12 Distribuciones bidimensionales

ACTIVIDADES INICIALES

12.I. Halla la ecuación de la recta que pasa por el punto $A(-2, 5)$ y tiene por pendiente $-\frac{1}{2}$. Calcula la ordenada en el origen y represéntala.

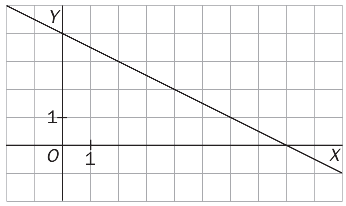
La ecuación de la recta es de la forma $y = -\frac{1}{2}x + b$.

La recta pasa por el punto $A(-2, 5)$; por tanto:

$5 = -\frac{1}{2} \cdot (-2) + b \Rightarrow b = 4$

La recta tiene por ecuación $y = -\frac{1}{2}x + 4$.

La ordenada en el origen es 4.



12.II. En cada caso, calcula la pendiente de la recta que pasa por los siguientes puntos.

a) $A(3, 2)$ y $B(1, 1)$

b) $A(-5, 4)$ y $B(-1, 0)$

a) $m = \frac{1 - 2}{1 - 3} = \frac{1}{2}$

b) $m = \frac{0 - 4}{-1 - (-5)} = \frac{-4}{4} = -1$

EJERCICIOS PROPUESTOS

12.1. La siguiente tabla proporciona la distribución conjunta de frecuencias absolutas de la variable X, que representa el número de tarjetas de crédito que posee una persona, y la variable Y, que representa el número de compras semanales realizadas con tarjeta de crédito.

Y \ X	1	2	3	4
1	20	16	2	0
2	10	4	6	0
3	8	2	8	4

a) Calcula las distribuciones marginales. ¿Cuántas personas tienen más de tres tarjetas?

b) ¿Cuál es el número más frecuente de tarjetas de crédito?

c) ¿Cuántas personas realizan dos o menos de dos compras semanales?

d) ¿Cuál es la media y la varianza del número de tarjetas que posee una persona?

e) ¿Cuál es la media y la varianza del número de compras semanales realizadas con tarjeta?

Construimos las siguientes tablas:

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
1	38	38	38
2	22	44	88
3	16	48	144
4	4	16	64
	80	146	334

y_i	f_i	$y_i f_i$	$y_i^2 f_i$
1	38	38	38
2	20	40	80
3	22	66	198
	80	146	316

a) Cuatro personas tienen más de tres tarjetas.

b) El número más frecuente de tarjetas de crédito es 1.

c) 20 + 38 = 58 personas realizan más de dos compras semanales.

d) $\bar{x} = \frac{146}{80} = 1{,}825$ tarjetas $\quad s_X^2 = \frac{334}{80} - 1{,}825^2 = 0{,}84 \quad s_x = 0{,}92$

e) $\bar{y} = \frac{144}{80} = 1{,}8$ compras $\quad s_Y^2 = \frac{316}{80} - 1{,}8^2 = 0{,}71 \quad s_Y = 0{,}84$

12.2. Las calificaciones de 39 alumnos en Filosofía e Historia han sido las siguientes:

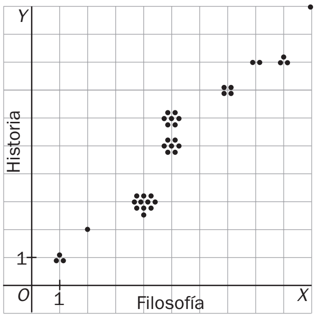
Filosofía (x_i)	1	2	4	5	6	7	8	9	10
Historia (y_i)	1	2	3	5	6	7	8	8	10
N.° de alumnos (f_i)	3	1	11	7	7	4	2	3	1

a) Representa el diagrama de dispersión.

b) A la vista del diagrama de dispersión, ¿se puede establecer que existe algún tipo de relación entre las calificaciones de Historia y Filosofía?

c) Calcula la nota media en Historia.

a)



c) Formamos la tabla:

y_i	f_i	$y_i f_i$
1	3	3
2	1	2
3	11	33
5	7	35
6	7	42
7	4	28
8	5	40
10	1	10
	39	193

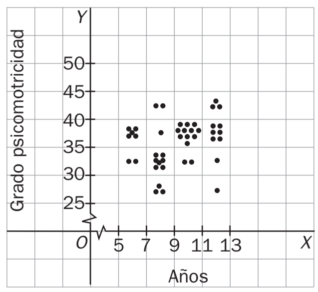
$$\bar{y} = \frac{193}{39} = 4{,}95$$

b) A mayor nota de Filosofía, mayor nota de Historia

12.3. En la siguiente tabla se recogen las edades y el grado de psicomotricidad de 44 niños:

X (años) / Y (psicom.)	[5, 7)	[7, 9)	[9, 11)	[11, 13)	
[25, 30)	4	3	0	1	8
[30, 35)	2	7	2	0	11
[35, 40)	1	1	11	1	14
[40, 45)	0	2	0	6	8
[45, 50)	0	0	0	3	3
	7	13	13	11	44

Representa el diagrama de dispersión.



12.4. La siguiente tabla muestra las calificaciones obtenidas por cinco alumnos en Bachillerato (X) y en las PAU (Y).

Bachillerato	5,4	6,8	5,3	7,4	4,3
PAU	5,8	4,8	5,9	7,4	4,2

A partir de ella, calcula:

a) Las medias y las varianzas de X y de Y.

b) La covarianza de (X, Y).

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
5,4	5,8	29,16	33,64	31,32
6,8	4,8	46,24	23,04	32,64
5,3	5,9	28,09	34,81	31,27
7,4	7,4	54,76	54,76	54,76
4,3	4,2	18,49	17,64	18,06
29,2	28,1	176,74	163,89	168,05

a) $\bar{x} = \dfrac{29{,}2}{5} = 5{,}84 \qquad s_X^2 = \dfrac{176{,}74}{5} - 5{,}84^2 = 1{,}2424$

$\bar{y} = \dfrac{28{,}1}{5} = 5{,}62 \qquad s_Y^2 = \dfrac{163{,}89}{5} - 5{,}62^2 = 1{,}1936$

b) $S_{XY} = \dfrac{168{,}05}{5} - 5{,}84 \cdot 5{,}62 = 0{,}7892$

12.5. En un depósito cilíndrico la altura del agua que contiene varía conforme pasa el tiempo según la siguiente tabla:

Tiempo (h)	8	22	27	33	50	70
Altura (m)	17	14	12	11	6	1

Halla:

a) Las medias de X y de Y.

b) Las varianzas de X y de Y.

c) La covarianza de (X, Y)

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	17	64	289	136
22	14	484	196	308
27	12	729	144	324
33	11	1089	121	363
50	6	2500	36	300
70	1	4900	1	70
210	61	9766	787	1501

a) $\bar{x} = \dfrac{210}{6} = 35$

$\bar{y} = \dfrac{61}{6} = 10{,}17$

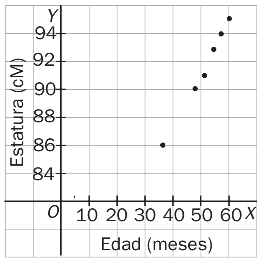
b) $s_X^2 = \dfrac{9766}{6} - 35^2 = 402{,}67$

$s_Y^2 = \dfrac{787}{6} - 10{,}17^2 = 27{,}74$

c) $S_{XY} = \dfrac{1501}{6} - 35 \cdot 10{,}17 = -105{,}78$

12.6. La tabla adjunta expresa los valores de la variable bidimensional edad, en meses, y estatura, en centímetros, de una niña entre los 3 y los 5 años. Representa la nube de puntos de esta variable e indica la relación existente entre la edad y la estatura.

Edad (meses)	36	48	51	54	57	60
Estatura (cm)	86	90	91	93	94	95

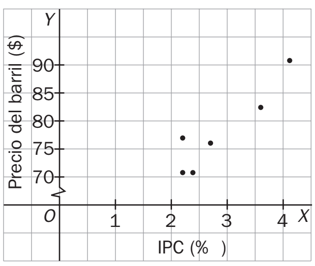


Según se observa en el diagrama de dispersión, existe una correlación lineal positiva fuerte.

12.7. En la siguiente tabla se recoge la evolución del IPC (índice de precios al consumo) y el precio del barril de petróleo (*brent*) durante el segundo semestre de 2007.

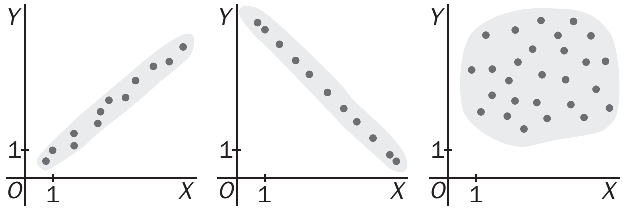
IPC (%)	2,4	2,2	2,2	2,7	3,6	4,1
Precio barril ($)	71,54	77,01	70,73	76,87	82,50	90,16

¿Se puede asegurar que la evolución del IPC está directamente relacionada con el precio del petróleo?



Sí, existe una correlación lineal positiva fuerte.

12.8. Los números 0, 0,8 y 1 son los valores absolutos del coeficiente de correlación de las distribuciones bidimensionales cuyas nubes de puntos se adjuntan:



Asigna a cada diagrama su coeficiente de correlación, cambiando el signo cuando sea necesario.

Primero: 0,8

Segundo: −1

Tercero: 0

12.9. (PAU) Los resultados de los exámenes de Inglés (X) y Matemáticas (Y) de 8 alumnos han sido los siguientes:

X	8	9	8,5	7	7	7,5	7,5	6,5
Y	7	7,5	8	6	6,5	7	6,5	2

a) Halla el coeficiente de correlación de las calificaciones en Inglés y Matemáticas de los siete primeros alumnos.

b) Calcula el coeficiente de correlación de esas dos variables para los ocho alumnos.

c) Explica la diferencia entre los resultados obtenidos.

a) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	7	64	49	56
9	7,5	81	56,25	67,5
8,5	8	72,25	64	68
7	6	49	36	42
7	6,5	49	42,25	45,5
7,5	7	56,25	49	52,5
7,5	6,5	56,25	42,25	48,75
54,5	48,5	427,75	338,75	380,25

$$\bar{x} = \frac{54,5}{7} = 7,79 \qquad \bar{y} = \frac{48,5}{7} = 6,93$$

$$s_X^2 = \frac{427,75}{7} - 7,79^2 = 0,42 \qquad s_X = \sqrt{0,42} = 0,65$$

$$s_Y^2 = \frac{338,75}{7} - 6,93^2 = 0,37 \qquad s_Y = \sqrt{0,37} = 0,61$$

$$S_{XY} = \frac{380,25}{7} - 7,79 \cdot 6,93 = 0,34$$

$$r = \frac{S_{XY}}{s_X s_Y} = \frac{0,34}{0,65 \cdot 0,61} = 0,86$$

b) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	7	64	49	56
9	7,5	81	56,25	67,5
8,5	8	72,25	64	68
7	6	49	36	42
7	6,5	49	42,25	45,5
7,5	7	56,25	49	52,5
7,5	6,5	56,25	42,25	48,75
6,5	2	42,25	4	13
61	50,5	470	342,75	393,25

$$\bar{x} = \frac{61}{8} = 7,635 \qquad \bar{y} = \frac{50,5}{8} = 6,4125$$

$$s_X^2 = \frac{470}{8} - 7,625^2 = 0,61 \qquad s_X = \sqrt{0,61} = 0,78$$

$$s_Y^2 = \frac{342,75}{8} - 6,3125^2 = 3 \qquad s_Y = \sqrt{3} = 1,73$$

$$S_{XY} = \frac{393,25}{8} - 7,625 \cdot 6,3125 = 1,02$$

$$r = \frac{S_{XY}}{s_X s_Y} = \frac{1,02}{0,78 \cdot 1,73} = 0,76$$

c) Mientras que los siete primeros alumnos tienen una nota pareja en las dos materias, el último no.

12.10. (PAU) En cierto país, el tipo de interés y el índice de la Bolsa en los seis últimos meses vienen dados por la siguiente tabla.

Tipo de interés (%)	8	7,5	7,2	6	5,5	5
Índice	120	130	134	142	150	165

Halla el índice previsto de la Bolsa en el séptimo mes, suponiendo que el tipo de interés en ese mes fue del 4,1%, y analiza la fiabilidad de la predicción, según el valor del coeficiente de correlación.

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	120	64	14 400	960
7,5	130	56,25	16 900	975
7,2	134	51,84	17 956	964,8
6	142	36	20 164	852
5,5	150	30,25	22 500	825
5	165	25	27 225	825
39,2	841	263,34	119 145	4501,8

$\bar{x} = \dfrac{39,2}{6} = 6,53$ $\bar{y} = \dfrac{841}{6} = 140,17$

$s_X^2 = \dfrac{263,34}{6} - 6,53^2 = 1,25$ $s_X = \sqrt{1,2491} = 1,12$

$s_Y^2 = \dfrac{119\,145}{6} - 140,17^2 = 209,8711$ $s_Y = \sqrt{209,8711} = 14,48$

$S_{XY} = \dfrac{5401,8}{6} - 6,53 \cdot 140,17 = -15,01$

La recta de regresión de Y sobre X es: $y - 140,17 = \dfrac{-15,01}{1,25}(x - 6,53) \Rightarrow y = -12,008x + 218,58.$

$y = -12,008 \cdot 4,1 + 218,58 = 169,35$ es el índice de Bolsa esperado para el siguiente mes.

Como $r = \dfrac{S_{XY}}{s_X s_Y} = \dfrac{-15,01}{1,12 \cdot 14,48} = -0,93$, el resultado obtenido es fiable.

12.11. (PAU) Como consecuencia de un estudio estadístico realizado sobre 100 universitarios se ha obtenido una estatura media de 155 cm, con una desviación típica de 15,5 cm. Además se obtuvo la recta de regresión: $y = 80 + 1,5x$ (siendo X el peso, e Y, la altura).

Determina el peso medio de estos 100 universitarios.

Las rectas de regresión se cortan en el punto $(\bar{x}, \bar{y})$: $155 = 80 + 1,5\bar{x} \Rightarrow \bar{x} = \dfrac{155 - 80}{1,5} = 50$ kilos.

12.12. (PAU) Un estudio sociológico proporcionó la siguiente tabla.

Nivel de estudios	1	2	3	4	5
Salario medio (€)	800	1000	1500	2000	3000

a) Calcula el coeficiente de correlación lineal entre el nivel de estudios y el salario medio, y, en función del valor obtenido, explica si se puede considerar que el salario medio está determinado por el nivel de estudios.

1 = estudios primarios 2 = estudios secundarios 3 = formación profesional
4 = técnicos de grado medio 5 = técnicos superiores 6 = doctores

b) Deduce el salario esperado para el nivel de estudios 6.

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	800	1	640 000	800
2	1000	4	1 000 000	2000
3	1500	9	2 250 000	4500
4	2000	16	4 000 000	8000
5	3000	25	9 000 000	15 000
15	8300	55	16 890 000	30 300

a) $\bar{x} = \dfrac{15}{5} = 3$ $s_X^2 = \dfrac{55}{5} - 3^2 = 2$ $s_X = 1,41$

$\bar{y} = \dfrac{8300}{5} = 1660$ $s_Y^2 = \dfrac{16\,890\,000}{5} - 1660 = 622\,400$ $s_Y = 788,92$

$S_{XY} = \dfrac{30\,300}{5} - 3 \cdot 1660 = 1080$

$r = \dfrac{S_{XY}}{s_X s_Y} = \dfrac{1080}{1,41 \cdot 788,92} = 0,97$

Se puede considerar que el salario es en función del nivel de estudios.

b) $y - 1660 = 540(x - 3) \Rightarrow y = 540x + 40 \Rightarrow y = 540 \cdot 6 + 40 = 3280$ euros.

12.13. Sea la variable bidimensional dada por la siguiente tabla.

X	1	2	3	4	5	6	7	8	9
Y	5	6	8	11	1	13	14	14	17

a) Halla la recta de Tukey.

b) Halla la recta de regresión de Y sobre *X*.

c) Representa la nube de puntos y las dos rectas obtenidas.

a) Dividimos el conjunto de datos en los grupos:

$G_1 = \{(1, 5)\ (2, 6)\ (3, 8)\}$

$G_2 = \{(4, 11)\ (5, 1)\ (6, 13)\}$

$G_3 = \{(7, 14)\ (8, 14)\ (9, 17)\}$

Mediana de las abscisas de G_1: $x_1 = 2 \quad \Rightarrow P_1(2, 6)$

Mediana de las ordenadas de G_1: $y_1 = 6$

Mediana de las abscisas de G_2: $x_2 = 5 \quad \Rightarrow P_2(5, 11)$

Mediana de las ordenadas de G_2: $y_2 = 11$

Mediana de las abscisas de G_3: $x_3 = 8 \quad \Rightarrow P_3(8, 14)$

Mediana de las ordenadas de G_3: $y_3 = 14$

Baricentro del triángulo P_1, P_2, P_3: $G\left(\frac{2 + 5 + 8}{3}, \frac{6 + 11 + 14}{3}\right) = \left(5, \frac{31}{3}\right)$

Pendiente de la recta que pasa por P_1 y P_3: $m = \frac{14 - 6}{8 - 2} = \frac{8}{6} = \frac{4}{3}$

Recta de Tukey: $y - \frac{31}{3} = \frac{4}{3}(x - 5) \Rightarrow y = \frac{4}{3}x + \frac{11}{3}$

b) Formamos la tabla:

x_i	y_i	x_i^2	$x_i\, y_i$
1	5	1	5
2	6	4	12
3	8	9	24
4	11	16	44
5	1	25	5
6	13	36	78
7	14	49	98
8	14	64	112
9	17	81	153
45	89	285	531

$\bar{x} = \frac{45}{9} = 5$

$\bar{y} = \frac{89}{9} = 9{,}89$

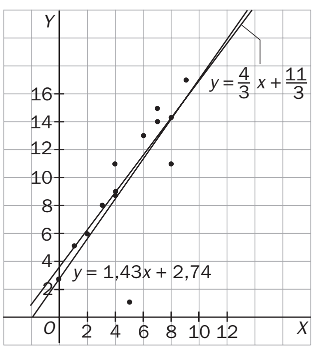
$s_X^2 = \frac{285}{9} - 5^2 = 6{,}67$

$S_{XY} = \frac{531}{9} - 5 \cdot 9{,}89 = 9{,}55$

La recta de regresión de *Y* sobre *X* es:

$y - 9{,}89 = \frac{9{,}55}{6{,}67}(x - 5) \Rightarrow y = 1{,}43x + 2{,}74$

c)



12.14. La siguiente tabla da los datos obtenidos para una variable bidimensional.

X	1	2	3	4	5	6	7	8	9
Y	14	4	18	16	13	18	15	10	11

a) Halla la recta de regresión de Y sobre X.

b) Calcula la recta de Tukey.

c) Representa la nube de puntos y las dos rectas obtenidas.

a) Formamos la tabla:

x_i	y_i	x_i^2	$x_i y_i$
1	14	1	14
2	4	4	8
3	18	9	54
4	16	16	64
5	13	25	65
6	18	36	108
7	15	49	105
8	10	64	80
9	11	81	99
45	119	285	597

$$\bar{x} = \frac{45}{9} = 5$$

$$\bar{y} = \frac{119}{9} = 13{,}22$$

$$s_X^2 = \frac{285}{9} - 5^2 = 6{,}67$$

$$s_{XY} = \frac{597}{9} - 5 \cdot 13{,}22 = 0{,}23$$

La recta de regresión de Y sobre X es:

$$y - 13{,}22 = \frac{0{,}23}{6{,}67}(x - 5) \Rightarrow y = 0{,}034x + 13{,}05$$

b) Formamos con los datos ordenados tres grupos:

$G_1 = \{(1, 14)\ (2, 4)\ (3, 18)\}$

$G_2 = \{(4, 16)\ (5, 13)\ (6, 18)\}$

$G_3 = G_3 = \{(7, 15)\ (8, 10)\ (9, 11)\}$

Para cada grupo G_i hallamos el punto $P_i(x_i, y_i)$:

$P_1(2, 14)$ $\quad P_2(5, 16)$ $\quad P_3(8, 11)$

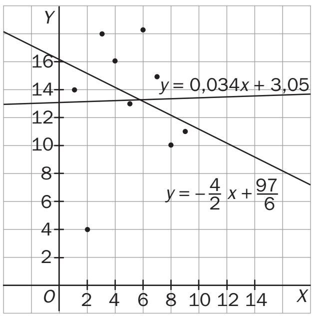
El baricentro del triángulo de vértices $P_1P_2P_3$ tiene por coordenadas:

$$x_G = \frac{2 + 5 + 8}{3} = 5$$

$$y_G = \frac{14 + 16 + 11}{3} = \frac{41}{3}$$

La pendiente P_1P_3 es: $m = \dfrac{11 - 14}{8 - 2} = -\dfrac{3}{6} = -\dfrac{1}{2}$

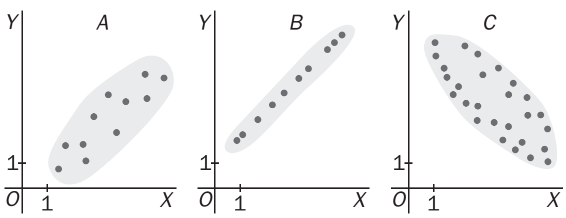
Recta de Tukey: $y - \dfrac{41}{3} = -\dfrac{1}{2}(x - 5) \Rightarrow y = -\dfrac{1}{2}x + \dfrac{97}{6}$



EJERCICIOS

Nube de puntos y correlación

12.15. Considera las siguientes nubes de puntos.



a) ¿En cuál de ellas los datos se ajustarán mejor a una recta?

b) Asigna a cada una de las nubes uno de los siguientes coeficientes de correlación, fijando el signo en cada caso.

$r_1 - \pm 0{,}99$

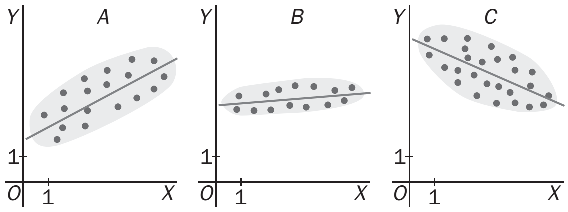
$r_2 = \pm 0{,}6$

$r_3 = \pm 0{,}8$

a) Se ajustará mejor a una recta la nube de puntos del apartado b.

b) A: $r = 0{,}8$ $\quad B$: $r = 0{,}99$ $\quad C$: $r = -0{,}6$

12.16. (PAU) En las gráficas siguientes se muestran las rectas de regresión obtenidas en tres estudios estadísticos.



a) ¿En cuál de las gráficas el coeficiente de correlación será mayor?

b) Indica en qué gráficas el coeficiente de correlación sería negativo.

Justifica las respuestas.

a) El de la gráfica *B*, ya que los puntos están más agrupados.

b) El de la gráfica *C*, ya que los puntos se agrupan en torno a una recta de pendiente negativa.

12.17. (PAU) En una empresa de transportes trabajan cuatro conductores. Los años de antigüedad de sus permisos de conducir y el número de infracciones cometidas en el último año por cada uno de ellos son los siguientes:

X: años de antigüedad	3	4	5	6
Y: infracciones	4	3	2	1

a) Representa gráficamente los datos anteriores. Razona si muestran correlación positiva o negativa.

b) Calcula el coeficiente de correlación e interprétalo en términos de la situación real.

a)

Relación positiva

b) $\bar{x} = \dfrac{3+4+5+6}{4} = \dfrac{18}{4} = 4,5$ $\qquad \bar{y} = \dfrac{4+3+2+1}{4} = \dfrac{10}{4} = 2,5$

$s_x^2 = \dfrac{9+16+25+36}{4} - 4,5^2 = 1,25 \qquad s_X = 1,12$

$s_y^2 = \dfrac{16+9+4+1}{4} - 2,5^2 = 1,25; \qquad s_Y = 1,12$

$s_{XY} = \dfrac{12+12+10+6}{4} - 4,5 \cdot 2,5 = 10 - 11,25 = -1,25;$

$r = \dfrac{s_{xy}}{s_x s_y} = \dfrac{-1,25}{1,12 \cdot 1,12} = -1.$

Existe dependencia funcional negativa.

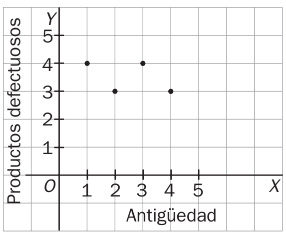
12.18. En una empresa trabajan cuatro obreros. La antigüedad y el número de productos defectuosos elaborados por ellos durante el último año vienen dados en la siguiente tabla.

Antigüedad	3	2	4	1
Productos defectuosos	4	3	3	4

a) Dibuja el diagrama de dispersión y justifica si los datos presentan correlación positiva o negativa.

b) Calcula el coeficiente de correlación.

a)



b) $\bar{x} = \dfrac{3+2+4+1}{4} = \dfrac{10}{4} = 2,5;$ $\qquad \bar{y} = \dfrac{4+3+3+4}{4} = \dfrac{14}{4} = 3,5$

$s_x^2 = \dfrac{9+4+16+1}{4} - 2,5^2 = 1,25 \qquad s_X = 1,12$

$s_y^2 = \dfrac{16+9+9+16}{4} - 3,5^2 = 0,25 \qquad s_Y = 0,5$

$S_{XY} = \dfrac{12+6+12+4}{4} - 2,5 \cdot 3,5 = -0,25;$

$r = \dfrac{s_{xy}}{s_x s_y} = \dfrac{-0,25}{1,12 \cdot 0,5} = -0,45$

Los datos no presentan relación.

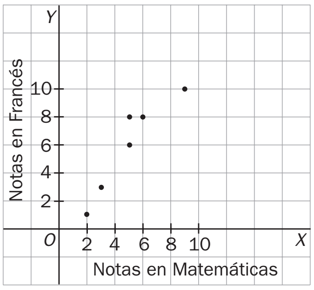
12.19. La tabla siguiente muestra las notas de Matemáticas, X, y de Francés, Y, de seis estudiantes.

X	2	3	5	5	6	9
Y	1	3	8	16	8	10

a) Representa gráficamente los valores de la tabla mediante una nube de puntos.

b) Halla el coeficiente de correlación. ¿Crees que las variables están fuertemente correlacionadas?

a)



b) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	1	4	1	2
3	3	9	9	9
5	8	25	64	40
5	6	25	36	30
6	8	36	64	48
9	10	81	100	90
30	36	180	274	219

$\bar{x} = \dfrac{30}{6} = 5 \qquad s_X^2 = \dfrac{180}{6} - 5^2 = 5 \qquad s_X = 2{,}24$

$\bar{y} = \dfrac{36}{6} = 6 \qquad s_Y^2 = \dfrac{274}{6} - 6^2 = 9{,}67 \qquad s_Y = 3{,}11$

$s_{XY} = \dfrac{219}{6} - 5 \cdot 6 = 6{,}5 \qquad r = \dfrac{s_{XY}}{s_X s_Y} = \dfrac{6{,}5}{2{,}24 \cdot 3{,}11} = 0{,}93$

Las variables están fuertemente correlacionadas y la correlación es positiva, es decir, cuando aumenta la variable X, aumenta también la variable Y.

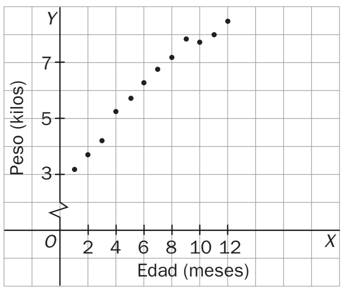
12.20. Durante su primer año de vida han pesado a Marta cada mes. En la tabla siguiente aparecen sus pesos. X representa la edad en meses, e Y, el peso en kilogramos.

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	3,2	3,7	4,2	5,3	5,7	6,5	6,8	7,2	7,9	7,7	8	8,5

a) Representa el diagrama de dispersión.

b) Calcula el coeficiente de correlación lineal e interprétalo.

a)



b) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	3,2	1	10,24	3,2
2	3,7	4	13,69	7,4
3	4,2	9	17,64	12,6
4	5,3	16	28,09	21,2
5	5,7	25	32,49	28,5
6	6,5	36	42,25	39
7	6,8	49	46,24	47,6
8	7,2	64	51,84	57,6
9	7,9	81	62,41	71,1
10	7,7	100	59,29	77
11	8	121	64	88
12	8,5	144	72,25	102
78	74,7	650	500,43	555,2

$\bar{x} = \dfrac{78}{12} = 6{,}5$

$s_X^2 = \dfrac{650}{12} - 6{,}5^2 = 11{,}9 \qquad s_X = 3{,}45$

$\bar{y} = \dfrac{74{,}7}{6} = 6{,}23$

$s_Y^2 = \dfrac{500{,}43}{12} - 6{,}23^2 = 2{,}89 \qquad s_Y = 1{,}7$

$s_{XY} = \dfrac{555{,}2}{12} - 6{,}5 \cdot 6{,}23 = 5{,}78$

$r = \dfrac{s_{XY}}{s_X s_Y} = \dfrac{5{,}78}{3{,}45 \cdot 1{,}7} = 0{,}99$

La correlación es positiva y fuerte: al aumentar el tiempo, aumenta el peso de Marta aumenta.

12.21. Se hizo una prueba a 10 estudiantes para ver la relación que había entre la expresión oral (X) y la destreza manual (Y), obteniéndose la siguiente tabla.

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

a) Calcula razonadamente la media y la desviación típica de X.

b) Calcula razonadamente la media y la desviación típica de Y.

c) ¿Qué distribución está más dispersa? Justifica la respuesta.

d) Calcula el coeficiente de correlación lineal e interprétalo.

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	5	64	25	40
7	5	49	25	35
6	6	36	36	36
5	7	25	49	35
4	8	16	64	32
3	7	9	49	21
7	4	49	16	28
6	5	36	25	30
9	3	81	9	27
5	5	125	25	25
60	55	390	323	309

a) $\bar{x} = \dfrac{60}{10} = 6$ $\qquad s_X^2 = \dfrac{390}{10} - 6^2 = 3$ $\qquad s_X = 1{,}73$

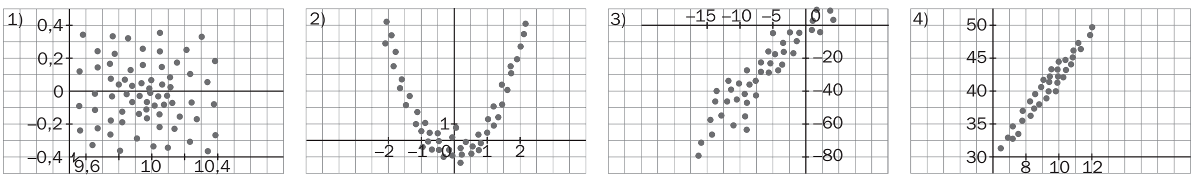
b) $\bar{y} = \dfrac{55}{10} = 5{,}5$ $\qquad s_Y^2 = \dfrac{323}{10} - 6^2 = 2{,}05$ $\qquad s_Y = 1{,}43$

c) Como la desviación típica de X es mayor que la desviación típica de Y, está más dispersa la distribución de X.

d) $s_{xy} = \dfrac{309}{10} - 6 \cdot 5{,}5 = -2{,}1$ $\qquad r = \dfrac{s_{xy}}{s_x s_y} = \dfrac{-2{,}1}{1{,}73 \cdot 1{,}43} = -0{,}85$

La correlación es inversa: a mejor expresión oral, peor destreza manual.

12.22. Considera las nubes de puntos de la figura.



a) Indica si hay relación de dependencia entre la variable X y la variable Y. En caso de haberla, ¿puede considerarse esta relación lineal?

b) Asigna a cada gráfico una de las siguientes rectas:

$$y = x \qquad y = 1 - 0{,}2x \qquad y = 2 + 4x.$$

a) Hay relación entre las variables 2, 3 y 4, siendo lineal en las dos últimas.

b) La recta $y = 2 + 4x$ es la más adecuada para reflejar la relación entre las variables X e Y de los gráficos 3 y 4. Esta recta tiene pendiente 4 y en ambas nubes de puntos se observa que el rango de y es aproximadamente 4 veces mayor que el rango de X.

Modelo de regresión lineal

12.23. Los datos siguientes corresponden a la altura sobre el nivel del mar (X) y la presión atmosférica (Y) de siete puntos.

X	11	14	16	15	16	18	20	21	14	20	19	11
Y	2	3	5	6	5	3	7	10	6	10	5	6

a) Halla la recta de regresión de Y sobre X.

b) ¿Qué presión atmosférica habría sobre Peña Vieja (2600 metros de altitud aproximadamente)?

a) Formamos la tabla siguiente:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
0	760	0	577 600	0
184	745	33 856	555 025	137 080
231	740	53 361	547 600	170 940
481	720	231 361	518 400	346 320
730	700	532 900	490 000	511 000
911	685	829 921	469 225	624 035
1550	650	2 402 500	422 500	1 007 500
4087	5000	390	3 580 350	2 796 875

$$\bar{x} = \frac{4087}{7} = 583{,}86 \qquad \bar{y} = \frac{5000}{7} = 714{,}29$$

$$s_x^2 = \frac{4\,083\,899}{7} - 583{,}86^2 = 242\,521{,}64$$

$$s_{xy} = \frac{2\,796\,875}{7} - 583{,}86 \cdot 714{,}29 = -17\,491{,}79$$

Recta de regresión de la presión respecto de la altura:

$$y - 714{,}29 = -\frac{17\,491{,}79}{242\,521{,}64}(x - 583{,}86)$$

$$y = -0{,}07x + 755{,}16$$

b) Para saber qué presión atmosférica habrá en Peña Vieja, que se encuentra situada a 2600 m de altitud, sustituiremos en la ecuación anterior $x = 2600$.

$$y = -0{,}07 \cdot 2600 + 755{,}16 = 573{,}16 \text{ mm de mercurio}$$

12.24. (PAU) La información estadística obtenida de una muestra de tamaño 12 sobre la relación existente entre la inversión realizada, X, y el rendimiento obtenido, Y, en miles de euros para explotaciones agropecuarias se muestra en la siguiente tabla.

X	0	184	231	481	730	911	1550
Y	760	745	740	720	700	685	650

a) Halla la recta de regresión de Y sobre X.

b) Determina la previsión de inversión que se obtendrá con un rendimiento de 7500 euros.

a) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
11	2	121	4	22
14	3	196	9	42
16	5	256	25	80
15	6	225	36	90
16	5	256	25	80
18	3	324	9	54
20	7	400	49	140
21	10	441	100	210
14	6	196	36	84
20	10	400	100	200
19	5	361	25	95
11	6	121	121	66
195	68	3297	454	1163

$$\bar{x} = \frac{195}{12} = 16{,}25; \qquad \bar{y} = \frac{68}{12} = 5{,}67$$

$$s_x^2 = \frac{3297}{12} - 16{,}25^2 = 274{,}75 - 264{,}06 = 10{,}68$$

$$s_y^2 = \frac{454}{12} - 5{,}67^2 = 37{,}83 - 32{,}15 = 5{,}68$$

$$s_{xy} = \frac{1163}{12} - 16{,}25 \cdot 5{,}67 = 96{,}92 - 92{,}14 = 4{,}78$$

Recta de regresión de Y sobre X

$$y - 5{,}67 = \frac{4{,}78}{10{,}68}(x - 16{,}25) \Rightarrow y = 0{,}45x - 1{,}64$$

b) La recta de regresión de X sobre Y es:

$$x - \bar{x} = \frac{s_{xy}}{s_y^2}(y - \bar{y});\ x - 16{,}25 = \frac{4{,}78}{5{,}68}(y - 5{,}67) \Rightarrow x = 0{,}84y + 11{,}49$$

Para $y = 7{,}5$, sustituimos este valor en la ecuación obtenida: $x = 0{,}84 \cdot 7{,}5 + 11{,}49 = 17{,}79$.

Por tanto, para un rendimiento de 7500 euros se prevé una inversión de 17 790.

12.25. Cinco niñas de 2, 3, 5, 7 y 8 años de edad pesan, respectivamente, 14, 20, 32, 42 y 44 kilos.

a) Halla la ecuación de la recta de regresión de edad sobre el peso.

b) ¿Cuál sería el peso aproximado de una niña de 6 años?

a) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	14	4	196	28
3	20	9	400	60
5	32	25	1024	160
7	42	49	1764	294
8	44	64	1936	352
60	152	151	5320	894

$\bar{x} = \dfrac{25}{5} = 5$

$\bar{y} = \dfrac{152}{5} = 30,4$

$s_x^2 = \dfrac{151}{5} - 5^2 = 5,2$

$s_y^2 = \dfrac{5320}{5} - 30,4^2 = 139,84$

$s_{xy} = \dfrac{894}{5} - 5 \cdot 30,4 = 26,8$

Recta de regresión de X sobre Y:

$x - 5 = \dfrac{26,8}{139,84}(y - 30,4)$

$x = 0,19y - 0,78$

b) Recta de regresión de Y sobre X: $y - 30,4 = \dfrac{26,8}{139,84}(x - 5)$; $y = 5,15x + 4,65$.

A una niña de 6 años le corresponde un peso de: $y = 5,15 \cdot 6 + 4,65 = 35,55$ kg.

12.26. La siguiente tabla ofrece los resultados de 6 pares de observaciones, realizadas para analizar el grado de relación existente entre dos variables, X e Y.

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

a) Encuentra la recta de regresión de Y sobre X.

b) Representa, sobre unos mismos ejes, la recta anterior y los pares de observaciones de la tabla.

c) ¿Qué grado de relación lineal existe entre ambas variables?

a) Formamos la tabla

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
2	0	4	0	0
2	1	4	1	2
3	1	9	1	3
3	2	9	4	6
3	4	9	16	12
4	3	16	9	12
17	11	51	31	35

$\bar{x} = \dfrac{17}{6} = 2,83$

$\bar{y} = \dfrac{11}{6} = 1,83$

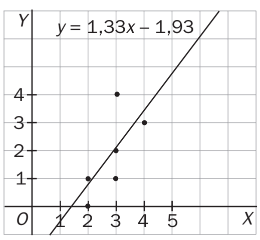
$s_x^2 = \dfrac{51}{6} - 2,83^2 = 0,49$

$s_{xy} = \dfrac{35}{6} - 2,83 \cdot 1,83 = 0,65$

Recta de regresión de Y sobre X:

$y - 1,83 = \dfrac{0,65}{0,49}(x - 2,83)$; $y = 1,33x - 1,93$

b)



c) $s_x = \sqrt{0,47} = 0,7$

$s_y^2 = \dfrac{31}{6} - 1,83^2 = 1,82$

$s_y = \sqrt{1,82} = 1,35$

$r = \dfrac{0,65}{0,7 \cdot 1,35} = 0,69$.

Existe relación positiva.

12.27. **Se sabe que entre el consumo de papel y el número de litros de agua por metro cuadrado que se recogen en una ciudad no existe relación. Responde razonadamente a las siguientes cuestiones.**

a) ¿Cuál es el valor de la covarianza de estas variables?

b) ¿Cuánto vale el coeficiente de correlación lineal?

c) ¿Qué ecuaciones tienen las dos rectas de regresión y cuál es su posición en el plano?

a) $s_{xy} = 0$

b) $r = 0$

c) Las ecuaciones de las rectas de regresión son: $y = \bar{y}$; $x = \bar{x}$

Por tanto, son paralelas a los ejes y, en consecuencia, perpendiculares.

12.28. **En una distribución bidimensional, la recta de regresión de Y sobre X es $y = \bar{y}$, siendo $\bar{y}$ la media de la distribución Y. ¿Cuál es la recta de regresión de X sobre Y? ¿Existe dependencia lineal entre Y y X? Razona las respuestas.**

Si la recta de regresión de Y sobre X es $y = \bar{y}$, la recta de regresión de X sobre Y será $x = \bar{x}$.

En estos casos no existe ningún tipo dependencia entre las variables X e Y; por tanto, están incorreladas.

12.29. **Dada la distribución bidimensional:**

X	5	6,5	8	4	3
Y	4,5	7	7,5	5	3,5

a) Calcula el coeficiente de correlación lineal, interpretando el resultado.

b) Determina la recta de regresión de Y sobre X.

c) Determina la recta de regresión de X sobre Y.

d) Halla el punto en que se cortan las dos rectas.

Formamos la tabla siguiente:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
5	4,5	25	20,25	22,5
6,5	7	42,25	49	45,5
8	7,5	64	56,25	60
4	5	16	25	20
3	3,5	9	12,25	10,5
26,5	27,5	156,25	162,75	158,5

$$\bar{x} = \frac{26,5}{5} = 5,3$$

$$s_x^2 = \frac{156,25}{5} - 5,3^2 = 3,16$$

$$s_x = \sqrt{3,16} = 1,78$$

$$\bar{y} = \frac{27,5}{5} = 5,5$$

$$s_y^2 = \frac{162,75}{5} - 5,5^2 = 2,3$$

$$s_y = \sqrt{2,3} = 1,52$$

$$s_{xy} = \frac{158,5}{5} - 5,3 \cdot 5,5 = 2,55$$

a) $r = \dfrac{2,55}{1,78 \cdot 1,52} = 0,95$. Al ser positivo y próximo a la unidad, se trata de una correlación fuerte y positiva.

b) $y - 5,5 = \dfrac{2,55}{3,16}(x - 5,3)$ $\qquad y = 0,81x + 1,2$

c) $x - 5,3 = \dfrac{2,55}{2,3}(y - 5,5)$ $\qquad x = 1,11y - 0,81$

d) El punto donde se cortan las dos rectas es el $(\bar{x}, \bar{y})$, es decir: (5,3; 5,5).

12.30. La tabla siguiente expresa el porcentaje de alcohol en sangre de 6 conductores y los segundos que tardan en reaccionar:

% de alcohol	0,08	0,11	0,12	0,14	0,15	0,16
Tiempo de reacción, en segundos	0,38	0,41	0,61	0,44	0,52	0,64

a) ¿Qué tipo de dependencia existe entre estas variables?

b) Estima cuál será el tiempo de reacción cuando el porcentaje de alcohol en sangre sea igual a 0,25.

a) Calculamos el coeficiente de correlación lineal.

Formamos la tabla siguiente:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
0,08	0,38	0,0064	0,1444	0,0304
0,11	0,41	0,0121	0,1681	0,0451
0,12	0,61	0,0144	0,3721	0,0732
0,14	0,44	0,0196	0,1936	0,0616
0,15	0,52	0,0225	0,2704	0,078
0,16	0,64	0,0256	0,4096	0,1024
0,76	3	0,1006	1,5582	0,3907

$\bar{x} = \frac{0,76}{6} = 0,127$ $\qquad \bar{y} = \frac{3}{6} = 0,5$

$s_x^2 = \frac{0,1006}{6} - 0,127^2 = 0,0009$ $\qquad s_x = \sqrt{0,0009} = 0,03$

$s_y^2 = \frac{1,5582}{6} - 0,5^2 = 0,0097$ $\qquad s_y = \sqrt{0,0097} = 0,098$

$s_{xy} = \frac{0,3907}{6} - 0,127 \cdot 0,5 = 0,0016$ $\qquad r = \frac{0,0016}{0,03 \cdot 0,098} = 0,54$

Por tanto, la dependencia es positiva y débil.

b) Recta de regresión de Y sobre X: $y - 0,5 = \frac{0,0016}{0,0009}(x - 0,127)$ $\qquad y = 1,78x + 0,27$

Para $x = 0,25 \Rightarrow y = 1,78 \cdot 0,25 + 0,27 = 0,715$ Luego el tiempo de reacción en segundos es 0,715.

12.31. La tabla siguiente muestra la altitud en metros y la temperatura en grados centígrados a medida que se asciende en una montaña.

Altitud (m)	1000	1100	1200	1300	1400	1500
Temperatura (°C)	12,5	11	10	9,8	8,5	8

a) ¿Qué tipo de dependencia existe entre estas variables?

b) Estima a qué altitud se alcanzarán los cero grados.

a) Calculamos el coeficiente de correlación lineal.

Formamos la tabla siguiente:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1000	12,5	1 000 000	156,25	12 500
1100	11	1 210 000	121	12 100
1200	10	1 440 000	100	12 000
1300	9,8	1 690 000	96,04	12 740
1400	8,5	1 960 000	72,25	11 900
1500	8	2 250 000	64	12 000
7500	59,8	9 550 000	609,54	73240

$\bar{x} = \frac{7500}{6} = 1250$ $\qquad \bar{y} = \frac{59,8}{6} = 9,967$

$s_x^2 = \frac{9\,550\,000}{6} - 1250^2 = 29166,7$

$s_x = \sqrt{29\,166,7} = 170,78$

$s_y^2 = \frac{609,54}{6} - 9,967^2 = 2,25$

$s_y = \sqrt{2,25} = 1,5$

$s_{xy} = \frac{73\,240}{6} - 1250 \cdot 9,967 = -252,08$

$r = \frac{-252,08}{170,78 \cdot 1,5} = -0,98$

Por tanto, la dependencia es negativa y fuerte.

b) Recta de regresión de X sobre Y: $x - 1250 = \frac{-252,08}{2,25}(y - 9,967)$ $\qquad x = 112,04x + 2366,66$

Para $y = 0 \Rightarrow x = 2366,66$. La altitud estimada es de 2366,66 metros.

Luego el tiempo de reacción en segundos es 0,715.

12.32. (PAU) Se midieron los valores de concentración de una sustancia A en suero fetal y los valores de su concentración en suero materno. Se obtuvieron los siguientes datos en una muestra de 6 embarazadas al final de la gestación.

Concentración suero madre (X)	8	4	12	2	7	9
Concentración suero feto (Y)	6	4	8	1	4	5

a) Calcula el coeficiente de correlación lineal.

b) Halla la expresión de la recta que permita estimar los valores fetales a partir de los maternos.

a) Calculamos el coeficiente de correlación lineal.

Formamos la tabla siguiente:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	6	64	36	48
4	4	16	16	16
12	8	144	64	96
2	1	4	1	2
7	4	49	16	28
9	5	81	25	45
42	28	358	158	235

$$\bar{x} = \frac{42}{6} = 7$$

$$\bar{y} = \frac{28}{6} = 4,67$$

$$s_x^2 = \frac{358}{6} - 7^2 = 10,67$$

$$s_x = \sqrt{10,67} = 3,27$$

$$s_y^2 = \frac{158}{6} - 4,67^2 = 4,52$$

$$s_y = \sqrt{2,25} = 2,13$$

$$s_{xy} = \frac{235}{6} - 7 \cdot 4,67 = 6,48$$

$$r = \frac{6,48}{3,27 \cdot 2,13} = 0,93$$

b) Recta de regresión de Y sobre X: $y - 4,67 = \dfrac{6,48}{10,67}(x - 7)$ $\qquad y = 0,607x + 3,41$

12.33. La tabla siguiente expresa los gastos en electricidad y los ingresos mensuales de 6 familias, en euros.

Gastos en electricidad	20	30	50	90	100	190
Ingresos	400	600	800	950	1200	2000

¿Qué gasto en electricidad se estima que tendrá una familia que percibe unos ingresos totales de 2500 euros?

Calculamos la recta de regresión de x (gastos en electricidad) sobre y (ingresos).

Formamos la tabla siguiente:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
20	400	400	160 000	8 000
30	600	900	360 000	18 000
50	800	2500	640 000	40 000
90	950	8100	902 500	85 500
100	1200	10 000	1 440 000	120 000
150	2000	22 500	4 000 000	300 000
440	5950	44 400	7 502 500	571 500

$$\bar{x} = \frac{440}{6} = 73,33$$

$$\bar{y} = \frac{5950}{6} = 991,67$$

$$s_y^2 = \frac{7\,502\,500}{6} - 991,67^2 = 267027,11$$

$$s_{xy} = \frac{571\,500}{6} - 73,33 \cdot 991,67 = 22\,530,84$$

$$x - 73,33 = \frac{22\,530,84}{267\,027,11}(y - 991,67)$$

$$x = 0,084y - 10,34$$

Para $y = 2500$ $\qquad x = 0,084 \cdot 2500 - 10,34 = 199,66$

Para una familia con unos ingresos mensuales de 2500 euros se espera un gasto en electricidad de 199,66 euros.

12.34. Sea la variable bidimensional dada por la tabla siguiente:

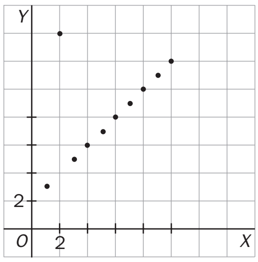
X	1	2	3	4	5	6	7	8	9	10
Y	3	14	5	6	7	8	9	10	11	12

a) Representa la nube de puntos.

b) Halla la recta de Tukey.

c) Halla la recta de regresión de Y sobre X.

a)



b) Formamos con los datos ordenados tres grupos:

$G_1 = \{(1, 3), (2, 14), (3, 5)\}$

$G_2 = \{(4, 6), (5, 7), (6, 8), (7, 9)\}$

$G_3 = \{(8, 10), (9, 11), (10, 12)\}$

Para cada grupo G_i hallamos el punto $P_i(x_i, y_i)$:

$P_1(2, 5)$

$P_2(5{,}5;\ 7{,}5)$

$P_3(9, 11)$

El baricentro del triángulo de vértices $P_1 P_2 P_3$ tiene por coordenadas:

$$x_G = \frac{2 + 5{,}5 + 9}{3} = 5{,}5$$

$$y_G = \frac{5 + 7{,}5 + 11}{3} = 7{,}83$$

La pendiente $P_1 P_3$ es: $m = \frac{11 - 5}{9 - 2} = \frac{6}{7} = 0{,}857$.

La recta de Tukey es: $y - 7{,}83 = 0{,}857\,(x - 5{,}5) \Rightarrow y = 0{,}857x + 3{,}12$

c) Formamos la tabla:

x_i	y_i	x_i^2	$x_i y_i$
1	3	1	3
2	14	4	28
3	5	9	15
4	6	16	24
5	7	25	35
6	8	36	48
7	9	49	63
8	10	64	80
9	11	81	99
10	12	100	120
55	85	385	515

$$\bar{x} = \frac{55}{10} = 5{,}5$$

$$\bar{y} = \frac{85}{10} = 8{,}5$$

$$s_X^2 = \frac{385}{10} - 5{,}5^2 = 8{,}25$$

$$s_{XY} = \frac{515}{10} - 5{,}5 \cdot 8{,}25 = 6{,}125$$

La recta de regresión de Y sobre X es:

$$y - 8{,}5 = \frac{6{,}125}{8{,}25}(x - 5{,}5) \Rightarrow y = 0{,}74x + 4{,}42$$

12.35. Dada la variable bidimensional cuyos datos se recogen en la siguiente tabla:

X	2	4	6	8	10	12	14	16	18	20
Y	3	25	26	20	11	23	24	27	28	24

a) Calcula la recta de Tukey.

b) Halla la recta de regresión de *Y* sobre *X*.

a) Formamos con los datos ordenados tres grupos:

$G_1 = \{(2, 23), (4, 25), (6, 26)\}$ $G_2 = \{(8, 20), (10, 11), (12, 23), (14, 24)\}$ $G_3 = \{(16, 27), (18, 28), (20, 24)\}$

Para cada grupo G_i hallamos el punto $P_i(x_i, y_i)$:

$P_1(4, 25)$

$P_2(11; 21{,}5)$

$P_3(18, 27)$

El baricentro del triángulo de vértices $P_1P_2P_3$ tiene por coordenadas:

$x_G = \dfrac{4 + 11 + 18}{3} = 7{,}67$ $y_G = \dfrac{25 + 21{,}5 + 27}{3} = 24{,}5$

La pendiente P_1P_3 es: $m = \dfrac{27 - 25}{18 - 4} = \dfrac{2}{14} = 0{,}14$.

La recta de Tukey es: $y - 24{,}5 = 0{,}14(x - 7{,}67) \Rightarrow y = 0{,}14x + 23{,}43$

b) Formamos la tabla:

x_i	y_i	x_i^2	$x_i y_i$
2	23	4	46
4	25	16	100
6	26	36	156
8	20	64	160
10	11	100	110
12	23	144	276
14	24	196	336
16	27	256	432
18	28	324	504
20	24	400	480
110	231	1540	2600

$\bar{x} = \dfrac{110}{10} = 11$ $\bar{y} = \dfrac{231}{10} = 23{,}1$

$s_X^2 = \dfrac{1540}{10} - 11^2 = 33$ $s_{XY} = \dfrac{2600}{10} - 11 \cdot 23{,}1 = 5{,}9$

La recta de regresión de *Y* sobre *X* es:

$$y - 23{,}1 = \frac{5{,}9}{33}(x - 11) \Rightarrow y = 0{,}18x + 21{,}13$$

12.36. Sea la variable bidimensional dada por la tabla siguiente:

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

a) Halla la recta de Tukey.

b) Calcula la recta de regresión de *Y* sobre *X*.

a) Formamos con los datos ordenados tres grupos:

$G_1 = \{(1, 7), (2, 6), (3, 6)\}$ $G_2 = \{(4, 8), (5, 1), (6, 10), (7, 9)\}$ $G_3 = \{(8, 7), (9, 6), (10, 11)\}$

Para cada grupo G_i hallamos el punto $P_i(x_i, y_i)$:

$P_1(2, 6)$

$P_2(5{,}5; 8)$

$P_3(9, 7)$

El baricentro del triángulo de vértices $P_1P_2P_3$ tiene por coordenadas:

$x_G = \dfrac{2 + 5{,}5 + 9}{3} = 5{,}5$

$y_G = \dfrac{6 + 8 + 7}{3} = 7$

La pendiente P_1P_3 es: $m = \dfrac{7 - 6}{9 - 2} = \dfrac{1}{7} = 0{,}14$.

La recta de Tukey es: $y - 7 = 0{,}14\,(x - 5{,}5) \Rightarrow y = 0{,}14x + 6{,}23$

b) Formamos la tabla:

x_i	y_i	x_i^2	$x_i y_i$
1	7	1	7
2	6	4	12
3	6	9	18
4	8	16	32
5	1	25	5
6	10	36	60
7	9	49	63
8	7	64	56
9	6	81	54
10	11	100	110
55	71	385	417

$\bar{x} = \frac{55}{10} = 5{,}5$ $\qquad \bar{y} = \frac{71}{10} = 7{,}1$

$s_X^2 = \frac{385}{10} - 5{,}5^2 = 8{,}25$ $\qquad s_{XY} = \frac{417}{10} - 5{,}5 \cdot 7{,}1 = 2{,}65$

La recta de regresión de Y sobre X es:

$$y - 7{,}1 = \frac{2{,}65}{8{,}25}(x - 5{,}5) \Rightarrow y = 0{,}32x + 5{,}33$$

PROBLEMAS

12.37. Una planta envasadora de frutos secos necesita adquirir una máquina empaquetadora de bolsas de 50 gramos lo más precisa posible, para lo que efectúa una prueba de 10 pesadas con cada una de las máquinas X e Y, obteniendo los siguientes resultados en gramos:

X	52	54	53	47	48	49	46	48	51	52
Y	51	54	41	46	49	49	48	49	51	52

a) Calcula la media y la desviación típica de cada una de las distribuciones X e Y. ¿Qué máquina se debe elegir y por qué?

b) Calcula la recta de regresión de Y sobre X. ¿Qué pesada se espera de la máquina Y en una nueva prueba si se sabe que X ha dado 54 gramos?

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
52	51	2704	2601	2652
54	54	2916	2916	2916
53	51	2809	2601	2703
47	46	2209	2116	2162
48	49	2304	2401	2352
49	49	2401	2401	2401
46	48	2116	2304	2208
48	49	2304	2401	2352
51	51	2601	2601	2601
52	52	2704	2704	2704
195	68	25 068	25 046	25 051

a) $\bar{x} = \frac{500}{10} = 50;$ $\qquad \bar{y} = \frac{500}{10} = 50$

$s_x^2 = \frac{25\,068}{10} - 50^2 = 6{,}8$ $\qquad s_x = \sqrt{6{,}8} = 2{,}61$

$s_y^2 = \frac{25\,046}{10} - 50^2 = 4{,}6$ $\qquad s_y = \sqrt{4{,}6} = 2{,}14$

Las dos distribuciones tienen igual media, pero la desviación típica de la máquina Y es más pequeña que la de la máquina X. Se debe elegir la máquina Y, ya que las pesadas están más concentradas alrededor de la media.

b) $s_{xy} = \frac{25\,051}{10} - 50 \cdot 50 = 5{,}1$

Recta de regresión de Y sobre X

$$y - 50 = \frac{5{,}1}{6{,}8}(x - 50) \quad \Rightarrow \quad y = 0{,}75x + 12{,}5$$

Para $x = 54$ $\quad y = 0{,}55 \cdot 54 + 12{,}5 = 53$ gramos

12.38. (PAU) A partir de los datos recogidos sobre la estatura (E) y el peso (P) en un grupo de 50 estudiantes se ha obtenido una estatura media de 165 cm y un peso medio de 61 kg.

Sabiendo que al aumentar la estatura aumenta también el peso, identifica, entre las siguientes, cuál podría ser la recta de regresión del peso en función de la estatura obtenida a través de los datos recogidos en ese grupo de estudiantes.

a) $P = 226 - E$ $\qquad$ b) $P = -104 + E$ $\qquad$ c) $P = 5 + \frac{1}{3}E$ $\qquad$ d) $P = 171 - \frac{2}{3}E$

La recta de regresión debe ser de pendiente positiva, ya que al aumentar la estatura, aumenta el peso. Por tanto, estudiaremos si las rectas b y c pasan por (165, 61).

b) $P = -104 + 165 = 61$. Cumple la condición.

c) $P = 5 + \frac{165}{3} = 60$. No cumple la condición.

La recta pedida podría ser $P = -104 + E$.

12.39. El número de horas dedicadas al estudio de una prueba y las respuestas correctas obtenidas en un test de 100 preguntas vienen en la siguiente tabla.

X: horas de estudio	20	16	34	23	27	32	18	22
Y: aciertos	65	60	85	70	90	95	75	80

a) Halla la recta de regresión de *Y* sobre *X*.

b) Calcula la calificación estimada para una persona que hubiese estudiado 28 horas.

Formamos la siguiente tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
20	65	400	4225	1300
16	60	256	3600	960
34	85	1156	7225	2890
23	70	529	4900	1610
27	90	729	8100	2430
32	95	1024	9025	3040
18	75	324	5625	1350
22	80	484	6400	1760
192	620	4902	49 100	15 340

a) $\bar{x} = \dfrac{192}{8} = 24;$

$\bar{y} = \dfrac{620}{8} = 77{,}5$

$s_x^2 = \dfrac{4902}{8} - 24^2 = 36{,}75$

$s_{xy} = \dfrac{15\,340}{8} - 24 \cdot 77{,}5 = 57{,}5$

$y - 77{,}5 = \dfrac{57{,}5}{36{,}75}(x - 24) \Rightarrow y = 1{,}56x + 39{,}95$

b) Si $x = 28$, $y = 1{,}56 \cdot 28 + 39{,}95 = 83{,}63$. Por tanto, si un alumno dedica al estudio 28 horas, se espera que responda correctamente a 84 preguntas.

12.40. En la siguiente tabla se consideran las puntuaciones en dos pruebas (*X*, *Y*) de 5 alumnos.

X	1	6	9	3	2
Y	2	3	9	6	1

a) Encuentra las ecuaciones de las rectas de regresión de *X* sobre *Y* y de *Y* sobre *X*.

b) Con los resultados obtenidos en el apartado anterior, determina el coeficiente de correlación.

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	2	1	4	2
6	3	36	9	18
9	9	81	81	81
3	6	9	36	18
2	1	4	1	2
21	21	131	131	121

a) $\bar{x} = \dfrac{21}{5} = 4{,}2$

$\bar{y} = \dfrac{21}{5} = 4{,}2$

$s_x^2 = \dfrac{131}{5} - 4{,}2^2 = 8{,}56$

$s_y^2 = \dfrac{131}{5} - 4{,}2^2 = 8{,}56$

$s_{xy} = \dfrac{121}{5} - 4{,}2^2 = 6{,}56$

Recta de regresión de *Y* sobre *X*: $y - 4{,}2 = \dfrac{6{,}56}{8{,}56}(x - 4{,}2)$; $y = 0{,}767x + 0{,}98$

Recta de regresión de *X* sobre *Y*: $x - 4{,}2 = \dfrac{6{,}56}{8{,}56}(y - 4{,}2)$; $x = 0{,}767y + 0{,}98$

b) El coeficiente de correlación lineal es igual a $r = \sqrt{0{,}767 \cdot 0{,}767} = 0{,}767$.

12.41. **(PAU) La siguiente tabla relaciona la inversión, en millones, y la rentabilidad obtenida, en tanto por ciento, de 6 inversores.**

Inversión	10	12	14	14	15	15
Rentabilidad (%)	4	4	5	4	5	5

a) Calcula la media y la desviación típica de las variables inversión y rentabilidad.

b) Halla el coeficiente de correlación e interprétalo.

c) Si un inversionista invierte 13,5 millones, ¿qué rentabilidad puede esperar?

d) Si un inversionista ha obtenido una rentabilidad del 5,5%, ¿qué capital se puede esperar que haya invertido?

Consideramos la inversión como variable X y la rentabilidad como variable Y.

Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
10	4	100	16	40
12	4	144	16	48
14	5	196	25	70
14	4	196	16	56
15	5	225	25	75
15	5	225	25	75
80	27	1086	123	364

a) $\bar{x} = \frac{80}{6} = 13{,}33$

$s_X^2 = \frac{1086}{6} - 13{,}33^2 = 3{,}31$

$s_X = \sqrt{3{,}31} = 1{,}82$

$\bar{y} = \frac{27}{6} = 4{,}5$

$s_Y^2 = \frac{123}{6} - 4{,}5^2 = 0{,}25$

$s_Y = \sqrt{0{,}25} = 0{,}5$

b) $s_{xy} = \frac{364}{6} - 13{,}33 \cdot 4{,}5 = 0{,}68$ $\qquad r = \frac{0{,}68}{1{,}82 \cdot 0{,}5} = 0{,}74$

Como el valor de r es próximo a 1, la correlación es directa y moderadamente fuerte. Por tanto, las variables están en dependencia aleatoria.

c) Hallamos la recta de regresión de Y sobre X: $y - 4{,}5 = \frac{0{,}68}{3{,}31}(x - 13{,}33)$ $\quad y = 0{,}21x + 1{,}76$

Por tanto, para x = 13,5 se obtiene: $y = 0{,}21 \cdot 13{,}5 + 1{,}76 = 4{,}59$. Así pues, si un inversionista invierte 13,5 millones, se espera que obtenga una rentabilidad del 4,57%.

d) Hallamos la recta de regresión de X sobre Y: $x - 13{,}33 = \frac{0{,}68}{0{,}25}(y - 4{,}5)$ $\quad y = 2{,}72x + 1{,}09$

Por tanto, para $y = 5{,}5$ se obtiene: $x = 2{,}72 \cdot 5{,}5 + 1{,}09 = 16{,}05$. Así pues, si un inversionista obtiene una rentabilidad del 5,5%, se supone que había invertido 16,05 millones.

12.42. **En una población, la media de los pesos es de 70 kg, y la de las estaturas, de 175 cm. Las desviaciones típicas son 5 kg y 10 cm, respectivamente, y la covarianza de ambas variables es 40.**

a) Estima el peso de una persona de esa población que mide 185 cm de estatura.

b) Usando el coeficiente de correlación lineal, explica hasta qué punto confiarías en la estimación que se ha hecho en el apartado a.

Si X es la variable peso e Y la variable estatura, el enunciado nos da los siguientes datos:

$$\bar{x} = 70 \text{ kg} \quad s_X = 5 \text{ kg} \quad \bar{y} = 175 \text{ cm} \quad s_Y = 10 \text{ cm} \quad s_{xy} = 40$$

a) Hay que hallar la recta de regresión de X sobre Y.

$$x - 70 = \frac{40}{100}(y - 175) \qquad x = 0{,}4y$$

Para una estatura de $y = 185$ cm, el peso esperado es $x = 0{,}4 \cdot 185 = 74$ kg.

b) El coeficiente de correlación es $r = \frac{40}{5 \cdot 10} = 0{,}8$.

Este valor de r indica que la correlación es directa y fuerte; por tanto, existe una alta confianza en las estimaciones obtenidas.

12.43. Las rectas de regresión de cuatro distribuciones bidimensionales son las siguientes:

a) $y = x + 2; \quad x = 4$

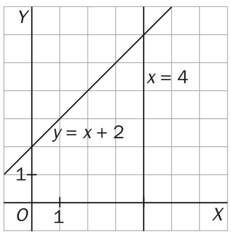
b) $y = \frac{4}{5}x + 2 \quad x = \frac{5}{6}y + 2$

c) $y = 3; \quad x = 2$

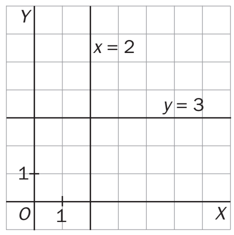
d) $y = x; \quad x = \frac{4}{5}y + 1$

Indica en qué casos es significativa la correlación lineal.

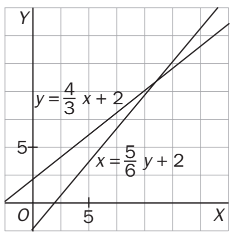
a)



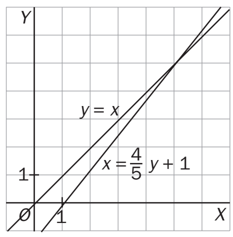
c)



b)



d)



El ángulo formado por las rectas es más pequeño en d y b. Por tanto, en esos casos es más significativa la correlación.

12.44. A partir de los datos recogidos sobre facturación y beneficios en un determinado año sobre un conjunto de 50 grandes empresas europeas se ha calculado una facturación media de 80 millones de euros y unos beneficios medios de 65 millones de euros.

a) Teniendo en cuenta esa información, determina la recta de regresión que permite obtener los beneficios en función de la facturación, sabiendo que a partir de ella se han calculado unos beneficios de 59 millones de euros para una empresa que ha facturado 75 millones de euros en 1998.

b) ¿Qué signo tendría el coeficiente de correlación lineal entre ambas variables?

a) Consideramos X como la variable facturación e Y como la variable beneficio.

La recta de regresión de Y sobre X es de la forma $y = mx + n$.

Como la recta pasa por los puntos (80, 65) y (75, 59), se tiene:

$65 = 80m + n$

$59 = 75m + n$

Resolviendo el sistema, resulta: $m = \frac{6}{5}$ y $n = -31$.

La recta de regresión de Y sobre X es: $y = \frac{6}{5}x - 31$.

b) El coeficiente de correlación tiene el mismo signo que la pendiente de la recta de regresión; por tanto, es positivo.

12.45. (PAU) La recta de regresión de una variable Y respecto de la variable X es $y = 0{,}3x + 1$. Los valores que ha tomado la variable x han sido {3, 4, 5, 6, 7}.

a) Determina el valor esperado de Y para el valor particular de $x = 3{,}5$.

b) Si los valores de la variable Y utilizados para la regresión se multiplican por 10 y se dejan los mismos valores para la variable X, determina razonadamente la nueva recta de regresión.

a) Para $x = 3{,}5 \Rightarrow y = 0{,}3 \cdot 3{,}5 + 1 = 2{,}05$

b) Si los valores de la variable Y se multiplican por 10, se tendrá:

$\bar{y}' = 10\bar{y}$, siendo $\bar{y}$ la media inicial

$$s'_{xy} = \frac{\sum 10x_iy_i}{N} - \bar{x}\ 10\,\bar{y} = 10\,s_{xy}$$

Con esto, la nueva recta será:

$$y - 10\,\bar{y} = \frac{10\ S_{xy}}{S_x^2}(x - \bar{x}) \Rightarrow y = \frac{10\ S_{xy}}{S_x^2}\,x + 10\left(\bar{y} - \frac{Sxy}{S_x^2}\,\bar{x}\right) \Rightarrow y = 3x + 10$$

12.46. (PAU) Cien alumnos prepararon un examen de Matemáticas. Se representa por X el número de problemas hechos por cada alumno en la preparación, y por Y, la calificación obtenida. Sabiendo que las medias aritméticas de esas variables fueron $\bar{x} = 9{,}2$ e $\bar{y} = 9{,}5$, que el coeficiente de correlación entre esas variables fue 0,7 y que la desviación típica de la variable Y fue el doble que la de la variable X, calcula las ecuaciones de las rectas de regresión.

Como la desviación típica de la variable Y fue el doble que la de la variable X, se tiene:

1) $r = \dfrac{s_{XY}}{s_Xs_Y} = \dfrac{s_{XY}}{s_X \cdot 2s_X} = \dfrac{s_{XY}}{2s_X^2} = 0{,}7 \Rightarrow \dfrac{s_{XY}}{s_X^2} = 0{,}7 \cdot 2 = 1{,}4$

La recta de regresión de Y sobre X es:

$y - 9{,}5 = 1{,}4\ (x - 9{,}2)$.

2) $\dfrac{s_{XY}}{s_Y^2} = \dfrac{s_{XY}}{(2s_X)^2} = \dfrac{s_{XY}}{4s_X^2} = \dfrac{1}{4} \cdot 1{,}4 = 0{,}35$

La recta de regresión de X sobre Y es:

$x - 9{,}2 = 0{,}35\ (y - 9{,}5)$.

PROFUNDIZACIÓN

12.47. (PAU) La tabla siguiente muestra los valores observados de dos variables X e Y en 5 individuos.

X	1	−1	x	2	3
Y	−2	−3	2	1	0

a) Halla el valor x para que el coeficiente de correlación sea nulo.

b) Suponiendo que $x = 4$, halla la recta de regresión de Y sobre X y estudia el valor de Y cuando X toma el valor −2.

a) $\bar{x} = \dfrac{5 + x}{5} \quad \bar{y} = -\dfrac{2}{5} \quad \Rightarrow \quad s_{xy} = \dfrac{-2 + 3 + 2x + 2}{5} - \dfrac{5 + x}{5} \cdot \left(\dfrac{-2}{5}\right) = \dfrac{12x + 25}{25}$

Como $r = \dfrac{s_{xy}}{s_xs_y} = 0$, $s_{xy} = 0$; $\dfrac{12x + 25}{25} = 0$, y, por tanto, $x = -\dfrac{25}{12}$

b) Si $x = 4$, $\bar{x} = \dfrac{9}{5}$; $\bar{y} = -\dfrac{2}{5}$; $s_{xy} = \dfrac{73}{25}$

$$s_x^2 = \frac{1 + 1 + 16 + 4 + 9}{5} - \frac{81}{25} = \frac{74}{25}$$

Recta de regresión de Y sobre X: $y - \dfrac{2}{5} = \dfrac{73}{74}\left(x - \dfrac{9}{5}\right)$; $y = 0{,}986x - 2{,}176$. Si $x = -2$, $y = -4{,}148$

12.48. Se considera la siguiente tabla estadística, donde a es una incógnita:

X	2	4	*a*	3	5
Y	1	2	1	1	3

a) Calcula el valor de a sabiendo que la media de X es 3.

b) Mediante la correspondiente recta de regresión lineal, predice el valor que se obtiene para Y cuando $X = 4{,}5$. Explica la fiabilidad de la predicción anterior.

a) $x = \dfrac{2 + 4 + a + 3 + 5}{5} = 3 \Rightarrow 14 + a = 15 \Rightarrow a = 1$

b) Formamos la tabla:

x_i	y_i	x_i^2	y_i^2	$x_i\,y_i$
2	1	4	1	2
4	2	16	4	8
1	1	1	1	1
3	1	9	1	3
5	3	25	9	15
15	8	55	16	29

$\bar{x} = \dfrac{15}{5} = 3 \qquad s_x^2 = \dfrac{55}{5} - 9^2 = 2 \qquad s_X = \sqrt{2} = 1{,}41$

$\bar{y} = \dfrac{8}{5} = 1{,}6 \qquad s_y^2 = \dfrac{16}{5} - 1{,}6^2 = 0{,}64 \qquad s_X = \sqrt{0{,}64} = 0{,}8$

$s_{xy} = \dfrac{29}{5} - 3 \cdot 1{,}6 = 1$

$r = \dfrac{1}{1{,}41 \cdot 0{,}8} = 0{,}89$

$y - 1{,}6 = 0{,}5(x - 3)$; $\quad y = 0{,}5\,x + 0{,}1$. Si $x = 4{,}5$, $y = 0{,}5 \cdot 4{,}5 = 2{,}35$

Como r es próximo a la unidad, permite obtener conclusiones muy fiables del comportamiento de Y estudiando la variable X.

12.49. Los siguientes pares de datos corresponden a las variables X (producto interior bruto en decenas de millones de euros) e Y (tasa de inflación):

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

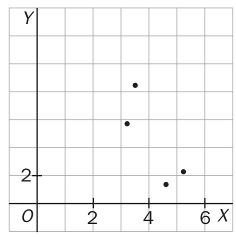
a) Dibuja el diagrama de dispersión de los datos.

b) Decide razonadamente cuál de las siguientes rectas es la de regresión de Y sobre X:

$$y = 16{,}26 + 2{,}88x \qquad y = 16{,}26 - 2{,}88x$$

c) Calcula el valor esperado de la tasa de inflación que corresponde a un producto interior bruto de 4,3 decenas de millones de euros.

a)



b) De ser alguna de esas dos rectas, será la de la pendiente negativa, $y = 16{,}26 - 2{,}88x$, pues así lo sugiere el diagrama de dispersión.

c) Si $x = 4{,}3$, sustituyendo en la recta dada se obtiene: $y = -2{,}88 \cdot 4{,}3 + 16{,}26 = 3{,}88$.

Es decir, para un producto interior bruto de 4,3 decenas de millones de euros se espera una tasa de inflación del 3,88.

12.50. Se ha solicitado a un grupo de 50 individuos información sobre el número de horas que dedican diariamente a dormir y a ver la televisión. Los resultados vienen dados por la siguiente tabla.

X: N.º de horas dormidas	6	7	8	9	10
Y: N.º de horas viendo la televisión	4	3	3	2	1
Frecuencias absolutas	3	16	20	10	1

a) Calcula el coeficiente de correlación entre *X* e *Y* e interprétalo en los términos del enunciado.

b) Calcula la ecuación de la recta de regresión de *Y* sobre *X*.

c) Si una persona duerme 8 horas y media, ¿cuántas horas cabe esperar que vea la televisión?

d) Sin calcular la recta de regresión de *X* sobre *Y*, ¿en qué punto se cortará esta recta con la calculada en el apartado b?

e) Si una persona ve la televisión 2 horas, ¿cuánto tiempo cabe esperar que duerma?

Formamos la siguiente tabla:

x_i	y_i	f_i	x_i^2	y_i^2	$f_i x_i$	$f_i y_i$	$f_i x_i^2$	$f_i y_i^2$	$f_i x_i y_i$
6	4	3	36	16	18	12	108	48	72
7	3	16	49	9	112	48	784	144	336
8	3	20	64	9	160	60	1280	180	480
9	2	10	81	4	90	20	810	40	180
10	1	1	100	1	10	1	100	1	10
		50			390	141	3082	413	1078

a) $\bar{x} = \dfrac{390}{50} = 7{,}8;$

$$s_x^2 = \frac{3082}{50} - (7{,}8)^2 = 0{,}80$$

$$s_x = \sqrt{0{,}8} = 0{,}89$$

$$\bar{y} = \frac{141}{50} = 2{,}82$$

$$s_y^2 = \frac{413}{50} - (2{,}82)^2 = 0{,}3076$$

$$\text{sY} = \sqrt{0{,}3076} = 0{,}55$$

$$s_{xy} = \frac{1078}{50} - (7{,}8)\ (2{,}82) = -0{,}436$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{-0{,}436}{0{,}89 \cdot 0{,}55} = -0{,}88$$

La correlación lineal entre ambas variables es grande e inversa.

b) $y - 2{,}82 = -\dfrac{0{,}436}{0{,}80}\ (x - 7{,}8) \quad \Rightarrow \quad y = -0{,}545x + 7{,}071$

c) Si $x = 8{,}5$: $y = -0{,}545 \cdot 8{,}5 + 7{,}071 = 2{,}44$

Si una persona duerme 8 horas y media, verá la televisión durante 2 horas 26,4 minutos.

d) La recta de regresión de *Y* sobre *X* y la recta de regresión de *X* sobre *Y* se cortan en el centro de gravedad de la nube de puntos, es decir, en el punto $(\bar{x}, \bar{y}) = (7{,}8;\ 2{,}82)$.

e) La recta de regresión de *X* sobre *Y* es: $x - 7{,}8 = \dfrac{-0{,}436}{0{,}3076}\ (y - 2{,}82)$; $\quad x = 11{,}787 - 1{,}417y.$

Si $y = 2$, $x = 11{,}787 - 1{,}41 \cdot 2 = 8{,}953$ horas, es decir, que si una persona ve la televisión durante 2 horas, se espera que duerma aproximadamente 9 horas.

12.51. Los valores de dos variables X e Y se distribuyen según la tabla siguiente.

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

Determina el coeficiente de correlación y la recta de regresión de Y sobre X.

Convertimos la tabla de doble entrada en tabla simple y efectuamos los siguientes cálculos:

x_i	y_i	f_i	$x_i f_i$	$x_i^2 f_i$	$y_i f_i$	$y_i^2 f_i$	$x_i y_i f_i$
0	1	2	0	0	2	2	0
0	2	1	0	0	2	4	0
0	3	2	0	0	6	18	0
2	1	1	2	4	1	1	2
2	2	4	8	16	8	16	16
2	3	5	10	20	15	45	30
4	1	3	12	48	3	3	12
4	2	2	8	32	4	8	16
		20	40	120	41	97	76

$\bar{x} = \dfrac{40}{20} = 2$ $\qquad s_x^2 = \dfrac{120}{20} - 2^2 = 2$ $\qquad s_x = \sqrt{2} = 1{,}41$

$\bar{y} = \dfrac{41}{20} = 2{,}05$ $\qquad s_y^2 = \dfrac{97}{20} - 2{,}05^2 = 0{,}65$ $\qquad s_y = \sqrt{0{,}65} = 0{,}81$

$$s_{xy} = \frac{76}{20} - 2 \cdot 2{,}05 = -0{,}3$$

$$r = \frac{-0{,}3}{1{,}41 \cdot 0{,}81} = -0{,}26$$

Recta de regresión de Y sobre X:

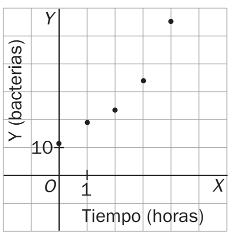
$$y - 2{,}05 = -0{,}15\,(x - 2);\ y = -0{,}15x + 2{,}35$$

12.52. El número de bacterias por unidad de volumen presentes en un cultivo después de cierto número de horas viene expresado por la siguiente tabla.

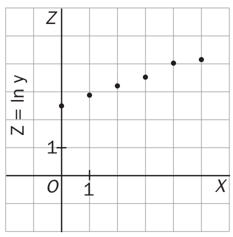
X: horas	0	1	2	3	4	5
Y: bacterias	12	19	23	24	56	62

¿Cuántas bacterias habrá al cabo de seis horas? Ayuda: realiza el cambio de variable $Z = \ln Y$

Dibujamos el diagrama de dispersión y observamos que existe una relación curvilínea.



X	Z = ln Y
0	2,48
1	2,94
2	3,14
3	3,53
4	4,03
5	4,13



Al realizar el cambio de variable $Z = \ln Y$, la nube de puntos se ajusta a una recta. Podemos calcular ahora la recta de regresión de Z sobre X.

x_i	z_i	x_i^2	z_i^2	$x_i z_i$
0	2,48	0	6,1504	0
1	2,94	1	8,6436	2,94
2	3,14	4	9,8596	6,28
3	3,53	9	12,4609	10,59
4	4,03	16	16,2409	16,12
5	4,13	25	17,0569	20,65
15	20,25	55	70,4123	56,58

$\bar{x} = \dfrac{15}{6} = 2{,}5; \qquad \bar{z} = \dfrac{20{,}25}{5} = 3{,}375$

$s_x^2 = \dfrac{55}{6} - 2{,}5^2 = 2{,}92 \Rightarrow s_x = \sqrt{2{,}92} = 1{,}71$

$s_z^2 = \dfrac{70{,}4123}{6} - 3{,}375^2 = 0{,}35 \Rightarrow s_z = \sqrt{0{,}35} = 0{,}59$

$s_{xz} = \dfrac{56{,}58}{6} - 2{,}5 \cdot 3{,}375 = 0{,}9925$

Como $r = \dfrac{0{,}9925}{1{,}71 \cdot 0{,}59} = 0{,}98$, nos indica que la relación existente entre X e Y es de tipo exponencial.

La recta de regresión de Z sobre X es: $z - 3{,}375 = \dfrac{0{,}9925}{2{,}92}(x - 2{,}5) \Rightarrow z = 0{,}34x + 2{,}53$

Como $Z = \ln Y$, se cumple que: $\ln y = 0{,}34x + 2{,}53$.

Tomando exponenciales en los dos miembros queda: $y = e^{0{,}34x + 2{,}53}$

El número de bacterias esperado al cabo de seis horas es: $y = e^{0{,}34 \cdot 6 + 2{,}53} = 96{,}54$