Estadística bidimensional

14

ACTIVIDADES

1. Considera estas variables bidimensionales, y escribe las variables unidimensionales correspondientes y tres pares de valores que las determinan.

a) Edad y sexo de los asistentes a un concierto.

b) Tamaño de un archivo informático y tiempo que se tarda en copiarlo.

c) Peso y talla de pies de los alumnos de una clase.

a) $X \longrightarrow$ Edad, en años, de los asistentes al concierto

$Y \longrightarrow$ Sexo de los asistentes

(20, mujer) (25, hombre) (28, mujer)

b) $X \longrightarrow$ Tamaño, en kb, del archivo informático

$Y \longrightarrow$ Tiempo, en s, que se tarda en copiarlo

(220, 35) (158, 24) (285, 42)

c) $X \longrightarrow$ Peso, en kilos, de los alumnos de una clase

$Y \longrightarrow$ Altura, en centímetros, de los alumnos de una clase

(61, 155) (76, 172) (56, 160)

2. En una clase con 27 alumnos se ha realizado un estudio sobre el número de horas diarias de estudio, *X*, y el número de suspensos, *Y*, obteniéndose estos resultados.

(2, 1) (0, 7) (1, 3) (3, 1) (3, 0) (1, 2) (1, 1) (2, 0)
(3, 0) (2, 7) (1, 0) (2, 1) (3, 1) (1, 4) (1, 2) (2, 1)
(2, 0) (3, 1) (2, 2) (1, 0) (1, 2) (2, 1) (0, 6) (2, 0)

Ordena estos datos en una tabla de doble entrada.

Y \ X	0	1	2	3	Total
0	0	2	3	2	7
1	0	1	4	3	8
2	0	3	1	0	4
3	0	1	0	0	1
4	0	1	0	0	1
5	0	0	0	0	0
6	1	0	0	0	1
7	1	0	1	0	2
Total	2	8	9	5	24

3. **Construye una tabla de doble entrada y sus tablas de frecuencias marginales de estos datos.**

(1, 2) (1, 3) (1, 1) (2, 2) (3, 1)   (1, 1) (3, 3) (2, 1) (2, 1) (3, 2)

Y \ X	1	2	3	Total
1	2	1	1	4
2	1	2	1	4
3	1	0	1	2
Total	4	3	3	10

X	1	2	3	Total
f_i	4	3	3	10

Y	1	2	3	Total
f_i	4	4	2	10

4. **Con la tabla de doble entrada de la actividad anterior, realiza las siguientes tablas condicionadas.**

a) $Y/X = 1$ b) $X/Y = 3$

a)

$Y/X = 1$	1	2	3	Total
f_i	2	1	1	4

b)

$X/Y = 3$	1	2	3	Total
f_i	1	0	1	2

5. **Construye la tabla de doble entrada y sus tablas marginales para los siguientes datos, (*X*, *Y*).**

(16, 5) (17, 4) (18, 6) (16, 6) (14, 8)
(17, 3) (14, 5) (13, 4) (14, 8) (15, 8)

Representa el diagrama de dispersión de la variable estadística bidimensional anterior.

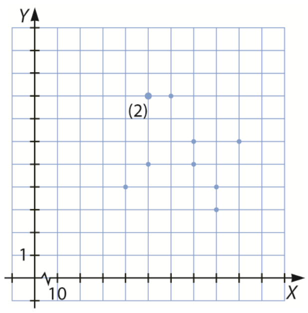
Y \ X	13	14	15	16	17	18	Total
3	0	0	0	0	1	0	1
4	1	0	0	0	1	0	2
5	0	1	0	1	0	0	2
6	0	0	0	1	0	1	2
8	0	2	1	0	0	0	3
Total	1	3	1	2	2	1	10

Tabla de frecuencias marginales de *X*

x_i	f_i
13	1
14	3
15	1
16	2
17	2
18	1
Total	10

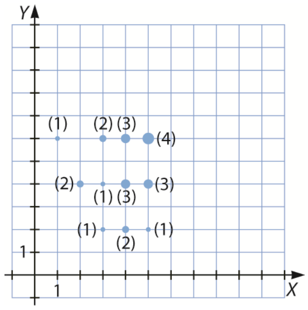
Tabla de frecuencias marginales de *Y*

y_i	f_i
3	1
4	2
5	2
6	2
8	3
Total	10



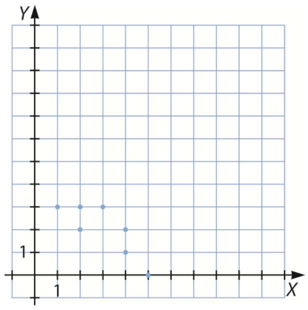
6. **Realiza un diagrama de dispersión teniendo en cuenta la frecuencia de cada pareja de datos.**

Y \ X	1	2	3	4	5
2	0	0	1	2	1
4	0	2	1	3	3
6	1	0	2	3	4



7. **Representa la nube de puntos y analiza la dependencia.**

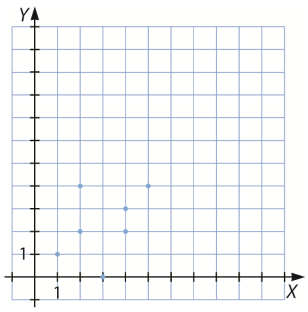
X	2	4	4	3	2	5	1
Y	2	2	1	3	3	0	3



La nube de puntos se aproxima poco a una recta y cuando la variable X crece, la variable Y decrece, por lo que existe una dependencia lineal débil y negativa entre ellas.

8. **Realiza la nube de puntos y analiza la dependencia.**

X	2	4	4	3	2	5	1
Y	4	2	3	0	2	4	1



La nube de puntos se aproxima poco a una recta y cuando la variable X crece, la variable Y crece, por lo que existe una dependencia lineal débil y positiva entre ellas.

9. **¿Tener mascota, *X*, influye para aprobar Matemáticas, *Y*?**

Y \ *X*	Sí	No
Aprobado	10	15
Suspenso	30	45

Completamos la tabla con los totales y analizamos si las filas y las columnas son proporcionales entre sí:

Y \ *X*	Sí	No	Total
Aprobado	10	15	25
Suspenso	30	45	75
Total	40	60	100

$$\frac{10}{40}=\frac{15}{60}=\frac{25}{100}=\frac{1}{4} \qquad \frac{30}{40}=\frac{45}{60}=\frac{75}{100}=\frac{3}{4} \qquad \frac{10}{25}=\frac{30}{75}=\frac{40}{100}=\frac{2}{5} \qquad \frac{15}{25}=\frac{45}{75}=\frac{75}{100}=\frac{3}{4}$$

Sí son proporcionales, luego tener mascota no influye en aprobar Matemáticas.

10. **Completa la siguiente tabla en tu cuaderno para que sean independientes.**

Y \ *X*	A	B	Total
C			175
D			325
Total	200	300	

El número total de datos es 500. La tabla es:

Y \ *X*	A	B	Total
C	x_{11}	x_{12}	175
D	x_{22}	x_{22}	325
Total	200	300	500

Las variables son independientes si las filas y las columnas son independientes entre sí, luego:

$$\frac{x_{11}}{200}=\frac{x_{12}}{300}=\frac{175}{500}\rightarrow\begin{cases}\dfrac{x_{11}}{200}=\dfrac{175}{500}\rightarrow x_{11}=\dfrac{175\cdot 200}{500}=70\\[2mm]\dfrac{x_{12}}{300}=\dfrac{175}{500}\rightarrow x_{12}=\dfrac{175\cdot 300}{500}=105\end{cases} \qquad \frac{x_{21}}{200}=\frac{x_{22}}{300}=\frac{325}{500}\rightarrow\begin{cases}\dfrac{x_{21}}{200}=\dfrac{325}{500}\rightarrow x_{21}=\dfrac{325\cdot 200}{500}=130\\[2mm]\dfrac{x_{22}}{300}=\dfrac{325}{500}\rightarrow x_{22}=\dfrac{325\cdot 300}{500}=195\end{cases}$$

Por tanto, la tabla quedaría así:

Y \ *X*	A	B	Total
C	70	105	175
D	130	195	325
Total	200	300	500

11. **Determina la covarianza de esta variable.**

X	8	10	11	9	13	12	9	14
Y	20	18	16	22	10	10	21	9

$$\sigma_{XY} = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{N} - \overline{x} \cdot \overline{y} = \frac{1279}{8} - 169{,}31 = -9{,}44$$

12. **Calcula la covarianza de esta variable.**

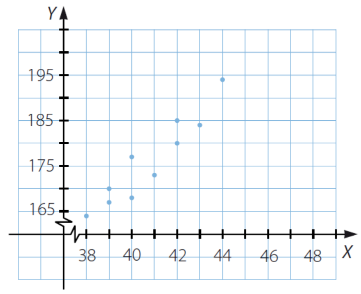
X	10	11	12	13	14	15	16
Y	20	25	32	30	33	34	34

$$\sigma_{XY} = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{N} - \overline{x} \cdot \overline{y} = \frac{2\,765}{7} - 386{,}30 = 8{,}70$$

13. **Representa el diagrama de dispersión y halla el coeficiente de correlación de esta variable.**

X	39	43	40	40	42	41	42	38	39	44
Y	167	184	177	168	185	173	180	164	170	194

¿Qué relación puedes describir entre ellas?



$$\overline{x} = \frac{408}{10} = 40{,}8 \qquad \overline{y} = \frac{1\,762}{10} = 176{,}2$$

$$\sigma_X = \sqrt{3{,}36} = 1{,}83 \qquad \sigma_Y = \sqrt{81{,}96} = 9{,}05$$

$$\sigma_{XY} = \frac{72\,046}{10} - 40{,}8 \cdot 176{,}25 = 13{,}6$$

$$r_{XY} = \frac{13{,}6}{1{,}83 \cdot 9{,}05} = 0{,}82$$

Existe una dependencia lineal fuerte, porque los valores, aunque no se ajustan a una recta, se encuentran muy próximos; el coeficiente de correlación es $r_{XY} = 0{,}82$, que es cercano a 1, y además $r_{XY} > 0$ por lo que es positiva.

14. **La tabla muestra la renta per cápita en miles de euros, *X*, y la esperanza de vida en años, *Y*, en 8 países.**

X	12	40	34	6	30	42	2	15	9
Y	65	79	75	63	74	82	60	62	62

¿Qué relación puedes describir entre ellas?

$\overline{x} = 21{,}11$ $\qquad \overline{y} = 69{,}11$

$\sigma_X = 15{,}40$ $\qquad \sigma_Y = 8{,}37$ $\qquad \sigma_{XY} = 126{,}11$

$r_{XY} = 0{,}98$

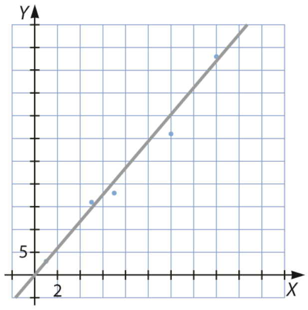
El coeficiente de correlación es $r_{XY} = 0{,}98$, que es muy cercano a 1 y determina una dependencia lineal fuerte, y además $r_{XY} > 0$, por lo que es positiva.

15. Halla la recta de regresión de *Y* sobre *X* y represéntala junto con el diagrama de dispersión de esta tabla.

X	1	5	7	10	12	16
Y	3	16	18	33	35	48

$\overline{x} = 8{,}5$ $\overline{y} = 25{,}5$ $\sigma_X^2 = 28{,}3$ $\sigma_{XY} = 85{,}3$

$$y - \overline{y} = \frac{\sigma_{XY}}{\sigma_X^2}(x - \overline{x}) \rightarrow y - 25{,}5 = \frac{85{,}3}{28{,}3}(x - 8{,}5) \rightarrow y = 3{,}01x - 0{,}12$$

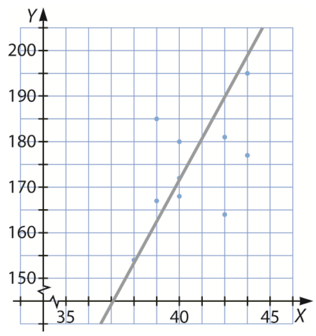


16. Representa el diagrama de dispersión y determina la recta de regresión de *Y* sobre *X* de esta tabla.

X	39	40	40	42	43	38	39	44	42	40
Y	167	168	180	164	177	154	185	195	183	172

$\overline{x} = 40{,}7$ $\overline{y} = 174{,}5$ $\sigma_X^2 = 3{,}79$ $\sigma_{XY} = 0{,}59$

$$y - \overline{y} = \frac{\sigma_{XY}}{\sigma_X^2}(x - \overline{x}) \rightarrow y - 174{,}5 = \frac{0{,}59}{3{,}79}(x - 40{,}7) \rightarrow y = 3{,}62x + 27{,}10$$



17. Determina las dos rectas de regresión e indica la relación que hay entre las siguientes variables.

a)

X	10	10	13	15	12
Y	6	5	2	3	5

b)

X	8	10	11	12	16	13	12	17	13	13
Y	15	10	15	10	20	15	10	25	10	15

a) $\overline{x} = \frac{60}{5} = 12$ $\overline{y} = \frac{21}{5} = 4{,}2$

$\sigma_X^2 = \frac{738}{5} - 12^2 = 3{,}6$ $\sigma_{XY} = \frac{241}{5} - 12 \cdot 4{,}2 = -2{,}2$

Recta de regresión de Y sobre X: $y - 4{,}2 = -\frac{2{,}2}{3{,}6}(x - 12) \rightarrow y = -0{,}61x + 11{,}52$

$\sigma_X^2 = \frac{99}{5} - 4{,}2^2 = 2{,}16$

Recta de regresión de X sobre Y: $x - 12 = -\frac{2{,}2}{2{,}16}(y - 4{,}2) \rightarrow x = -1{,}02y + 16{,}28$

$\sigma_X = \sqrt{3{,}6} = 1{,}89$ $\sigma_Y = \sqrt{2{,}16} = 1{,}47$

$r_{XY} = -\frac{2{,}2}{1{,}89 \cdot 1{,}47} = -0{,}79 \rightarrow$ La dependencia es débil y negativa.

b) $\overline{x} = \frac{125}{10} = 12{,}5$ $\overline{y} = \frac{145}{10} = 14{,}5$

$\sigma_X^2 = \frac{1\,625}{10} - 12{,}5^2 = 6{,}25$ $\sigma_{XY} = \frac{1\,890}{10} - 12{,}5 \cdot 14{,}5 = 7{,}75$

Recta de regresión de Y sobre X: $y - 14{,}5 = \frac{7{,}75}{6{,}25}(x - 12{,}5) \rightarrow y = 1{,}24x - 1$

$\sigma_X^2 = \frac{2\,325}{10} - 14{,}5^2 = 22{,}25$

Recta de regresión de X sobre Y: $x - 12{,}5 = \frac{7{,}75}{22{,}25}(y - 14{,}5) \rightarrow x = 0{,}35y + 7{,}43$

$\sigma_X = \sqrt{6{,}25} = 2{,}5$ $\sigma_Y = \sqrt{22{,}25} = 4{,}72$

$r_{XY} = \frac{7{,}75}{2{,}5 \cdot 4{,}72} = 0{,}66 \rightarrow$ La dependencia es débil y positiva.

18. La tabla siguiente muestra el tiempo empleado por 6 atletas en una carrera y las horas de entrenamiento que dedicaron en el último mes.

Tiempo	11,03	12,91	10,98	11,26	11,12	11,18
Horas de entrenamiento	175	35	201	150	190	100

Determina las dos rectas de regresión e indica la relación que hay entre las variables.

$\overline{x} = 40{,}7$ $\overline{y} = 174{,}5$ $\sigma_X^2 = 3{,}79$ $\sigma_{XY} = 0{,}59$

$$\begin{cases} y - \overline{y} = \frac{\sigma_{XY}}{\sigma_X^2}(x - \overline{x}) \rightarrow y - 141{,}83 = -\frac{40{,}77}{0{,}55}(x - 11{,}41) \rightarrow y = 991{,}42 - 74{,}44x \\ x - \overline{x} = \frac{\sigma_{XY}}{\sigma_Y^2}(y - \overline{y}) \rightarrow x - 11{,}41 = -\frac{40{,}77}{4\,030{,}17}(y - 141{,}83) \rightarrow x = 12{,}85 - 0{,}01y \end{cases}$$

Las variables son linealmente dependientes, con una dependencia débil y negativa.

19. En un estudio estadístico, el coeficiente de correlación entre dos variables X e Y es $-0{,}8$. Se sabe que $\overline{x} = 20$, $\sigma_X = 4$; $\overline{y} = 8$ y $\sigma_Y = 1$.

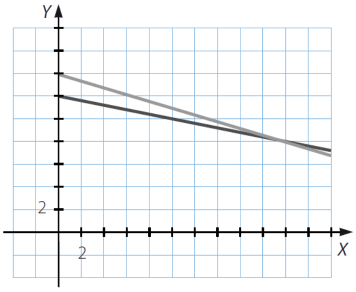
a) Determina las dos rectas de regresión, represéntalas y analiza la correlación que existe entre las variables.

b) Si $x = 30$, ¿cuál es la estimación de y?

a) $-0{,}8 = \dfrac{\sigma_{XY}}{4 \cdot 1} \rightarrow \sigma_{XY} = -3{,}2$

Recta de regresión de Y sobre X: $y - 8 = -\dfrac{3{,}2}{16}(x - 20) \rightarrow y = -0{,}2x + 12$

Recta de regresión de X sobre Y: $x - 20 = -\dfrac{3{,}2}{1}(y - 8) \rightarrow x = -3{,}2y + 45{,}6$



La dependencia es fuerte y negativa.

b) $y = -0{,}2 \cdot 30 + 12 = 6$

20. La tabla muestra el peso, X, y la altura, Y, de 10 personas. Estima el peso de otra persona que mide 1,90 m.

X	85	65	80	56	60	67	78	65	60	92
Y	1,80	1,61	1,71	1,57	1,65	1,73	1,79	1,67	1,59	1,83

$\overline{x} = 70{,}8$ $\quad \overline{y} = 1{,}70$ $\quad \sigma_Y^2 = 0{,}001$ $\quad \sigma_{XY} = 1{,}02$

$$x - \overline{x} = \frac{\sigma_{XY}}{\sigma_Y^2}(y - \overline{y}) \rightarrow x - 70{,}8 = -\frac{1{,}02}{0{,}001}(y - 1{,}70) \rightarrow x = 120y - 132{,}6$$

Para $y = 1{,}90$ m la estimación dada por la recta vendrá por:

$x = 120 \cdot 1{,}90 - 132{,}6 = 228 - 132{,}6 = 95{,}40$ kg

SABER HACER

21. Utiliza la calculadora para estudiar la correlación entre estas dos variables estadísticas.

X	2	3	5	7	1	4	2	8	10	1	9	8
Y	−2	0	−11	−16	1	−2	1	−16	−25	0	−21	−17

Establecemos el *Modo Regresión Lineal* en la calculadora.

Introducimos los datos y la calculadora halla la covarianza de las variables y sus desviaciones típicas marginales:

$\sigma_{XY} = -1{,}1\widehat{6}$ $\quad \sigma_X = 3{,}136$ $\quad \sigma_Y = 9{,}247$

Calculamos la correlación:

$$r_{XY} = \frac{-1{,}1\widehat{6}}{3{,}136 \cdot 9{,}247} = -0{,}04$$

22. **Una empresa quiere realizar un estudio estadístico sobre su gasto en publicidad (*X*, en miles de euros) y las ventas de productos que lanzan al mercado (*Y*, en miles de unidades).**

(20, 23)	(17, 20)	(11, 24)	(24, 4)	(23, 27)	(3, 21)	(15, 19)	(21, 1)	(19, 7)	(12, 12)
(19, 24)	(7, 7)	(21, 8)	(12, 19)	(11, 25)	(8, 13)	(17, 14)	(2, 21)	(9, 2)	(21, 11)
(3, 14)	(12, 18)	(21, 1)	(8, 12)	(12, 12)	(6, 29)	(12, 19)	(14, 21)	(3, 12)	(15, 15)

Agrupa el gasto en publicidad y las ventas de productos en intervalos con la misma amplitud 10.

Y \ X	[2, 11]	[12, 21]	[22, 31]	Total
[1, 10]	2	4	2	8
[11, 20]	4	10	0	14
[21, 30]	4	3	1	8
Total	10	17	3	30

23. **La tabla que aparece a continuación muestra un estudio estadístico en el que se analiza la relación entre las variables *X* e *Y*.**

Construye las tablas de frecuencias marginales de cada una de las variables.

Y \ *X*	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)
[0, 5)	0	1	0	0	3
[5, 10)	0	1	0	1	1
[10, 15)	2	2	2	1	1
[15, 20)	0	0	3	2	0
[20, 25)	2	0	2	2	1
[25, 30)	0	1	1	0	1
[30, 35)	1	3	0	1	1
[35, 40)	2	2	0	2	0

X	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	Total
f_i	7	10	8	9	8	42

Y	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35)	[35, 40)	Total
f_i	4	3	8	5	7	3	6	6	42

24. Los resultados de un estudio sobre la relación de dos variables estadísticas *X* e *Y*, son los siguientes.

Y \ X	[10, 12)	[12, 14)	[14, 16)	[16, 18)	[18, 20)
[0, 100)	12	3	5	11	7
[100, 200)	12	9	7	12	9
[200, 300)	15	1	12	5	6
[300, 400)	9	0	15	72	21

a) ¿Cuántos datos hay con *X* menor que 16?

b) ¿Cuántos valores de *Y* son menores que 100?

c) ¿Cuántos datos hay con *X* entre 12 y 18?

a) 104 b) 42 c) 152

25. Un agricultor quiere relacionar la cantidad de agua (en m³) y la cosecha de trigo obtenida (en toneladas). Para ello ha anotado los datos de las últimas 5 cosechas.

Determina la recta de regresión utilizando la calculadora.

Agua (m³)	4	7	12	6	14
Toneladas de trigo	17,9	24,3	38,5	22,5	50,1

Se establece el *Modo Regresión Lineal* en la calculadora.

Se introducen los datos y la calculadora halla directamente la media y la desviación típica.

$\overline{x} = 8,6$ $\overline{y} = 30,66$ $\sigma_{XY} = -30,82$ $\sigma_X = 3,28$ $\sigma_Y = 9,66$

$$\left.\begin{matrix} a = 3,11 \\ b = 3,88 \end{matrix}\right\} \rightarrow y = 3,11x + 3,88$$

26. Estudia la correlación entre estas variables utilizando la calculadora.

Determina la recta de regresión y razona si tiene sentido estimar el valor de *Y* si la variable *X* toma el valor 18.

X	14	16	17	14	15	12	13	13	14	16
Y	32	34	36	34	32	34	31	36	38	32

Se establece el *Modo Regresión Lineal* en la calculadora.

Se introducen los datos y la calculadora halla directamente la media y la desviación típica:

$\overline{x} = 14,4$ $\overline{y} = 33,9$

$\sigma_{XY} = 0,16$ $\sigma_X = 1,58$ $\sigma_Y = 2,23 \rightarrow r_{XY} = 0,04$

$$\left.\begin{matrix} a = 0,06 \\ b = 33 \end{matrix}\right\} \rightarrow y = 0,06x + 33$$

No tiene mucho sentido estimar el valor de *Y* para cuando *X* toma el valor 18, debido a que la correlación es muy cercana a 0, por lo que casi no existe dependencia.

27. **La recta de regresión que relaciona dos variables estadísticas es $y = 3{,}2x - 1{,}7$. Si se tiene que $\overline{y} = 2{,}34$, ¿cuál es el valor de $\overline{x}$?**

$$y = 3{,}2x - 1{,}7 \to \overline{y} = 3{,}2\overline{x} - 1{,}7 \to \frac{\overline{x} + 1{,}7}{3{,}2} = \overline{x} \to \overline{x} = 1{,}26$$

28. **La recta de regresión que relaciona dos variables es $y = 3{,}2x - 1{,}7$. Si se sabe que $\overline{x} = 2{,}87$, ¿se puede afirmar que cuanto mayor sea X mayor será Y ?**

La recta de regresión indica en este caso que existe una dependencia lineal creciente, porque la pendiente de la recta de regresión es positiva.

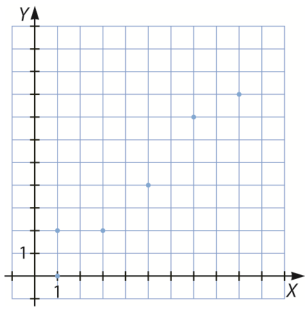
Se ve, por tanto, que si se eligen valores grandes para X, el valor para la Y también crecerá.

Por ejemplo:

$$\overline{y} = 3{,}2\overline{x} - 1{,}7 \to \overline{y} = 3{,}2 \cdot 2{,}87 - 1{,}7 = 7{,}48$$

29. **Dibuja la nube de puntos y analiza la dependencia.**

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



La nube de puntos se aproxima bastante a una recta y cuando la variable X crece, la variable Y también crece, por lo que existe una dependencia lineal fuerte y positiva entre ellas.

30. **Calcula la recta de regresión y el valor esperado para $x = 4$.**

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

Calculamos la recta de regresión de X sobre Y:

$$\overline{x} = \frac{25}{5} = 5 \qquad \overline{y} = \frac{23}{5} = 4{,}6 \qquad \sigma_X{}^2 = 8 \qquad \sigma_X = 2{,}83 \qquad \sigma_Y = 2{,}498 \qquad \sigma_{XY} = 6{,}8$$

$$r_{XY} = \frac{6{,}8}{2{,}83 \cdot 2{,}498} = 0{,}96$$

La ecuación de la recta de regresión sería:

$$y - 4{,}6 = \frac{6{,}8}{8}(x - 5) \to y = 0{,}85x + 0{,}35$$

Por tanto, el valor esperado para $x = 4$ sería:

$y = 0{,}85 \cdot 4 + 0{,}35 = 3{,}75$

Como r_{XY} está muy cerca de 1, la aproximación es bastante buena.

ACTIVIDADES FINALES

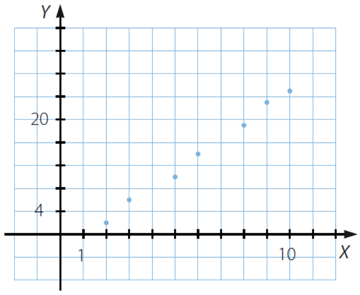
31. **Representa la nube de puntos asociada a estas variables bidimensionales y decide si hay dependencia entre las variables que las forman; si es así, indica de qué tipo es.**

a) (2, 2) (3, 6) (5, 10) (6, 14) (8, 19) (9, 23) (10, 25)

b) (5, 2) (6, 0) (8, −2) (10, −7) (11, −9) (13, −13) (15, −17)

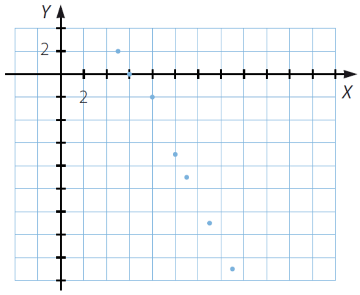
c) (120, 60) (122, 75) (126, 60) (128, 90) (130, 50) (132, 100) (136, 70)

a)



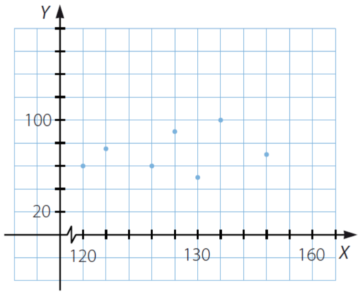
Depend. lineal positiva fuerte

b)



Depend. lineal negativa fuerte.

c)



Depend. lineal positiva débil.

32. **Representa la nube de puntos asociada a estas variables bidimensionales y decide si hay dependencia entre las variables que las forman.**

En caso afirmativo, califícala.

a)

A	6	8	9	11	13	15	16	18
B	8	13	13	16	21	26	28	33

b)

C	1	3	6	7	10	13	17	18
D	25	21	18	20	12	15	8	6

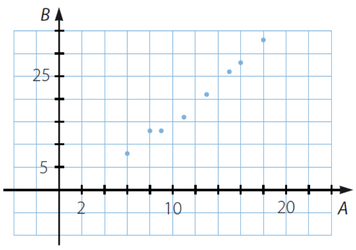
c)

E	110	112	115	116	118	120	121	124
F	40	45	35	40	60	70	45	33

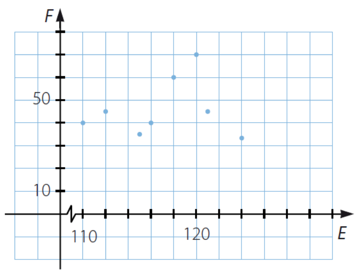
d)

G	26	24	23	22	18	15	14	12
H	8	12	14	7	10	11	9	13

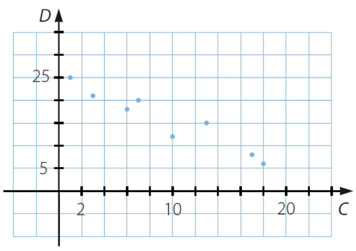
a) La dependencia es fuerte y positiva.



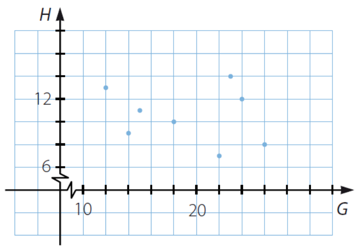
c) No se aprecia dependencia entre las variables *E* y *F*.



b) La dependencia es fuerte y negativa.



d) No se aprecia dependencia entre las variables *G* y *H*.



33. Halla las frecuencias marginales asociadas a estas tablas de doble entrada y calcula las medidas estadísticas de cada variable por separado.

a)

Y \ X	1	2	3	4
1	2	4	3	3
2	1	1	1	3
3	1	0	1	2
4	4	3	3	3

b)

Y \ X	[0, 2)	[2, 4)	[4, 6)	[6, 8)
10	10	13	5	8
20	8	4	3	9
30	5	5	7	9
40	13	12	11	12
50	4	5	7	11
60	3	6	9	11

c)

Y \ X	1	2	3	4	5	6	7	8	9	10
[0, 10)	1	5	6	4	8	5	2	1	3	4
[10, 20)	2	5	4	1	8	3	6	9	5	4
[20, 30)	0	2	1	5	4	7	8	3	2	1
[30, 40)	2	2	3	7	7	3	5	4	9	4

a)

X	1	2	3	4	Total
f_i	8	8	8	11	35

Y	1	2	3	4	Total
f_i	12	6	4	13	35

$\overline{x} = \frac{92}{35} = 2{,}629$ $\overline{y} = \frac{88}{35} = 2{,}514$

b)

X	[0, 2)	[2, 4)	[4, 6)	[6, 8)	Total
x_i	1	3	5	7	
f_i	43	45	42	60	190

Y	10	20	30	40	50	60	Total
f_i	36	24	26	48	27	29	190

$\overline{x} = \frac{808}{190} = 4{,}253$ $\overline{y} = \frac{6630}{190} = 34{,}895$

c)

X	1	2	3	4	5	6	7	8	9	10	**Total**
f_i	5	14	14	17	27	18	21	17	19	13	165

Y	[0, 10)	[10, 20)	[20, 30)	[30, 40)	**Total**
y_i	5	15	25	35	
f_i	39	47	44	46	165

$\overline{x} = \dfrac{970}{165} = 5,\widehat{878}$ $\qquad$ $\overline{y} = \dfrac{3\,335}{165} = 20,\widehat{212}$

34. **Las notas en Lengua e Inglés de los 30 alumnos de una clase en la última evaluación han sido las siguientes.**

Lengua:	**3 7 8 7 5**	**2 5 9 5 4**	**3 5 3 6 3**
	8 5 7 7 6	**2 4 9 4 9**	**7 6 7 1 7**
Inglés:	**2 6 10 6 4**	**2 5 9 5 5**	**2 4 1 5 1**
	10 4 7 8 4	**2 5 9 5 9**	**8 5 7 0 7**

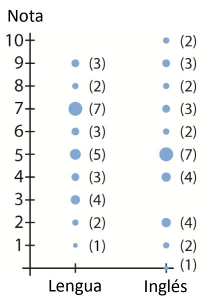
a) Construye la tabla de doble entrada con estos datos.

b) Realiza el diagrama de dispersión.

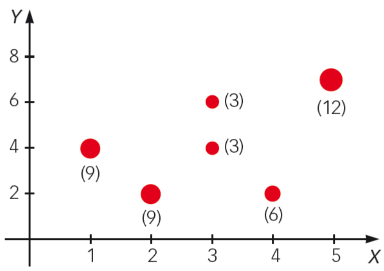
a)

Asignatura \ Notas	0	1	2	3	4	5	6	7	8	9	10	Total
Lengua	0	1	2	4	3	5	3	7	2	3	0	30
Inglés	1	2	4	0	4	7	2	3	2	3	2	30
Total	1	3	6	4	7	12	5	10	4	6	2	60

b)

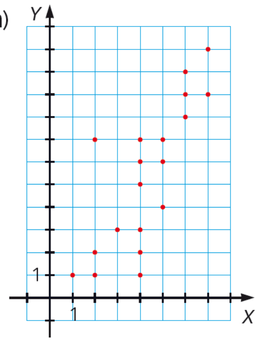


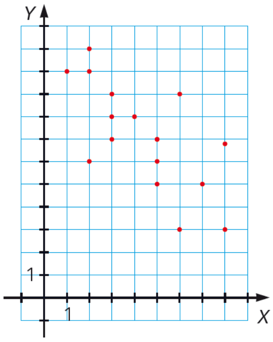
35. **Construye la tabla de doble entrada correspondiente a partir del diagrama de dispersión, teniendo en cuenta la frecuencia de los datos que figuran entre paréntesis.**



X / Y	1	2	3	4	5	Total
2	0	9	0	6	0	15
4	9	0	3	0	0	12
6	0	0	3	0	0	3
7	0	0	0	0	12	12
Total	9	9	6	6	12	42

36. A partir de estos diagramas de dispersión, construye sus correspondientes tablas de doble entrada.

a) 

b) 

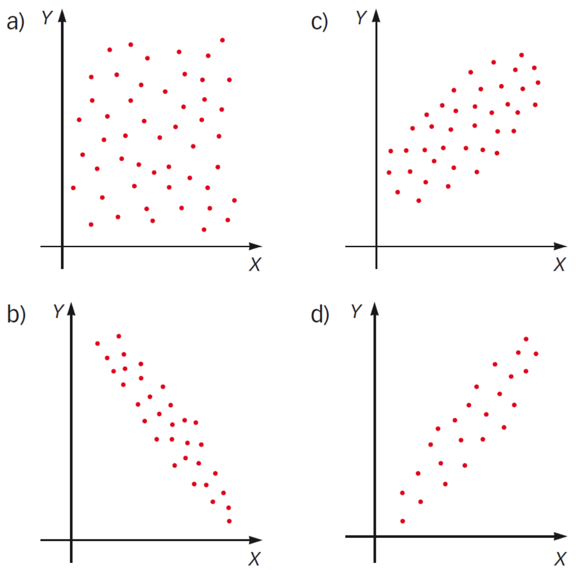
a)

X / Y	1	2	3	4	5	6	7	Total
1	1	1	0	1	0	0	0	3
2	0	1	0	1	0	0	0	2
3	0	0	1	1	0	0	0	2
4	0	0	0	0	1	0	0	1
5	0	0	0	1	0	0	0	1
6	0	0	0	1	1	0	0	2
7	0	1	0	1	1	0	0	3
8	0	0	0	0	0	1	0	1
9	0	0	0	0	0	1	1	2
10	0	0	0	0	0	1	0	1
11	0	0	0	0	0	0	1	1
Total	1	3	1	6	3	3	2	19

b)

Y \ X	1	2	3	4	5	6	7	8	Total
3	0	0	0	0	0	1	0	1	2
5	0	0	0	0	1	0	1	0	2
6	0	1	0	0	1	0	0	0	2
7	0	0	1	0	1	0	0	1	3
8	0	0	1	1	0	0	0	0	2
9	0	0	1	0	0	1	0	0	2
10	1	1	0	0	0	0	0	0	2
11	0	1	0	0	0	0	0	0	1
Total	1	3	3	1	3	2	1	2	16

37. **A partir de los diagramas de dispersión, decide si hay o no dependencia lineal entre las variables y, en su caso, si es fuerte o débil, y si es positiva o negativa.**



a) No hay dependencia lineal.

b) La dependencia lineal es débil y positiva.

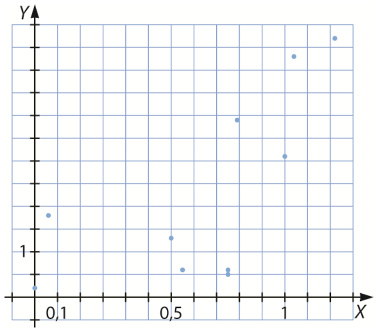
c) La dependencia lineal es débil y positiva.

d) La dependencia lineal es fuerte y positiva.

38. **Se quiere realizar un estudio para determinar la influencia de la velocidad del viento en la permanencia de una plaga en un tipo de planta. Para ello se registraron la velocidad en metros por segundo, *X*, y el tiempo en horas, *Y*. La tabla muestra los valores ordenados de forma ascendente respecto del valor de *X*.**

X	0,00	0,06	0,50	0,55	0,75	0,75	0,79	1,00	1,04	1,22
Y	0,2	1,8	1,3	0,6	0,5	0,6	3,9	3,3	5,3	5,7

Realiza el diagrama de dispersión, indica si hay dependencia entre las variables y, en caso afirmativo, señala de qué tipo es.



La nube de puntos se aproxima poco a una recta y cuando la variable *X* crece, la variable *Y* crece, por lo que existe una dependencia lineal débil y positiva entre ellas.

39. **La tabla muestra el número de cuadros que han pintado los alumnos de un taller sobre paisajes y bodegones.**

Bodegones \ Paisajes	4	5	6	7	8
4	2	1	0	0	0
5	4	4	3	0	1
6	2	5	4	2	0
8	0	0	3	2	1

a) Determina las tablas de frecuencias marginales de paisajes y bodegones.
b) Calcula las medias y las desviaciones típicas de cada una de las variables.
c) Realiza el diagrama de dispersión correspondiente a la variable bidimensional.
d) A partir de su diagrama de dispersión, indica si hay una dependencia entre las variables.

a) Tabla de frecuencias marginales de los paisajes

x_i	f_i
4	8
5	10
6	10
7	4
8	2
Total	34

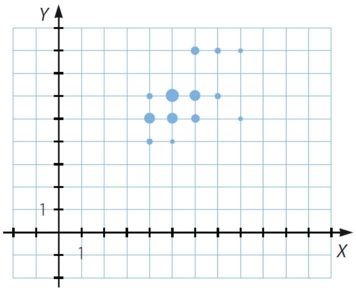
Tabla de frecuencias marginales de los bodegones

y_i	f_i
4	3
5	12
6	13
8	6
Total	34

b) $\overline{x} = 5{,}47$ $\quad \overline{y} = 5{,}82$

$\sigma_X = 1{,}15$ $\quad \sigma_Y = 1{,}19$

c)



d) Observando el diagrama de dispersión, podemos deducir que existe una dependencia lineal débil positiva entre las variables.

40. La siguiente tabla muestra los ingresos familiares mensuales de una familia en cientos de euros, *X*, y los metros cuadrados de la vivienda familiar, *Y*.

Y \ X	[0,5)	[5, 10)	[10,15)	[15,20)	[20,25)
[0, 50)	20	18	2	1	0
[50, 100)	25	40	30	2	1
[100, 200)	5	10	15	25	3
[200, 250)	0	5	15	20	8
[250, 300)	0	1	2	7	10

a) Indica las distribuciones marginales de las variables.
b) Halla las medias y las desviaciones típicas de cada una de las variables.
c) Realiza el diagrama de dispersión correspondiente a la variable bidimensional.
d) A partir de su diagrama de dispersión, indica si hay una dependencia entre las variables.

a)

X	[0, 50)	[50, 100)	[100, 200)	[200, 250)	[250, 300)	**Total**
f_i	41	98	58	48	20	265

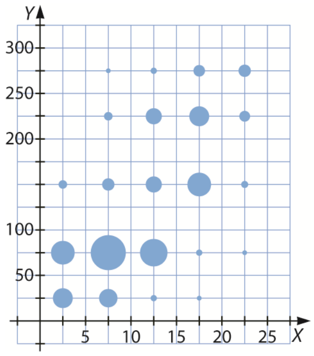
Y	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	**Total**
f_i	50	74	64	55	22	265

b) $$\overline{x} = \frac{41 \cdot 25 + 98 \cdot 75 + 58 \cdot 150 + 48 \cdot 225 + 20 \cdot 275}{265} = \frac{33375}{265} = 125,94$$

$$\overline{y} = \frac{50 \cdot 2,5 + 74 \cdot 7,5 + 64 \cdot 12,5 + 55 \cdot 17,5 + 22 \cdot 2,25}{265} = \frac{2937,5}{265} = 11,08$$

$$\sigma_X = \frac{\sum_{i=1}^{N} f_i \cdot x_i^2}{N} - \overline{x}^2 = \frac{5\,819\,506}{265} - 125,94 = 21\,960,4 - 125,94 = 21\,834,46$$

$$\sigma_Y = \frac{\sum_{i=1}^{N} f_i \cdot y_i^2}{N} - \overline{y}^2 = \frac{42\,456,25}{265} - 11,08 = 160,21 - 11,08 = 149,13$$

c) 

d) La nube de puntos es dispersa y los puntos no están muy pegados entre sí. Existe una dependencia lineal débil y creciente.

41. Calcula la covarianza y el coeficiente de correlación para las variables bidimensionales indicadas en estas tablas.

P	0	1	2	3	4	5	6	7
Q	20	18	17	15	12	10	7	4

R	90	80	70	60	50	40	30
S	−5	−7	−8	−11	−13	−16	−17

$\sigma_{PQ} = -7{,}22$ $\quad r_{PQ} = -0{,}11$ $\quad \sigma_{RS} = 84{,}29$ $\quad r_{RS} = 0{,}99$

42. Halla la covarianza y el coeficiente de correlación correspondientes a estas variables estadísticas.

T	−12	−14	−15	−16	−18	−20	−22
U	8	5	3	12	20	10	6

V	2,4	2,8	3,2	3,6	4	4,4	4,8	5,2
W	100	150	220	270	340	400	460	520

$\sigma_{TU} = -3{,}69$ $\quad r_{TU} = -0{,}22$ $\quad \sigma_{VW} = 127{,}5$ $\quad r_{VW} = 0{,}99$

43. Una compañía de seguros quiere relacionar el número de vehículos que circulan por una determinada autopista a más de 115 km/h, *X*, y el número de accidentes que ocurren en ella, *Y*. Durante una semana obtuvo los siguientes resultados.

N.º vehículos	150	180	100	80	200	170	400
N.º accidentes	5	7	2	1	9	6	15

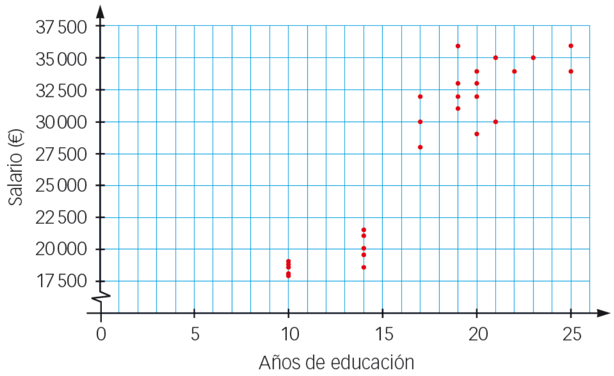
a) Halla las medias y desviaciones típicas de las variables número de vehículos y número de accidentes.
b) Calcula la covarianza de la variable bidimensional.
c) Halla el coeficiente de correlación e interpreta su valor.

a) $\begin{cases} \overline{x} = 182{,}85 \\ \sigma_X = 105 \end{cases}$ $\qquad \begin{cases} \overline{y} = 6{,}43 \\ \sigma_Y = 4{,}69 \end{cases}$

b) $\sigma_{XY} = 480{,}24$

c) $r_{XY} = 0{,}98$ es casi 1. Existe una dependencia lineal fuerte y, al ser mayor que 0, positiva.

44. **Se ha realizado una encuesta a 30 empleados de una empresa. Entre los datos recogidos consta el salario anual, en miles de euros, y los años de educación. Al realizar el diagrama de dispersión se obtiene la siguiente nube de puntos.**



Analiza, justificando la respuesta, cómo debe ser el signo de la covarianza y el del coeficiente de correlación.

Tanto la covarianza como el coeficiente de correlación tienen que ser positivos, porque la nube de puntos es creciente cuando crecen las variables.

45. **Se toman 6 muestras de suero y se anotan en una tabla el tiempo que llevaban preparadas y el número de bacterias encontradas en un mililitro.**

N.º de horas (X)	1	1	1	2	3	3
N.º de bacterias (Y)	10	21	23	35	54	60

a) Halla el valor de las medias y desviaciones típicas de las variables, el número de horas y el de bacterias.
b) Calcula la covarianza de la variable bidimensional.
c) Halla el coeficiente de correlación e interpreta su valor.

a) $\begin{cases} \overline{X} = 1{,}83 \\ \sigma_X = 0{,}83 \end{cases}$ $\begin{cases} \overline{Y} = 33{,}83 \\ \sigma_Y = 19{,}71 \end{cases}$ b) $\sigma_{XY} = 18{,}77$ c) $r_{XY} = 0{,}97$

46. **En una clase se realiza una encuesta sobre la cantidad de trabajos hechos y el grado de satisfacción, de 1 a 5, con el profesor de la asignatura.**

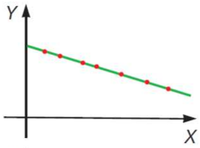
		Trabajos realizados					
		1	**2**	**3**	**4**	**5**	**Total**
Satisfacción con el profesor	**1**	3	2	1	0	2	8
	2	5	4	1	2	2	14
	3	2	2	8	2	3	17
	4	1	3	4	6	1	15
	5	0	1	3	7	9	20
	Total	11	12	17	17	17	74

Analiza, justificando la respuesta, cómo debe ser el signo de la covarianza y el del coeficiente de correlación.

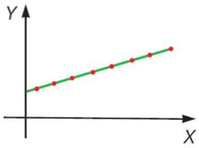
El signo de la covarianza debe ser positivo, y el coeficiente de correlación es positivo y cercano a 0, porque crece la variable "satisfacción" cuando crece el trabajo realizado, pero es cercano a 0 porque los datos están muy dispersos.

47. Relaciona cada diagrama de dispersión con el coeficiente de correlación.

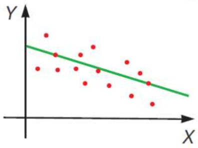
a) 0 b) 0, 3 c) −1 d) 1 e) 0 f) −0,6



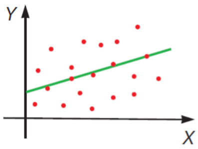
c) porque la pendiente es negativa y los puntos están sobre la recta de regresión.



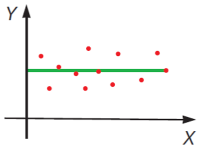
d) porque la pendiente es positiva y los puntos están sobre la recta de regresión.



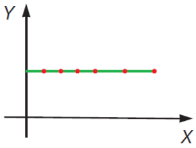
f) porque la pendiente es negativa y los puntos no están sobre la recta de regresión.



b) porque la pendiente es positiva y los puntos no están sobre la recta de regresión.

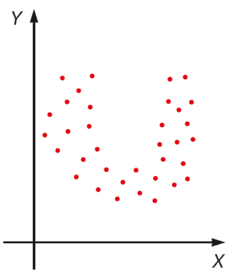


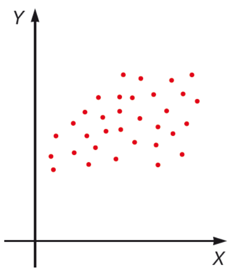
a) porque la pendiente es 0.

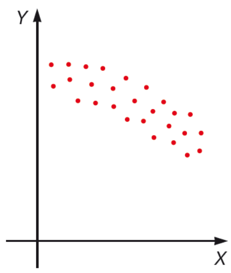


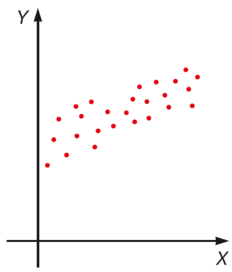
e) porque la pendiente es 0.

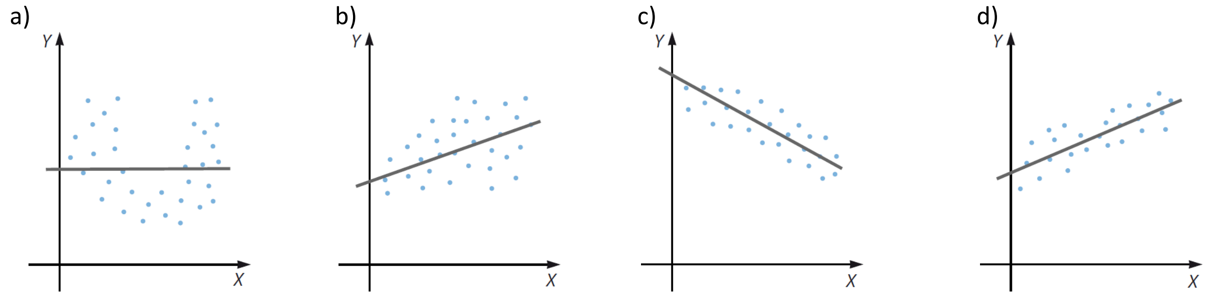
48. Traza a mano alzada, y sin realizar cálculos, la recta de regresión de las siguientes variables bidimensionales.

a)


b)


c)


d)




49. Determina la recta de regresión de *Y* sobre *X* y la recta de regresión de *X* sobre *Y* correspondientes a estas tablas.

a)

X	0	3	5	6	7
Y	2	−7	−14	−16	−9

b)

X	7,5	8	9	10	10,5	12	13
Y	20	20,5	23	24	25	27	28

a) $\begin{cases} \overline{x} = 4{,}2 & \overline{y} = -3{,}2 \\ \sigma_X^2 = 7{,}7 & \sigma_Y^2 = 133{,}7 \end{cases} \rightarrow \sigma_{XY} = -10{,}7$

Recta de regresión de *Y* sobre *X*:

$$y + 3{,}2 = -\frac{10{,}7}{7{,}7}(x - 4{,}2) \rightarrow y = -1{,}39x + 2{,}64$$

Recta de regresión de *X* sobre *Y*:

$$x - 4{,}2 = -\frac{10{,}7}{133{,}7}(y + 3{,}2) \rightarrow x = -0{,}08y + 3{,}94$$

b) $\begin{cases} \overline{x} = 10 & \overline{y} = 23{,}93 \\ \sigma_X^2 = 4{,}08 & \sigma_Y^2 = 9{,}20 \end{cases} \rightarrow \sigma_{XY} = 6{,}08$

Recta de regresión de *Y* sobre *X*:

$$y - 23{,}93 = \frac{6{,}08}{4{,}08}(x - 10) \rightarrow y = 1{,}49x + 9{,}03$$

Recta de regresión de *X* sobre *Y*:

$$x - 10 = \frac{6{,}08}{9{,}20}(y - 23{,}93) \rightarrow x = 0{,}66y - 5{,}82$$

50. Encuentra cinco puntos de la recta $y = 4x + 6$.

a) Calcula el coeficiente de correlación correspondiente y explica el resultado.

b) Halla las dos rectas de regresión.

Respuesta abierta.

x	−2	−1	0	1	2
y	−2	2	6	10	14

a) $\overline{x} = 0 \qquad \overline{y} = 6 \qquad \sigma_X = \sqrt{2} = 1{,}41 \qquad \sigma_Y = \sqrt{32} = 5{,}66 \qquad \sigma_{XY} = 8$

$r_{XY} = 1 \rightarrow$ La dependencia es lineal.

b) Recta de regresión de *Y* sobre *X*:

$$y - 6 = \frac{8}{2}(x - 0) \rightarrow y = 4x + 6$$

Recta de regresión de *X* sobre *Y*:

$$x - 0 = \frac{8}{32}(y - 6) \rightarrow x = \frac{1}{4}y - \frac{3}{2}$$

51. **Sabiendo que $\bar{x} = 3$, $\sigma_X^2 = 6$, $\sigma_Y^2 = 8$ y que la recta de regresión de *Y* sobre *X* es $y = 4 - 0{,}68x$, halla la recta de regresión de *X* sobre *Y*.**

Para calcular la recta de regresión de *X* sobre *Y* se necesitan aún $\bar{y}$ y σ_{XY}, que se obtienen de la recta de regresión de *Y* sobre *X* y de los datos del enunciado.

$$-0{,}68 = \frac{\sigma_{XY}}{s_X^2} \rightarrow -0{,}68 = \frac{\sigma_{XY}}{6} \rightarrow \sigma_{XY} = 6\cdot(-0{,}68) = -4{,}08$$

$$-4 = \frac{\sigma_{XY}}{s_X^2}\bar{x} - \bar{y} \rightarrow \bar{y} = \frac{\sigma_{XY}}{s_X^2}\bar{x} + 4 = -0{,}68\cdot 3 + 4 = 1{,}96$$

La recta de regresión de *X* sobre *Y* es: $x - 3 = -\frac{4{,}08}{8}(y - 1{,}96) \rightarrow x = 4 - 0{,}51y$

52. **Sabiendo que $\bar{x} = 3{,}2$; $\bar{y} = 1{,}2$ y la recta de regresión pasa por el punto (3,9; 3,8), halla la recta de regresión de *Y* sobre *X*.**

$$y - \bar{y} = \frac{\sigma_{XY}}{\sigma_X^2}(x - \bar{x}) \rightarrow y - 1{,}2 = \frac{\sigma_{XY}}{\sigma_X^2}(x - 3{,}2) \rightarrow y = \frac{\sigma_{XY}}{\sigma_X^2}x - (3{,}2)\frac{\sigma_{XY}}{\sigma_X^2} + 1{,}2$$

Como pasa por el punto (3,9; 3,8), se sustituyen y se obtiene el valor de $\frac{\sigma_{XY}}{\sigma_X^2}$:

$$3{,}8 = (3{,}9)\frac{\sigma_{XY}}{\sigma_X^2} - (3{,}2)\frac{\sigma_{XY}}{\sigma_X^2} + 1{,}2 \rightarrow 2{,}6 = (0{,}7)\frac{\sigma_{XY}}{\sigma_X^2} \rightarrow \frac{\sigma_{XY}}{\sigma_X^2} = 3{,}71$$

Con lo que la recta de regresión será: $y = 3{,}71x - 10{,}69$

53. **Se está estudiando la relación entre el número de meses que una persona está apuntada a un centro deportivo y el nivel de satisfacción con las instalaciones del centro, medido de 0 a 10. Para ello se eligen 10 personas al azar y se anotan los siguientes resultados.**

Meses	8	7	10	3	6	13	4	10	12	5
Satisfacción	6	6	8	4	5	8	3	8	9	4

a) Halla el valor del coeficiente de correlación.

b) Calcula la recta de regresión.

a) $\left\{\begin{array}{l}\sigma_{XY} = 6{,}69\\ \sigma_X = 3{,}39\\ \sigma_Y = 2{,}08\end{array}\right. \rightarrow r_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{6{,}69}{3{,}39\cdot 2{,}08} = 0{,}95$

b) $\left\{\begin{array}{l}\bar{x} = 7{,}8\\ \bar{y} = 6{,}1\end{array}\right. \rightarrow y - \bar{y} = \frac{\sigma_{XY}}{\sigma_X^2}(x - \bar{x}) \rightarrow y = 0{,}58x + 1{,}57$

54. **Una compañía discográfica ha recopilado información sobre 42 de sus grupos musicales, el número de conciertos y las ventas de discos, en miles, obteniendo la siguiente tabla.**

N.º de discos \ Conciertos	[0, 15)	[15, 30)	[30, 45)	[45, 60)
[0,5)	1	3	2	1
[5,10)	2	1	4	1
[10, 15)	2	2	2	6
[15, 30)	1	2	5	7

Obtén la recta de regresión y explica si hay, o no, dependencia lineal.

$$\begin{cases}\overline{x} = 35,71\\ \overline{y} = 13,45\end{cases} \to \begin{cases}\sigma_X^2 = 248,59\\ \sigma_Y^2 = 56,24\end{cases} \to \sigma_{XY} = -436,55$$

Por lo que la recta de regresión es: $y - 13,45 = -\frac{436,55}{248,59}(x - 35,71) \to y = 76,16 - 1,76x$

Hay dependencia lineal negativa, porque la pendiente es distinta de 0.

55. Una cofradía de pescadores registra la cantidad de sardinas que llegan al puerto, en kilogramos, y el precio de la subasta en la lonja, en euros por kilo, obteniendo los siguientes resultados.

X (kg)	1 000	1 200	1 250	1 500	1 400	1 350	1 550
Y (€/kg)	1,80	1,68	1,65	1,32	1,44	1,50	1,20

a) Halla el coeficiente de correlación lineal e interpreta su valor.
b) Escribe la ecuación de la recta de regresión.
c) Estima cuál sería el precio de las sardinas si un día se pescasen 1 300 kilogramos.

a) $r_{XY} = -0,98$

Existe dependencia lineal negativa, y es cercana a – 1 porque sus valores están muy cercanos a la recta de regresión.

b) La recta de regresión tiene por fórmula $y = 2,96 - 0,01x$.

c) La estimación se obtiene al calcular en la recta de regresión el valor de *y* cuando se introduce como valor de $x = 1\,300$: $y = 2,96 - 0,001 \cdot 1300 = 2,96 - 1,3 = 1,66$

56. Un equipo de baloncesto ha publicado la siguiente estadística.

- **N.º de jugadores: 10**
- **Estaturas: $\overline{x} = 1,96$ cm, $\sigma_x = 7,06$**
- **Pesos: $\overline{y} = 90,1$ kg, $\sigma_y = 6,92$**
- **Covarianza: $\sigma_{xy} = 36,7$**

a) Halla el coeficiente de correlación.
b) Calcula la recta de regresión de *Y* sobre *X*.
c) Si el equipo ficha un jugador que mide 205 cm, ¿cuál es el peso que cabría esperar de él? Justifica tu respuesta.

a) $r_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{36,7}{7,06 \cdot 6,92} = 0,75$

b) $y - 90,1 = \frac{36,7}{49,84}(x - 1,96) \to y = 0,73x + 88,66$

c) La estimación se obtiene al calcular en la recta de regresión el valor de *y* cuando se introduce como valor de $x = 2,05$: $y = 0,73 \cdot (2,05) + 88,66 = 90,15$

57. **María y Diego viven en la misma calle, pero en aceras opuestas. Los dos tienen un termómetro en su balcón y, como María cree que el suyo está estropeado, deciden tomar la temperatura exterior, en °C, durante una semana y a la misma hora del día.**

Han anotado los resultados en una tabla.

Diego	22	24	25	27	18	20	21
María	18	20	18	17	20	21	16

a) ¿Crees que las dos variables están relacionadas? ¿Y opinas que deberían estarlo?

b) Razona si con estos datos se puede obtener alguna conclusión sobre el termómetro de María.

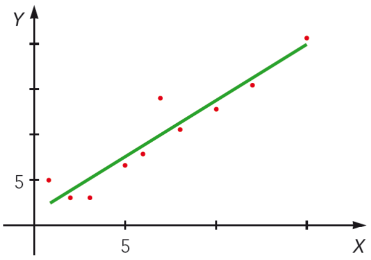
a) $\overline{x} = 22{,}43$ $\overline{y} = 18{,}57$ $\sigma_X = 2{,}86$ $\sigma_Y = 1{,}69$ $\sigma_{XY} = -2{,}097$

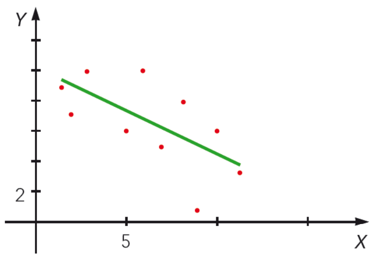
$r_{XY} = -0{,}43 \rightarrow$ La dependencia es débil y negativa.

Las dos variables están poco relacionadas, pues al estar los termómetros en lados opuestos de la acera reciben distinta exposición solar.

b) Como la dependencia es débil no se puede concluir nada sobre el termómetro de María.

58. **Tenemos dos variables bidimensionales representadas por estas nubes de puntos.**





a) Elige los coeficientes de correlación de ambas y razónalo.

$-0{,}92$ $0{,}95$ $0{,}6$ $-0{,}65$

b) ¿Cuáles son las ecuaciones de las dos rectas de regresión correspondientes? Justifica la respuesta.

$y = 3x + 0{,}2$ $y = -0{,}6x + 10$

$y = 1{,}3x + 0{,}9$ $y = -2x + 12{,}6$

a) El coeficiente de correlación de las variables representadas en el gráfico I es 0,95; porque la nube de puntos muestra una dependencia entre las variables fuerte y positiva. El coeficiente de correlación de las variables representadas en el gráfico II es $-0{,}65$; por ser la dependencia entre las variables débil y negativa.

b) La recta de regresión del gráfico I es $y = 1{,}3x + 0{,}9$; ya que la pendiente de la recta dibujada es un valor próximo a 1. La recta de regresión del gráfico II es $y = -0{,}6x + 10$, puesto que el valor de la ordenada de la recta representada es 10.

59. **Se tiene la siguiente variable bidimensional.**

X	3	5	8	9	10	12	15
Y	2	3	7	4	8	5	8

Investiga lo que sucede con la covarianza y el coeficiente de correlación en cada caso.

a) Sumamos 10 a todos los valores de la variable *X*.

b) Sumamos 10 a todos los valores de *X* y de *Y*.

c) Multiplicamos por 4 todos los valores de la variable *X*.

d) Multiplicamos por 4 todos los valores de *X* y de *Y*.

$\overline{x} = 8{,}86$ $\overline{y} = 5{,}29$ $\sigma_{XY} = 6{,}42$

$\sigma_X = \sqrt{14{,}07} = 3{,}75$ $\sigma_Y = \sqrt{5{,}02} = 2{,}24$ $r_{XY} = 0{,}76$

a)

X	13	15	18	19	20	22	25
Y	2	3	7	4	8	5	8

$\overline{x} = 8{,}86 + 10 = 18{,}86$ $\sigma_{XY} = 6{,}42$

$\sigma_X = 3{,}75$ $r_{XY} = 0{,}76$

b)

X	13	15	18	19	20	22	25
Y	12	13	17	14	18	15	18

$\overline{y} = 5{,}29 + 10 = 15{,}29$ $\sigma_{XY} = 6{,}42$

$\sigma_Y = 2{,}24$ $r_{XY} = 0{,}76$

c)

X	12	20	32	36	40	48	60
Y	2	3	7	4	8	5	8

$\overline{x} = 8{,}86 \cdot 4 = 35{,}44$ $\sigma_{XY} = 6{,}42 \cdot 4 = 25{,}68$

$\sigma_X = 3{,}75 \cdot 4 = 15$ $r_{XY} = 0{,}76$

d)

X	12	20	32	36	40	48	60
Y	8	12	28	16	32	20	32

$\overline{y} = 5{,}29 \cdot 4 = 21{,}16$ $\sigma_{XY} = 6{,}42 \cdot 16 = 102{,}72$

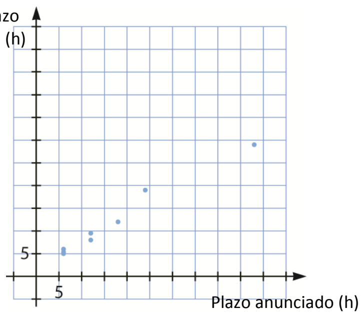
$\sigma_Y = 2{,}24 \cdot 4 = 8{,}96$ $r_{XY} = 0{,}76$

60. Una empresa de mensajería ha anotado el plazo de entrega, en horas, que anunciaba en sus envíos y el plazo real de entrega, también en horas, obteniendo la siguiente tabla.

Plazo anunciado	6	12	6	24	18	6	12	48
Plazo real	5	8	5,5	19	12	6	9,5	29

a) Representa los datos en una nube de puntos.
b) Dibuja aproximadamente la recta de regresión.
c) Calcula el coeficiente de correlación.
d) Halla la recta de regresión, dibújala y compárala con el valor obtenido en la aproximación.

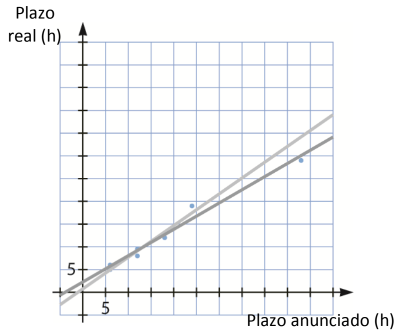
a) y b)



c) $r_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} = 0{,}99$

d) $\overline{x} = 16{,}5$ $\overline{y} = 11{,}75$ $\sigma_X{}^2 = 177{,}75$ $\sigma_{XY} = 102{,}75$

$$y - 11{,}75 = \frac{102{,}75}{177{,}75}(x - 16{,}5) \rightarrow y = 0{,}58x + 2{,}18$$



61. **Se desea estudiar la repercusión que tiene la lluvia en el número de visitas a un parque de atracciones. Para ello se observan, durante los últimos diez años, el número de días de lluvia durante la temporada en la que está abierto, *X*, y el número de visitas en la temporada, *Y*.**

X	9	13	15	16	19	20	21	22	23	18
Y	41	40	39	37	34	31	30	27	25	35

a) Calcula el coeficiente de correlación e interpreta el valor obtenido.

b) Halla la recta de regresión que explica el número de visitantes en función de los días de lluvia.

c) ¿Cuál es la previsión de visitas para el próximo año si hay una predicción de ser un año poco lluvioso con solo 12 días de lluvia en la temporada de apertura?

d) Determina la recta de regresión de *X* sobre *Y*.

e) Representa la nube de puntos y las dos rectas de regresión en los mismos ejes.

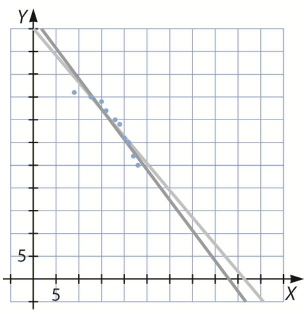
a) $r_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} = -0,94$ $\rightarrow$ Hay una correlación lineal fuerte, y el signo negativo implica que cuando crece la variable *X*, la *Y* decrece.

b) $y - 33,9 = -\dfrac{22,82}{19,16}(x - 17,6) \rightarrow y = 54,87 - 1,19x$

c) Con solo doce días lluviosos la previsión vendrá dada por la fórmula $y = 54,87 - 1,19 \cdot 12 = 40,59$.

d) $x = 42,93 - 0,75y$

e)



62. **La siguiente tabla muestra los datos obtenidos al preguntar por los ingresos familiares mensuales, *X*, en euros, y por los metros cuadrados de sus viviendas, *Y*.**

X \ *Y*	[30, 50)	[50, 70)	[70, 90)	[90, 110)	[110, 130)
[0,500)	20	18	2	1	0
[500, 1 000)	25	40	30	2	1
[1 000, 1 500)	5	10	15	25	3
[1 500, 2 000)	0	5	15	20	8
[2 000, 2 500)	0	1	2	7	10

a) Calcula las distribuciones marginales.

b) Calcula la covarianza, el coeficiente de correlación y las dos rectas de regresión.

a)

X	[0, 500)	[500, 1 000)	[1 000, 1 500)	[1 500, 2 000)	[2 000, 2 500)	**Total**
Frecuencias	41	98	58	48	20	**265**
Dist. Marginal	0,15	0,37	0,22	0,18	0,08	**1**

Y	[30, 50)	[50, 70)	[70, 90)	[90, 110)	[110, 130)	**Total**
Frecuencias	50	74	64	55	22	**265**
Dist. Marginal	0,19	0,28	0,24	0,21	0,08	**1**

b) $\sigma_{XY} = 25\,000$ $\qquad r_{XY} = 1$

Rectas de regresión: $\begin{cases} y = 30 + 0{,}04x \\ x = 25y - 750 \end{cases}$

63. Una marca de bicicletas de competición ofrece la siguiente comparativa entre el precio y el peso de sus bicicletas.

Peso (kg)	6,45	6,50	6,70	7,35	7,15	7,25
Precio (€)	6 000	4 700	4 200	3 300	2 700	2 600

Halla la covarianza y el coeficiente de correlación. ¿Cómo es la relación entre las dos variables?

$X \longrightarrow$ Peso en kilogramos de las bicicletas

$Y \longrightarrow$ Precio en euros de las bicicletas

$\sigma_{XY} = -470$ $\qquad r_{XY} = -0{,}90$

Existe una relación de dependencia lineal fuerte y negativa entre las dos variables, porque cuando la variable X crece, la Y decrece.

64. Las notas en idiomas, Francés e Inglés, de 10 estudiantes elegidos al azar de un centro escolar han sido las siguientes.

Inglés	9	6	3	6	7	6	9	8	10	6
Francés	8	6	4	3	8	7	8	7	10	5

a) Representa la nube de puntos correspondiente a esta distribución. ¿Qué hipótesis se puede hacer a la vista de la representación?

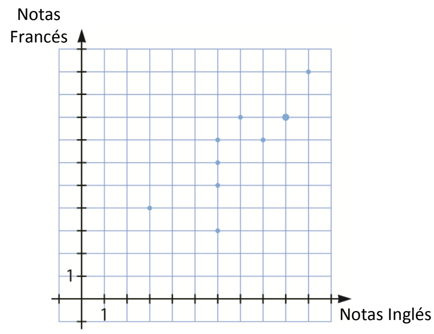
b) Halla los parámetros de la recta de regresión Y/X y la de X/Y e interpreta los coeficientes calculados.

c) Representa las dos rectas de regresión junto con la nube de puntos.

d) Para un alumno que haya obtenido un 4 en Inglés, ¿qué nota se le pronostica en Francés?

e) Para un alumno que ha tenido un 9 en Francés, ¿qué nota se le pronostica en Inglés?

a)

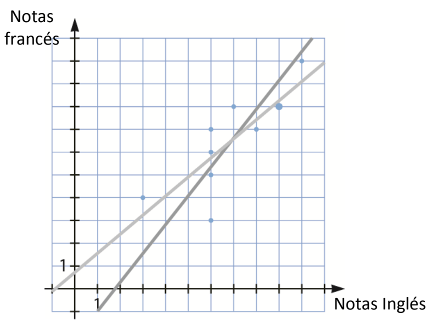


Hipótesis: la nota en Inglés va a ser parecida a la de Francés.

b) Rectas de regresión: $\begin{cases} y = 0{,}71 + 0{,}84x \\ x = 0{,}79y + 1{,}77 \end{cases}$

Como las pendientes de las rectas son muy parecidas, ello implica que los datos están muy pegados a las rectas regresoras.

c)



d) Si un alumno ha obtenido una 4 en Inglés, se espera que en Francés tenga un $y = 0{,}71 + 0{,}84 \cdot 4 = 4{,}07$.

e) Si un alumno ha obtenido una 9 en Francés, se espera que en Inglés tenga un $x = 0{,}79 \cdot 9 + 1{,}77 = 8{,}88$.

65. Un inversor bursátil quiere predecir la evolución que va a tener el Índice de la Bolsa de Madrid (IBEX).

Ha concluido que lo que sucede con el IBEX un día es lo que le ocurre a la cotización de la empresa AW&B el día anterior.

Investiga si esto es correcto, a partir de sus cotizaciones durante una semana y los valores alcanzados por el IBEX al día siguiente.

Día	1.º	2.º	3.º	4.º	5.º	6.º	7.º
AW&B	21,8	23,4	19,6	19,4	18,4	19,9	19,2

Día	2.º	3.º	4.º	5.º	6.º	7.º	8.º
IBEX	12 560	12 720	11 580	11 420	10 930	11 450	11 480

a) ¿Qué cotización tendrá AW&B el día anterior en que el IBEX alcance los 14 000 puntos?

b) Si un día AW&B tiene una cotización de 24 €, ¿qué valor podemos esperar que alcance el IBEX al día siguiente?

$\overline{x} = 20{,}24$ $\overline{y} = 11\,734{,}29$

$\sigma_X = \sqrt{2{,}77} = 1{,}66$ $\sigma_Y = \sqrt{366\,809{,}62} = 605{,}65$

$\sigma_{XY} = 977{,}26$

$r_{XY} = 0{,}97 \rightarrow$ La dependencia es fuerte y positiva.

a) Recta de regresión de X sobre Y:

$$x - 20{,}24 = \frac{977{,}26}{366\,809{,}62}(y - 11\,734{,}29) \rightarrow x = 0{,}0027y - 11{,}44$$

$y = 14\,000 \rightarrow x = 0{,}0027 \cdot 14\,000 - 11{,}44 = 26{,}36$

b) Recta de regresión de Y sobre X:

$$y - 11\,734{,}29 = \frac{977{,}26}{2{,}77}(x - 20{,}24) \rightarrow y = 352{,}8x - 4\,593{,}62$$

$x = 24 \rightarrow y = 352{,}8 \cdot 24 - 4\,593{,}62 = 3\,873{,}58$

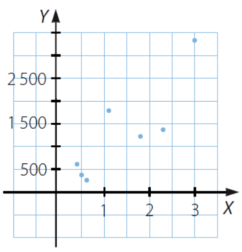
66. Se está estudiando imponer un impuesto a las empresas químicas que sea proporcional a sus emisiones de azufre a la atmósfera. Se ha experimentado con varios procedimientos para medir dichas emisiones, pero no se ha encontrado ninguno fiable. Finalmente, se ha decidido investigar algún método indirecto.

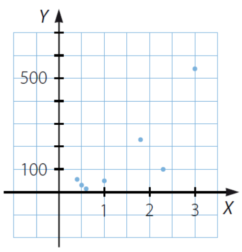
Se cree que la emisión de azufre puede estar relacionada con el consumo eléctrico, con el consumo de agua o con el volumen de las chimeneas de las fábricas. Para valorarlo se ha realizado un estudio en un medio controlado. Los resultados pueden verse en la tabla.

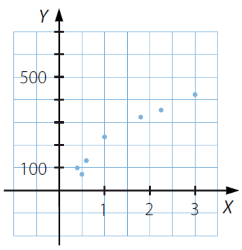
Cantidad de azufre (t)	2,3	1,8	1	0,4	0,6	3	0,5
Consumo eléctrico (kWh)	1 400	1 250	1 850	600	300	3 400	400
Consumo de agua (L)	100	230	45	50	10	540	22
Volumen de las chimeneas (m^3)	18	16	12	5	6	21	4

¿Cuál de las variables estadísticas se relaciona de forma más evidente con las emisiones de azufre? Justifícalo.

El volumen de las chimeneas es la variable que más se relaciona con la cantidad de emisiones de azufre.







67. **Se ha realizado un test de memoria, *X*, y otro test de atención, *Y*, a varios alumnos con estos resultados.**

Y \ X	[0, 10)	[10, 20)	[20, 30)	[30, 40)	[40, 50)
[0, 10)					
[10, 20)	Beatriz	Jesús	Marta		
[20, 30)		Daniel	María Esther	Miguel	
[30, 40)			Elena	Jacinto Carmen	Inés
[40, 50)				Diego	

a) Calcula la covarianza y el coeficiente de correlación.

b) Determina las dos rectas de regresión.

c) Si es factible, estima qué puntuación obtendrá un alumno que ha sacado 33 en atención.

d) Si es factible, estima qué puntuación obtendrá un alumno que ha conseguido 27 en memoria.

x_i \ y_j	5	15	25	35	45	Total
5	0	0	0	0	0	0
15	1	1	1	0	0	3
25	0	1	2	1	0	4
35	0	0	1	1	1	3
45	0	0	0	1	0	1
Total	1	2	4	3	1	11

a) $\overline{x} = 25{,}91$ $\quad \overline{y} = 26{,}82$ $\quad \sigma_{XY} = 71{,}0038$

$\sigma_X = \sqrt{117{,}31} = 10{,}83$ $\quad \sigma_Y = \sqrt{87{,}51} = 9{,}35$ $\quad r_{XY} = 0{,}7$

b) Recta de regresión de Y sobre X:

$$y - 26{,}82 = \frac{71{,}0038}{117{,}31}(x - 25{,}91) \rightarrow y = 0{,}61x + 11{,}01$$

Recta de regresión de X sobre Y:

$$x - 25{,}91 = \frac{71{,}0038}{87{,}51}(y - 26{,}82) \rightarrow x = 0{,}81y + 4{,}19$$

c) $y = 33 \rightarrow x = 0{,}81 \cdot 33 + 4{,}19 = 30{,}92$

d) $x = 27 \rightarrow y = 0{,}61 \cdot 27 + 11{,}01 = 27{,}48$

68. **Para una distribución bidimensional conocemos los siguientes datos:**

$$r = 0{,}7 \qquad \sigma_x = 1{,}2 \qquad \overline{Y} = 4$$

Recta de regresión *X* sobre *Y*: $x = 0{,}44y + 0{,}6$

Calcula los siguientes valores.

a) La media de X.

b) La recta de regresión de Y/X.

c) La varianza de X.

d) La covarianza de X e Y.

a) $\overline{x} = 0{,}44\overline{y} + 0{,}6 = 0{,}44 \cdot 4 + 0{,}6 = 2{,}36$

b) De la recta de regresión obtenemos que $0{,}7 = \dfrac{\sigma_{XY}}{\sigma_Y^2}$.

Por el coeficiente de correlación tenemos que $r_{XY} = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y} = 0{,}44$.

Por lo que al hallar σ_Y se puede construir la recta de regresión:

$$\left.\begin{array}{r} \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y} = 0{,}44 \\ 0{,}7 = \dfrac{\sigma_{XY}}{\sigma_Y^2} \end{array}\right\} \to \left.\begin{array}{l} \sigma_{XY} = 0{,}44 \cdot \sigma_X\sigma_Y \\ \sigma_{XY} = 0{,}7 \cdot \sigma_Y^2 \end{array}\right\} \to 0{,}44 \cdot \sigma_X\sigma_Y = 0{,}7 \cdot \sigma_Y^2 \to \sigma_Y = \frac{0{,}44 \cdot \sigma_X}{0{,}7} \to \sigma_Y = 0{,}75$$

Por tanto, la covarianza será $\sigma_{XY} = 0{,}25$ y la recta de regresión es $y = 0{,}17x + 3{,}59$.

c) La varianza de X es $\sigma_X^2 = 1{,}44$.

d) La covarianza de X e Y es $\sigma_{XY} = 0{,}25$.

69. Encuentra el coeficiente de correlación de la variable bidimensional cuyas rectas de regresión son:

Recta de Y sobre X: $2x - y - 1 = 0$

Recta de X sobre Y: $9x - 4y - 9 = 0$

a) Halla la media aritmética de cada una de las variables.

b) ¿Podrías calcular la desviación típica de Y sabiendo que la de la variable X es $\sqrt{2}$?

a) Las medias se calculan fácilmente al solucionar el sistema de ecuaciones del enunciado:

$$\left.\begin{array}{r} 2\overline{x} - \overline{y} - 1 = 0 \\ 9\overline{x} - 4\overline{y} - 9 = 0 \end{array}\right\} \to \left.\begin{array}{r} 8\overline{x} - 4\overline{y} - 4 = 0 \\ 9\overline{x} - 4\overline{y} - 9 = 0 \end{array}\right\} \to \left.\begin{array}{r} -\overline{x} + 5 = 0 \\ \overline{y} = 2\overline{x} - 1 \end{array}\right\} \to \left\{\begin{array}{l} \overline{x} = 5 \\ \overline{y} = 9 \end{array}\right.$$

b) De las dos rectas se obtiene:

$$\left.\begin{array}{r} \dfrac{\sigma_{XY}}{\sigma_X^2} = 2 \\ \dfrac{\sigma_{XY}}{\sigma_Y^2} = \dfrac{4}{9} \end{array}\right\} \to \sqrt{\frac{\sigma_{XY}}{\sigma_X^2}} \cdot \sqrt{\frac{\sigma_{XY}}{\sigma_Y^2}} = \sqrt{\frac{\sigma_{XY}}{\sigma_X^2} \cdot \frac{\sigma_{XY}}{\sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \sqrt{2} \cdot \frac{2}{3} = \frac{2\sqrt{2}}{3} = r_{XY}$$

Solucionando el sistema anterior, se obtiene la desviación típica de Y:

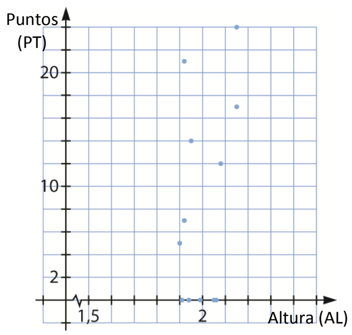
$$\left.\begin{array}{r} \dfrac{\sigma_{XY}}{4} = 2 \\ \dfrac{\sigma_{XY}}{\sigma_Y^2} = \dfrac{4}{9} \end{array}\right\} \to \left.\begin{array}{l} \sigma_{XY} = 8 \\ \dfrac{9}{4}\sigma_{XY} = \sigma_Y^2 \end{array}\right\} \to \sigma_Y^2 = 36 \to \sigma_Y = 6$$

70. **En las Olimpiadas de Londres 2012, España jugó la final en baloncesto. Las estadísticas de los jugadores de España fueron las siguientes.**

Jugador	AL	MJ	AS	RB	PT
Rudy Fernández	1,95	36	0	5	14
Pau Gasol	2,15	33	3	8	24
Juan C. Navarro	1,92	33	1	2	21
Marc Gasol	2,15	18	1	1	17
José M. Calderón	1,91	16	0	0	0
Serge Ibaka	2,08	23	0	0	12
Sergio Llull	1,90	17	2	9	5
Sergio Rodríguez	1,91	14	3	0	7
Felipe Reyes	2,06	7	0	2	0
Víctor Sada	1,93	1	0	2	0
Fernando San Emeterio	1,99	1	0	0	0
Víctor Claver	2,05	1	0	0	0

a) Representa la nube de puntos de la variable puntos marcados (PT) frente a la altura de los jugadores (AL).
b) Calcula la recta de regresión de la variable altura de los jugadores (AL) frente a los puntos marcados (PT).
c) Es razonable pensar que la altura explica los puntos marcados. Justifica la respuesta.
d) Representa la nube de puntos de la variable puntos marcados (PT) frente a los minutos jugados (MJ).
e) Calcula la recta de regresión de la variable minutos jugados (MJ) frente a los puntos marcados (PT).
f) Es razonable pensar que el número de minutos explica los puntos marcados. Justifica la respuesta.

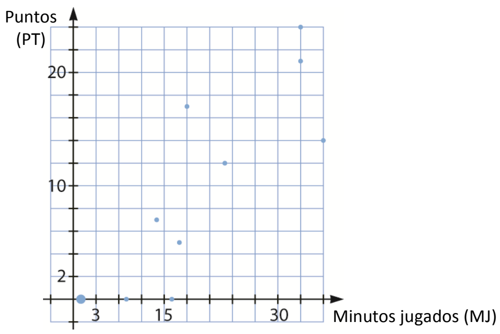
a)



b) La recta de regresión buscada será $y = 36{,}89x - 65{,}44$.

c) La recta de regresión da una aproximación lineal de los puntos *y* de un jugador de altura *x* en un partido. Tiene pendiente positiva, así que será creciente y da la idea de que a mayor altura, más puntos de meten.

d)



e) La recta de regresión buscada será $y = 0{,}61x - 1{,}81$.

f) La recta regresora devuelve una aproximación lineal de los puntos *y* de un jugador que juega *x* minutos en un partido. Es una recta con pendiente positiva, así que será siempre creciente y da la idea de que un jugador, cuanto más juegue, más puntos meterá para el equipo.

PARA PROFUNDIZAR

71. Elige la respuesta adecuada. (Concurso de Primavera)

En la primera fase de un concurso de matemáticas, la media de las puntuaciones fue de 76 sobre 100. La nota media de los estudiantes que se clasificaron para la segunda fase fue de 83 y la media de los que no se clasificaron fue de 55. ¿Qué porcentaje de los estudiantes se clasificó para la segunda fase?	44 %	66 %	68 %	72 %	75 %
En una ciudad, el cociente entre el número de mujeres y el de hombres es 11/10. Si la media de las edades de las mujeres es 34 años y la media de las edades de los hombres es 32 años, la media de las edades de toda la población, en años, es:	$\frac{329}{10}$	$\frac{692}{21}$	33	$\frac{694}{21}$	$\frac{331}{10}$
En un centro se hizo la misma prueba del Concurso de Primavera a un grupo de alumnos muy buenos de ESO y a todos los de Bachillerato. La media global fue de 84 puntos. Los de ESO, que eran solamente el 10%, obtuvieron todos la misma puntuación y la media de los de Bachillerato fue 83 puntos. ¿Cuál fue la puntuación de cada estudiante de ESO?	85	88	93	94	98

- Sean *C* y *D* el número total de estudiantes clasificados y desclasificados, respectivamente. Consideramos las siguientes definiciones:

$$\overline{C} = \frac{\sum C_i}{C} = 83 \qquad \overline{D} = \frac{\sum D_i}{D} = 55$$

Queremos calcular el porcentaje de clasificados, es decir, el coeficiente, *A*, por el que hay que multiplicar el número total de estudiantes para conseguir los que se han clasificado:

$$C = A(C+D) \rightarrow C - AC = AD \rightarrow D = \frac{1-A}{A}C$$

Sea $\overline{x}$ la media de las puntuaciones de todos los estudiantes. Entonces:

$$\overline{x} = \frac{\sum C_i + \sum D_i}{C+D} = \frac{\sum C_i}{C+D} + \frac{\sum D_i}{C+D} = \frac{\sum C_i}{C + \frac{1-A}{A}C} + \frac{\sum D_i}{\frac{A}{1-A}D + D} = A \cdot \frac{\sum C_i}{C} + (1-A) \cdot \frac{\sum D_i}{D} = A\overline{C} + (1-A)\overline{D} \rightarrow$$

$$76 = 83A + (1-A)55 \rightarrow 28A = 21 \rightarrow A = \frac{21}{28} = 0{,}75 = 75\%$$

- Sean *m* y *h* el número total de mujeres y hombres, respectivamente. Consideramos las siguientes definiciones:

$$\overline{m} = \frac{\sum m_i}{m} = 34 \qquad \overline{h} = \frac{\sum h_i}{h} = 32$$

Además tenemos que $\frac{11}{10} = \frac{m}{h}$.

Sea $\overline{x}$ la media de edad de la población total. Entonces:

$$\overline{x} = \frac{\sum m_i + \sum h_i}{m+h} = \frac{\sum m_i}{m+h} + \frac{\sum h_i}{m+h} = \frac{\sum m_i}{m + \frac{10}{11}m} + \frac{\sum h_i}{\frac{11}{10}h + h} = \frac{11 \cdot \sum m_i}{21m} + \frac{10 \cdot \sum h_i}{21h} = \frac{11}{21} \cdot \overline{m} + \frac{10}{21} \cdot \overline{h} =$$

$$= \frac{11 \cdot 34 + 10 \cdot 32}{21} = \frac{694}{21}$$

□ Sean B y E el número de alumnos presentados al examen de Bachillerato y ESO, respectivamente, y P la puntuación que han conseguido todos los alumnos de la ESO. Consideramos las siguientes definiciones:

$$\overline{B}=\frac{\sum B_i}{B}=83 \qquad \overline{E}=\frac{\sum E_i}{E}=\frac{E\cdot P}{E}=P$$

Además tenemos que 10% de $(B+E)=E\rightarrow\frac{10}{100}\cdot(B+E)=E\rightarrow B=9E$.

$$\overline{X}=\frac{\sum B_i+\sum E_i}{B+E}=\frac{\sum B_i}{B+E}+\frac{\sum E_i}{B+E}=\frac{\sum B_i}{B+\frac{1}{9}B}+\frac{\sum E_i}{9E+E}=\frac{9\cdot\sum B_i}{10B}+\frac{\sum h_i}{10E}=\frac{9\cdot\overline{B}+\overline{E}}{10}\rightarrow$$

$\rightarrow 84=\frac{9\cdot 83+P}{10}\rightarrow 84\cdot 10-9\cdot 83=P\rightarrow P=93$ es la puntuación de los alumnos de ESO.

72. Halla la relación existente entre el coeficiente de correlación lineal de una distribución bidimensional y las pendientes de sus rectas de regresión.

Las pendientes de las rectas de regresión son:

$$\begin{cases} m_X=\dfrac{\sigma_{XY}}{\sigma_X{}^2}\rightarrow\sigma_X=\sqrt{\dfrac{\sigma_{XY}}{m_x}} \\ m_Y=\dfrac{\sigma_{XY}}{\sigma_Y{}^2}\rightarrow\sigma_Y=\sqrt{\dfrac{\sigma_{XY}}{m_Y}} \end{cases}$$

Entonces, resulta que: $r_{XY}=\dfrac{\sigma_{XY}}{\sigma_X\cdot\sigma_Y}=\dfrac{\sigma_{XY}}{\sqrt{\dfrac{\sigma_{XY}}{m_X}}\cdot\sqrt{\dfrac{\sigma_{XY}}{m_Y}}}=m_X\cdot m_Y$

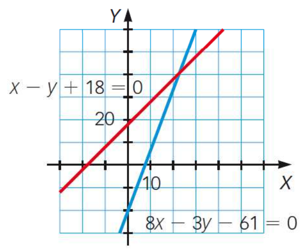
73. Discute si es posible que la recta de regresión de *X* sobre *Y* y la recta de regresión de *Y* sobre *X* sean paralelas. ¿Y perpendiculares?

No es posible que sean paralelas, ya que tienen siempre un punto común: $(\overline{x},\overline{y})$

Son perpendiculares si la correlación es nula.

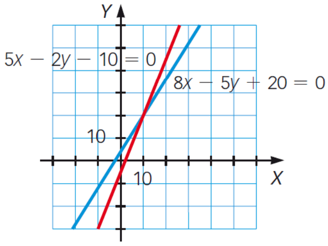
74. En dos estudios estadísticos realizados sobre los datos de una variable bidimensional, las rectas de regresión fueron las siguientes.

En el primer estudio, la recta de regresión de *Y* sobre *X* es: $8x-3y-61=0$ y la de *X* sobre *Y* es: $x-y+18=0$.



Y en el otro estudio, las rectas de regresión son, respectivamente:

$$8x - 5y + 20 = 0 \qquad 5x - 2y - 10 = 0$$



Si conocemos $\overline{x} = 23$, $\overline{y} = 41$ y $r = 0{,}8$, comprueba cuál de los dos estudios es válido.

$$\left.\begin{array}{r} 8x - 3y - 61 = 0 \\ x - y + 18 = 0 \end{array}\right\} \rightarrow x = 23, y = 41 \qquad \left.\begin{array}{r} 8x - 5y + 20 = 0 \\ 5x - 2y - 10 = 0 \end{array}\right\} \rightarrow x = 10,\ y = 20$$

El primer estudio es el correcto, ya que las rectas se cortan en el punto $(\overline{x}, \overline{y})$.

75. Sean dos variables estadísticas, *X* e *Y*. Sabemos que:

- **La recta de regresión de *Y* sobre *X* pasa por los puntos (1, 3) y (2, 5).**
- **La recta de regresión de *X* sobre *Y* tiene pendiente 3 y su ordenada en el origen es 2.**
- **La varianza de *Y* es 3.**

Calcula las medidas estadísticas de cada una de las variables y el coeficiente de correlación.

La recta que pasa por los puntos (1, 3) y (2, 5) tiene como ecuación: $y = 2x + 1$

La ecuación de la otra recta es: $y = 3x + 2$

$$\left.\begin{array}{r} 2x - y + 1 = 0 \\ 3x - y + 2 = 0 \end{array}\right\} \rightarrow x = -1,\ y = -1$$

Entonces, resulta que: $\begin{cases} \overline{x} = -1 \\ \overline{y} = -1 \end{cases}$

El coeficiente de correlación es igual a la raíz cuadrada del producto de la pendiente de la recta de regresión de *Y* sobre *X* por la inversa de la pendiente de la recta de regresión de *X* sobre *Y*:

$$r = \sqrt{m \cdot \frac{1}{m'}} = \sqrt{\frac{\sigma_{XY}}{\sigma_X^2} \cdot \frac{\sigma_{XY}}{\sigma_Y^2}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

Por tanto, tenemos que: $r = \sqrt{2 \cdot \frac{1}{3}} = 0{,}8164$

El sistema de ecuaciones formado por las dos rectas de regresión es:

$$\left.\begin{array}{r} \dfrac{\sigma_{XY}}{\sigma_X^2} = 2 \\ \dfrac{\sigma_Y^2}{\sigma_{XY}} = 3 \end{array}\right\} \rightarrow \frac{\sigma_Y^2}{\sigma_X^2} = 2 \cdot 3 \rightarrow \sigma_Y^2 = 6\sigma_X^2 \rightarrow \sigma_Y = \sqrt{6}\,\sigma_X$$

Como la varianza de *Y* es 3: $\sigma_X^2 = \dfrac{1}{2}$

76. Investiga sobre cómo varía el coeficiente de correlación entre dos variables estadísticas cuando multiplicamos los datos relativos a una de ellas por una cantidad constante, *k*.

¿Y si multiplicamos las dos por la misma constante?

¿Qué sucedería si multiplicamos cada variable por una constante distinta?

Al multiplicar los datos de una variable por una cantidad constante k, sus medidas estadísticas verifican que:

$$\frac{\sum_{i=1}^{n} f_i \cdot kx_i}{N} = \frac{k \cdot \sum_{i=1}^{n} f_i \cdot x_i}{N} = k \cdot \overline{x}$$

$$\frac{\sum_{i=1}^{n} f_i \cdot (kx_i - k\overline{x})^2}{N} = \frac{\sum_{i=1}^{n} f_i \cdot k^2(x_i - \overline{x})^2}{N} = \frac{k^2 \cdot \sum_{i=1}^{n} f_i \cdot (x_i - \overline{x})^2}{N} = k^2 \cdot \sigma_X^2$$

$$\sqrt{k^2 \cdot \sigma_X^2} = k \cdot \sigma_X$$

Entonces la covarianza entre las dos variables es:

$$\frac{\sum_{i=1}^{n} f_i \cdot kx_i \cdot y_i}{N} - k\overline{x} \cdot \overline{y} = \frac{k \cdot \sum_{i=1}^{n} f_i \cdot x_i \cdot y_i}{N} - k\overline{x} \cdot \overline{y} = k\left(\frac{\sum_{i=1}^{n} f_i \cdot x_i \cdot y_i}{N} - \overline{x} \cdot \overline{y}\right) = k \cdot \sigma_{XY}$$

Así, el coeficiente de correlación es:

$$\frac{k \cdot \sigma_{XY}}{k \cdot \sigma_X \cdot \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = r_{XY}$$

Si se multiplican los datos de las dos variables por la misma constante k, entonces el coeficiente de correlación es:

$$\frac{k^2 \cdot \sigma_{XY}}{k \cdot \sigma_X \cdot k \cdot \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = r_{XY}$$

Y si multiplicamos la segunda variable por un constante m:

$$\frac{k \cdot m \cdot \sigma_{XY}}{k \cdot \sigma_X \cdot m \cdot \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = r_{XY}$$

MATEMÁTICAS EN TU VIDA

1. Explica qué entiendes por estacionalidad.

La estacionalidad, en este caso, es la variación de los datos en función de la época del año, que se se repite cíclicamente año tras año.

2. ¿En cuál de los gráficos estadísticos que aparecen en el texto se ve la dependencia de dos variables? ¿Cómo se ve?

En el segundo gráfico. Mediante un diagrama de dispersión que relaciona la edad de los clientes y el gasto en euros que realizan.

3. ¿Qué tipo de gráficos estadísticos aparecen en el texto?

El primero y el tercero son polígonos de frecuencia y el segundo es un diagrama de dispersión.

4. Pon un ejemplo de estacionalidad.

Respuesta abierta. Por ejemplo, el caudal del río Duero a su paso por Zamora.

5. **Unos días al año, en marzo o en abril, la ocupación hotelera se dispara aunque no siempre es en los mismos días. Razona por qué crees que pasa esto.**

Porque en estos meses tienen lugar las vacaciones de Semana Santa, cuya fecha varía debido a que se celebra la semana anterior al primer domingo posterior a la primera luna llena tras el equinocio de marzo.

6. **Pon un ejemplo donde se dé un fenómeno de estacionalidad.**

Respuesta abierta. Por ejemplo, la migración de las cigüeñas.