

001 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.

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

$Y \rightarrow$ Sexo de los asistentes

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

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

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

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

005 Construye la tabla de doble entrada y las tablas marginales correspondientes.

X	16	17	18	16	14	17	14	13	14	15
Y	5	4	6	6	8	3	5	4	8	8

	X	13	14	15	16	17	18	Total
Y	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

006 Determina la covarianza para los datos que aparecen en la siguiente tabla.

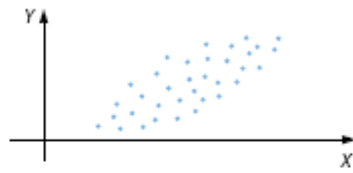
X	8	10	11	9	13	12	9	14
Y	20	18	16	22	10	10	21	9

$$\bar{x} = \frac{86}{8} = 10,75$$

$$\bar{y} = \frac{126}{8} = 15,75$$

$$\sigma_{xy} = \frac{1,279}{8} - 10,75 \cdot 15,75 = -9,44$$

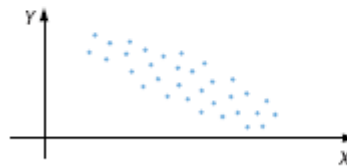
008 Indica la dependencia entre estas variables.



Dependencia lineal débil y positiva.

009 Describe el grado de correlación entre las dos variables representadas.

La correlación lineal es débil y negativa.



010 Si el signo de la covarianza entre dos variables es negativa, ¿qué podemos decir del signo del coeficiente de correlación?

¿Y si la covarianza es positiva?

Si la covarianza es negativa, el coeficiente de correlación es negativo.

Y si la covarianza es positiva, el coeficiente de correlación es también positivo.

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

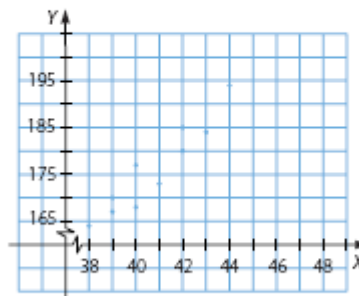
X	39	43	40	40	42	41	42	38	39	44
Y	167	184	177	168	185	173	180	164	170	194

¿Qué relación puedes describir entre ellos?

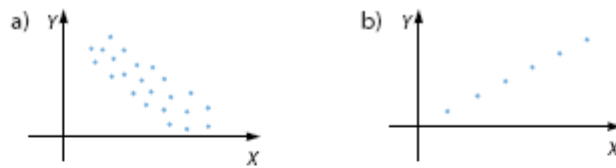
$$\bar{x} = \frac{408}{10} = 40,8 \quad \bar{y} = \frac{1.762}{10} = 176,2$$

$$\sigma_x = \sqrt{3,36} = 1,83 \quad \sigma_y = \sqrt{81,96} = 9,05$$

$$\sigma_{xy} = \frac{72.046}{10} - 40,8 \cdot 176,25 = 13,6 \quad r_{xy} = \frac{13,6}{1,83 \cdot 9,05} = 0,82$$



012 Razona qué valor tomará el coeficiente de correlación.



- a) El coeficiente de correlación tomará un valor relativamente cercano a -1 , porque la nube de puntos se aproxima bastante a una recta con pendiente negativa y la correlación es fuerte.
- b) El coeficiente de correlación es 1 , ya que la nube de puntos coincide con una recta de pendiente positiva.

013 Halla la recta de regresión de Y sobre X .

X	2	5	6	8	9
Y	4	13	16	22	25

$$\bar{x} = \frac{30}{5} = 6$$

$$\bar{y} = \frac{80}{5} = 16$$

$$\sigma_x^2 = \frac{30}{5} = 6$$

$$\sigma_{xy} = \frac{570}{5} - 6 \cdot 16 = 18$$

$$\text{Recta de regresión de } Y \text{ sobre } X: y - 16 = \frac{18}{6}(x - 6) \rightarrow y = 3x - 2$$

014 Determina la recta de regresión correspondiente.

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

$$\bar{x} = \frac{407}{10} = 40,7$$

$$\bar{y} = \frac{1.745}{10} = 174,5$$

$$\sigma_x^2 = \frac{16.599}{10} - 40,7^2 = 3,41$$

$$\sigma_{xy} = \frac{71.145}{10} - 40,7 \cdot 174,5 = 12,35$$

$$\text{Recta de regresión de } Y \text{ sobre } X: y - 174,5 = \frac{12,35}{3,41}(x - 40,7) \rightarrow y = 3,62x + 27,17$$

017 En un estudio sobre los ingresos mensuales, X , y la superficie de las viviendas, Y , resulta: $y = 0,02x + 47,96$.

a) Halla la estimación de la superficie de la vivienda de una familia cuyos ingresos mensuales son de 3.200 €.

b) Si una familia vive en una casa de 90 m², ¿cuáles serán sus ingresos mensuales?

a) $y = 0,02 \cdot 3.200 + 47,96 = 111,96 \text{ m}^2$

b) $0,02x + 47,96 = 90 \rightarrow x = 2.102 \text{ €}$

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

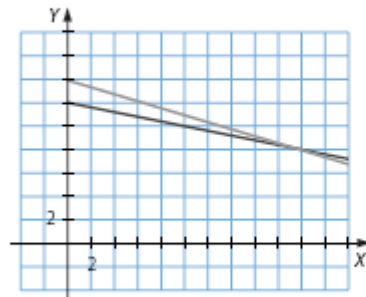
a) Determina las dos rectas de regresión, represéntalas y analiza la correlación que existe entre las variables.

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

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

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

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

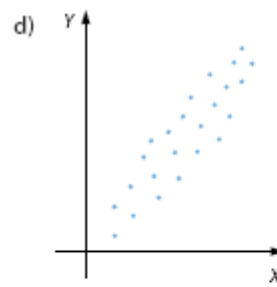
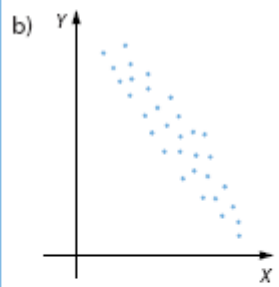
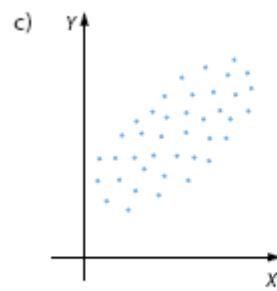
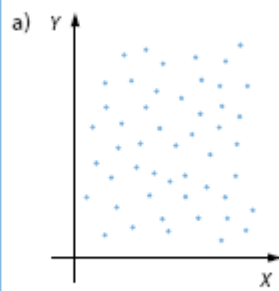


La dependencia es fuerte y negativa.

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

023
•○○

A partir de los diagramas de dispersión, decide si hay o no dependencia lineal 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 fuerte y negativa.
- c) La dependencia lineal es débil y positiva.
- d) La dependencia lineal es fuerte y positiva.

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

Bodegones \ Paisajes	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

- Determina las tablas de frecuencias marginales de paisajes y bodegones.
- Calcula las medias y las desviaciones típicas de cada una de las variables.
- Usa el coeficiente de variación para decidir cuál de las dos variables es más dispersa.
- Realiza el diagrama de dispersión correspondiente a la variable bidimensional.



- a) Tabla de frecuencias marginales de los paisajes

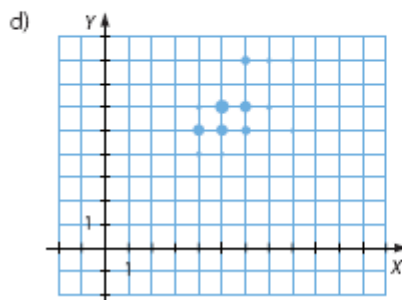
x_i	f_i
4	8
5	10
6	10
7	4
8	2
Total	34

- Tabla de frecuencias marginales de los bodegones

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

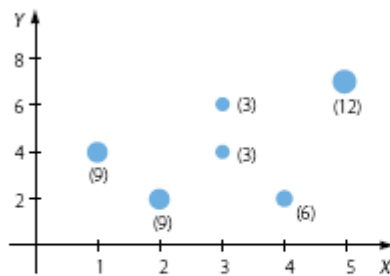
- b) $\bar{x} = 5,47$ $\bar{y} = 5,82$
 $\sigma_x = 1,15$ $\sigma_y = 1,19$
- c) $CV_x = 0,21$ $CV_y = 0,204$

La variable de los paisajes es un poco más dispersa que la de los bodegones.



028
•••

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



Y \ X	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

029
•••

Calcula la covarianza y el coeficiente de correlación para las variables bidimensionales indicadas en las siguientes 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

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

030
•••

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

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

049
•••

Daniel afirma que si una nube de puntos es de una recta, el coeficiente de correlación siempre vale 1 o -1 . Como Eva no está de acuerdo, Daniel prueba con los puntos de la recta cuya ecuación es $y = -5x + 20$, y Eva hace lo mismo con los puntos de $y = 2x - x^2$.

- ¿Quién tiene razón? ¿Por qué?
- Si la hipótesis de Daniel no resulta cierta, ¿podrías formularla de forma que se verifique siempre?

- Si $y = -5x + 20$, entonces algunos de los puntos son:

X	-2	-1	0	1	2
Y	30	25	20	15	10

$$\bar{x} = 0 \quad \bar{y} = 20 \quad \sigma_x = 1,41 \quad \sigma_y = 7,07 \quad \sigma_{xy} = -10$$

$$r_{xy} = -1 \rightarrow \text{La dependencia es lineal.}$$

Si $y = 2x - x^2$, no es una recta, y algunos de los puntos son:

X	-2	-1	0	1	2
Y	-8	-3	0	1	0

$$\bar{x} = 0 \quad \bar{y} = -2 \quad \sigma_x = 1,41 \quad \sigma_y = 3,29 \quad \sigma_{xy} = 4$$

$$r_{xy} = 0,86 \rightarrow \text{La dependencia es débil; por tanto, Eva no tiene razón.}$$

- Es cierta.

El alcalde de un pueblo ha constatado una reducción del número de nacimientos de niños, y ha encargado realizar un estudio.

Año	86	89	92	95	98	01	04	07
Nacimientos	50	54	40	33	34	23	21	17

- a) ¿Puede establecerse, de forma fiable, una fórmula que relacione el año con el número de nacimientos?
- b) ¿Cuántos nacimientos pueden estimarse en 2008? ¿Y en 2010? ¿Qué puede estimarse para 2050?
- c) ¿Es fiable esta última estimación? Razona la respuesta.

a)

X	0	3	6	9	12	15	18	21
Y	50	54	40	33	34	23	21	17

$$\bar{x} = 10,5 \quad \bar{y} = 34 \quad \sigma_x = 6,87 \quad \sigma_y = 12,61 \quad \sigma_{xy} = -83,63$$

$r_{xy} = -0,97 \rightarrow$ La dependencia es fuerte y negativa, por lo que puede utilizarse la recta de regresión para relacionar las dos variables.

- b) Recta de regresión de Y sobre X:

$$y - 34 = -\frac{83,63}{47,25}(x - 10,5) \rightarrow y = -1,77x + 52,59$$

En el año 2008 se estiman: $x = 22 \rightarrow y = -1,77 \cdot 22 + 52,59 = 13,65$ nacimientos

En el año 2010 se estiman: $x = 64 \rightarrow y = -1,77 \cdot 64 + 52,59 = -60,69$ nacimientos

Para el año 2050 se estiman -60 nacimientos.

- c) No es fiable, ya que el año 2050 está muy alejado del rango de años estudiados en la regresión.