12 Estadística bidimensional

LITERATURA Y MATEMÁTICAS

La caverna

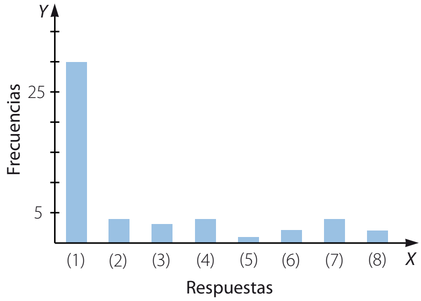
Buenas tardes, señor Algor, Buenas tardes, señor, Supongo que imagina por qué motivo le estoy telefoneando hoy, Supone bien, señor, dígame, Tengo ante mí los resultados y las conclusiones del sondeo acerca de sus artículos, [...] Y esos resultados cuáles son, señor, preguntó Cipriano Algor, Lamento informarle de que no fueron tan buenos cuanto desearíamos, Si es así nadie lo lamentará más que yo, Temo que su participación en la vida de nuestro Centro ha llegado al final, [...] Vaya tomando nota de los resultados, Dígamelos, El universo de los clientes sobre el que incidiría el sondeo quedó definido desde el principio por la exclusión de las personas que por edad, posición social, educación y cultura, y también por sus hábitos conocidos de consumo, fuesen previsible y radicalmente contrarias a la adquisición de artículos de este tipo, es bueno que sepa que si tomamos esta decisión, señor Algor, fue para no perjudicarlo de entrada, Muchas gracias, señor, Le doy un ejemplo, si hubiéramos seleccionado cincuenta jóvenes modernos, cincuenta chicos y chicas de nuestro tiempo, puede tener la certeza, señor Algor, de que ninguno querría llevarse a casa uno de sus muñecos, o si se lo llevase sería para usado en algo así como tiro al blanco, Comprendo, Escogimos veinticinco personas de cada sexo, de profesiones e ingresos medios, personas con antecedentes familiares modestos, todavía apegadas a gustos tradicionales, y en cuyas casas la rusticidad del producto no desentonaría demasiado, E incluso así, Es verdad, señor Algor, incluso así los resultados fueron malos, Qué le vamos a hacer, señor, Veinte hombres y diez mujeres respondieron que no les gustaban los muñecos de barro, cuatro mujeres dijeron que quizá los compraran si fueran más grandes, tres podrían comprarlos si fuesen más pequeños, de los cinco hombres que quedaban, cuatro dijeron que ya no estaban en edad de jugar y otro protestó por el hecho de que tres de las figurillas representasen extranjeros, para colmo exóticos, y en cuanto a las ocho mujeres que todavía faltan por mencionar, dos se declararon alérgicas al barro, cuatro tenían malos recuerdos de esta clase de objetos, y sólo las dos últimas respondieron agradeciendo mucho la posibilidad que les había sido proporcionada de decorar gratuitamente su casa con unos muñequitos tan simpáticos, hay que añadir que se trata de personas de edad que viven solas, Me gustaría conocer los nombres y las direcciones de esas señoras para darles las gracias, dijo Cipriano Algor, Lo lamento, pero no estoy autorizado a revelar datos personales de los encuestados, es una condición estricta de cualquier sondeo de este tipo, respetar el anonimato de las respuestas. [...] Buenas tardes, Buenas tardes.

JOSÉ SARAMAGO

Resume los datos en una tabla de frecuencias y represéntalos en un gráfico estadístico. ¿Puedes calcular alguna medida de centralización?

Se ha elegido una muestra de 50 personas con características personales y profesionales similares.

Respuestas	f_i
No les gustan los muñecos (1)	30
Los comprarían si fueran más grandes (2)	4
Los comprarían si fueran más pequeños (3)	3
No están en edad de jugar (4)	4
Protestan por figurillas de extranjeros (5)	1
Alérgicas al barro (6)	2
Malos recuerdos (7)	4
Agradecidas por ser gratuitas (8)	2



No es posible calcular ninguna medida de centralización porque la variable no es cuantitativa.

ANTES DE COMENZAR… RECUERDA

001 En una revista leemos que el pastor alemán tiene una alzada media de 55 cm. ¿Crees que han medido a todos los pastores alemanes del planeta? Explica cómo crees que han llegado a esta conclusión.

No los han medido. Se elige una muestra representativa de la población de pastores alemanes y se estudia el valor de la media en dicha muestra.

002 Indica el tipo de variable estadística que estamos estudiando.

a) El programa favorito de los miembros de tu familia.
b) El número de calzado de los alumnos de un IES.
c) La temperatura media diaria de tu provincia.
d) La edad de los habitantes de un país.
e) El sexo de los habitantes de un pueblo.
f) El dinero gastado a la semana por tus amigos.
g) Los efectos de un nuevo medicamento en el ser humano.
h) El color del pelo de tus compañeros de clase.

a) Cualitativa
b) Cuantitativa discreta
c) Cuantitativa continua
d) Cuantitativa discreta
e) Cualitativa
f) Cuantitativa discreta
g) Cualitativa
h) Cualitativa

003 El número de horas diarias de estudio de 30 alumnos es:

3 4 3 5 5 1 1 1 1 2 3 4 5 0 2 0 3 2 2 1 2 1 3 2 0 1 2 1 4 3

a) Organiza los resultados en una tabla de frecuencias.
b) ¿Qué significan las frecuencias acumuladas?

a)

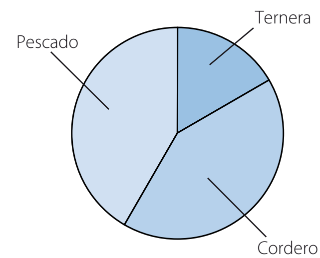
Horas	f_i	h_i	F_i	H_i
0	3	0,1	3	0,1
1	8	0,27	11	0,37
2	7	0,23	18	0,6
3	6	0,2	24	0,8
4	3	0,1	27	0,9
5	3	0,1	30	1
	$N = 30$	$\sum h_i = 1$		

b) Las frecuencias acumuladas indican el número de alumnos que estudian como máximo el número de horas correspondiente. Por ejemplo, la frecuencia acumulada para el valor 2 es 18, es decir, hay 18 alumnos que estudian 0, 1 o 2 horas.

004 **De los 30 asistentes a una cena, el 20 % comió ternera, el 40 % cordero y el resto tomó pescado. Indica la variable estadística y organiza los resultados en una tabla de frecuencias; después, representa los datos en un diagrama de sectores.**

La variable estadística es el plato elegido en la cena.

Plato	f_i	h_i	F_i	H_i
Ternera	6	0,2	6	0,2
Cordero	12	0,4	18	0,6
Pescado	12	0,4	30	1
	$N = 30$	$\sum h_i = 1$		



ACTIVIDADES

001 **Pon dos ejemplos de variables estadísticas unidimensionales.**

Respuesta abierta. Por ejemplo: la calificación de los alumnos de una clase en un examen y la estatura de los miembros de un equipo de baloncesto.

002 **Organiza estos datos en una tabla de frecuencias absolutas y relativas.**

0	5	1	1	4	4
0	6	2	2	3	6
1	4	7	1	7	5
3	6	8	2	8	6

x_i	f_i	h_i	F_i	H_i
0	2	0,08	2	0,08
1	4	0,17	6	0,25
2	3	0,13	9	0,38
3	2	0,08	11	0,46
4	3	0,13	14	0,59
5	2	0,08	16	0,67
6	4	0,17	20	0,84
7	2	0,08	22	0,92
8	2	0,08	24	1
	$N = 24$	$\sum h_i = 1$		

003 **La tabla muestra la estatura, en centímetros, de un grupo de personas.**

Estatura (cm)	[165, 175)	[175, 185)	[185, 195)
N.º de personas	40	85	25

a) Elabora una tabla de frecuencias.

b) ¿Qué porcentaje de personas miden entre 165 cm y 175 cm? ¿Y menos de 185 cm?

a)

Estatura	x_i	f_i	h_i	F_i	H_i
[165, 175)	170	40	0,27	40	0,27
[175, 185)	180	85	0,57	125	0,83
[185, 195)	190	25	0,17	150	1
		$N = 150$	$\sum h_i = 1$		

b) El porcentaje de personas que miden entre 165 cm y 175 cm es del 27 %.
Y el porcentaje de personas que miden menos de 185 cm es del 83 %.

004 **A partir de los datos, construye la tabla de frecuencias, y calcula las medidas de centralización.**

23	10	25	12	13	24	17	22
16	20	26	23	22	13	21	18
16	19	14	17	11	17	15	26

x_i	f_i	h_i	F_i	H_i
10	1	0,04	1	0,04
11	1	0,04	2	0,08
12	1	0,04	3	0,13
13	2	0,08	5	0,21
14	1	0,04	6	0,25
15	1	0,04	7	0,29
16	2	0,08	9	0,38
17	3	0,13	12	0,5
18	1	0,04	13	0,54
19	1	0,04	14	0,58
20	1	0,04	15	0,63
21	1	0,04	16	0,67
22	2	0,08	18	0,75
23	2	0,08	20	0,83
24	1	0,04	21	0,88
25	1	0,04	22	0,92
26	2	0,08	24	1
	$N = 24$	$\sum h_i = 1$		

$\overline{x} = \dfrac{440}{24} = 18{,}33 \rightarrow$ El valor medio es 18,33.

$Mo = 17 \rightarrow$ El valor más frecuente es 17.

$Me = \dfrac{17 + 18}{2} = 17{,}5$

Hay tantos valores menores que 17,5 como mayores.

005 **Obtén e interpreta las medidas de centralización correspondientes a los datos de esta tabla.**

Peso (kg)	[50, 65)	[65, 80)	[80, 95)
N.º de personas	75	140	80

Peso	x_i	f_i	h_i
[50, 65)	57,5	75	75
[65, 80)	72,5	140	215
[80, 95)	87,5	80	295
		$N = 295$	

$$\overline{x} = \frac{21.462,5}{295} = 72,75$$

El peso medio es de 72,75 kg.

El intervalo modal es [65, 80); su marca de clase: 72,5 es la moda.

Lo más frecuente es que el peso esté comprendido entre 65 kg y 80 kg.

El intervalo mediano es [65, 80); su marca de clase: 72,5 es la mediana.

Hay tantas personas que pesan menos de 72,5 kg como personas que pesan más.

006 **Calcula las medidas de dispersión para estos datos.**

Clases	[5, 15)	[15, 25)	[25, 35)	[35, 45)
Frecuencias	35	15	25	45

Clases	x_i	f_i	$\vert x_i - \overline{x}\vert$	$(x_i - \overline{x})^2$
[5, 15)	10	35	16,67	277,89
[15, 25)	20	15	6,67	44,49
[25, 35)	30	25	3,33	11,09
[35, 45)	40	45	13,33	177,69
		$N = 120$		

$$\overline{x} = \frac{3.200}{120} = 26,67$$

Rango: $R = 45 - 5 = 40$

Desviación media: $DM = \frac{1.366,6}{120} = 11,39$

Varianza: $\sigma^2 = \frac{18.666,8}{120} = 155,56$

Desviación típica: $\sigma = 12,47$

Coeficiente de variación: $CV = \frac{12,47}{26.67} = 0,47$

007 **Compara las edades, en años, de los jugadores de estos equipos de baloncesto, utilizando las medidas estadísticas.**

A:	18	26	20	26	22	26	23	27	25	25
B:	20	21	20	21	22	23	23	24	25	25

$$\overline{x_A} = \frac{238}{10} = 23{,}8 \qquad \sigma_A = \frac{79{,}6}{10} = 7{,}96 \qquad \sigma_A = 2{,}82 \qquad CV_A = \frac{2{,}82}{23{,}8} = 0{,}12$$

$$\overline{x_B} = \frac{224}{10} = 22{,}4 \qquad \sigma_B^2 = \frac{32{,}4}{10} = 3{,}24 \qquad \sigma_B = 1{,}8 \qquad CV_B = \frac{1{,}8}{22{,}4} = 0{,}08$$

La media de las edades del equipo *A* es superior, pero también es mayor el coeficiente de variación de este equipo, por lo que hay más diferencias entre sus jugadores.

008 **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) $X \longrightarrow$ 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)

009 **Ordena estos datos en una tabla de doble entrada.**

X	*Y*
0	18
0	12
2	8

X	*Y*
1	14
2	23
1	17

Y \ *X*	0	1	2	Total
8	0	0	1	1
12	1	0	0	1
14	0	1	0	1
17	0	1	0	1
18	1	0	0	1
23	0	0	1	1
Total	2	2	2	6

010 **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

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

011 **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

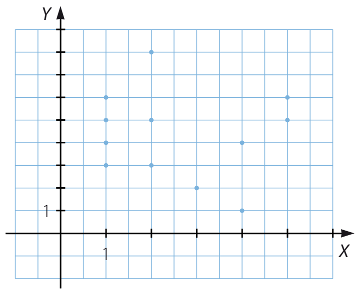
$$\overline{x} = \frac{86}{8} = 10{,}75$$

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

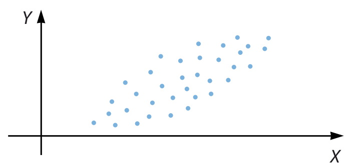
$$\sigma_{XY} = \frac{1.279}{8} - 10{,}75 \cdot 15{,}75 = -9{,}44$$

012 **Representa la nube de puntos correspondiente a la siguiente variable estadística bidimensional.**

X	1	1	3	5	2	4	5	2	5	2	4	3	2	1	1
Y	4	5	2	5	5	4	5	3	6	5	1	2	8	6	3

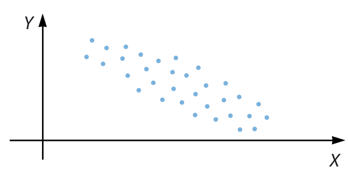


013 **Indica la dependencia entre estas variables.**



Dependencia lineal débil y positiva.

014 **Describe el grado de correlación entre las dos variables representadas.**



La correlación lineal es débil y negativa.

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

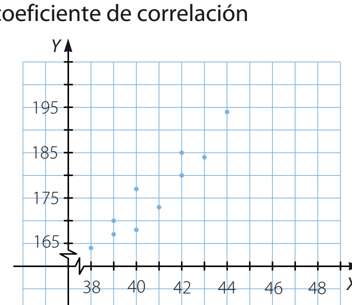
¿Y si la covarianza es positiva?

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

016 **Represente 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?

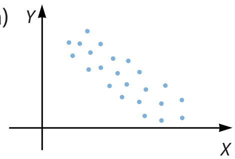


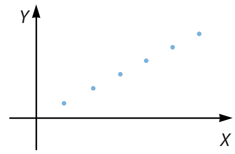
$\overline{x} = \frac{408}{10} = 40,8 \qquad \overline{y} = \frac{1.762}{10} = 176,2$

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

$\sigma_{XY} = \frac{72.046}{10} - 40,8 \cdot 176,25 = 13,6 \qquad r_{XY} = \frac{13,6}{1.83 \cdot 9.05} = 0,82$

017 **Razona qué valor tomará el coeficiente de correlación.**

a) 

b) 

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.

018 **Halla la recta de regresión de *Y* sobre *X*.**

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

$\overline{x} = \frac{30}{5} = 6$ $\overline{y} = \frac{80}{5} = 16$

$\sigma_X^2 = \frac{30}{5} = 6$ $\sigma_{XY} = \frac{570}{5} - 6 \cdot 16 = 18$

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

019 **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

$\overline{x} = \frac{407}{10} = 40{,}7$ $\overline{y} = \frac{1.745}{10} = 174{,}5$

$\sigma_X^2 = \frac{16.599}{10} - 40{,}7^2 = 3{,}41$ $\sigma_{XY} = \frac{71.145}{10} - 40{,}7 \cdot 174{,}5 = 12{,}35$

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

020 **Determina las dos rectas de regresión, e indica la relación que hay entre las 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) $\overline{x} = \dfrac{60}{5} = 12$ $\qquad$ $\overline{y} = \dfrac{21}{5} = 4,2$

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

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

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

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

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

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

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

$\sigma_X^2 = \dfrac{1.625}{10} - 12,5^2 = 6,25$ $\qquad$ $\sigma_{XY} = \dfrac{1.890}{10} - 12,5 \cdot 14,5 = 7,75$

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

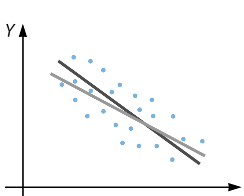
$\sigma_X^2 = \dfrac{2.325}{10} - 14,5^2 = 22,25$

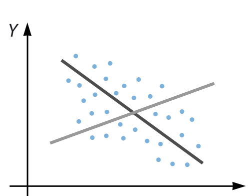
Recta de regresión de X sobre Y: $x - 12,5 = \dfrac{7,75}{22,25}(y - 14,5) \rightarrow x = 0,35y + 7,43$

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

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

021 **Razona cuál es el grado de dependencia entre las variables en cada caso.**

a) 

b) 

a) La dependencia es fuerte y negativa.

b) La dependencia es débil y negativa.

022 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 m^2, ¿cuáles serán sus ingresos mensuales?

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

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

023 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 $\sigma_Y = 1$.

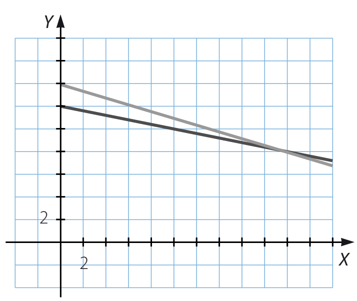
a) Determina las dos rectas de regresión, represéntalas y analiza la correlación que existe entre las variables.

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

024 Utiliza la calculadora para determinar todas las medidas estadísticas.

a)

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

b)

X	24	27	22	23	24	26	27	28	22	23
Y	2	1	2	4	5	2	3	4	1	2

a) $\overline{x} = 2{,}93$ $\overline{y} = 6{,}73$
$\sigma_X^2 = 1{,}82$ $\sigma_Y^2 = 1{,}97$
$\sigma_X = 1{,}35$ $\sigma_Y = 1{,}4$
$\sigma_{XY} = 0{,}35$
$r_{XY} = 0{,}19$

b) $\overline{x} = 24{,}6$ $\overline{y} = 2{,}6$
$\sigma_X^2 = 4{,}44$ $\sigma_Y^2 = 1{,}64$
$\sigma_X = 2{,}11$ $\sigma_Y = 1{,}28$
$\sigma_{XY} = 0{,}44$
$r_{XY} = 0{,}16$

025 **Estudia la correlación entre estas variables, utilizando la calculadora para realizar las operaciones.**

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

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

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

$\sigma_X^2 = 2{,}24$ $\qquad \sigma_Y^2 = 4{,}49$

$\sigma_X = 2{,}11$ $\qquad \sigma_Y = 2{,}12$

$\sigma_{XY} = 0{,}14$

$r_{XY} = 0{,}03$

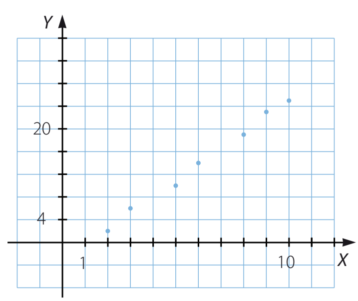
Recta de regresión de Y sobre X: $y - 33{,}9 = \dfrac{0{,}14}{2{,}24}(x - 14{,}4) \rightarrow y = 0{,}06x + 33$

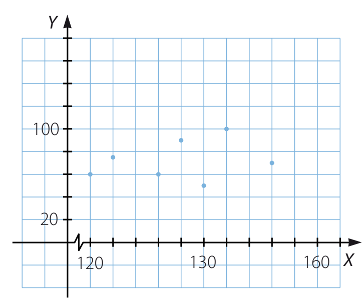
Como la correlación es casi nula, no tiene sentido estimar el valor de y para $x = 18$.

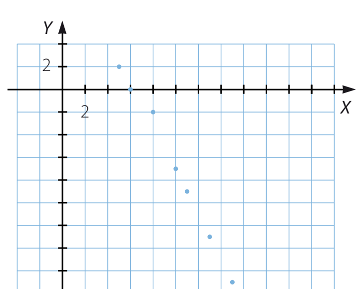
026 **Representa la nube de puntos asociada a las siguientes distribuciones bidimensiones.**

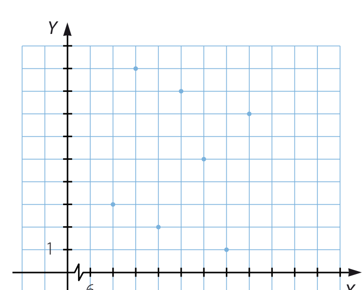
a) (2, 2) (3, 6) (5, 10) (6, 14) (8, 19) (9, 23) (10, 25)
b) (5, 2) (6, 0) (8, −2) (10, −7) (11, −9) (13, −13) (15, −17)
c) (120, 60) (122, 75) (126, 60) (128, 90) (130, 50) (132, 100) (136, 70)
d) (7, 3) (8, 9) (9, 2) (10, 8) (11, 5) (12, 1) (13, 7)

Decide si existe dependencia entre las variables y de qué tipo es.

a)


c)


b)


d)


027 ●○○ 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

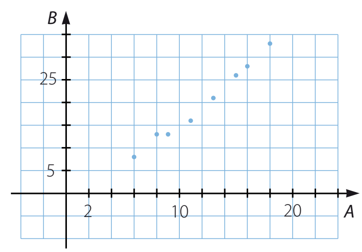
c)

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

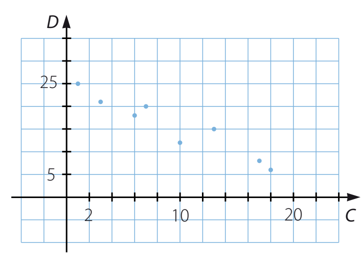
d)

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

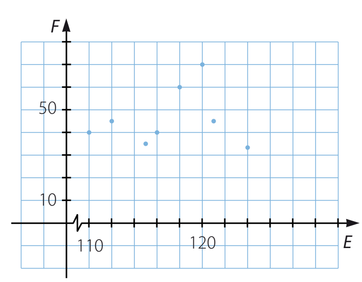
a) La dependencia es fuerte y positiva.



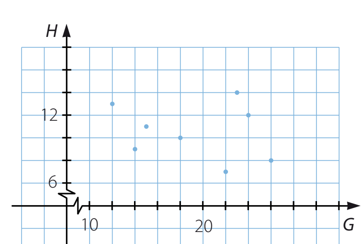
b) La dependencia es fuerte y negativa.



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



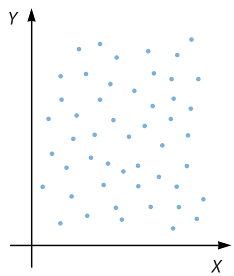
d) No se aprecia dependencia entre las variables *G* y *H*.



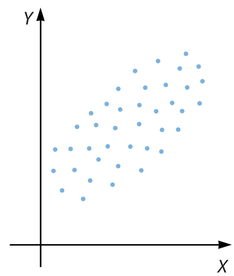
028 ●○○

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.

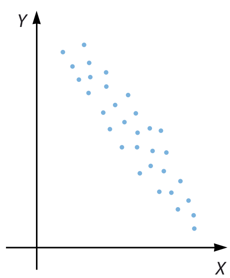
a)



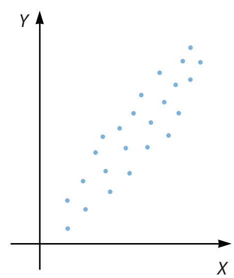
c)



b)



d)



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.

029 ●○○

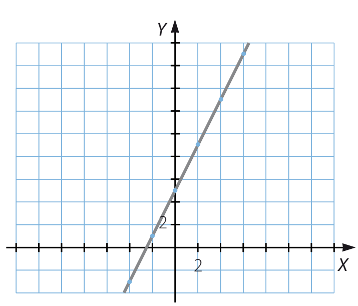
Representa las nubes de puntos correspondientes a las variables bidimensionales definidas por estas fórmulas.

a) $y = 2x + 5$

b) $y = x^2 + 3x$

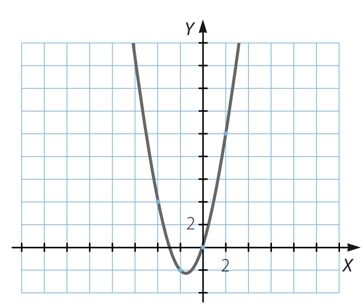
¿Qué tipo de dependencia presentan?

a)



La dependencia es lineal.

b)



La dependencia es funcional.

030 ●○○ 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) Usa el coeficiente de variación para decidir cuál de las dos variables es más dispersa.
d) 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

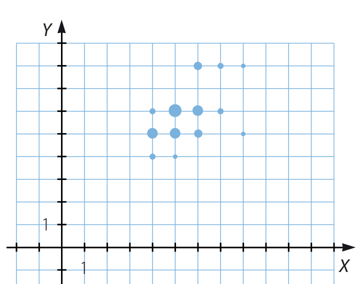
b) $\overline{x} = 5,47$ $\qquad \overline{y} = 5,82$

$\sigma_X = 1,15$ $\qquad \sigma_Y = 1,19$

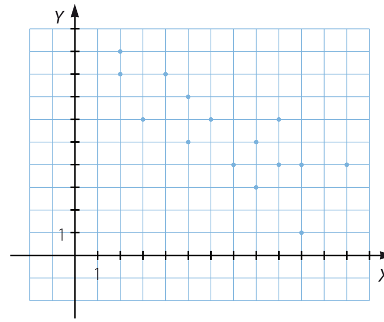
c) $CV_X = 0,21$ $\qquad CV_Y = 0,204$

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

d)

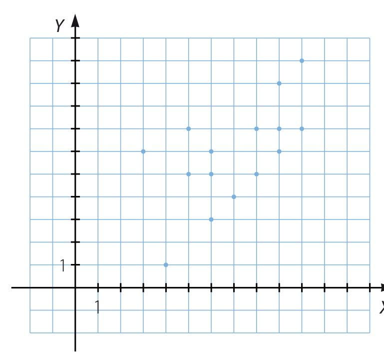


031 Construye la tabla de doble entrada que corresponde a esta variable bidimensional, representada mediante el diagrama de dispersión.



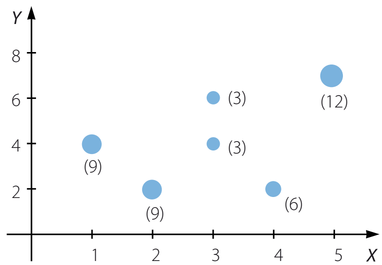
Y \ X	2	3	4	5	6	7	8	9	10	12	Total
1	0	0	0	0	0	0	0	0	1	0	1
3	0	0	0	0	0	0	1	0	0	0	1
4	0	0	0	0	0	1	0	1	1	1	4
5	0	0	0	1	0	0	1	0	0	0	2
6	0	1	0	0	1	0	0	1	0	0	3
7	0	0	0	1	0	0	0	0	0	0	1
8	1	0	1	0	0	0	0	0	0	0	2
9	1	0	0	0	0	0	0	0	0	0	1
Total	2	1	1	2	1	1	2	2	2	1	15

032 A partir de este diagrama de dispersión, construye la tabla de doble entrada correspondiente.



Y \ X	3	4	5	6	7	8	9	10	Total
1	0	1	0	0	0	0	0	0	1
3	0	0	0	1	0	0	0	0	1
4	0	0	0	0	1	0	0	0	1
5	0	0	1	1	0	1	0	0	3
6	1	0	0	1	0	0	1	0	3
7	0	0	1	0	0	1	1	1	4
9	0	0	0	0	0	0	1	0	1
10	0	0	0	0	0	0	0	1	1
Total	1	1	2	3	1	2	3	2	15

033 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**

034 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

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

035 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

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

036 Representa la variable bidimensional cuyos pares de valores son:

(8, 2) (12, 6) (10, 4) (12, 2) (8, 6)

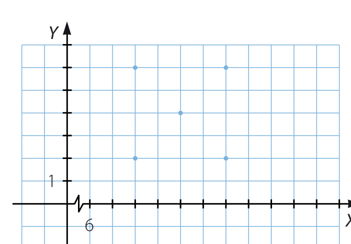
a) Calcula su covarianza y razona el resultado.

b) Elimina un punto de manera que se mantenga la correlación.

a) $\sigma_{XY} = 0$

No hay dependencia entre las variables, por lo que la covarianza es nula.

b) Al eliminar el punto (10, 4), la correlación no varía.



037 Construye el diagrama de dispersión correspondiente a la variable bidimensional determinada por los siguientes pares de datos.

(10, 20) (16, 30) (10, 30) (16, 20)

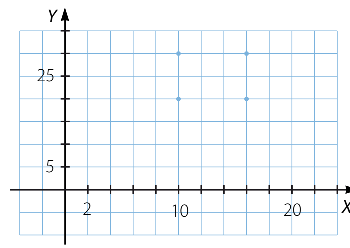
a) Calcula su covarianza y explica a qué se debe el resultado.

b) Añade un punto de manera que se mantenga la correlación.

a) $\sigma_{XY} = 0$

No hay dependencia entre las variables, por lo que la covarianza es nula.

b) Al añadir el punto (13, 15), la correlación no varía.



038 En la tabla se presentan datos climatológicos referidos a una ciudad: la temperatura, en °C; la humedad relativa del aire, en %, y la velocidad del viento, en km/h.

Días	L	M	X	J	V	S	D
Temperatura	22	24	25	24	23	21	20
Humedad	78	90	80	92	88	74	80
Velocidad del viento	1	3	6	4	4	1	0

Determina la covarianza y el coeficiente de correlación de las siguientes variables bidimensionales.

a) *Temperatura–Humedad.*

b) *Temperatura–Velocidad del viento.*

c) *Humedad–Velocidad del viento.*

a) $\sigma_{TH} = 6{,}46$ $\quad r_{TH} = 0{,}59$

b) $\sigma_{TV} = 3{,}17$ $\quad r_{TV} = 0{,}93$

c) $\sigma_{HV} = 6{,}404$ $\quad r_{HV} = 0{,}507$

039 Se ha hecho una encuesta a personas que han tenido un accidente de tráfico, preguntando por el número de meses transcurridos e incluyendo el grupo de edad.

Las respuestas han sido:

Carmen, 35: [60, 70)
Teresa, 15: [50, 60)
Pilar, 12: [50, 60)
Esther, 6: [20, 30)
Juan, 8: [40, 50)
Jacinto, 15: [30, 40)
Jesús, 24: [50, 60)
Marta, 12: [30, 40)
José, 28: [40, 50)
Andrés, 3: [20, 30)
María Jesús, 20: [40, 50)
Beatriz, 16: [30, 40)

a) Construye la tabla correspondiente a la variable bidimensional.

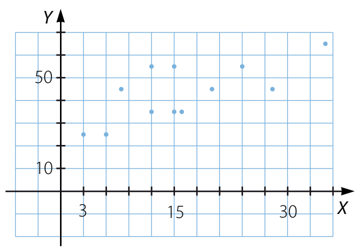
b) Representa el diagrama de dispersión.

c) Estudia si hay correlación entre ambas variables, y determina su coeficiente de correlación lineal.

a)

Y \ X	3	6	8	12	15	16	20	24	28	35	Total
[20, 30)	1	1	0	0	0	0	0	0	0	0	2
[30, 40)	0	0	0	1	1	1	0	0	0	0	3
[40, 50)	0	0	1	0	0	0	1	0	1	0	3
[50, 60)	0	0	0	1	1	0	0	1	0	0	3
[60, 70)	0	0	0	0	0	0	0	0	0	1	1
Total	1	1	1	2	2	1	1	1	1	1	12

b) La correlación es débil y positiva.



c) $\sigma_{XY} = 89,5$
$r_{XY} = 0,73$

040 En la siguiente tabla se han perdido dos datos.

x_1	23	24	25	27	28	29	33	34	36
2	4	3	5	y_5	6	7	9	6	8

Se sabe que la media de la primera variable es 28 y la media de la segunda variable es 5,8. Completa la tabla y determina el coeficiente de correlación.

$$\overline{x} = 28 \rightarrow \frac{x_1 + 259}{10} = 28 \rightarrow x_1 = 21 \qquad \overline{y} = 5,8 \rightarrow \frac{y_5 + 50}{10} = 5,8 \rightarrow y_5 = 8$$

$r_{XY} = 0,802$

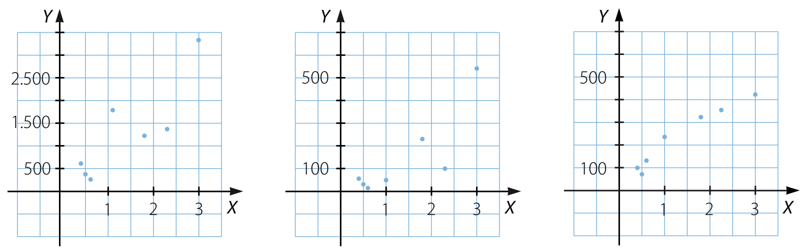
041 Se está estudiando imponer un impuesto a las empresas químicas que sea proporcional a sus emisiones de azufre a la atmósfera.
Se ha experimentado con varios procedimientos para medir dichas emisiones, pero no se ha encontrado ninguno fiable. Finalmente, se ha decidido investigar algún método indirecto.

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.

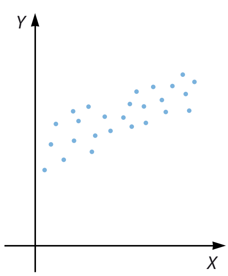
Cantidad de azufre (t)	2,3	1,8	1	0,4	0,6	3	0,5
Consumo eléctrico (kWh)	1.400	1.250	1.850	600	300	3.400	400
Consumo de agua (ℓ)	100	230	45	50	10	540	22
Volumen de las chimeneas (m^3)	18	16	12	5	6	21	4

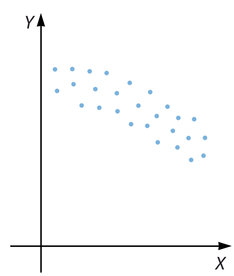
¿Cuál de las medidas estadísticas se relaciona de forma más evidente con las emisiones de azufre? Justifica la respuesta.

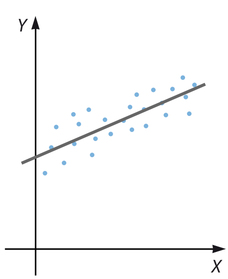
El volumen de las chimeneas es la variable que más se relaciona con la cantidad de emisiones de azufre.

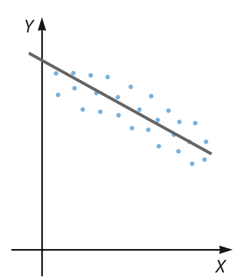


042 Traza a mano alzada, y sin realizar cálculos, la recta de regresión de las siguientes variables bidimensionales.

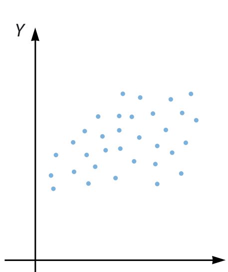
a) 

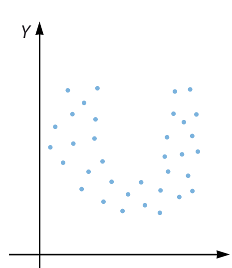
b) 

a) 

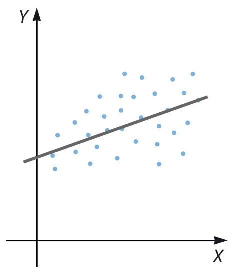
b) 

043 Representa, sin hallar su ecuación, la recta de regresión correspondiente a estas variables.

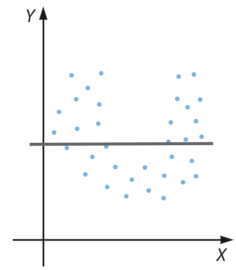
a) 

b) 

a)



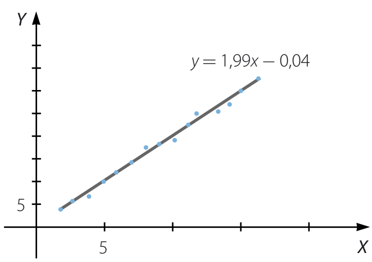
b)



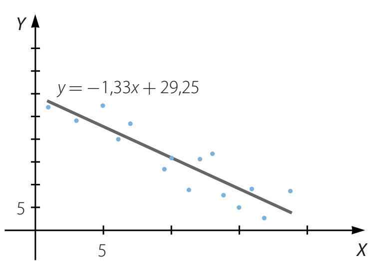
044

Para las variables bidimensionales representadas a continuación, hemos ajustado diferentes rectas de regresión a las nubes de puntos correspondientes. Estima el valor que tendrá y en cada una de ellas para un valor de $x = 12$.

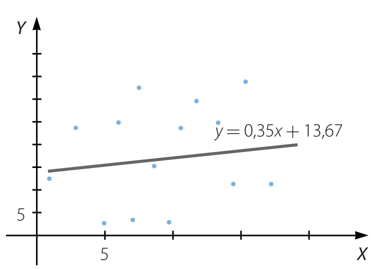
a)



b)



c)



¿Cuál de las estimaciones te parece más fiable?

a) $y = 1,99 \cdot 12 - 0,04 = 23,84$

b) $y = -1,83 \cdot 12 + 29,25 = 7,29$

c) $y = 0,35 \cdot 12 + 13,67 = 17,87$

La estimación más fiable es la del apartado a).

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

a)

X	10	11	12	13	14	15	16	17
Y	20	24	28	30	36	32	42	40

b)

X	60	70	80	90	100	110	120
Y	−5	−8	−12	−15	−16	−24	−20

c)

X	−3	−4	−5	−6	−9	−10	−13
Y	80	92	100	88	76	70	60

d)

X	0,2	0,4	0,5	0,7	0,8	0,9	1	1,2
Y	40	50	120	70	40	40	60	50

a) $\overline{x} = 13,5$ $\quad \overline{y} = 31,5$ $\quad \sigma_{XY} = 15,5$
$\sigma_X^2 = 5,25$ $\quad \sigma_Y^2 = 50,75$

Recta de regresión de Y sobre X:

$$y - 31,5 = \frac{15,5}{5,25}(x - 13,5) \rightarrow y = 2,95x - 8,33$$

Recta de regresión de X sobre Y:

$$x - 13,5 = \frac{15,5}{50,75}(y - 31,5) \rightarrow x = 0,31y + 3,74$$

b) $\overline{x} = 90$ $\quad \overline{y} = -14,29$ $\quad \sigma_{XY} = -115,33$
$\sigma_X^2 = 400$ $\quad \sigma_Y^2 = 37,22$

Recta de regresión de Y sobre X:

$$y + 14,29 = -\frac{115,33}{400}(x - 90) \rightarrow y = -0,29x + 11,81$$

Recta de regresión de X sobre Y:

$$x - 90 = -\frac{115,33}{37,22}(y + 14,29) \rightarrow x = -3,099y + 45,72$$

c) $\overline{x} = -7,14$ $\quad \overline{y} = 80,86$ $\quad \sigma_{XY} = 34,48$
$\sigma_X^2 = 11,31$ $\quad \sigma_Y^2 = 159,37$

Recta de regresión de Y sobre X:

$$y - 80,86 = \frac{34,48}{11,31}(x + 7,14) \rightarrow y = 3,049x + 102,63$$

Recta de regresión de X sobre Y:

$$x + 7,14 = \frac{34,48}{159,37}(y - 80,86) \rightarrow x = 0,22y - 24,93$$

d) $\overline{x} = 0,71$ $\quad \overline{y} = 58,75$ $\quad \sigma_{XY} = -1,088$
$\sigma_X^2 = 0,099$ $\quad \sigma_Y^2 = 635,94$

Recta de regresión de Y sobre X:

$$y - 58,75 = -\frac{1,088}{0,099}(x - 0,71) \rightarrow y = -10,99x + 66,55$$

Recta de regresión de X sobre Y:

$$x - 0,71 = -\frac{1,088}{635,94}(y - 58,75) \rightarrow x = -0,0017y + 0,81$$

046 Encuentra cinco puntos que pertenecen a 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) $\overline{x} = 0$ $\quad \overline{y} = 6$ $\quad \sigma_X = \sqrt{2} = 1{,}41$ $\quad \sigma_Y = \sqrt{32} = 5{,}66$ $\quad \sigma_{XY} = 8$

$r_{XY} = 1 \rightarrow$ 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:

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

047 Obtén cinco puntos que pertenecen a la recta.

$$y = -20x + 10$$

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

b) Halla las dos rectas de regresión. Razona los resultados obtenidos.

Respuesta abierta.

x	-2	-1	0	1	2
y	50	30	10	-10	-30

a) $\overline{x} = 0$ $\quad \overline{y} = 10$ $\quad \sigma_X = \sqrt{2} = 1{,}41$ $\quad \sigma_Y = \sqrt{800} = 28{,}28$ $\quad \sigma_{XY} = -40$

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

b) Recta de regresión de Y sobre X:

$$y - 10 = -\frac{40}{2}(x - 0) \rightarrow y = -20x + 10$$

Recta de regresión de X sobre Y:

$$x - 0 = -\frac{40}{800}(y - 10) \rightarrow x = -\frac{1}{20}y + \frac{1}{2}$$

048 Se cree que el número de zorros en una finca está relacionado con el número de conejos.

En los últimos años se han realizado ocho censos de ambos animales, resultando estos datos.

N.º de zorros	20	32	16	18	25	30	14	15
N.º de conejos	320	500	260	300	400	470	210	240

Si la correlación es fuerte:

a) Determina las dos rectas de regresión.
b) Estima la cantidad de conejos que habría si hubiera 10 zorros.
c) ¿Cuántos zorros serían si hubiéramos contado 350 conejos?
d) ¿Cuál de las dos estimaciones es más fiable?

a) $\overline{x} = 21{,}25$ $\overline{y} = 337{,}5$ $\sigma_X = \sqrt{42{,}19} = 6{,}5$ $\sigma_Y = \sqrt{10.168{,}75} = 100{,}84$

$\sigma_{XY} = 653{,}13$ $r_{XY} = 0{,}99 \rightarrow$ La dependencia es fuerte y positiva.

Recta de regresión de Y sobre X:

$$y - 337{,}5 = \frac{653{,}13}{42{,}19}(x - 21{,}25) \rightarrow y = 15{,}48x + 8{,}55$$

Recta de regresión de X sobre Y:

$$x - 21{,}25 = \frac{653{,}13}{10.168{,}75}(y - 337{,}5) \rightarrow x = 0{,}064y - 0{,}35$$

b) $x = 10 \rightarrow y = 15{,}48 \cdot 10 + 8{,}55 = 163{,}35$
En este caso habría 163 conejos.

c) $y = 350 \rightarrow x = 0{,}064 \cdot 350 - 0{,}35 = 22{,}05$
En este caso serían 22 zorros.

d) Como el coeficiente de correlación es muy próximo a 1, las dos estimaciones son bastante fiables.

049 A lo largo de un día se han medido la tensión y el pulso cardíaco de una persona, tratando de decidir si ambas variables tienen alguna relación.

Los datos obtenidos se han reflejado en la tabla.

Nivel mínimo de tensión	6	5	9	4	10	8	6	9
N.º de pulsaciones por minuto	60	55	80	40	95	75	55	90

a) Calcula la covarianza, el coeficiente de correlación y las dos rectas de regresión.
b) Si la correlación es fuerte, estima las pulsaciones que tendrá la persona cuando su nivel mínimo de tensión sea 15.
c) ¿Qué nivel mínimo de tensión se estima cuando las pulsaciones cardíacas por minuto son 70?
d) ¿Cuál de las dos estimaciones es más fiable?
e) Dibuja la nube de puntos y la recta de regresión correspondientes.

a) $\overline{x} = 7{,}13$ $\quad \overline{y} = 68{,}75$ $\quad \sigma_X = \sqrt{4{,}04} = 2{,}01$ $\quad \sigma_Y = \sqrt{323{,}44} = 17{,}98$

$\sigma_{XY} = 35{,}44$ $\quad r_{XY} = 0{,}98 \rightarrow$ La dependencia es fuerte y positiva.

Recta de regresión de Y sobre X:

$$y - 68{,}75 = \frac{35{,}44}{4{,}04}(x - 7{,}13) \rightarrow y = 8{,}77x + 6{,}22$$

Recta de regresión de X sobre Y:

$$x - 7{,}13 = \frac{35{,}44}{323{,}44}(y - 68{,}75) \rightarrow x = 0{,}11y - 0{,}43$$

b) $x = 15 \rightarrow y = 8{,}77 \cdot 15 + 6{,}22 = 137{,}77$

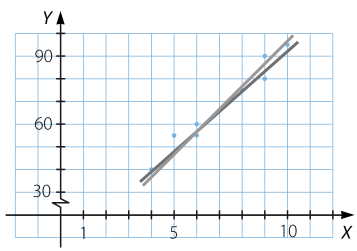
En este caso tendría 138 pulsaciones por minuto.

c) $y = 70 \rightarrow x = 0{,}11 \cdot 70 - 0{,}43 = 7{,}27$

Se estima que tendría un nivel mínimo de 7.

d) Las dos estimaciones son muy fiables, porque el coeficiente de correlación es bastante cercano a 1.

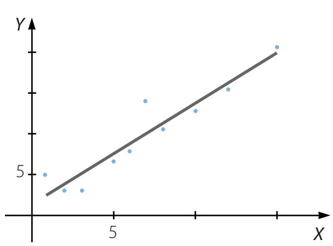
e)



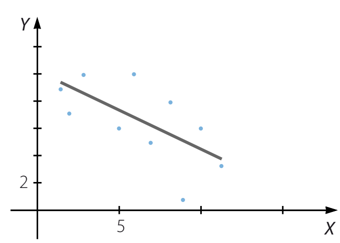
050

Tenemos dos variables bidimensionales representadas por estas nubes de puntos.

(I)



(II)



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

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

b) Ahora decide cuáles son las ecuaciones de las dos rectas de regresión correspondientes.

$y = 3x + 0{,}2$ $\quad y = 1{,}3x + 0{,}9$ $\quad y = -0{,}6x + 10$ $\quad y = -2x + 12{,}6$

Justifica la respuesta.

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.

051 Una empresa está investigando la relación entre sus gastos en publicidad y sus beneficios (en millones de euros).

Este es un resumen del estudio.

Año	98	99	00	01	02	03	04	05	06	07
Gastos	2	2,4	2	2,8	3	3,2	3,2	3,3	3,5	4
Beneficios	12	15	13	15	18	19	19	20	20	22

a) Comprueba si existe relación entre las magnitudes y, si es posible, estima los beneficios que se obtendrán en el año 2008, si se van a invertir 4,2 millones de euros en publicidad.

b) ¿Qué inversión sería necesaria para alcanzar 30 millones de euros de beneficios?

a) $\overline{x} = 2,94$ $\overline{y} = 17,3$ $\sigma_X = \sqrt{0,38} = 0,61$ $\sigma_Y = \sqrt{10,01} = 3,16$ $\sigma_{XY} = 1,89$

$r_{XY} = 0,98 \rightarrow$ La dependencia es fuerte y positiva.

Recta de regresión de Y sobre X:

$$y - 17,3 = \frac{1,89}{0,38}(x - 2,94) \rightarrow y = 4,97x + 2,69$$

$x = 4,2 \rightarrow y = 4,97 \cdot 4,2 + 2,69 = 23,56$

Los beneficios serían de 23,56 millones de euros.

b) Recta de regresión de X sobre Y:

$$x - 2,94 = \frac{1,89}{10,01}(y - 17,3) \rightarrow x = 0,19y - 0,35$$

$y = 30 \rightarrow y = 0,19 \cdot 30 - 0,35 = 5,35$

La inversión tendría que ser de 5,35 millones de euros.

052 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.

053 Se ha medido el peso, X, y la estatura, Y, de los alumnos de una clase. Su peso medio ha sido de 56 kg, con una desviación típica de 2,5 kg.

La ecuación de la recta de regresión que relaciona la estatura y el peso es: $y = 1{,}8x + 62$

a) ¿Qué estatura puede estimarse en un alumno que pesa 64 kg?
b) Y si un alumno pesara 44 kg, ¿cuál sería su altura?
c) ¿Cuál es la estatura media de los alumnos de esa clase?
d) La pendiente de esa recta es positiva. ¿Qué significa esto?

a) $x = 64 \rightarrow y = 1{,}8 \cdot 64 + 62 = 177{,}2$
El alumno medirá 1,77 m.

b) $x = 44 \rightarrow y = 1{,}8 \cdot 44 + 62 = 141{,}2$
En este caso medirá 1,41 m.

c) $y = 1{,}8 \cdot 56 + 62 = 162{,}8$
La estatura media es 1,63 m.

d) Si la pendiente es positiva, entonces la correlación entre las variables también es positiva, es decir, cuando los valores de una variable aumentan, los valores de la otra variable también lo hacen.

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

X	−2	−1	0	1	2
Y	30	25	20	15	10

$\overline{x} = 0$ $\overline{y} = 20$ $\sigma_X = 1{,}41$ $\sigma_Y = 7{,}07$ $\sigma_{XY} = -10$

$r_{XY} = -1 \rightarrow$ 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

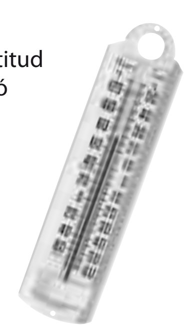
$\overline{x} = 0$ $\overline{y} = -2$ $\sigma_X = 1{,}41$ $\sigma_Y = 3{,}29$ $\sigma_{XY} = 4$

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

055 Un equipo de alpinistas que escaló una montaña, midió la altitud y la temperatura cada 200 metros de ascensión. Luego reflejó los datos en estas tablas.

Altitud (m)	800	1.000	1.200	1.400	1.600	1.800	2.000
Temperatura (°C)	22	20	17	15	11	9	8

Altitud (m)	2.200	2.400	2.600	2.800	3.000	3.200
Temperatura (°C)	5	3	2	2	2	1



a) Toma las diez primeras mediciones y, si la correlación es fuerte, calcula la recta de regresión de la temperatura sobre la altitud.

b) Estima la temperatura que habrá a los 1.900 metros de altitud.

c) ¿Qué temperatura se estima a los 3.200 metros? ¿Cómo explicas las diferencias?

a) $\overline{x} = 1.700$ $\overline{y} = 11,2$ $\sigma_X = 574,46$ $\sigma_Y = 6,69$ $\sigma_{XY} = -3.820$

$r_{XY} = -0,99 \rightarrow$ La dependencia es fuerte y negativa.

Recta de regresión de Y sobre X:

$$y - 11,2 = -\frac{3.820}{330.000}(x - 1.700) \rightarrow y = -0,012x + 31,6$$

b) $x = 1.900 \rightarrow y = -0,012 \cdot 1.900 + 31,6 = 8,8$

La temperatura estimada es de 8,8 °C.

c) $x = 3.200 \rightarrow y = -0,012 \cdot 3.200 + 31,6 = -6,8$

La diferencia se debe a que el valor no está incluido en el intervalo [800, 2.600], formado por los datos que se han utilizado para calcular la recta de regresión.

056 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

$\overline{x} = 10,5$ $\overline{y} = 34$ $\sigma_X = 6,87$ $\sigma_Y = 12,61$ $\sigma_{XY} = -83,63$

$r_{XY} = -0,97 \rightarrow$ 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.

057 ●○○ En una empresa se está estudiando el número de días de baja por enfermedad, Y, de cada uno de sus empleados en el último año. Para compararlo con la antigüedad, X, de los empleados dentro de la empresa, se ha elaborado la siguiente tabla.

Y \ X	1	2	3	4	5
0	6	12	8	3	0
2	4	5	3	2	1
3	0	1	3	2	0
5	0	0	2	2	1
9	0	0	0	0	1

a) Calcula las medias y las desviaciones típicas de las distribuciones marginales.
b) Determina la covarianza y el coeficiente de correlación.
c) Halla la recta de regresión de Y sobre X y estima, si es fiable, el número de días de baja que puede esperarse en un empleado con 6 años de antigüedad en la empresa.

a) Tabla de frecuencias marginales de los años de antigüedad

x_i	f_i
1	10
2	18
3	16
4	9
5	3
Total	56

Tabla de frecuencias marginales de los días de baja

y_i	f_i
0	29
2	15
3	6
5	5
9	1
Total	56

$\overline{x} = 2,59$ $\qquad$ $\overline{y} = 1,46$

$\sigma_X = 1,11$ $\qquad$ $\sigma_Y = 1,89$

b) $\sigma_{XY} = 1,02$ $\qquad$ $r_{XY} = 0,49$

c) Recta de regresión de Y sobre X: $y - 1,46 = \dfrac{1,02}{1,23}(x - 2,59) \rightarrow y = 0,83x - 0,69$

La dependencia es débil, por lo que la estimación no es fiable.

058 ●●○ 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 sucede 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 al día en que el IBEX alcance los 14.000 puntos?

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

$\overline{x} = 20{,}24 \qquad \overline{y} = 11.734{,}29$

$\sigma_X = \sqrt{2{,}77} = 1{,}66 \qquad \sigma_Y = \sqrt{366.809{,}62} = 605{,}65$

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

$r_{XY} = 0{,}97 \rightarrow$ 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:

$$y - 11.734{,}29 = \frac{977{,}26}{2{,}77}(x - 20{,}24) \rightarrow y = 352{,}8x - 4.593{,}62$$

$x = 24 \rightarrow y = 352{,}8 \cdot 24 - 4.593{,}62 = 3.873{,}58$

059 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 $\sqrt{2}$?

$$2x - y - 1 = 0 \rightarrow y = 2x - 1 \rightarrow \frac{\sigma_{XY}}{\sigma_X^2} = 2 \rightarrow \sigma_X = \sqrt{\frac{\sigma_{XY}}{2}}$$

$$9x - 4y - 9 = 0 \rightarrow x = \frac{4}{9}y + 1 \rightarrow \frac{\sigma_{XY}}{\sigma_Y^2} = \frac{4}{9} \rightarrow \sigma_Y = \frac{3\sqrt{\sigma_{XY}}}{2}$$

$$r_{XY} = \frac{\sigma_{XY}}{\sqrt{\frac{\sigma_{XY}}{2} \cdot \frac{3\sqrt{\sigma_{XY}}}{2}}} = \frac{2\sqrt{2}}{3} = 0{,}94$$

a) Las rectas de regresión se cortan en el punto $(\overline{x}, \overline{y})$.

$$\left.\begin{array}{l} 2x - y - 1 = 0 \\ 9x - 4y - 9 = 0 \end{array}\right\} \rightarrow x = 5,\ y = 9$$

Entonces, resulta que: $\overline{x} = 5, \overline{y} = 9$

b) $\sigma_X = \sqrt{2} \rightarrow \dfrac{\sigma_{XY}}{2} = 2 \rightarrow \sigma_{XY} = 4 \rightarrow \sigma_Y = \dfrac{3 \cdot 2}{2} = 3$

060 Se tiene la siguiente variable bidimensional.
Investiga lo que sucede con la covarianza y el coeficiente de correlación en cada caso.

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

a) Sumamos 10 a todos los valores de la variable X.
b) Sumamos 10 a todos los valores de la variable X y de la variable Y.
c) Multiplicamos por 4 todos los valores de la variable X.
d) Multiplicamos por 4 todos los valores de la variable X y de la variable 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

$\overline{y} = 5,29 \cdot 4 = 21,16$ $\sigma_{XY} = 6,42 \cdot 16 = 102,72$
$\sigma_Y = 2,24 \cdot 4 = 8,96$ $r_{XY} = 0,76$

061 Investiga sobre las siguientes cuestiones.
a) ¿Es cierto que el signo de las pendientes de las dos rectas de regresión de una variable bidimensional es siempre igual?
b) ¿Qué sucede si las dos rectas de regresión tienen la misma pendiente? ¿Cómo es la correlación?

a) Es cierto, porque el signo de las pendientes de las rectas de regresión coincide con el signo de la covarianza en ambas.
b) Como las dos rectas pasan por el punto $(\overline{x}, \overline{y})$, si tienen la misma pendiente, entonces son coincidentes. Por tanto, la dependencia entre las dos variables unidimensionales es lineal.
La correlación es igual a 1 o 0.

062 El ángulo que forman las dos rectas de regresión de una distribución bidimensional es mayor cuanto menor sea el coeficiente de correlación.

Vamos a comprobarlo estudiando las dos magnitudes en estas distribuciones.

10	12	14	16	18
3	8	1	9	2

10	12	14	16	18
3	6	8	6	7

10	12	14	16	18
5	6	6,5	8,5	9

$\overline{x} = 14$ $\quad\overline{y} = 4,6$ $\quad\sigma_{XY} = -0,4$

$\sigma_X = \sqrt{8} = 2,83$ $\quad\sigma_Y = \sqrt{1,55} = 1,24$ $\quad r_{XY} = -0,11$

Recta de regresión de Y sobre X: $y - 4,6 = -\frac{0,4}{8}(x - 14) \rightarrow y = -0,05x + 5,3$

Recta de regresión de X sobre Y: $x - 14 = -\frac{0,4}{1,55}(y - 4,6) \rightarrow x = -0,26y + 15,2$

El ángulo que forman las rectas es:

$$\cos\alpha = \frac{0,26 + 0,05}{\sqrt{(-1)^2 + 0,05^2} \cdot \sqrt{(-0,26)^2 + 1^2}} = 0,29 \rightarrow \alpha = 72° \, 33' \, 48''$$

$\overline{y} = 6$ $\quad\sigma_Y = \sqrt{2,8} = 1,67$ $\quad\sigma_{XY} = 3,2$ $\quad r_{XY} = 0,68$

Recta de regresión de Y sobre X: $y - 6 = \frac{3,2}{8}(x - 14) \rightarrow y = 0,4x + 0,4$

Recta de regresión de X sobre Y: $x - 14 = \frac{3,2}{2,8}(y - 6) \rightarrow x = 1,14y + 7,16$

El ángulo que forman las rectas es:

$$\cos\alpha = \frac{1,14 - 0,4}{\sqrt{1^2 + 0,4^2} \cdot \sqrt{1,14^2 + (-1)^2}} = 0,45 \rightarrow \alpha = 63° \, 3' \, 30''$$

$\overline{y} = 7$ $\quad\sigma_Y = \sqrt{2,4} = 1,55$ $\quad\sigma_{XY} = 4,2$ $\quad r_{XY} = 0,96$

Recta de regresión de Y sobre X: $y - 7 = \frac{4,2}{8}(x - 14) \rightarrow y = 0,53x - 0,42$

Recta de regresión de X sobre Y: $x - 14 = \frac{4,2}{2,4}(y - 7) \rightarrow x = 1,75y + 1,75$

El ángulo que forman las rectas es:

$$\cos\alpha = \frac{1,75 + 0,53}{\sqrt{1^2 + 0,53^2} \cdot \sqrt{1,75^2 + 1^2}} = 0,99 \rightarrow \alpha = 1° \, 49' \, 16''$$

063 ●●● Se ha realizado un test de memoria, X, y otro test de atención, Y, a varios alumnos y se han reflejado los resultados en esta tabla.

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 Andrés en memoria, si ha obtenido 33 en atención.

d) Si es factible, estima qué puntuación obtendrá Eva en atención, si ha obtenido 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) $y = 33 \rightarrow x = 0{,}81 \cdot 33 + 4{,}19 = 30{,}92$

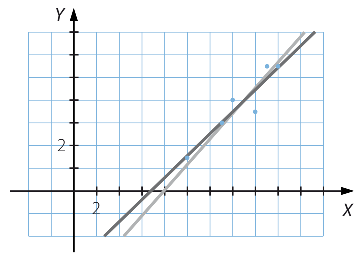
d) $x = 27 \rightarrow y = 0{,}61 \cdot 27 + 11{,}01 = 27{,}48$

PARA FINALIZAR…

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

Comprueba el resultado obtenido para estos datos.

X	10	13	16	14	17	18
Y	3	6	7	8	11	11



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

$\overline{x} = 14{,}67$ $\qquad$ $\overline{y} = 7{,}67$ $\qquad$ $\sigma_{XY} = 6{,}98$

$\sigma_X = \sqrt{7{,}12} = 2{,}67$ $\qquad$ $\sigma_Y = \sqrt{7{,}84} = 2{,}8$ $\qquad$ $r_{XY} = 0{,}93$

Recta de regresión de Y sobre X: $y - 7{,}67 = \dfrac{6{,}98}{7{,}12}(x - 14{,}67) \rightarrow m_X = 0{,}98$

Recta de regresión de X sobre Y: $x - 14{,}67 = \dfrac{6{,}98}{7{,}84}(y - 7{,}67) \rightarrow m_Y = 0{,}89$

$m_X \cdot m_Y = 0{,}93$

065 **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.

066 **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 las multiplicamos 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:

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

067 **Demuestra que el coeficiente de correlación de dos variables estadísticas no varía si a cada valor de las dos variables se les suma o resta un mismo número.**

Utiliza esta propiedad para calcular el coeficiente de correlación de las siguientes variables estadísticas.

X	2.001	2.002	2.003	2.004	2.005
Y	7.390	7.350	7.240	7.210	7.110

Si se suma un valor c a cada valor de una variable estadística, entonces la media de los datos obtenidos es:

$$\frac{\sum_{i=1}^{n} f_i \cdot (x_i + c)}{N} = \frac{\sum_{i=1}^{n} f_i \cdot x_i + \sum_{i=1}^{n} f_i \cdot c}{N} = \frac{\sum_{i=1}^{n} f_i \cdot x_i + c \cdot \sum_{i=1}^{n} f_i}{N} =$$

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

La varianza de estos datos verifica que:

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

Por tanto, la desviación típica también coincide.

La covarianza entre las dos variables es:

$$\frac{\sum_{i=1}^{n} f_i \cdot (x_i + c) \cdot y_i}{N} - (\overline{x} + c) \cdot \overline{y} = \frac{\sum_{i=1}^{n} f_i \cdot x_i \cdot y_i + \sum_{i=1}^{n} f_i \cdot c \cdot y_i}{N} - \overline{x} \cdot \overline{y} - c \cdot \overline{y} =$$

$$= \frac{\sum_{i=1}^{n} f_i \cdot x_i \cdot y_i + c \cdot \sum_{i=1}^{n} f_i \cdot y_i}{N} - \overline{x} \cdot \overline{y} - c \cdot \overline{y} =$$

$$= \frac{\sum_{i=1}^{n} f_i \cdot x_i \cdot y_i}{N} + c \cdot \overline{y} - \overline{x} \cdot \overline{y} - c \cdot \overline{y} = \sigma_{XY}$$

Así, el coeficiente de correlación es igual que el de las variables iniciales.
Del mismo modo, si se suma o se resta un mismo número a las dos variables el coeficiente no varía.

068 **En dos estudios 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 recta 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 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 $(\overline{x}, \overline{y})$.

069 **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 $m = 3$ y su ordenada en el origen es 2.**
- **La varianza de Y es 3.**

Calcula las medidas estadísticas de cada una de las variables estadísticas 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: $r = \sqrt{2 \cdot \dfrac{1}{3}} = 0{,}8164$

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: $\sigma_X^2 = \dfrac{1}{2}$