

## REFLEXIONA Y RESUELVE

## Relación funcional y relación estadística

En cada uno de los siguientes casos debes decir si, entre las dos variables que se citan, hay relación funcional o relación estadística (correlación) y, en este último caso, indicar si es positiva o negativa:

- En un conjunto de familias: *estatura media de los padres – estatura media de los hijos.*

Correlación positiva.

- *Temperatura a la que calentamos una barra de hierro – longitud alcanzada.*

Funcional.

- Entre los países del mundo respecto a España: *volumen de exportación – volumen de importación.*

Correlación negativa.

- Entre los países del mundo: *índice de mortalidad infantil – número de médicos por cada 1 000 habitantes.*

Correlación negativa.

- En las viviendas de una ciudad: *kWh consumidos durante enero – coste del recibo de la luz.*

Funcional.

- *Número de personas que viven en cada casa – coste del recibo de la luz.*

Correlación positiva.

- Equipos de fútbol: *lugar que ocupan al finalizar la liga – número de partidos perdidos.*

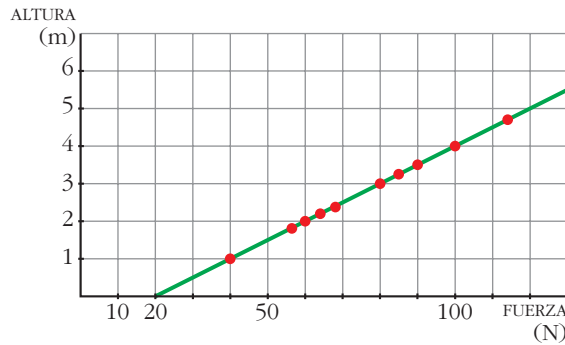
Correlación positiva.

- Equipos de fútbol: *lugar que ocupan al finalizar la liga – número de partidos ganados.*

Correlación negativa.

## Ejemplo de relación funcional

Distintas personas lanzan hacia arriba una misma piedra de 2 kg de masa, que alcanza más o menos altura según la fuerza con que ha sido impulsada. (La fuerza actúa en un tramo de 1 m).



- a) ¿Qué altura, por encima de la mano, alcanzará la piedra si se impulsa con una fuerza de 110 newton?
- b) ¿Podríamos escribir una fórmula que dé directamente la altura que alcanza la piedra, desde el momento en que se la suelta, en función de la fuerza con que es impulsada hacia arriba?

a) 4,5 m

b)  $\text{Altura} = \frac{F}{20} - 1$  para  $F \geq 20$

*Obtención física de la fórmula:*

La fórmula en la que se basa todo el desarrollo posterior es:

$$v = \sqrt{2ad}$$

donde  $v$ : Aumento de la velocidad en el tramo  $d$ .

$a$ : Aceleración constante con la que se mueve el móvil.

$d$ : Espacio que recorre con la aceleración  $a$ .

Así, la velocidad con que sale de la mano es:

$$v_s = \sqrt{2a \cdot 1} = \sqrt{2a}$$

Además:

$$F = m(a + g) \rightarrow a = \frac{F}{m} - g = \frac{F}{2} - 10$$

Luego:

$$v_s = \sqrt{2\left(\frac{F}{2} - 10\right)} = \sqrt{F - 20}$$

Por otra parte, si se deja caer una piedra desde una altura  $h$ , adquiere una velocidad:

$$v_s = \sqrt{2gh}$$

O bien, si se empuja una piedra hacia arriba de modo que salga con una velocidad  $v_s$ , alcanza una altura  $h$ .

En este caso:

$$v_s = \sqrt{2 \cdot 10 \cdot h} = \sqrt{20h}$$

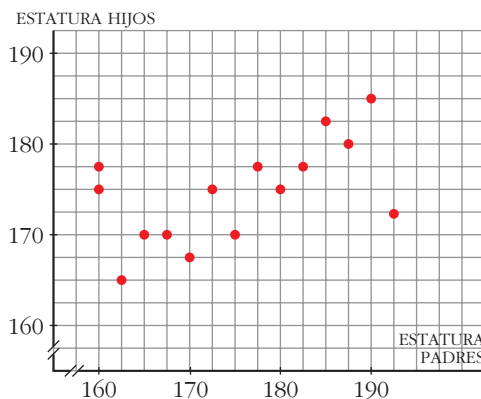
Igualando:

$$\sqrt{F - 20} = \sqrt{20h} \rightarrow h = \frac{F}{20} - 1$$

Para que  $h \geq 0$ , debe ser  $F \geq 20$ .

## Ejemplo de relación estadística

En la siguiente gráfica, cada punto corresponde a un chico. La abscisa es la estatura de su padre, y la ordenada, su propia altura.



- Identifica a Guille y Gabriel, hermanos de buena estatura, cuyo padre es bajito.
- Identifica a Sergio, de estatura normalita, cuyo padre es un gigantón.
- ¿Podemos decir que hay una cierta relación entre las estaturas de estos 15 chicos y las de sus padres?

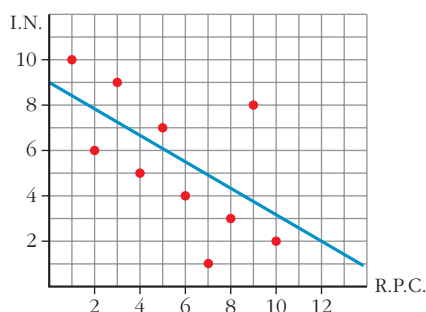
- Guille y Gabriel están representados por los puntos (160, 175) y (160, 177,5)
- Sergio está representado por el punto (192,5; 172,5).
- En general, sí.

## Página 333

1. La tabla de la derecha muestra cómo se ordenan entre sí diez países, A, B, C..., según dos variables, R.P.C. (*renta per cápita*) e I.N. (*índice de natalidad*). Representa los resultados en una nube de puntos, traza la recta de regresión y di cómo te parece la correlación.

PAÍSES	A	B	C	D	E	F	G	H	I	J
R.P.C.	1	2	3	4	5	6	7	8	9	10
I.N.	10	6	9	5	7	4	1	3	8	2

La correlación es negativa y moderadamente alta ( $-0,62$ ).



## Página 335

1. Obtén mediante cálculos manuales los coeficientes de correlación de las distribuciones de la página 332:

*Matemáticas – Filosofía*

*Distancia – Número de encestes*

Hazlo también con una calculadora con MODO LR.

*Matemáticas-Filosofía:*

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
2	2	4	4	4
3	5	9	25	15
4	2	16	4	8
4	7	16	49	28
5	5	25	25	25
6	4	36	16	24
6	6	36	36	36
7	6	49	36	42
7	7	49	49	49
8	5	64	25	40
10	5	100	25	50
10	9	100	81	90
72	63	504	375	411

$$\bar{x} = \frac{72}{12} = 6$$

$$\bar{y} = \frac{63}{12} = 5,25$$

$$\sigma_x = \sqrt{\frac{504}{12} - 6^2} = 2,45$$

$$\sigma_y = \sqrt{\frac{375}{12} - 5,25^2} = 1,92$$

$$\sigma_{xy} = \frac{411}{12} - 6 \cdot 5,25 = 2,75$$

$$\text{Por tanto: } r = \frac{2,75}{2,45 \cdot 1,92} = 0,58$$

*Distancia-Número de encestes:*

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$
1	9	1	81	9
2	10	4	100	20
3	6	9	36	18
4	4	16	16	16
5	2	25	4	10
6	0	36	0	0
7	1	49	1	7
8	0	64	0	0
36	32	204	238	80

$$\bar{x} = \frac{36}{8} = 4,5 \quad \bar{y} = \frac{32}{8} = 4$$

$$\sigma_x = \sqrt{\frac{204}{8} - 4,5^2} = 2,29$$

$$\sigma_y = \sqrt{\frac{238}{8} - 4^2} = 3,71$$

$$\sigma_{xy} = \frac{80}{8} - 4,5 \cdot 4 = -8$$

$$\text{Por tanto: } r = \frac{-8}{2,29 \cdot 3,71} = -0,94$$

## EJERCICIOS Y PROBLEMAS PROPUESTOS

### PARA PRACTICAR

#### Sin fórmulas

**1** Para cada uno de los siguientes casos indica:

- Cuáles son las variables que se relacionan.
- Si se trata de una relación funcional o de una relación estadística y, en estos casos, el signo de la correlación.

- Renta mensual de una familia-gasto en electricidad.
- Radio de una esfera-volumen de esta.
- Litros de lluvia recogidos en una ciudad-tiempo dedicado a ver la televisión por sus habitantes.
- Longitud del trayecto recorrido en una línea de cercanías-precio del billete.
- Peso de los alumnos de 1º de Bachillerato-número de calzado que usan.
- Toneladas de tomate recogidas en una cosecha-precio del kilo de tomate en el mercado.

a) Renta (€), gasto (€).

Correlación positiva.

b) Relación funcional.

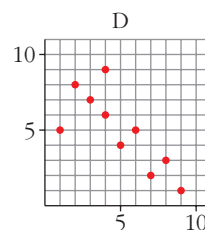
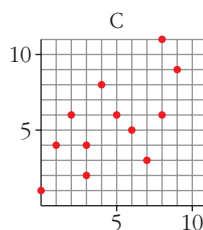
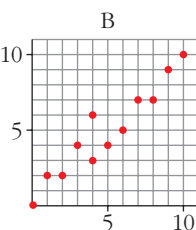
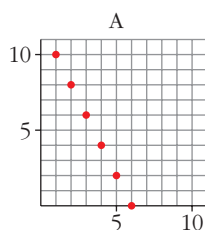
c) Relación estadística. Seguramente muy débil. Positiva (¿cabe pensar que cuanto más llueva más tiempo pasarán en casa y, por tanto, más verán la televisión?).

d) Aunque lo parezca a priori, seguramente la relación no es funcional. Es una correlación positiva fuerte.

e) Correlación positiva.

f) Correlación negativa (cuanto mayor sea la cosecha, más baratos están los tomates).

**2** a) Traza, a ojo, la recta de regresión en cada una de estas distribuciones bidimensionales:

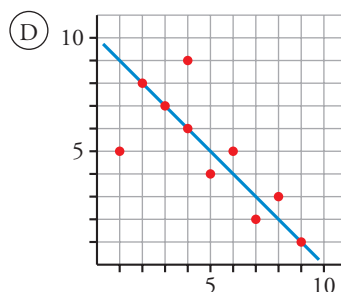
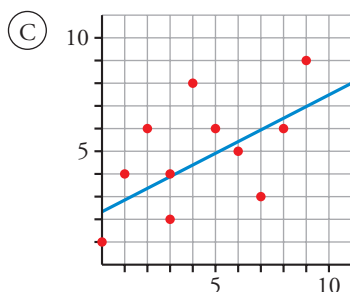
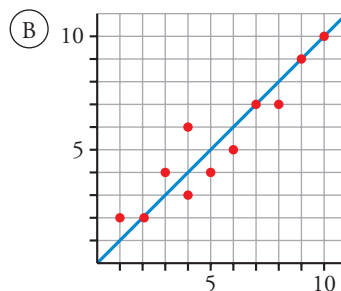
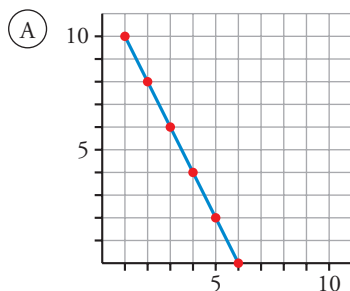


b) ¿Cuáles de ellas tienen correlación positiva y cuáles tienen correlación negativa?

c) Una de ellas presenta relación funcional. ¿Cuál es? ¿Cuál es la expresión analítica de la función que relaciona las dos variables?

d) Ordena de menor a mayor las correlaciones.

a)



b) B y C tienen correlación positiva; A y D, negativa.

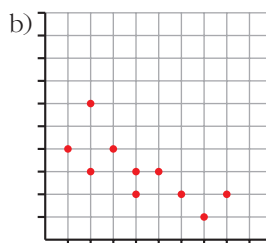
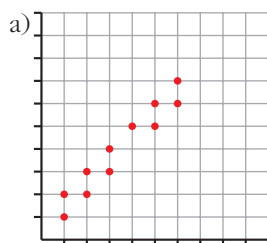
c) La A es relación funcional:  $y = 12 - 2x$ .

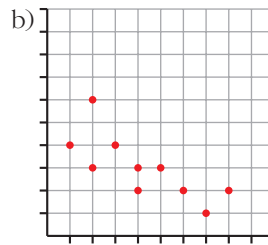
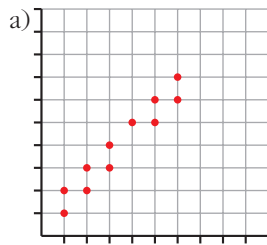
d) C, D, B, A (prescindiendo del signo).

**3** Los coeficientes de correlación de las distribuciones bidimensionales que aparecen a continuación son, en valor absoluto, los siguientes:

0,55      0,75      0,87      0,96

Asigna a cada uno el suyo, cambiando el signo cuando proceda:





a)  $r = 0,96$

b)  $r = -0,75$

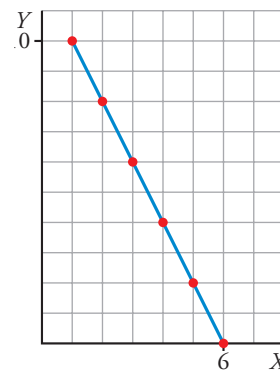
c)  $r = 0,55$

d)  $r = -0,87$

**4 Representa la nube de puntos correspondiente a esta distribución y di cuánto vale el coeficiente de correlación.**

x	1	2	3	4	5	6
y	10	8	6	4	2	0

El coeficiente de correlación vale  $-1$ .



**5 Representa la nube de puntos de esta distribución y estima cuál de estos tres puede ser el coeficiente de correlación:**

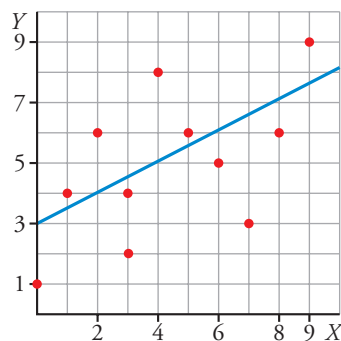
a)  $r = 0,98$

b)  $r = -0,87$

c)  $r = 0,5$

x	0	1	2	3	4	5	6	7	8	9
y	1	4	6	2	4	8	6	5	3	6

c)  $r = 0,5$





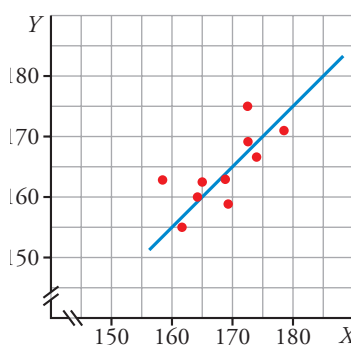
- 6 Las estaturas de 10 chicas y las de sus respectivas madres son:

$x_i$	158	162	164	165	168	169	172	172	174	178
$y_i$	163	155	160	161	164	158	175	169	166	172

Representa los valores, sobre papel cuadrulado, mediante una nube de puntos.

Traza a ojo la recta de regresión y di si la correlación es positiva o negativa y si es más o menos fuerte de lo que esperabas.

La correlación es positiva y fuerte.



## Página 345

### Con fórmulas

- 7 Esta es la distribución bidimensional dada en el ejercicio 2B) mediante una nube de puntos:

$x$	0	1	2	3	4	4	5	6	7	8	9	10
$y$	0	2	2	4	3	6	4	5	7	7	9	10

Halla:

a)  $\bar{x}$ ,  $\bar{y}$ ,  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_{xy}$ .

b) El coeficiente de correlación,  $r$ . Interpretalo.

c) Las dos rectas de regresión.

$$n = 12, \quad \Sigma x = 59 \quad \Sigma y = 59$$

$$\Sigma x^2 = 401 \quad \Sigma y^2 = 389 \quad \Sigma xy = 390$$

a)  $\bar{x} = 4,92 \quad \bar{y} = 4,92$

$$\sigma_x = 3,04 \quad \sigma_y = 2,87 \quad \sigma_{xy} = 8,33$$

b)  $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = 0,95$ . Se trata de una correlación fuerte y positiva.

c) Recta de regresión de  $Y$  sobre  $X$ :

$$\frac{\sigma_{xy}}{\sigma_x^2} = 0,90 \rightarrow y = 4,92 + 0,9(x - 4,92)$$

Recta de regresión de  $X$  sobre  $Y$ :

$$\frac{\sigma_{xy}}{\sigma_y^2} = 1,01 \rightarrow y = 4,92 + \frac{1}{1,01}(x - 4,92) \rightarrow y = 4,92 + 0,99(x - 4,92)$$

## 8 Observa la distribución D del ejercicio 2.

a) Descríbela mediante una tabla de valores.

b) Realiza los cálculos para obtener su coeficiente de correlación.

c) Representa los puntos en tu cuaderno. Halla la ecuación de la recta de regresión de  $Y$  sobre  $X$  y represéntala.

a)

<b>x</b>	1	2	3	4	4	5	6	7	8	9
<b>y</b>	5	8	7	6	9	4	5	2	3	1

b)  $n = 10$        $\Sigma x = 49$        $\bar{x} = \frac{49}{10} = 4,9$

$\Sigma y = 50$        $\bar{y} = \frac{50}{10} = 5$

$\Sigma x^2 = 301$        $\sigma_x = \sqrt{\frac{301}{10} - 4,9^2} = 2,47$

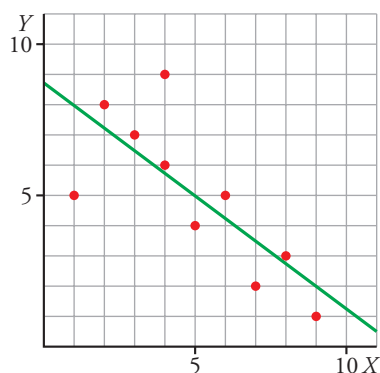
$\Sigma y^2 = 310$        $\sigma_y = \sqrt{\frac{301}{10} - 5^2} = 2,45$

$\Sigma xy = 199$        $\sigma_{xy} = \frac{199}{10} - 4,9 \cdot 5 = -4,6$

$r = \frac{4,6}{2,47 \cdot 2,45} = -0,76$

c) Recta de regresión de  $Y$  sobre  $X$ :

$$y = 5 - \frac{4,6}{6,1}(x - 4,9) \rightarrow y = 8,675 - 0,75x$$



9 a) Representa la siguiente distribución bidimensional:

x	0	1	2	3	3	4	5	6	7	8	9
y	1	4	6	2	4	8	6	5	3	6	9

b) Comprueba con la calculadora que sus parámetros son:

$$\bar{x} = 4,4 \quad \bar{y} = 4,9 \quad \sigma_{xy} = 3,67$$

$$\sigma_x = 2,77 \quad \sigma_y = 2,31 \quad r = 0,57$$

c) Halla las ecuaciones de las dos rectas de regresión,  $X$  sobre  $Y$  e  $Y$  sobre  $X$ , y represéntalas junto con la nube de puntos.

a) Representada en el ejercicio 5.

b) Se comprueba.

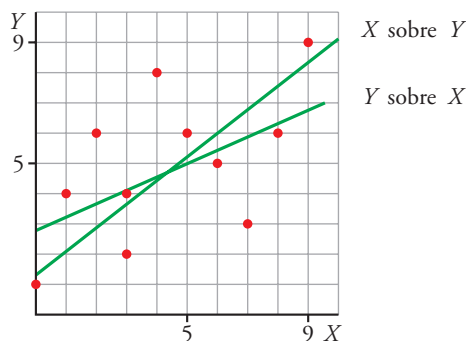
c) • Recta de regresión de  $Y$  sobre  $X$ :

$$m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{3,67}{2,77^2} = 0,48 \rightarrow y = 4,9 + 0,48(x - 4,4) \rightarrow y = 0,48x + 2,79$$

• Recta de regresión de  $X$  sobre  $Y$ :

$$m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{3,67}{2,31^2} = 0,69 \rightarrow \frac{1}{m_{xy}} = 1,45 \rightarrow y = 4,9 + 1,45(x - 4,4) \rightarrow$$

$$\rightarrow y = 1,45x - 1,48$$



10 Una distribución bidimensional en la que los valores de  $x$  son 12, 15, 17, 21, 22 y 25, tiene una correlación  $r = 0,99$  y su recta de regresión es  $y = 10,5 + 3,2x$ .

Calcula  $\hat{y}(13)$ ,  $\hat{y}(20)$ ,  $\hat{y}(30)$ ,  $\hat{y}(100)$ .

¿Cuáles de las estimaciones anteriores son fiables, cuál poco fiable y cuál no se debe hacer?

Expresa los resultados en términos adecuados. (Por ejemplo:  $\hat{y}(13) = 52,1$ . Para  $x = 13$  es muy probable que el valor correspondiente de  $y$  sea próximo a 52).

$$\hat{y}(13) = 52,1; \quad \hat{y}(20) = 74,5; \quad \hat{y}(30) = 106,5; \quad \hat{y}(100) = 330,5$$

Son fiables  $\hat{y}(13)$  e  $\hat{y}(20)$ , porque 13 y 20 están en el intervalo de valores utilizados para obtener la recta de regresión.

$\hat{y}(30)$  es menos fiable, pues 30 está fuera del intervalo, aunque cerca de él.

$\hat{y}(100)$  es una estimación nada fiable, pues 100 está muy lejos del intervalo [12, 25].

## PARA RESOLVER

- 11** La siguiente tabla muestra el número de gérmenes patógenos por centímetro cúbico de un determinado cultivo según el tiempo transcurrido:

N.º DE HORAS	0	1	2	3	4	5
N.º DE GÉRMESES	20	26	33	41	47	53

a) Calcula la recta de regresión para predecir el número de gérmenes por centímetro cúbico en función del tiempo.

b) ¿Qué cantidad de gérmenes por centímetro cúbico cabe esperar que haya a las 6 horas? ¿Es buena esta estimación?

a)  $y = 19,81 + 6,74x$ , donde:  $x \rightarrow$  número horas,  $y \rightarrow$  número de gérmenes

b)  $\hat{y}(6) = 60,25 \approx 60$  gérmenes.

Es una buena predicción, puesto que  $r = 0,999$  (y 6 está cercano al intervalo de valores considerado).

- 12** La media de los pesos de los individuos de una población es de 65 kg, y la de sus estaturas, 170 cm. Sus desviaciones típicas son 5 kg y 10 cm. La covarianza es 40 kg · cm. Halla:

a) Coeficiente de correlación.

b) La recta de regresión de los pesos respecto de las estaturas.

c) Estima el peso de un individuo de 180 cm de estatura perteneciente a ese colectivo.

a)  $r = 0,8$

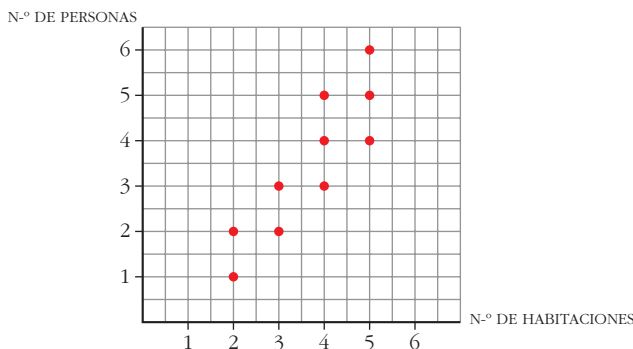
b)  $y = 65 + 0,4(x - 170) = 0,4x - 3 \rightarrow \begin{cases} x: \text{ estaturas en cm} \\ y: \text{ pesos en kg} \end{cases}$

c)  $\hat{y}(180) = 69$  kg

- 13** En una zona residencial se ha tomado una muestra para relacionar el número de habitaciones que tiene cada piso ( $h$ ) con el número de personas que viven en él ( $p$ ). Estos son los resultados:

<b>h</b>	2	2	3	3	4	4	4	5	5	5
<b>p</b>	1	2	2	3	3	4	5	4	5	6

Represéntalos mediante una nube de puntos. Calcula el coeficiente de correlación e interprétalo.



$h$ : número de habitaciones

$p$ : número de personas

$$n = 10 \quad \Sigma h = 37 \quad \bar{h} = \frac{37}{10} = 3,7$$

$$\Sigma p = 35 \quad \bar{p} = \frac{35}{10} = 3,5$$

$$\Sigma h^2 = 149 \quad \sigma_h = \sqrt{\frac{149}{10} - 3,7^2} = 1,1$$

$$\Sigma p^2 = 145 \quad \sigma_p = \sqrt{\frac{145}{10} - 3,5^2} = 1,5$$

$$\Sigma hp = 144 \quad \sigma_{hp} = \frac{144}{10} - 3,7 \cdot 3,5 = 1,45$$

$$r = \frac{1,45}{1,1 \cdot 1,5} = 0,88$$

Es una correlación positiva y fuerte (a más habitaciones, más personas en el piso).

**14** La tabla adjunta relaciona el número atómico de varios metales con su densidad:

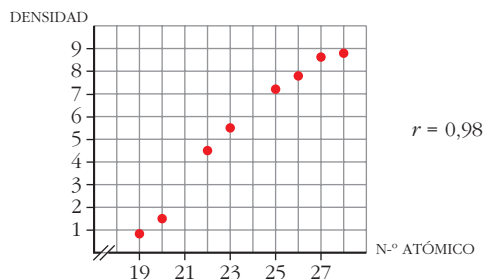
<b>Elemento</b>	K	Ca	Ti	V	Mn	Fe	Co	Ni
<b>Nº atómico</b>	19	20	22	23	25	26	27	28
<b>Densidad</b>	0,86	1,54	4,50	5,60	7,11	7,88	8,70	8,80

a) Representa los puntos y halla el coeficiente de correlación.

b) Mediante una recta de regresión, estima la densidad del cromo si su número atómico es 24: Cr (24).

c) Estima la densidad del escandio: Sc (21).

a)



b) y c)  $\hat{y} = -16,5 + 0,93x$        $\hat{y}(24) = 5,86$        $\hat{y}(21) = 3,06$

Las densidades del Cr y del Sc son, aproximadamente, 5,86 y 3,01. (Los valores reales de estas densidades son 7,1 y 2,9.)

## Página 346

**15** En una cofradía de pescadores, las capturas registradas de cierta variedad de pescados, en kilogramos, y el precio de subasta en lonja, en euros/kg, fueron los siguientes:

x (kg)	2 000	2 400	2 500	3 000	2 900	2 800	3 160
y (euros/kg)	1,80	1,68	1,65	1,32	1,44	1,50	1,20

a) ¿Cuál es el precio medio registrado?

b) Halla el coeficiente de correlación lineal e interprétalo.

c) Estima el precio que alcanzaría en lonja el kilo de esa especie si se pesasen 2 600 kg.

a)  $\bar{y} = 1,51$  euros

b)  $r = -0,97$ . La relación entre las variables es fuerte y negativa. A mayor cantidad de pescado, menor es el precio por kilo.

c) La recta de regresión es  $y = 2,89 - 0,0005x$ .

$\hat{y}(2\,600) = 1,59$  euros.

- 16** Durante 10 días, hemos realizado mediciones sobre el consumo de un coche (litros consumidos y kilómetros recorridos). Los datos obtenidos han sido los siguientes:

$x$ (km)	100	80	50	100	10	100	70	120	150	220
$y$ (l)	6,5	6	3	6	1	7	5,5	7,5	10	15

a) Halla el coeficiente de correlación y la recta de regresión de  $Y$  sobre  $X$ .

b) Si queremos hacer un viaje de 190 km, ¿qué cantidad de combustible debemos poner?

a)  $r = 0,99$ ;  $y = 0,157 + 0,066x$

b)  $\hat{y}(190) = 12,697$  litros. Debemos poner, como mínimo, unos 13 litros.

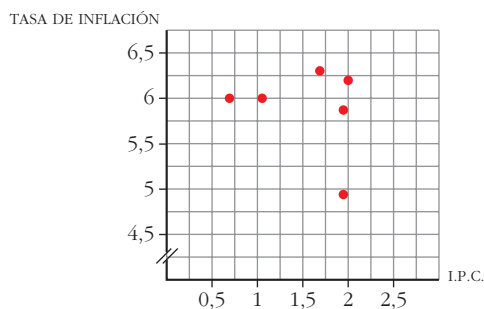
- 17** La evolución del IPC (índice de precios al consumo) y de la tasa de inflación en 1987 fue:

	ENERO	FEBRERO	MARZO	ABRIL	MAYO	JUNIO
IPC	0,7	1,1	1,7	2	1,9	1,9
TASA DE INFLACIÓN	6	6	6,3	6,2	5,8	4,9

a) Representa la nube de puntos.

b) Calcula el coeficiente de correlación entre el IPC y la tasa de inflación.

c) ¿Se puede estimar la tasa de inflación a partir del IPC?



$r = -0,24$ . La nube de puntos es muy dispersa. No se puede estimar de forma fiable la tasa de inflación a partir del IPC (pues  $|r|$  es muy bajo).

## CUESTIONES TEÓRICAS

- 18** El coeficiente de correlación de una distribución bidimensional es 0,87.

Si los valores de las variables se multiplican por 10, ¿cuál será el coeficiente de correlación de esta nueva distribución?

El mismo, puesto que  $r$  no depende de las unidades; es adimensional.

- 19 Hemos calculado la covarianza de una cierta distribución y ha resultado negativa.**

**Justifica por qué podemos afirmar que tanto el coeficiente de correlación como las pendientes de las dos rectas de regresión son números negativos.**

Hay que tener en cuenta que  $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ ;  $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}$ ;  $m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2}$  y que  $\sigma_x \geq 0$ ,  $\sigma_y \geq 0$  siempre.

Luego  $r$ ,  $m_{yx}$ ,  $m_{xy}$  tienen el mismo signo que  $\sigma_{xy}$ . (Además, suponemos  $\sigma_x \neq 0$  y  $\sigma_y \neq 0$ .)

- 20 ¿Qué punto tienen en común las dos rectas de regresión?**

El centro de gravedad de la distribución,  $(\bar{x}, \bar{y})$ .

- 21 ¿Qué condición debe cumplir  $r$  para que las estimaciones hechas con la recta de regresión sean fiables?**

$|r|$  debe estar próximo a 1.

- 22 Prueba que el producto de los coeficientes de regresión  $m_{yx}$  y  $m_{xy}$  es igual al cuadrado del coeficiente de correlación.**

$$m_{yx} \cdot m_{xy} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot \frac{\sigma_{xy}}{\sigma_y^2} = \left( \frac{\sigma_{xy}}{\sigma_x \sigma_y} \right)^2 = r^2$$

- 23 De una distribución bidimensional  $(x, y)$  conocemos los siguientes resultados:**

- Recta de regresión de  $Y$  sobre  $X$ :

$$y = 8,7 - 0,76x$$

- Recta de regresión de  $X$  sobre  $Y$ :

$$y = 11,36 - 1,3x$$

**a) Calcula el centro de gravedad de la distribución.**

**b) Halla el coeficiente de correlación.**

a) El centro de gravedad,  $(\bar{x}, \bar{y})$ , es el punto de corte entre las dos rectas:

$$\left. \begin{array}{l} y = 8,7 - 0,76x \\ y = 11,36 - 1,3x \end{array} \right\}$$

$$8,7 - 0,76x = 11,36 - 1,3x$$

$$0,54x = 2,66$$

$$x = 4,93$$

$$y = 4,95$$

El centro de gravedad es  $(\bar{x}, \bar{y}) = (4,93; 4,95)$ .



b) Para hallar  $r$  tenemos en cuenta el ejercicio anterior:

$$r^2 = m_{yx} \cdot m_{xy} = -0,76 \cdot \frac{1}{-1,3} = 0,58 \rightarrow r = 0,76$$

- 24** La estatura media de 100 escolares de cierto curso de ESO es de 155 cm con una desviación típica de 15,5 cm.

La recta de regresión de la estatura respecto al peso es:

$$y = 80 + 1,5x \quad (x: \text{peso}; y: \text{estatura})$$

a) ¿Cuál es el peso medio de esos escolares?

b) ¿Cuál es el signo del coeficiente de correlación entre peso y estatura?

a) La recta de regresión es:

$$\begin{aligned} y &= \bar{y} + m(x - \bar{x}) = 155 + 1,5(x - \bar{x}) = 155 + 1,5x - 1,5\bar{x} = (155 - 1,5\bar{x}) + 1,5x = \\ &= 80 + 1,5x \rightarrow 155 - 1,5\bar{x} = 80 \rightarrow \bar{x} = 50 \text{ kg} \end{aligned}$$

b) Positivo (igual que el signo de la pendiente de la recta de regresión).

## Página 347

### PARA PROFUNDIZAR

- 25** En una muestra de 64 familias se han estudiado el número de miembros en edad laboral,  $x$ , y el número de ellos que están en activo,  $y$ . Los resultados son los de la tabla. Calcula el coeficiente de correlación lineal entre ambas variables e interprétalo.

$r = 0,31$ . La relación entre las variables es débil.

$x \backslash y$	1	2	3
1	6	0	0
2	10	2	0
3	12	5	1
4	16	8	4

- 26** Una compañía discográfica ha recopilado la siguiente información sobre el número de conciertos dados, durante el verano, por 15 grupos musicales y las ventas de discos de estos grupos (expresados en miles de CD):

CONCIERTOS (y) \ CD (x)	10 - 30	30 - 40	40 - 80
1 - 5	3	0	0
5 - 10	1	4	1
10 - 20	0	1	5

- a) Calcula el número medio de CD vendidos.
- b) ¿Cuál es el coeficiente de correlación?
- c) Obtén la recta de regresión de  $Y$  sobre  $X$ .
- d) Si un grupo musical vende 18 000 CD, ¿qué número de conciertos se prevé que dé?

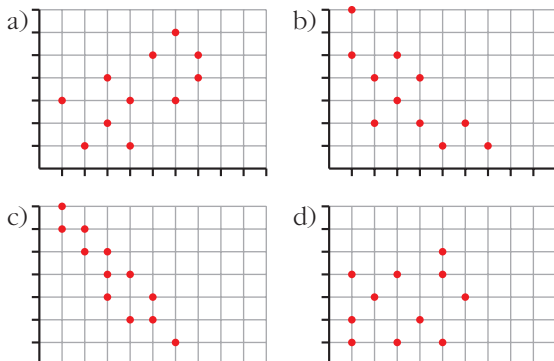
$x \rightarrow$  CD;  $y \rightarrow$  Conciertos

- a)  $\bar{x} = 9,6 \approx 10$
- b)  $r = 0,814$
- c)  $y = 13,51 + 2,86x$
- d)  $\hat{y}(18) = 64,99 \approx 65$  conciertos

## Página 347

### AUTOEVALUACIÓN

#### 1. Observa estas distribuciones bidimensionales:



Asigna razonadamente uno de los siguientes coeficientes de correlación a cada gráfica: 0,2; -0,9; -0,7; 0,6.

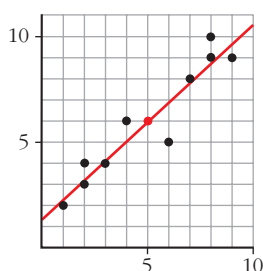
La correlación de a) es positiva, y las de b) y c), negativas. En d) no se aprecia correlación. La correlación de c) es más fuerte que la de b). Por tanto:

- a)  $\rightarrow 0,6$
- b)  $\rightarrow -0,7$
- c)  $\rightarrow -0,9$
- d)  $\rightarrow 0,2$

## 2. Representa esta distribución bidimensional:

x	1	2	2	3	4	6	7	8	8	9
y	2	4	3	4	6	5	8	9	10	9

- Calcula los parámetros  $\bar{x}$ ,  $\bar{y}$ ,  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_{xy}$ .
- Halla el coeficiente de correlación.
- Halla la recta de regresión de  $Y$  sobre  $X$ .
- Estima el valor de  $y$  para  $x = 5$  y para  $x = 10$ . ¿Son “buenas” estas estimaciones?



- $\bar{x} = 5$ ,  $\bar{y} = 6$   
 $\sigma_x = 2,8$ ;  $\sigma_y = 2,7$ ;  $\sigma_{xy} = 7,1$
- $r = 0,95$
- $y = 0,91x + 1,45$
- $\hat{y}(5) = 6$ ,  $\hat{y}(10) = 10,55$

Las estimaciones son muy fiables porque  $r = 0,95$  es un valor muy alto. Si se tratase de “notas” (de 0 a 10), la segunda estimación habría que “hacerla real” y darle el valor 10.

## 3. La recta de regresión de $Y$ sobre $X$ de una cierta distribución bidimensional es $y = 1,6x - 3$ . Sabemos que $\bar{x} = 10$ y $r = 0,8$ .

- Calcula  $\bar{y}$ .
- Estima el valor de  $y$  para  $x = 12$  y para  $x = 50$ . ¿Qué estimación te parece más fiable?

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

- Puesto que la recta pasa por  $(\bar{x}, \bar{y})$ :

$$\bar{y} = 1,6\bar{x} - 3 = 1,6 \cdot 10 - 3 = 13$$

$$b) \hat{y}(12) = 1,6 \cdot 12 - 3 = 16,2$$

$$\hat{y}(50) = 1,6 \cdot 50 - 3 = 77$$

La primera estimación es aceptable por ser 12 próximo a  $\bar{x} = 10$  (carecemos de información sobre los valores que toma  $x$ ). La segunda estimación es muy poco significativa, pues 50 se separa demasiado de  $\bar{x}$ .

- Conociendo  $r = 0,8$  y el coeficiente de regresión de  $Y$  sobre  $X$  (pendiente de la recta), 1,6:

$$(\text{Coef. } Y \text{ sobre } X) \cdot (\text{Coef. } X \text{ sobre } Y) = r^2$$

$$\text{Coef. } X \text{ sobre } Y = \frac{0,8^2}{1,6} = 0,4$$

Por tanto, la pendiente de la recta de regresión de  $X$  sobre  $Y$  es  $m_{xy} = \frac{1}{0,4} = 2,5$ .

Ecuación de la recta de regresión de  $X$  sobre  $Y$ :  $y = 6 + 2,5(x - 5)$

- 4. El consumo de energía per cápita  $y$  en miles de kWh y la renta per cápita  $x$  en miles de euros de seis países son:**

	A	B	C	D	E	F
$x$	11,1	8,5	11,3	4,5	9,9	6,5
$y$	5,7	5,0	5,1	2,7	4,6	3,1

- Calcula la recta de regresión de  $Y$  sobre  $X$ .
- Halla el coeficiente de correlación entre el consumo y la renta.
- ¿Qué predicción podemos hacer sobre el consumo de energía per cápita de un país cuya renta per cápita es de 4,4 miles de euros?

$$\bar{x} = 8,63, \bar{y} = 4,37$$

$$\sigma_x = 2,46, \sigma_y = 1,09, \sigma_{xy} = 2,51$$

a) Recta de regresión de  $Y$  sobre  $X$ :

$$y = 4,37 + \frac{2,51}{2,46^2}(x - 8,63) \rightarrow y = 0,79 + 0,41x$$

b) Coeficiente de correlación:

$$r = \frac{2,51}{1,09 \cdot 2,46} = 0,93$$

c) Para  $x = 4,4$ , estimamos el valor de  $y$ :

$$\hat{y}(4,4) = 0,79 + 0,41 \cdot 4,4 = 2,59$$

Se le estima un consumo de energía de 2,59 miles de Kw/h por habitante.