EJERCICIOS PROPUESTOS

15.1 Copia y completa la siguiente tabla.

Y \ X	A	B	C	Total
a	2		1	4
b		2	2	5
c			0	7
Total	7	6		

$\Rightarrow$

Y \ X	A	B	C	Total
a	2	**1**	1	4
b	**1**	2	2	5
c	**4**	**3**	0	7
Total	7	6	**3**	**16**

a) ¿Qué porcentaje de datos presentan la característica *B* en la variable unidimensional *X*?

b) ¿Qué porcentaje de datos presentan la característica *c* en la variable unidimensional *Y*?

c) ¿Qué porcentaje de datos presentan la característica (*B*, *c*) en la variable bidimensional (*X*, *Y*)?

a) Porcentaje de datos que presenta la característica *B* en la variable unidimensional $X = \dfrac{6}{16} \cdot 100 = 37{,}5\,\%$

b) Porcentaje de datos que presenta la característica *c* en la variable unidimensional $Y = \dfrac{7}{16} \cdot 100 = 43{,}75\,\%$

c) Porcentaje de datos que presenta la característica (*B*, *c*) en la variable bidimensional $(X, Y) = \dfrac{3}{16} \cdot 100 = 18{,}75\,\%$

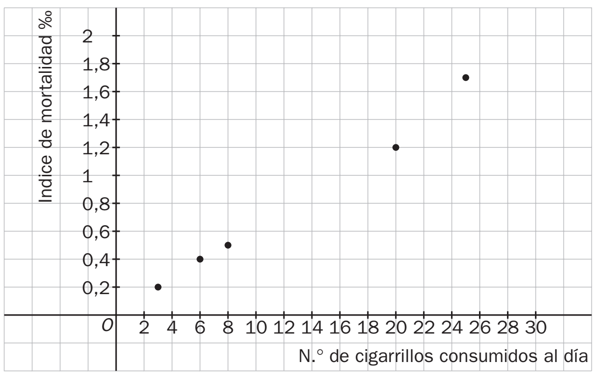
15.2 Observa la siguiente variable bidimensional.

N.° de cigarrillos consumidos al día	3	6	8	20	25
Índice de mortalidad	0,2	0,4	0,5	1,2	1,7

a) Representa la nube de puntos.

b) Indica el tipo de correlación.

a) La nube de puntos es la siguiente:



b) Como al aumentar el número de cigarrillos consumidos aumenta el índice de mortalidad, la correlación es positiva.

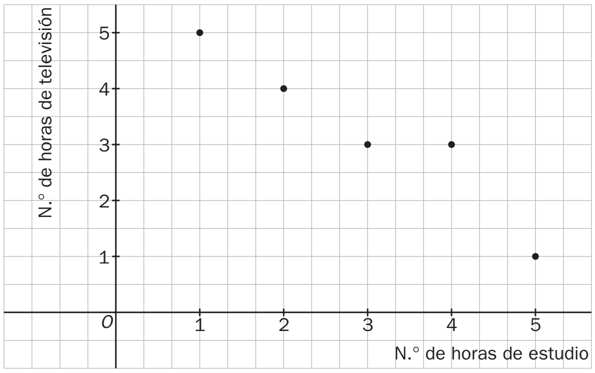
15.3 Una variable bidimensional viene dada por la siguiente tabla.

N.º de horas de estudio	1	2	3	4	5
N.º de horas de televisión	5	4	3	3	1

a) Representa el diagrama de dispersión.

b) Indica el tipo de correlación.

a) La nube de puntos es la siguiente:



b) Como al aumentar el número de horas de estudio disminuye el número de horas de televisión, la correlación es negativa.

15.4 Una variable bidimensional viene dada por la siguiente tabla.

X	3	6	8	20	25
Y	0,2	0,4	0,5	1,2	1,7

a) Calcula las medias y las desviaciones típicas de las variables *X* e *Y*.

b) Calcula la covarianza de la variable (*X, Y*).

Consideramos la siguiente tabla:

N.º de cigarrillos x_i	Índice de mortalidad y_i	x_i^2	y_i^2	x_iy_i
3	0,2	9	0,04	0,6
6	0,4	36	0,16	2,4
8	0,5	64	0,25	4
20	1,2	400	1,44	24
25	1,7	625	2,89	42,5
62	4	1134	4,78	73,5

a) $\overline{x} = \dfrac{62}{5} = 12{,}40$

$\overline{y} = \dfrac{4}{5} = 0{,}8$

$s_X = \sqrt{\dfrac{1134}{5} - 12{,}4^2} = 8{,}5463$

$s_Y = \sqrt{\dfrac{4{,}78}{5} - 0{,}8^2} = 0{,}5621$

b) $s_{XY} = \dfrac{73{,}5}{5} - 12{,}4 \cdot 0{,}8 = 4{,}78$

15.5 Dados los siguientes valores de una variable bidimensional:

X	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Y	**5**	**4**	**3**	**3**	**1**	**3**	**4**	**6**	**3**

a) Halla las medias y las desviaciones típicas de las variables *X* e *Y*.

b) Calcula la covarianza de la variable (*X*, *Y*).

Consideramos la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	x_iy_i
1	5	1	25	5
2	4	4	16	8
3	3	9	9	9
4	3	16	9	12
5	1	25	1	5
6	3	36	9	18
7	4	49	16	28
8	6	64	36	48
9	3	81	9	27
45	32	285	130	160

a) $\overline{x} = \frac{45}{9} = 5$

$\overline{y} = \frac{32}{9} = 3,\widehat{5}$

$s_X = \sqrt{\frac{285}{9} - 5^2} = 2,582$

$s_Y = \sqrt{\frac{130}{9} - 3,\widehat{5}^2} = 1,3426$

b) $s_{XY} = \frac{160}{9} - 5 \cdot 3,\widehat{5} = 0$

15.6 Dada la siguiente variable bidimensional, calcula el coeficiente de correlación.

X	**3**	**5**	**6**	**7**	**8**	**9**	**10**
Y	**2**	**4**	**10**	**5**	**2**	**6**	**4**

Consideramos la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	x_iy_i
3	2	9	4	6
5	4	25	16	20
6	10	36	100	60
7	5	49	25	35
8	2	64	4	16
9	6	81	36	54
10	4	100	16	40
48	33	364	201	231

$\overline{x} = \frac{48}{7} = 6,8571$

$\overline{y} = \frac{33}{7} = 4,7143$

$s_X = \sqrt{\frac{364}{7} - 6,8571^2} = 2,2316$

$s_Y = \sqrt{\frac{201}{7} - 4,7143^2} = 2,5475$

$s_{XY} = \frac{231}{7} - \overline{x} \cdot \overline{y} = 0,6736$

$r = \frac{s_{XY}}{s_X s_Y} = 0,1185$

15.7 Calcula el coeficiente de correlación e interprétalo.

X	6	8	11	14	15	18	20
Y	3	4	2	5	6	7	8

Consideramos la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	x_iy_i
6	3	36	9	18
8	4	64	16	32
11	2	121	4	22
14	5	196	25	70
15	6	225	36	90
18	7	324	49	126
20	8	400	64	160
92	35	1366	203	518

$$\overline{x} = \frac{92}{7} = 13{,}1429$$

$$\overline{y} = \frac{35}{7} = 5$$

$$s_X = \sqrt{\frac{1366}{7} - 13{,}1429^2} = 4{,}7336$$

$$s_Y = \sqrt{\frac{203}{7} - 5^2} = 2$$

$$s_{XY} = \frac{518}{7} - \overline{x} \cdot \overline{y} = 8{,}2855$$

$$r = \frac{s_{XY}}{s_X s_Y} = 0{,}8752$$

Como $r = 0{,}8752 \Rightarrow$ las variables tienen una correlación positiva media.

15.8 Una variable bidimensional viene dada por la siguiente tabla.

X	13	14	11	13	14	14	15	22
Y	54	52	54	53	53	50	49	42

a) Calcula el coeficiente de correlación.

b) Halla la recta de regresión.

c) Si $x = 12$, ¿cuánto valdrá y?

d) ¿Es fiable esta predicción? Justifícalo.

x_i	y_i	x_i^2	y_i^2	x_iy_i
13	54	169	2916	702
14	52	196	2704	728
11	54	121	2916	594
13	53	169	2809	689
14	53	196	2809	742
14	50	196	2500	700
15	49	225	2401	735
22	42	484	1764	924
116	407	1756	20 819	5814

a) $$\overline{x} = \frac{116}{8} = 14{,}5$$

$$\overline{y} = \frac{407}{8} = 50{,}875$$

$$s_X = \sqrt{\frac{1756}{8} - 14{,}5^2} = 3{,}0414$$

$$s_Y = \sqrt{\frac{20\,819}{8} - 50{,}875^2} = 3{,}7562$$

$$s_{XY} = \frac{5814}{8} - \overline{x} \cdot \overline{y} = -10{,}9375$$

$$r = \frac{s_{XY}}{s_X s_Y} = -0{,}9574$$

b) $y - \overline{y} = \frac{s_{XY}}{s_X^2}(x - \overline{x}) \Rightarrow y - 50{,}875 = \frac{-10{,}9375}{9{,}2501}(x - 14{,}5) \Rightarrow y = 50{,}875 - 1{,}1824(x - 14{,}5)$

c) Para $x = 12 \Rightarrow y = 50{,}875 - 1{,}1824(12 - 14{,}5) \Rightarrow y = 53{,}831$

d) La predicción es fiable pues el valor del coeficiente de correlación, $-0{,}9574$, está muy próximo a -1.

15.9 Halla, usando la recta de regresión, los valores de Y para 2, 4, 10, 12, 14, 16 y 18 años. Comprueba que se aproximan lo suficiente a los valores reales y calcula el error relativo cometido.

La recta de regresión es: $y = 138{,}29 + 5{,}605(x - 10{,}86)$.

Para $x = 2 \Rightarrow y = 88{,}6297 \Rightarrow$ Error relativo $= \dfrac{|90 - 88{,}6297|}{90} = 0{,}0152$

Para $x = 4 \Rightarrow y = 99{,}8397 \Rightarrow$ Error relativo $= \dfrac{|100 - 99{,}8397|}{100} = 1{,}603 \cdot 10^{-3}$

Para $x = 10 \Rightarrow y = 133{,}4697 \Rightarrow$ Error relativo $= \dfrac{|130 - 133{,}4697|}{130} = 0{,}0267$

Para $x = 12 \Rightarrow y = 144{,}6797 \Rightarrow$ Error relativo $= \dfrac{|145 - 144{,}6797|}{145} = 2{,}2090 \cdot 10^{-3}$

Para $x = 14 \Rightarrow y = 155{,}8897 \Rightarrow$ Error relativo $= \dfrac{|155 - 155{,}8897|}{155} = 5{,}74 \cdot 10^{-3}$

Para $x = 16 \Rightarrow y = 167{,}0997 \Rightarrow$ Error relativo $= \dfrac{|170 - 167{,}0997|}{170} = 0{,}0171$

Para $x = 18 \Rightarrow y = 178{,}3097 \Rightarrow$ Error relativo $= \dfrac{|178 - 178{,}3097|}{178} = 1{,}7399 \cdot 10^{-3}$

15.10 ¿Qué ocurre al tomar pocos valores para construir la recta de regresión? Halla la recta usando sólo las tallas correspondientes a los 10 y 12 años. Comprueba si la aproximación es mejor o peor que en el caso anterior.

Al tomar pocos valores obtendremos una aproximación peor.

Calculemos la recta de regresión para las tallas correspondientes a los 10 y 12 años.

Para ello, consideremos la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	x_iy_i
10	130	100	16 900	1300
12	145	144	21 025	1740
22	275	244	37 925	3040

$\overline{x} = \dfrac{22}{11} = 11$

$\overline{y} = \dfrac{275}{2} = 137{,}5$

$s_{XY} = \dfrac{3040}{2} - \overline{x} \cdot \overline{y} = 7{,}5$

$s_X^2 = \dfrac{244}{2} - 11^2 = 1$

Luego la ecuación de la recta de regresión será:

$$y - \overline{y} = \frac{s_{XY}}{s_X^2}(x - \overline{x}) \Rightarrow y - 137{,}5 = \frac{7{,}5}{1}(x - 11) \Rightarrow y = 137{,}5 + 7{,}5(x - 11)$$

Si hallamos los valores de Y para 2, 4, 10, 12, 14, 16 y 18 usando la ecuación de esta recta de regresión, comprobaremos que, efectivamente, las aproximaciones son peores que en el caso anterior.

E J E R C I C I O S   P A R A   E N T R E N A R S E

Variables bidimensionales, dependencia y diagramas de dispersión

15.11 Sea la siguiente tabla de doble entrada.

$Y \backslash X$	2	3	5	7	Total
4	3	1	4	1	9
7	1	3	1	1	6
11	1	1	3	5	10
13	1	3	1	2	7
Total	6	8	9	9	32

Copia y completa las siguientes frases.

a) La frecuencia absoluta de (5, 11) es…

b) El número de puntos del tipo (3, y) es…

c) El número de puntos del tipo (x, 13) es…

d) El punto de mayor frecuencia absoluta es…

e) El punto .... tiene frecuencia 4.

a) 3 b) 8 c) 7 d) (7, 11) e) (5, 4)

15.12 Elabora una tabla de doble entrada a partir de la siguiente variable bidimensional.

X	1	3	6	10	12
Y	5	10	18	22	25
f_i	1	2	4	3	2

a) ¿Qué porcentaje de datos representa el valor (6, 18) dentro del conjunto de datos de la variable (X, Y)?

b) Calcula $\overline{x}$, s_X.

$Y \backslash X$	1	3	6	10	12	Total
5	1					1
10		2				2
18			4			4
22				3		3
25					2	2
Total	1	2	4	3	2	12

a) $\frac{4}{12} \cdot 100 = 33,\widehat{3}\%$

b) $\overline{x} = \frac{\sum x_i f_i}{N} = \frac{1 + 6 + 24 + 30 + 24}{12} = \frac{85}{12} = 7,08$

$$s_X = \sqrt{\frac{\sum x_i^2 f_i}{N} - \overline{x}^2} = \sqrt{\frac{1 + 18 + 144 + 300 + 288}{12} - 7,08^2} = \sqrt{62,58 - 50,13} = \sqrt{12,45} = 3,53$$

15.13 Determina la media y la desviación típica de las variables *X* e *Y*, y representa la nube de puntos de la siguiente distribución.

Y \ X	0	2	4	6	8
1	3		4	1	
3		3	3		1
5	6	2		4	1
7	5		4		2
9	4		4		1

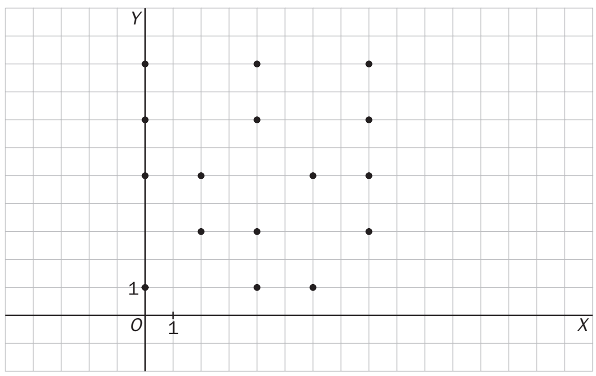
¿Qué tipo de relación existe entre ambas variables?

$$\overline{x} = \frac{\sum x_i f_i}{N} = \frac{0 + 10 + 60 + 30 + 40}{48} = 2{,}9$$

$$s_X = \sqrt{\frac{\sum x_i^2 f_i}{N} - \overline{x}^2} = \sqrt{\frac{0 + 20 + 240 + 180 + 320}{48} - 2{,}9^2} = \sqrt{15{,}8 - 8{,}41} = \sqrt{7{,}39} = 2{,}7$$

$$\overline{y} = \frac{\sum y_i f_i}{N} = \frac{8 + 21 + 65 + 77 + 81}{48} = 5{,}25$$

$$s_Y = \sqrt{\frac{\sum y_i^2 f_i}{N} - \overline{y}^2} = \sqrt{\frac{8 + 63 + 325 + 539 + 729}{48} - 5{,}25^2} = \sqrt{34{,}7 - 27{,}6} = 2{,}7$$



Como se observa en el diagrama de dispersión, las variables no están relacionadas. Estamos ante un caso de independencia.

15.14 Indica qué tipo de relación tienen las variables bidimensionales (X, Y_1); (X, Y_2), y (X, Y_3).

X	−3	−1	2	4	5	7
Y_1	9	5	−1	−5	−7	−11
Y_2	4	3	2	1	0	−1
Y_3	−2	2	−1	1	5	3

(X, Y_1) $\rightarrow$ Dependencia funcional $\Rightarrow y = -2x + 3$.

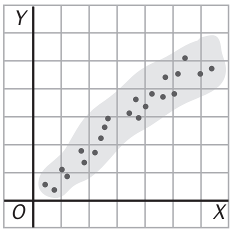
(X, Y_2) $\rightarrow$ Correlación negativa y fuerte.

(X, Y_3) $\rightarrow$ Correlación positiva y débil.

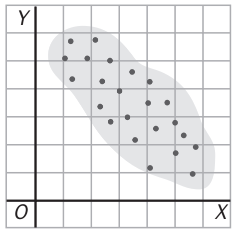
Covarianza y correlación

15.15 Asocia cada índice de correlación con el diagrama de dispersión correspondiente.

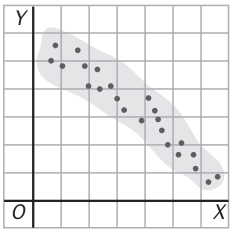
$r = 1$ **$r = 0,92$** **$r = -0,25$** **$r = -0,78$**

a) 

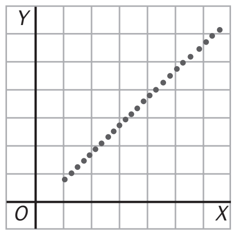
$r = 0,92$

c) 

$r = -0,25$

b) 

$r = -0,78$

d) 

$r = 1$

15.16 La siguiente tabla muestra los valores de una variable bidimensional.

X	**0,25**	**1,32**	**1,24**	**0,17**	**−0,12**
Y	**0,33**	**0,63**	**1,55**	**0,46**	**0,21**

a) Calcula el coeficiente de correlación.

b) Indica el tipo de correlación que existe entre ambas variables.

a) $r = \frac{s_{XY}}{s_X s_Y}$

$$\overline{x} = \frac{2,86}{5} = 0,5720$$

$$\overline{y} = \frac{3,18}{5} = 0,636$$

$$s_{XY} = \frac{0,08 + 0,83 + 1,92 + 0,078 - 0,025}{5} - 0,572 \cdot 0,636 = 0,5778 - 0,3638 = 0,2140$$

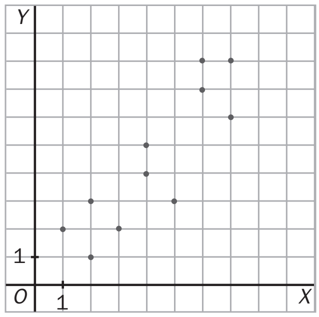
$$s_X = \sqrt{\frac{0,0625 + 1,7424 + 1,5376 + 0,0289 + 0,0144}{5} - 0,57^2} = \sqrt{0,6772 - 0,3272} = 0,5916$$

$$s_Y = \sqrt{\frac{0,1089 + 0,3969 + 2,4025 + 0,2116 + 0,0441}{5} - 0,636^2} = \sqrt{0,6328 - 0,4047} = 0,4778$$

$$r = \frac{0,2140}{0,5916 \cdot 0,4778} = 0,7572$$

b) Entre las dos variables existe una correlación positiva media.

15.17 Dado el siguiente diagrama de dispersión:



a) Elabora una tabla de doble entrada.

b) ¿Qué tipo de correlación tienen las dos variables? ¿Fuerte o débil? ¿Positiva o negativa?

c) ¿Qué coeficiente de correlación de los indicados se ajustaría mejor a la nube de puntos $r = -0{,}91$, $r = 0{,}35$, $r = 0{,}92$? Compruébalo calculando numéricamente dicho coeficiente.

a)

Y \ X	1	2	3	4	5	6	7	Total
1		1						1
2	1		1					2
3		1						1
4				1	1			2
5				1				1
6							1	1
7						1		1
8						1	1	2
Total	1	2	1	2	1	2	2	11

b) Correlación positiva y fuerte.

c) Se ajusta más $r = 0{,}92$.

$$r = \frac{s_{XY}}{s_X s_Y}$$

$$\overline{x} = \frac{1 + 4 + 3 + 8 + 5 + 12 + 14}{11} = 4{,}3$$

$$\overline{y} = \frac{1 + 4 + 3 + 8 + 5 + 6 + 7 + 16}{11} = 4{,}5$$

$$s_{XY} = \frac{2 + 2 + 6 + 6 + 16 + 20 + 20 + 42 + 48 + 42 + 56}{11} - 4{,}3 \cdot 4{,}5 = \frac{260}{11} - 19{,}35 = 4{,}29$$

$$s_X = \sqrt{\frac{1 + 8 + 9 + 32 + 25 + 72 + 98}{11} - 4{,}3^2} = \sqrt{22{,}27 - 18{,}49} = \sqrt{3{,}8} = 1{,}9$$

$$s_Y = \sqrt{\frac{1 + 8 + 9 + 32 + 25 + 36 + 49 + 128}{11} - 4{,}5^2} = \sqrt{26{,}18 - 20{,}25} = \sqrt{5{,}93} = 2{,}43$$

$$r = \frac{4{,}29}{1{,}9 \cdot 2{,}43} = \frac{4{,}29}{4{,}617} = 0{,}929 \text{ (como ya se había intuido)}$$

15.18 La relación entre dos variables viene dada en la siguiente tabla.

X	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Y	**4**	**9**	**16**	**24**	**30**	**34**	**38**	**42**

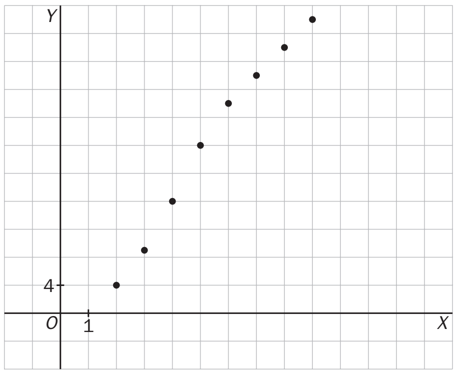
a) Dibuja la nube de puntos asociada a la tabla.

b) Elabora una tabla de doble entrada.

c) Halla $\overline{x}$, $\overline{y}$, s_{XY}.

d) Calcula el coeficiente de correlación lineal. ¿Cómo es la correlación?

a)



b)

Y \ X	2	3	4	5	6	7	8	9	Total
4	1								1
9		1							1
16			1						1
24				1					1
30					1				1
34						1			1
38							1		1
42								1	1
Total	1	1	1	1	1	1	1	1	8

c) $\overline{x} = \dfrac{44}{8} = 5,5$

$\overline{y} = \dfrac{197}{8} = 24,625$

$s_{XY} = \dfrac{8 + 27 + 64 + 120 + 180 + 238 + 304 + 378}{8} - 5,5 \cdot 24,625 = \dfrac{1319}{8} - 135,4375 = 29,4375$

d) $r = \dfrac{s_{XY}}{s_X s_Y}$

$s_X = \sqrt{\dfrac{4 + 9 + 16 + 25 + 36 + 49 + 64 + 81}{8} - 5,5^2} = \sqrt{35,5 - 30,25} = \sqrt{5,25} = 2,2913$

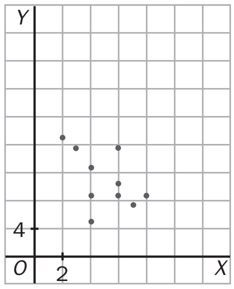
$s_Y = \sqrt{\dfrac{16 + 81 + 256 + 576 + 900 + 1156 + 1444 + 1764}{8} - 24,625^2} = \sqrt{167,7344} = 12,9512$

$r = \dfrac{29,4375}{2,2913 \cdot 12,9512} = 0,9920$

La correlación es positiva y muy fuerte.

Rectas de regresión y estimaciones

15.19 **Observa el siguiente diagrama de dispersión y calcula la recta de regresión.**



Consideramos la siguiente tabla:

x_i	y_i	x_iy_i	x_i^2	y_i^2
2	17	34	4	289
4	5	20	16	25
4	13	52	16	169
6	11	66	36	121
7	7	49	49	49
3	15	45	9	225
4	9	36	16	81
6	9	54	36	81
6	15	90	36	225
8	9	72	64	81
50	110	518	282	1346

$\overline{x} = \frac{50}{10} = 5$

$\overline{y} = \frac{110}{10} = 11$

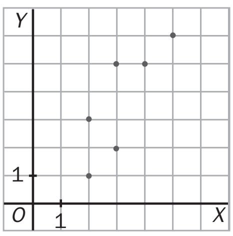
$s_X^2 = \frac{282}{10} - 5^2 = 3,2$

$s_{XY} = \frac{518}{10} - \overline{x} \cdot \overline{y} = -3,2$

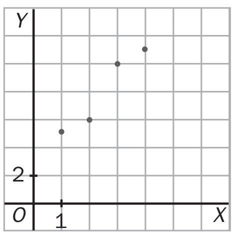
Recta de regresión: $y - 11 = -1(x - 5) \Rightarrow y = 11 - (x - 5)$

15.20 **¿A qué nube de puntos se ajustaría mejor la recta de regresión de ecuación $y = 3x + 1$? Justifica la respuesta.**

a)



b)



Se ajusta mejor a la distribución *B*. Para comprobarlo, consideramos la siguiente tabla:

X	1	2	3	4
Y	5	6	10	11
$y = 3x + 1$	4	7	10	13

Obsérvese que las *Y* de la distribución *B* son muy semejantes a las que proporciona la recta de regresión.

15.21 En la siguiente variable bidimensional:

X	1	2	3	4	3	7	6	3	4	5
Y	45	30	30	25	25	10	20	15	10	15

a) Halla su centro de masas y su covarianza.

b) Calcula su coeficiente de correlación lineal.
¿Tiene sentido calcular su recta de regresión y realizar predicciones?

c) Calcula su recta de regresión.

d) Si el valor de la variable X es 15, ¿cuál es el valor estimado de la variable Y?

e) Si el valor de la variable Y es 13, ¿cuál es el valor estimado de la variable X?

a) Centro de masas $\equiv (\overline{x}, \overline{y})$

$$\overline{x} = \frac{38}{10} = 3{,}8$$

$$\overline{y} = \frac{225}{10} = 22{,}5$$

$$s_{XY} = \frac{45 + 60 + 90 + 100 + 75 + 70 + 120 + 45 + 40 + 75}{10} - 3{,}8 \cdot 22{,}5 = \frac{720}{10} - 85{,}5 = -13{,}5$$

b) $$s_X = \sqrt{\frac{1 + 4 + 9 + 16 + 9 + 49 + 36 + 9 + 16 + 25}{10} - 3{,}8^2} = 1{,}7205$$

$$s_Y = \sqrt{\frac{2025 + 900 + 900 + 625 + 625 + 100 + 400 + 225 + 100 + 225}{10} - 22{,}5^2} = 10{,}3078$$

$$r = \frac{-13{,}5}{1{,}7205 \cdot 10{,}3078} = -0{,}7612$$

Es una correlación media, por lo que sí tendría sentido calcular predicciones.

c) Recta de regresión:

$$y - \overline{y} = \frac{s_{XY}}{s_X^2}(x - \overline{x}) \Rightarrow y - 22{,}5 = \frac{-13{,}5}{2{,}9601}(x - 3{,}8) \Rightarrow y = 22{,}5 + 4{,}5608(x - 3{,}8)$$

d) $x = 15 \Rightarrow y = 22{,}5 - 4{,}5608 \cdot (15 - 3{,}8) = -28{,}5811$

e) $y = 13 \Rightarrow 13 - 22{,}5 = -4{,}5608 \cdot (x - 3{,}8) \Rightarrow -9{,}5 = -4{,}5608x + 17{,}33104 \Rightarrow x = 5{,}8830$

15.22 Contesta verdadero (V) o falso (F) a las siguientes afirmaciones.

a) Si $r = 0{,}34$, las estimaciones con la recta de regresión son poco fiables.

b) Dos variables no correlacionadas X e Y tienen un coeficiente de correlación de 0,98.

c) Dos variables X e Y relacionadas por la ecuación $y = 6x - 7$ tienen un coeficiente de correlación lineal $r = 0{,}85$.

d) Una variable bidimensional de centro de masas (3; 5,5) tiene una recta de regresión que pasa por el punto (3, −2).

a) Verdadero

b) Falso

c) Falso

d) Falso

15.23 Si la covarianza de una distribución es negativa, ¿qué podemos afirmar tanto del coeficiente de correlación como de la pendiente de la recta de regresión?

Signo de coeficiente de correlación:

$$r = \frac{s_{XY}}{s_X s_Y} = \frac{-}{(+)(+)} = -$$

Signo de la pendiente:

$$m = \frac{s_{XY}}{s_X^2} = \frac{-}{+} = -$$

15.24 ¿Un coeficiente de correlación negativo significa una correlación débil de las variables?

No necesariamente. Lo que significa es una nube de puntos decreciente.

15.25 Si dos variables bidimensionales (X, Y), (Z, T) tienen coeficientes de correlación $r_{XY} = -0{,}989$ y $r_{ZT} = 0{,}989$, ¿en cuál de ellas será más fiable hacer una estimación?

En cualquiera de las dos, ya que el valor absoluto del coeficiente de correlación es el mismo.

15.26 El coeficiente de correlación de una distribución bidimensional es 0,85. Si los valores de las variables se multiplican por 10, ¿cuál será el coeficiente de correlación de esa nueva distribución?

$$r_b = \frac{100 s_{XY}}{10 s_X 10 s_Y} = r_a$$

15.27 En una distribución de 40 datos, la covarianza vale 2,605 y $\overline{x} = 5{,}45$, $\overline{y} = 5{,}6$.

Calcula el valor de: $\sum f_{ij} x_i y_j$

$$s_{XY} = \frac{\sum x_i y_i}{n} - \overline{x} \cdot \overline{y} \Rightarrow \sum x_i y_i = (s_{XY} + \overline{x} \cdot \overline{y}) \cdot n = (2{,}605 + 5{,}45 \cdot 5{,}6) \cdot 40 = 1325$$

15.28 **Sabiendo que *m* es la pendiente de la recta de regresión, justifica si la siguiente igualdad es cierta.**

$$\frac{m}{r} = \frac{s_Y}{s_X}$$

La igualdad es cierta:

$$\frac{m}{r} = \frac{s_Y}{s_X} \Rightarrow \frac{m}{r} = \frac{\frac{s_{XY}}{s_X^2}}{\frac{s_{XY}}{s_X s_Y}} = \frac{s_X s_Y}{s_X^2} = \frac{s_Y}{s_X}$$

15.29 **La variable bidimensional (*X*, *Y*) tiene como recta de regresión $y = 3x - 2$ y $r = 0{,}75$.**

La variable bidimensional (*T*, *P*) tiene como recta de regresión $p = 3t - 2$ y $r = 0{,}96$.

¿En cuál de las dos rectas es más fiable estimar −12,5? ¿A qué valor corresponde la estimación realizada?

Es más fiable en (*T*, *P*) al tener una correlación más fuerte.

$$-12{,}5 = 3t - 2 \Rightarrow 3t = -10{,}5 \Rightarrow t = -3{,}5$$

PROBLEMAS PARA APLICAR

15.30 **Escoge una variable de las dos primeras tablas y, a partir de la tercera, indica el tipo de relación que existe. Por ejemplo:**

2. 'Número de horas sentado al cabo del día'.

***A*: 'Peso del individuo'.**

***Y*: correlación positiva media.**

1. Número de cigarros fumados al día.
2. Número de horas sentado al cabo del día.
3. Velocidad a la que voy en el coche.
4. Nota en una asignatura.

***A*.** Peso del individuo.
***B*.** Número de horas diarias de móvil.
***C*.** Capacidad pulmonar.
***D*.** Espacio que recorro.

***W*.** Dependencia funcional.
***X*.** Correlación negativa fuerte.
***Y*.** Correlación positiva media.
***Z*.** Correlación positiva medio-fuerte.

Podemos hacer las siguientes relaciones: 1*CX*, 2*AZ*, 3*DW*, 4*BY*

15.31 En una encuesta a 30 jóvenes sobre el número de libros que leen al cabo de un año han respondido lo siguiente.

X	0	1	2	3	4	5	6	$\geqslant 7$
f_i	5	6	8	3	3	2	2	1

Y sobre el número de películas vistas en un año:

Y	3	5	7	9	12	15	18	$\geqslant 20$
f_i	5	6	8	3	3	2	2	1

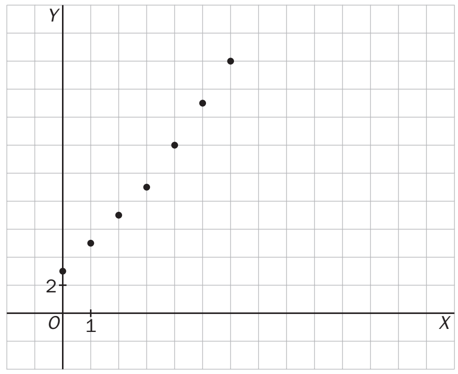
a) ¿Está relacionado el número de libros que leen los jóvenes con las películas que visualizan? Considera para ello la variable bidimensional (X, Y) construida a través de los pares (x_i, y_i), y elabora una tabla de doble entrada.

b) Dibuja el diagrama de dispersión. Indica qué tipo de correlación tienen.

a)

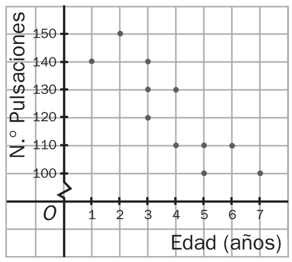
Y \ X	0	1	2	3	4	5	6	$\geqslant 7$	Total
3	5								5
5		6							6
7			8						8
9				3					3
12					3				3
15						2			2
18							2		2
$\geqslant 20$								1	1
Total	5	6	8	3	3	2	2	1	30

b)



Tienen una correlación positiva y fuerte.

15.32 El siguiente diagrama expresa la relación entre la edad y el número de pulsaciones de 12 personas.



Calcula la recta de regresión.

Consideramos la siguiente tabla:

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	160	160	1	25 600
1	140	140	1	19 600
2	150	300	4	22 500
3	140	420	9	19 600
3	130	390	9	16 900
3	120	360	9	14 400
4	130	520	16	16 900
4	110	440	16	12 100
5	110	550	25	12 100
5	100	500	25	10 000
6	110	660	36	12 100
7	100	700	49	10 000
44	1500	5140	200	191 800

$$\overline{x} = \frac{44}{12} = 3,\widehat{6}$$

$$\overline{y} = \frac{1500}{12} = 125$$

$$s_X^2 = \frac{200}{12} - 3,\widehat{6}^2 = 3,\widehat{2}$$

$$s_{XY} = \frac{5140}{12} - \overline{x} \cdot \overline{y} = -30$$

Recta de regresión:

$$y - 125 = -9,3103(x - 3,\widehat{6}) \Rightarrow y = 125 - 9,3103(x - 3,\widehat{6})$$

15.33 Una empresa realiza un estudio de los efectos de la publicidad sobre sus ventas. Los resultados de ese estudio en un determinado producto son los siguientes.

Gasto	1	2	3	4	5	6
N.º productos	9	18	32	27	40	46

Los datos están dados en miles de euros.

a) Calcula su coeficiente de correlación.

b) Obtén su recta de regresión.

c) Si se invierten 15 000 euros en publicidad, ¿cuántas ventas del producto se estima que se producirán?

d) Si en un determinado año se consiguen vender 60 000 unidades del producto, ¿cuánto se estima que se ha invertido en publicidad ese año?

a) $r = \frac{s_{XY}}{s_X s_Y}$

$\overline{x} = 3{,}5$ mil €

$\overline{y} = 28{,}\widehat{6}$ mil €

$$s_{XY} = \frac{9 + 36 + 96 + 108 + 200 + 276}{6} - 3{,}5 \cdot 28{,}\widehat{6} = 120{,}8 - 100{,}\widehat{3} = 20{,}5$$

$$s_X = \sqrt{\frac{1 + 4 + 9 + 16 + 25 + 36}{6} - 3{,}5^2} = \sqrt{15{,}1\widehat{6} - 12{,}25} = 1{,}7078$$

$$s_Y = \sqrt{\frac{81 + 324 + 1024 + 729 + 1600 + 2116}{6} - 28{,}\widehat{6}^2} = 12{,}5388$$

$$r = \frac{20{,}5}{1{,}7078 \cdot 12{,}5388} = 0{,}9573$$

b) $y - \overline{y} = \frac{s_{XY}}{s_X^2}(x - \overline{x}) \Rightarrow y - 28{,}\widehat{6} = \frac{20{,}5}{2{,}9167}(x - 3{,}5) \Rightarrow y = 7{,}0284x + 4{,}0673$

c) $x = 15\,000 = 15$ mil € $\Rightarrow y = 7{,}0284 \cdot 15 + 4{,}0673 = 109{,}4933$ mil € $= 109\,493{,}3$ €

d) $y = 60\,000 = 60$ mil € $\Rightarrow 60 = 7{,}0284x + 4{,}0673 \Rightarrow x = 7{,}9581$ mil € $= 7958{,}1$ €

15.34 La media de los pesos de los individuos de una población es de 72 kilogramos, y la de sus estaturas, 173 centímetros. Las desviaciones típicas son 4 kilogramos y 9 centímetros, y la covarianza es 35.

a) ¿Cuál es el coeficiente de correlación?

b) Calcula la recta de regresión del peso respecto a la estatura.

c) ¿Cuál es el peso estimado para un individuo de 182 centímetros?

a) Consideremos la variable bidimensional (X, Y) = (Peso, Estatura) $\Rightarrow r = \frac{s_{XY}}{s_X s_Y} = \frac{35}{4 \cdot 9} = \frac{35}{36} = 0{,}9722$

b) $y - 173 = \frac{35}{16} \cdot (x - 72) \Rightarrow y - 173 = 2{,}1875 \cdot (x - 72) \Rightarrow y = 2{,}1875x + 15{,}5$

c) Si $y = 182$ cm $\Rightarrow x = \frac{182 - 15{,}5}{2{,}1875} = 76{,}1143$ kg

15.35 En un curso de Bachillerato de 40 alumnos se ha querido estudiar la correlación de notas de las asignaturas de Lengua (X) e Historia (Y) como materias clave en itinerarios de letras. Los resultados han sido los siguientes.

Y \ X	3	4	5	6	7	8	10
2	4						
5		7	11				
6				5	3		
7				5	2		
9						1	
10							2

El coeficiente de correlación es $r = 0{,}919$, y su recta de regresión, $y = x + 0{,}15$.

a) Comprueba que la recta pasa por $(\overline{x}, \overline{y})$.

b) Se define la desviación de cada punto, d, como la diferencia entre el valor real (y_r) y el valor estimado (y_e); es decir, $d = y_r - y_e$.

Calcula todas las desviaciones de los datos y halla su suma, comprobando que se van equilibrando unas con otras.

a) $(\overline{x}, \overline{y}) = (5{,}45;\ 5{,}6) \Rightarrow$ Sustituyendo en la recta de regresión se comprueba que $5{,}6 = 5{,}45 + 0{,}15$.

b) Consideramos la siguiente tabla:

x_i	y_i	y_e	$d = y_r - y_e$
3	2	3,15	−1,15
4	5	4,15	0,85
5	5	5,15	−0,15
6	6	6,15	−0,15
6	7	6,15	0,85
7	6	7,15	−1,15
7	7	7,15	−0,15
8	9	8,15	0,85
10	10	10,15	−0,15
$\sum(y_r - y_e)$			−0,35

Como se observa en la tabla, $\sum(y_r - y_e) = -0{,}35$, que es un valor muy pequeño.

Esto nos muestra que las desviaciones de los datos se van equilibrando unas con otras.

Diagramas de dispersión. Correlación

15.36 Dada la siguiente distribución bidimensional:

a) Dibuja la nube de puntos.

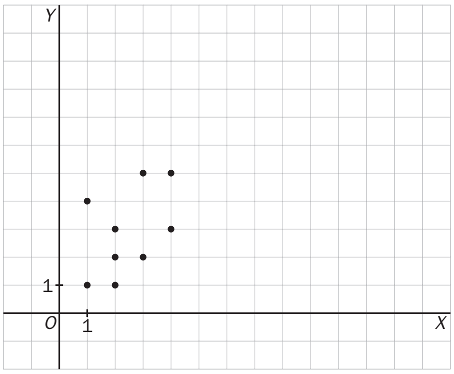
b) ¿Qué porcentaje de veces aparece (2, 2)?

c) ¿Cuál es la frecuencia absoluta del valor (3, *y*)?

d) ¿Cuál es la frecuencia absoluta del valor (*x*, 5)?

Y \ X	1	2	3	4	Total
1	1	1			2
2		3	3		6
3		1		1	2
4	1				1
5			2	2	4
Total	2	5	5	3	15

a)



b) $\frac{3}{15} \cdot 100 = 20\,\%$

c) 5

d) 4

15.37 Los valores de una variable bidimensional (*X*, *Y*) son los que siguen.

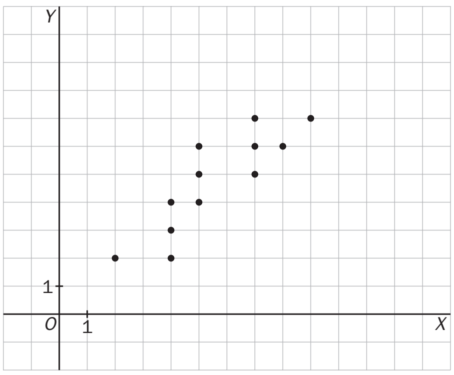
(2, 2); (4, 2); (4, 4); (4, 3); (7, 5); (7, 7); (7, 6); (5, 6); (5, 5); (5, 4); (8, 6); (9, 7)

a) Dibuja el diagrama de dispersión.

b) Halla el coeficiente de correlación. Interpreta el resultado.

c) Indica el tipo de dependencia entre ambas variables.

a)



$r = 0{,}8605 \Rightarrow$ Las variables tienen una correlación positiva media.

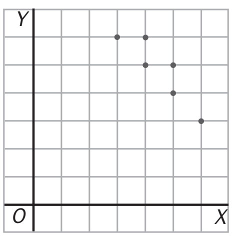
c) La dependencia entre ambas variables es aleatoria.

b) Consideramos la siguiente tabla:

x_i	y_i	x_iy_i	x_i^2	y_i^2
2	2	4	4	4
4	2	8	16	4
4	4	16	16	16
4	3	12	16	9
7	5	35	49	25
7	7	49	49	49
7	6	42	49	36
5	6	30	25	36
5	5	25	25	25
5	4	20	25	16
8	6	48	64	36
9	7	63	81	49
67	57	352	419	305

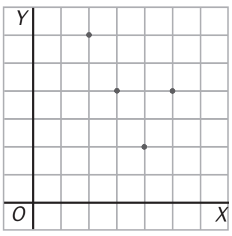
15.38 Asocia cada coeficiente de correlación con su gráfica correspondiente: −0,26; −0,81; −0,95; 0,71

a)



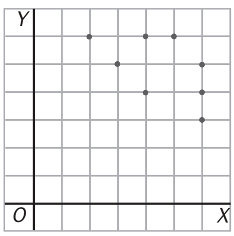
$r = -0{,}95$

b)



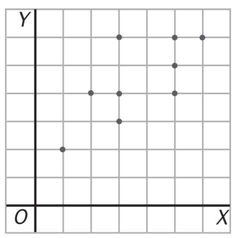
$r = -0{,}26$

c)



$r = -0{,}81$

d)



$r = 0{,}71$

Recta de regresión y estimaciones

15.39 La relación entre dos variables (X, Y) viene dada por la siguiente tabla.

X	1	2	2	3	4	4	5	6	6	7
Y	1	1	2	4	4	6	7	6	7	7

a) Calcula el coeficiente de correlación lineal.

b) Halla la recta de regresión.

c) Si $x = 8$, ¿cuánto valdría y?

d) ¿Es buena esta predicción? Justifica la respuesta.

a)

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	1	1	1	1
2	1	2	4	1
2	2	4	4	4
3	4	12	9	16
4	4	16	16	16
4	6	24	16	36
5	7	35	25	49
6	6	36	36	36
6	7	42	36	49
7	7	49	49	49
40	45	221	196	257

$$\overline{x} = \frac{40}{10} = 4$$

$$\overline{y} = \frac{45}{10} = 4{,}5$$

$$s_X = \sqrt{\frac{196}{10} - 4^2} = 1{,}8974$$

$$s_Y = \sqrt{\frac{257}{10} - 4{,}5^2} = 2{,}3345$$

$$s_{XY} = \frac{221}{10} - \overline{x} \cdot \overline{y} = 4{,}1$$

Luego $r = \dfrac{s_{XY}}{s_X s_Y} = 0{,}9256$

b) $y - 4{,}5 = \dfrac{4{,}1}{3{,}6} \cdot (x - 4) \Rightarrow y - 4{,}5 = 1{,}1389 \cdot (x - 4) \Rightarrow y = 4{,}5 + 1{,}1389(x - 4)$

c) Si $x = 8 \Rightarrow y = 4{,}5 + 1{,}1389(8 - 4) = 9{,}0556$

d) La predicción es buena, ya que el coeficiente de correlación está muy próximo a 1 y el valor pedido no está alejado del resto de valores de X.

AMPLIACIÓN

15.40 Se ha medido experimentalmente el área de distintos triángulos equiláteros de lados 1, 2, 3 decímetros, sucesivamente, y se han obtenido los siguientes resultados.

(Lado)^2	2	4	9	16	25
Área	0,42	1,65	3,7	6,5	10,2

a) Calcula el coeficiente de correlación lineal entre el cuadrado del lado y el área del triángulo. ¿Qué tipo de correlación existe?

b) ¿Debería haber una relación funcional? ¿A qué se debe que la relación no llegue a ser funcional?

a) La correlación es positiva y fuerte. Veámoslo:

$r = \dfrac{s_{XY}}{s_X s_Y}$ $\qquad \overline{x} = 11 \qquad \overline{y} = 4{,}5$

$$s_{XY} = \frac{0{,}84 + 6{,}6 + 33{,}3 + 104 + 255}{5} - 11 \cdot 4{,}5 = 30{,}4$$

$$s_X = \sqrt{\frac{4 + 16 + 81 + 256 + 625}{5} - 11^2} = 8{,}7$$

$$s_Y = \sqrt{\frac{0{,}18 + 2{,}72 + 13{,}7 + 42{,}25 + 104{,}04}{5} - 4{,}5^2} = 3{,}5$$

$r = \dfrac{30{,}4}{8{,}7 \cdot 3{,}5} = 0{,}99 \Rightarrow$ La correlación es positiva y fuerte.

b) Sí debería haber relación funcional, pero el coeficiente de correlación no es 1 debido al redondeo que se hace al medir experimentalmente.

15.41 Una variable bidimensional viene dada por la siguiente tabla.

X	**2**	**3**	**5**	***a***
Y	**1**	**25**	***b***	**3**

Sabiendo que $s_{XY} = 1$ y $s_X^2 = 3$ y que *a* es el valor máximo de la variable *X*, calcula *a* y *b*.

Planteamos el siguiente sistema:

$$\begin{cases} 1 = s_{XY} \Rightarrow 1 = \dfrac{2 + 75 + 5b + 3a}{4} - \left(\dfrac{10 + a}{4}\right) \cdot \left(\dfrac{29 + b}{4}\right) \\ 3 = s_X^2 \Rightarrow 3 = \dfrac{4 + 9 + 25 + a^2}{4} - \left(\dfrac{10 + a}{4}\right)^2 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} 2(5b + 9) - a(b + 17) = 16 \\ 3a^2 - 20a + 4 = 0 \end{cases} \Rightarrow \begin{cases} a = 6{,}4603 \Rightarrow b = 30{,}46 \\ a = 0{,}2064 \Rightarrow b = 0{,}1538 \end{cases}$$

Como nos dicen que *a* es el valor máximo de la variable *X*, la única solución válida es $a = 6{,}4603$ y $b = 30{,}46$.

15.42 En una variable bidimensional (*X*, *Y*), su coeficiente de correlación lineal es 0,48 y la pendiente de su recta de regresión es 1,34.

Sabiendo que la suma de las desviaciones típicas de *X* e *Y* es 7,33, calcula cada una de ellas y la covarianza de la variable bidimensional.

$$r = 0{,}48 = \frac{s_{XY}}{s_X s_Y}$$

$$m = \frac{s_{XY}}{s_X^2} = 1{,}34$$

$$s_X + s_Y = 7{,}33$$

$$\begin{cases} \dfrac{s_{XY}}{s_X s_Y} = 0{,}48 \\ \dfrac{s_{XY}}{s_X^2} = 1{,}34 \\ s_X + s_Y = 7{,}33 \end{cases} \Rightarrow \begin{cases} s_{XY} = 0{,}48 \cdot s_X s_Y \\ s_{XY} = 1{,}34 \cdot s_X^2 \\ s_Y = 7{,}33 - s_X \end{cases} \Rightarrow \begin{cases} 0{,}48 \cdot s_X \cdot (7{,}33 - s_X) = 1{,}34 \cdot s_X^2 \\ 3{,}52 \cdot s_X - 0{,}48 \cdot s_X^2 = 1{,}34 \cdot s_X^2 \\ 1{,}82 \cdot s_X^2 - 3{,}52 \cdot s_X = 0 \Rightarrow s_X = \dfrac{3{,}52}{1{,}82} = 1{,}93 \end{cases} \Rightarrow$$

$$\Rightarrow s_Y = 7{,}33 - 1{,}93 = 5{,}4 \Rightarrow s_{XY} = 1{,}34 \cdot s_X^2 = 5$$

15.43 Encuentra la recta de regresión de la variable (*X*, *Y*) sabiendo que es paralela a la recta

$$2x - 4y = -17$$

y que su centro de masas es un punto que comparte con las rectas:

$$x + 3y = 19 \qquad\qquad 4x - y = 2$$

La recta de regresión es paralela a $2x - 4y = -17$; por tanto, tienen la misma pendiente: $m = \dfrac{1}{2}$.

$$(\overline{x},\ \overline{y}) \Rightarrow \begin{cases} x + 3y = 19 \\ 4x - y = 2 \end{cases} \Rightarrow \begin{cases} x + 3y = 19 \\ 12x - 3y = 6 \end{cases} \Rightarrow 13x = 25 \Rightarrow x = \frac{25}{13} = 1{,}9 \Rightarrow y = 5{,}6$$

$(\overline{x},\ \overline{y}) = (1{,}9;\ 5{,}6)$

La recta de regresión es: $y - 5{,}6 = \dfrac{1}{2} \cdot (x - 1{,}9)$.

15.44 Una variable bidimensional (*X*, *Y*) tiene de coeficiente de correlación *r* = 0,78, y las medias de las distribuciones marginales son $\overline{x} = 2$ e $\overline{y} = 9$. Razona cuál de las siguientes rectas se ajusta más a dicha variable.

$$y = -3x + 12 \qquad y = 1{,}5x + 6 \qquad y = -2{,}5x + 14$$

Descartamos la recta $y = -3x + 12$ porque el punto $(\overline{x}, \overline{y})$ no pertenece a ella.

Descartamos también la recta $y = -2{,}5x + 14$, pues el coeficiente de correlación es $r = 0{,}78$, que es número positivo.

Por tanto, la recta que más se ajusta a dicha variable es $y = 1{,}5x + 6$.

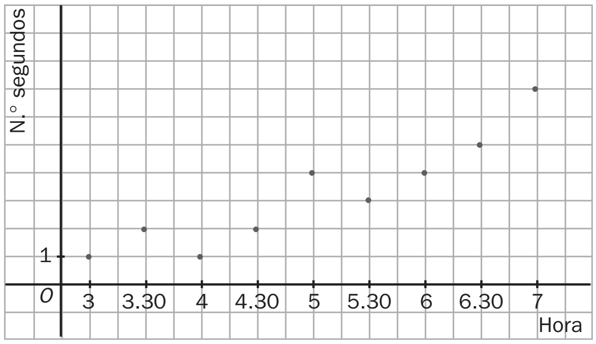
PARA INTERPRETAR Y RESOLVER

15.45 La página del instituto

Juan se ha molestado en observar los segundos que ha tardado en acceder a la página web de su instituto en diferentes momentos del día.

El siguiente gráfico de dispersión muestra los tiempos empleados en nueve momentos comprendidos entre las tres y las siete de la tarde.

Indica cuál de las siguientes opciones (en las que se incluyen la recta de regresión, el coeficiente de correlación y el tiempo esperado cuando se conecta a las 7.30) se corresponde con los datos.



A	*B*	*C*
$y = 3{,}45x - 1{,}35$	$y = 1{,}35x - 3{,}45$	$y = x - 3$
$r = 0{,}90$	$r = 0{,}92$	$r = -0{,}92$
$x = 7{,}5 \Rightarrow y = 7$	$x = 7{,}5 \Rightarrow y = 6{,}6$	$x = 7{,}5 \Rightarrow y = 6$

La opción que se corresponde con los datos es la *B*.

15.46 Diferentes rectas

La tabla siguiente muestra el crecimiento de los beneficios obtenidos por una empresa en cada uno de los ocho trimestres de los últimos dos años.

Primer año				
Trimestre	**1.°**	**2.°**	**3.°**	**4.°**
Beneficios (%)	**7**	**6**	**7**	**5**

Segundo año				
Trimestre	**1.°**	**2.°**	**3.°**	**4.°**
Beneficios (%)	**6**	**4**	**3**	**2**

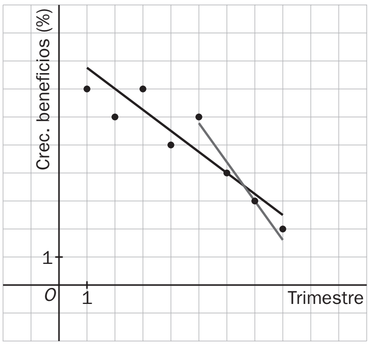
a) Representa los datos en un diagrama de dispersión.

b) Calcula y dibuja la recta de regresión considerando los ocho pares de datos. ¿Qué crecimiento se estima para el primer trimestre del tercer año?

c) Calcula y dibuja la recta de regresión considerando únicamente los datos correspondientes al segundo año. ¿Qué crecimiento se estima para el primer trimestre del tercer año?

d) Indica alguna ventaja y algún inconveniente al utilizar la segunda recta en vez de la primera para realizar la estimación solicitada.

a) y b)



Se numeran los trimestres de forma correlativa: 1, 2, 3… 8

Considerando *x* como el trimestre e *y* como el crecimiento del beneficio, se obtiene la recta de regresión $y = -0{,}69x + 8{,}1$.

Así pues, para el primer trimestre del tercer año se espera un crecimiento de $-0{,}69 \cdot 9 + 8{,}1 = 1{,}89\,\%$.

c) Si se consideran solo los datos del segundo año, se obtiene la recta $y = -1{,}3x + 12{,}2$. Así, para el primer trimestre del tercer año se espera un crecimiento de $-1{,}3 \cdot 9 + 12{,}2 = 0{,}5\,\%$.

d) Se utilizan datos más cercanos en el tiempo, pero en menor cantidad.

15.A1 Sea la siguiente tabla simple de una variable bidimensional.

X	−4	−1	0	2	3	5	6	9
Y	9	7	6	4	4	2	1	3

a) Elabora una tabla de doble entrada.

b) Dibuja su nube de puntos.

c) Elige el coeficiente de correlación que crees que se puede ajustar más a la nube y justifica la respuesta.

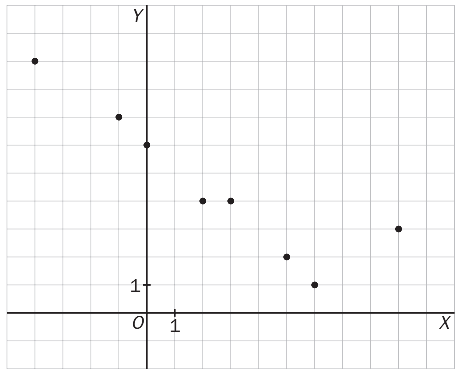
$$r = 0{,}95 \qquad r = -1{,}2 \qquad r = -0{,}25 \qquad r = -0{,}85$$

d) Calcula r y contrasta tu suposición.

a)

Y \ X	−4	−1	0	2	3	5	6	9
1							1	
2						1		
3								1
4				1	1			
6			1					
7		1						
9	1							

b)



c) $r = -0{,}85$, al tener una correlación fuerte y negativa (decreciente).

d) $r = \dfrac{s_{XY}}{s_X s_Y}$

$\overline{x} = 2{,}5$

$\overline{y} = 4{,}5$

$$s_{XY} = \frac{-36 - 7 + 8 + 12 + 10 + 6 + 27}{8} - 2{,}5 \cdot 4{,}5 = 2{,}5 - 11{,}25 = -8{,}75$$

$$s_X = \sqrt{\frac{16 + 1 + 4 + 9 + 25 + 36 + 81}{8} - 2{,}5^2} = \sqrt{21{,}5 - 6{,}25} = 3{,}9$$

$$s_Y = \sqrt{26{,}5 - 20{,}25} = 2{,}5$$

$$r = \frac{-8{,}75}{3{,}9 \cdot 2{,}5} = \frac{-8{,}75}{9{,}75} = -0{,}89$$

15.A2 Relaciona en tu cuaderno cada coeficiente de correlación con el tipo de correlación.

a) $r = -1$	**I. Correlación positiva fuerte**
b) $r = 0{,}95$	**II. Independientes**
c) $r = 0{,}26$	**III. Dependencia funcional**
d) $r = -0{,}89$	**IV. Correlación negativa débil**
e) $r = -0{,}62$	**V. Correlación negativa fuerte**

$r = -1 \longrightarrow$ Dependencia funcional

$r = 0{,}95 \longrightarrow$ Correlación positiva fuerte

$r = 0{,}26 \longrightarrow$ Independientes

$r = -0{,}89 \longrightarrow$ Correlación negativa fuerte

$r = -0{,}62 \longrightarrow$ Correlación negativa débil

15.A3 En una encuesta realizada a 25 personas, se les ha preguntado su edad y el número de horas de ejercicio que realizan al día. Las respuestas han sido las siguientes.

(15, 3); (17; 2,5); (25, 2); (35, 1); (40; 0,5); (45, 1); (50, 1); (60, 2); (65, 3); (16, 4); (20, 2); (25, 2); (32; 1,5); (52; 1,5); (47, 1); (52, 2); (68; 2,5); (16, 4); (24; 2,5); (40, 30); (25, 2); (31; 2,5); (45, 50); (62; 2,5); (15; 3,5).

Calcula la recta de regresión.

Realizamos una tabla de doble entrada.

Y \ X	15	16	17	20	24	25	31	32	35	40	45	47	50	52	60	62	65	68	Total
20′																			0
30′										2									2
50′											1								1
1 h									1		1	1	1						4
1,5 h								1						1					2
2 h				1		3								1	1				6
2,5 h			1		1		1									1		1	5
3 h	1																1		2
3,5 h	1																		1
4 h		2																	2
Total	2	2	1	1	1	3	1	1	1	2	2	1	1	2	1	1	1	1	25

La recta de regresión es de la forma:

$$y - \overline{y} = \frac{s_{XY}}{s_X^2}(x - \overline{x})$$

$$\overline{x} = \frac{30 + 32 + 17 + 20 + 24 + 75 + 31 + 32 + 35 + 80 + 90 + 47 + 50 + 104 + 60 + 62 + 65 + 68}{25} = 36{,}88 \text{ años}$$

$$\overline{y} = \frac{60 + 50 + 240 + 180 + 720 + 750 + 360 + 210 + 480}{25} = 122 \text{ min}$$

$$s_{XY} = \frac{103\,200}{25} - 36{,}88 \cdot 122 = -371{,}36$$

$$s_X^2 = \frac{40\,972}{25} - 36{,}88^2 = 278{,}7456$$

$$y - 122 = \frac{-371{,}36}{278{,}7456} \cdot (x - 36{,}88) \Rightarrow y - 122 = -1{,}33(x - 36{,}88) \Rightarrow y = -1{,}33x + 171{,}05$$

15.A4 Se ha preguntado a los alumnos de un centro el número de horas de estudio diario, *X*, y el número de asignaturas aprobadas al final del curso, *Y*. A la nube de puntos resultado de la encuesta se ha ajustado la recta de regresión $y = 3{,}8x + 0{,}2$.

a) Para aprobar 4 asignaturas, ¿cuánto tiempo de estudio deberían emplear?

b) Y para superar las 11 asignaturas, es decir, todas, ¿cuál sería la recomendación de horas de estudio?

a) Si $y = 4 \Rightarrow 4 = 3{,}8x + 0{,}2 \Rightarrow x = \dfrac{4 - 0{,}2}{3{,}8} = 1$ hora

b) Si $y = 11 \Rightarrow 11 = 3{,}8x + 0{,}2 \Rightarrow x = \dfrac{11 - 0{,}2}{3{,}8} = 2{,}84$ horas (aproximadamente 2 h 50 min)

MATETIEMPOS

¿Metros y kilogramos?

Un profesor ha realizado un estudio sobre la altura y el peso medios de los alumnos de una clase de 4.º de ESO, obteniendo los siguientes valores:

Variable	Media	Desviación típica
Altura (m)	1,72	0,29
Peso (kg)	65,4	4,68

¿Qué varía más, la altura o el peso? ¿Por qué?

La única forma de comparar las magnitudes es mediante el coeficiente de variación definido como $CV = \frac{s}{\overline{x}}$.

De este modo se eliminan las unidades, y el resultado se expresa en porcentaje de variación que compara el grado de dispersión entre las distribuciones. En nuestro caso:

Altura: $CV_A = \frac{0{,}29}{1{,}72} \cdot 100 = 16{,}86\,\%$

Peso: $CV_P = \frac{4{,}68}{65{,}4} \cdot 100 = 7{,}15\,\%$

Luego hay más dispersión de valores en la altura que en el peso.