, donde:

$$a' = \frac{S_{XY}}{S_Y^2} = \frac{30}{325} = 0'089$$

$$\boxed{S_Y^2 = \left(\frac{1}{n}.\sum y_j^2.n_{\bullet j}\right) - (\bar{y})^2 = \left(\frac{1}{9}.3240\right) - 5^2 = 325}$$

$$b' = \bar{x} - a'.\bar{y} = 5 - 0'089.5 = 4'552$$

El **coeficiente de determinación** es el producto de los coeficientes de regresión:

$$R^2 = 5.0'089 = 0'445$$

Varianza residual:

$$S_e^2 = (1 - R^2).S_Y^2$$

FONEMATO 10

Se sabe que la recta de regresión de "Y" sobre "X" para un conjunto de 10 datos es $y^* = 0'74.x + 0'84$, siendo $S_e^2 = 3'218$ la correspondiente varianza residual. También se sabe que $\sum y_i = 82'4$ y $\sum x_i^2 = 3340$.

Determine la recta de regresión de "X" sobre "Y", y el coeficiente de determinación.

SOLUCIÓN

Si $y^* = a.x + b \equiv 0'74.x + 0'84$ es la recta de regresión de "Y" sobre "X", entonces:

$$b = \overline{y} - a.\overline{x} \Rightarrow 0'84 = 8'24 - 0'74.\overline{x} \Rightarrow \overline{x} = 10$$

$$\boxed{b = 0'84 \ ; \ a = 0'74 \ ; \ \overline{y} = \frac{1}{10}.\sum y_i = \frac{82'4}{10} = 8'24}$$

Es:

$$S_X^2 = \left(\frac{1}{n}.\sum x_i^2\right) - (\overline{x})^2 = \frac{3340}{10} - 10^2 = 234$$

Como $a = 0'74 = \dfrac{S_{XY}}{S_X^2}$ y $S_X^2 = 234$, es $S_{XY} = 234.0'74 = 173'16$.

La recta de regresión de "X" sobre "Y" es $x^* = a'.y + b'$, donde:

$$a' = \frac{S_{XY}}{S_Y^2} = \frac{173'16}{131'356} = 1'318$$

$$\boxed{\begin{array}{c} S_Y^2 = S_{Y^*}^2 + S_e^2 = 128'131 + 3'218 = 131'356 \\ \boxed{\begin{array}{c} y^* = 0'74.x + 0'84 \Rightarrow S_{Y^*}^2 = 0'74^2.S_X^2 = 0'74^2.234 = 128'131 \\ S_e^2 = 3'218 \end{array}} \end{array}}$$

$$b' = \overline{x} - a'.\overline{y} = 10 - 1'318.8'24 = 0'86$$

El **coeficiente de determinación** es el producto de los coeficientes de regresión:

$$R^2 = 0'074.1'1318 = 0'9753$$

FONEMATO 11

Se conocen los siguientes datos relativos a 5 observaciones de la producción "X" y el coste total "Y" de una industria:

$$\sum_{i=1}^{5} x_i = 64 \;\; ; \;\; \sum_{i=1}^{5} y_i = 247 \;\; ; \;\; \sum_{i=1}^{5} x_i . y_i = 3199 \;\; ; \;\; \sum_{i=1}^{5} x_i^2 = 828 \;\; ; \;\; \sum_{i=1}^{5} y_i^2 = 12363$$

1) Determine la recta de regresión de "Y" sobre "X"
2) Estímese el coste si la producción es 15, valorando su bondad.

SOLUCIÓN

1) Regresión de "Y" sobre "X":

$$y^* = a.x + b = 4'25.x - 5$$

$$\boxed{\begin{array}{l} \bullet\; a = \dfrac{S_{XY}}{S_X^2} = \dfrac{\dfrac{3199}{5} - \dfrac{64}{5}.\dfrac{247}{5}}{\dfrac{828}{5} - (\dfrac{64}{5})^2} = 4'25 \\[2ex] \bullet\; b = \overline{y} - a.\overline{x} = \dfrac{247}{5} - 4'25.\dfrac{64}{5} = -5 \end{array}}$$

2) Si $x = 15 \Rightarrow y^* = 4'25.15 - 5 = 58'75$ ... y la estimación es muy fiable, porque el coeficiente de determinación R^2 es muy próximo a 1:

$$R^2 = \frac{S_{XY}^2}{S_X^2 . S_Y^2} = \frac{\dfrac{3199}{5} - \dfrac{64}{5}.\dfrac{247}{5}}{\left(\dfrac{828}{5} - (\dfrac{64}{5})^2\right).\left(\dfrac{12363}{5} - (\dfrac{247}{5})^2\right)} = 0'986$$

FONEMATO 12

En un mercado municipal se hacen 6 observaciones relativas al precio "X" de los malocotones y a la cantidad "Y" vendida de éstos, resultando:

$$\sum_{i=1}^{6} x_i = 84 \;;\; \sum_{i=1}^{6} y_i = 21 \;;\; \sum_{i=1}^{6} x_i . y_i = 273 \;;\; \sum_{i=1}^{6} x_i^2 = 1202 \;;\; \sum_{i=1}^{6} y_i^2 = 91$$

1) Determine la recta de regresión de "Y" sobre "X"
2) Calcule la varianza de "Y", descomponiéndola en suma de la varianza explicada por la regresión y la varianza residual.
3) Calcule el coeficiente de determinación.
4) Estime la cantidad vendida si el precio es 10, valorando su bondad.

SOLUCIÓN

1) Regresión de "Y" sobre "X":

$$y^* = a.x + b = -0'814.x + 14'8$$

- $a = \dfrac{S_{XY}}{S_X^2} = \dfrac{-3'5}{4'3} = -0'814$

$$S_{XY} = \left(\frac{1}{6}.\sum_{i=1}^{6} x_i . y_i\right) - \bar{x}.\bar{y} = \frac{273}{6} - \frac{84}{6}.\frac{21}{6} = -3'5$$

$$S_X^2 = \left(\frac{1}{6}.\sum_{i=1}^{6} x_i^2\right) - (\bar{x})^2 = \frac{1202}{6} - (\frac{84}{6})^2 = 4'3$$

- $b = \bar{y} - a.\bar{x} = \dfrac{21}{6} - (-0'14).\dfrac{84}{6} = 14'8$

2) Es $S_Y^2 = \left(\dfrac{1}{6}.\sum_{i=1}^{6} y_i^2\right) - (\bar{y})^2 = \dfrac{91}{6} - (\dfrac{21}{6})^2 = 2'912$

- Siendo $y^* = a.x + b$, es:

$$S_{Y*}^2 = a^2.S_X^2 = (-0'814)^2.\left(\frac{1202}{6} - (\frac{84}{6})^2\right) = 2'8$$

- Es $S_e^2 = S_Y^2 - S_{Y*}^2 = 0'112$.

3) Es $R^2 = \dfrac{S_{XY}^2}{S_X^2 . S_Y^2} = \dfrac{(-3'5)^2}{4'4.2'912} = 0'982.$

4) Si $x = 10 \Rightarrow y^* = -0'814.x + 14'8 = -0'814.10 + 14'8 = 6'6$ ... y la estimación es muy fiable, porque el coeficiente de determinación R^2 es muy próximo a 1.

FONEMATO 13

De una distribución bidimensional de frecuencias se sabe que:

$$a_{01} = 5 \ ; \ a_{20} = 200 \ ; \ S_Y^2 = 5 \ ; \ S_X = 10 \ ; \ S_{XY} = 10$$

1) Determine regresión de "Y" sobre "X" y el coeficiente de correlación lineal.
2) Calcule la varianza residual de la anterior regresión de dos formas distintas.

SOLUCIÓN

1) La recta de regresión de "Y" sobre "X" es $y^* = a.x + b$, donde:

$$a = \frac{S_{XY}}{S_X^2} = \frac{10}{10^2} = 0'1$$

Al exigir que $S_X^2 = 100$, resulta $\overline{x} = 10$ ó $\overline{x} = -10$:

$$S_X^2 = 100 = a_{20} - (\overline{x})^2 \Rightarrow 100 = 200 - (\overline{x})^2 \Rightarrow \overline{x} = \pm 10$$

Si $\overline{x} = 10$ es $b = \overline{y} - a.\overline{x} = 5 - 0'1.10 = 4$, por lo que $y^* = 0'1.x + 4$.

Si $\overline{x} = -10$ es $b = \overline{y} - a.\overline{x} = 5 - 0'1.(-10) = 6$, por lo que $y^* = 0'1.x + 6$.

2) Es $r^2 = \dfrac{S_{XY}^2}{S_X^2.S_Y^2} = \dfrac{10^2}{10^2.5} = 0'2$, por lo que el ajuste no es bueno.

3) Es $S_e^2 = (1 - R^2).S_Y^2 = (1 - 0'2).5 = 4$.

Siendo $e_i = y_i - (a.x_i + b)$, es:

$$S_e^2 = S_Y^2 + a^2.S_X^2 - 2.a.S_{XY} = 5 + 0'1^2.100 - 2.0'1.10 = 4$$

FONEMATO 14

De una distribución bidimensional de frecuencias se sabe que:

$$m_{11} = 4011'23 \;\; ; \;\; a_{10} = 214'48 \;\; ; \;\; a_{11} = 37761'8$$

$$r^2 = 0'92 \;\; ; \;\; x^* = 0'497.y + k$$

1) Determine la media y la varianza de "Y".
2) Determine "k".
3) Calcule la regresión de "Y" sobre "X" y la correspondiente varianza residual.

SOLUCIÓN

1) Se tiene que:

$$m_{11} = a_{11} - a_{10}.a_{01} \Rightarrow a_{01} = \frac{a_{11} - m_{11}}{a_{10}} = \frac{37761'8 - 4011'23}{214'48} = 157'36$$

Es:

$$0'497 = \frac{S_{XY}}{S_Y^2} \equiv \frac{m_{11}}{S_Y^2} \Rightarrow S_Y^2 = \frac{4011'23}{0'497} = 8070'88$$

2) Si $x^* = 0'497.y + k \Rightarrow k = \overline{x} - 0'497.\overline{y} = 214'48 - 0'497.157'36 = 136'27$.

3) La recta de regresión de "X" sobre "Y" es $y^* = a.x + b$, donde:

$$a = \frac{S_{XY}}{S_X^2} = \frac{4011'23}{2166'9} = 1'85$$

$$\boxed{r^2 = 0'92 = \frac{S_{XY}^2}{S_X^2.S_Y^2} \Rightarrow S_X^2 = \frac{S_{XY}^2}{0'92.S_Y^2} = \frac{(4011'23)^2}{0'92.8070'88} = 2166'9}$$

$$b = \overline{y} - a.\overline{x} = 157'36 - 1'85.214'48 = -239'43$$

- Es $S_e^2 = (1 - r^2).S_Y^2 = (1 - 0'92).8070'88 = 645'67$.

FONEMATO 15

Determine la recta de regresión de "Y" sobre "X" sabiendo que:

$$n = 26 \ ; \ \sum_{i=1}^{26} x_i = -124 \ ; \ \sum_{i=1}^{26} y_i = 64 \ ; \ \sum_{i=1}^{26} x_i.y_i = 17365$$

$$\sum_{i=1}^{26} x_i^2 = 25038 \ ; \ \sum_{i=1}^{26} y_i^2 = 14728$$

SOLUCIÓN

- Regresión de "Y" sobre "X":

$$y^* = a.x + b = 0'7228.x + 5'9$$

- $a = \dfrac{S_{XY}}{S_X^2} = \dfrac{679'62}{940'25} = 0'7228$

$$S_{XY} = \left(\frac{1}{26}.\sum_{i=1}^{26} x_i.y_i\right) - \overline{x}.\overline{y} = \frac{17365}{26} - (-\frac{124}{26}).\frac{64}{26} = 679'62$$

$$S_X^2 = \left(\frac{1}{26}.\sum_{i=1}^{26} x_i^2\right) - (\overline{x})^2 = \frac{25038}{26} - (-\frac{124}{26})^2 = 940'25$$

- $b = \overline{y} - a.\overline{x} = \dfrac{64}{26} - 0'7228.(-\dfrac{124}{26}) = 5'9$

- Es $S_Y^2 = \left(\dfrac{1}{26}.\displaystyle\sum_{i=1}^{26} y_i^2\right) - (\overline{y})^2 = \dfrac{14728}{26} - (\dfrac{64}{26})^2 = 560'4$

- Es $R^2 = \dfrac{S_{XY}^2}{S_X^2.S_Y^2} = \dfrac{679'62}{940'25.560'4} = 0'876.$

FONEMATO 16

Las rectas de regresión de dos variables son

$$x + 4.y = 1 \;\; ; \;\; x + 5.y = 2$$

Calcule las medias de dichas variables y el coeficiente de determinación.

SOLUCIÓN

El valor absoluto de la pendiente de $x + 4.y = 1$ es $1/4$, y el valor absoluto de la pendiente $x + 5.y = 2$ es $1/5$.

La recta con menor pendiente en valor absoluto (o sea, la recta $x + 5.y = 2$) es la de regresión de "Y" sobre "X":

$$y^* = -\frac{1}{5}.x + \frac{2}{5}$$

La recta de regresión de "X" sobre "Y" es $x^* = -4.y + 1$.

El punto $(\overline{x}\,;\overline{y})$ es la solución del sistema que formas las ecuaciones de las dos rectas; resulta $\overline{x} = -3$, $\overline{y} = 1$.

El coeficiente de determinación R^2 es el producto de las pendientes de las rectas de regresión:

$$R^2 = (-\frac{1}{5}).(-4) = \frac{4}{5}$$

FONEMATO 17

Analice si son posibles las siguientes situaciones:

1) $r = -0'5,\ y = x + 6$

2) $S_{XY} = 100,\ S_x^2 = 25,\ S_Y = 20,\ S_{Y*}^2 = S_Y^2$

3) $y = 5.x + 8,\ y = 9 + (x/5),\ r = 0'2$

4) $2.y = x + 8,\ y = x - 4,\ \overline{x} = 16,\ \overline{y} = 12$

SOLUCIÓN

1) No es posible, pues como $y = x + 6$ tiene pendiente positiva, ha de ser $r > 0$.

2) Es posible, pues siendo $y* = a.x + b$, es:

$$S_{Y*}^2 = a^2.S_x^2 = \left(\frac{S_{XY}}{S_X^2}\right)^2.S_x^2 = \frac{S_{XY}^2}{S_X^2} = \frac{100^2}{25} = 400 = S_Y^2$$

3) No es posible.

La pendiente de $y = 5.x + 8$ es 5, y la de $y = 9 + (x/5)$ es 1/5. La de menor pendiente en valor absoluto $y = 9 + (x/5)$ es la de regresión de "Y" sobre "X":

$$y* = 9 + \frac{x}{5} \Rightarrow a = \frac{1}{5}$$

La recta de regresión de "X" sobre "Y" es $y = 5.x + 8$; o sea:

$$x* = \frac{1}{5}.y - 8 \Rightarrow a' = \frac{1}{5}$$

Así, es $a.a' = 1 = r^2 \Rightarrow r = \pm\sqrt{1} \neq 0'2$

4) Es posible, pues la solución del sistema $\left\{\begin{matrix} 2.y = x + 8 \\ y = x - 4 \end{matrix}\right\}$ es $x = 16,\ y = 12$.

FONEMATO 18

Demuestre que si $Y*$ es el valor teórico obtenido mediante la recta de regresión de "Y" sobre "X", sucede que $|r_{Y*Y}| = |r_{XY}|$.

SOLUCIÓN

$$|r_{Y*Y}| = \left|\frac{S_{Y*Y}}{S_Y.S_{Y*}}\right| \underset{\uparrow}{=} \left|\frac{a.S_{XY}}{S_Y.(|a|.S_x)}\right| = \left|\frac{S_{XY}}{S_Y.S_x)}\right| = |r_{XY}|$$

$$\boxed{Y* = a.X + b \Rightarrow \begin{cases} S_{Y*} = |a|.S_x \\ S_{Y*Y} = a.S_{XY} \end{cases}}$$

FONEMATO 19

Analice si son posibles las siguientes situaciones:

1) $y = 2.x + 4,\ \ y = 3.x + 2,\ \ r = 2/3$

2) $S_x^2 = 20,\ S_{Y*}^2 = 30,\ \ y* = 2.x + b$

3) $y = 2.x + 3,\ \ r = 0$

4) $r = -0'4,\ \ y = 2.x + 3$

SOLUCIÓN

1) No es posible.

La pendiente de $y = 2.x + 4$ es 2, y la de $y = 3.x + 2$ es 3. La de menor pendiente en valor absoluto, $y = 2.x + 4$, es la de regresión de "Y" sobre "X":

$$y* = 2.x + 4 \Rightarrow a = 2$$

La recta de regresión de "X" sobre "Y" es $y = 3.x + 2$ o sea:

$$x* = \frac{1}{3}.y - \frac{2}{3} \Rightarrow a' = \frac{1}{3}$$

Así, es $a.a' = 2/3 = r^2 \Rightarrow r = \pm\sqrt{2/3} \neq 2/3$

2) Es posible, pues siendo $y* = a.x + b = 2.x + b$, es:

$$S_{Y*}^2 = a^2.S_x^2 = 2^2.S_x^2 = 2^2.20 = 80 \neq 30$$

2) No es posible: si $r = 0 \Rightarrow a = a' = 0$.

4) Es posible, pues pendiente de $y = 2.x + 3$ tiene distinto signo que $r = -0'4$.

FONEMATO 20

1) ¿Es posible que $y^* = (2 - x)/3$, $x^* = 1 - 4.y$?

2) Si $2.x - y = 1$ en una recta de regresión, puede ser negativo "r"?

3) Si $x + 2.y = 1$ y $3.x + 5.y = 2$ son las rectas de regresión, calcule $\overline{x}$ e $\overline{y}$

4) Si entre "X" e "Y" hay correlación positiva, ¿cómo es la correlación entre $U = 3 - 2.X$ y $V = 4 + 3.Y$?

5) ¿Es cierto que $S_e^2 = S_Y^2 - a.S_{XY}$?

SOLUCIÓN

1) No es posible.

Si $a = -1/3$ y $a' = -4$ es $a.a' = r^2 = 4/3 > 1$, lo que no puede ser, pues siempre sucede que $|r| \leq 1$.

2) No es posible: como $2.x - y = 1$ tiene pendiente positiva, es $r > 0$.

3) Es el punto $(-1;1)$ de intersección de las rectas.

4) Negativa, pues $r_{UV} = \dfrac{S_{UV}}{S_U.S_V} = \dfrac{(-2).3.S_{XY}}{(2.S_X).(3.S_Y)} = -\dfrac{S_{XY}}{S_X.S_Y} = -r_{XY}$.

- Sí: $$S_e^2 = S_Y^2 - S_{Y*}^2 = S_Y^2 - a^2.S_X^2 = S_Y^2 - a.a.S_X^2 =$$ $$= S_Y^2 - a.\frac{S_{XY}}{S_X^2}S_X^2 = S_Y^2 - a.S_{XY}$$

FONEMATO 21

Se sabe que la recta de regresión de "Y" sobre "X" es $y^* = a.x + 4$, y su coeficiente de determinación es 0'8; además, $S_{Y*}^2 = 16$ y las rectas de regresión se cortan en el punto (1;2). Determine las varianzas de "X" y de "Y", y la covarianza. Estime el valor de "X" si $Y = 2$.

SOLUCIÓN

- $R^2 = \dfrac{S_{Y*}^2}{S_Y^2} \Rightarrow S_Y^2 = \dfrac{S_{Y*}^2}{R^2} = \dfrac{16}{0'8} = 20.$
- Es $\overline{x} = 1$ e $\overline{y} = 2$; así, si $y^* = a.x + 4$ es: $b \equiv 4 = \overline{y} - a.\overline{x} \Rightarrow 4 = 2 - a \Rightarrow a = -2$
- Si $y^* = -2.x + 4$, es $S_{Y*}^2 = (-2)^2.S_X^2 \Rightarrow 16 = 4.S_X^2 \Rightarrow S_X^2 = 4.$
- $a = -2 = S_{XY}/S_X^2 \Rightarrow S_{XY} = -2.S_X^2 = -8.$
- Siendo $a' = \dfrac{S_{XY}}{S_Y^2} = \dfrac{-8}{20} = -0'4$ y $b' = \overline{x} - a'.\overline{y} = 1 - (-0'4).2 = 1'8$, es: $$x^* = a'.y + b' = -0'4.y + 1'8$$

Si $y = 2$ la estimación es $x^* = -0'4.2 + 1'8 = 1$ .... y estamos matando moscas a cañonazos, pues dicen que las rectas de regresión se cortan en el punto (1;2).