

# 13 DISTRIBUCIONES BIDIMENSIONALES

C.E.: CE 1.3. (EA 1.3.1.-EA 1.3.2.) CE 1.6. (EA 1.6.1.-EA 1.6.2.) CE 1.13. (EA 1.13.1.) CE 5.3. (EA 5.3.1.)

Página 355

## Resuelve

### Relación funcional y relación estadística

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

a) En un conjunto de familias:

*Estatura media de los padres-Estatura media de los hijos*

b) Entre los países del mundo respecto a España:

*Volumen de exportación-Volumen de importación*

c) En los países del mundo:

*Tasa de mortalidad infantil-Médicos por cada 1 000 habitantes*

d) En las viviendas de una ciudad:

*kWh consumidos durante enero-Coste del recibo de la luz*

*Número de personas en cada casa-Coste del recibo de la luz*

e) En los equipos de fútbol:

*Posición al finalizar la liga-Número de partidos perdidos*

*Posición al finalizar la liga-Número de partidos ganados*

a) Estadística, porque la estatura media de los padres no nos permite saber exactamente la estatura media de los hijos. Hay correlación positiva. Normalmente, los hijos de padres altos son altos.

b) Estadística, porque el volumen de exportación no nos permite saber exactamente el volumen de importación. Hay correlación negativa. Normalmente, los países que exportan mucho, importan poco.

c) Estadística, porque la tasa de mortalidad infantil no nos permite saber exactamente el número de médicos por cada 1 000 habitantes. Hay correlación negativa. Normalmente, los países que tienen una tasa de mortalidad infantil grande, tienen pocos médicos por cada 1 000 habitantes.

d) *kWh consumidos durante enero - Coste del recibo de la luz* → Funcional; si conocemos los kWh consumidos durante enero, podemos calcular el coste del recibo de la luz.

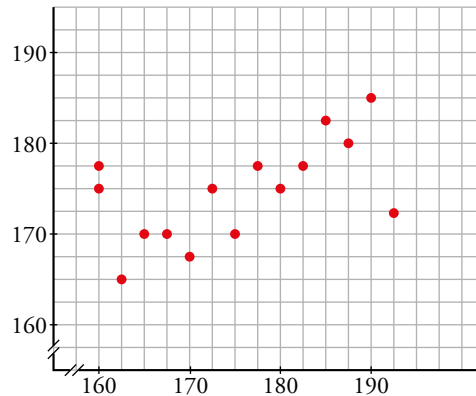
*Número de personas en cada casa - Coste del recibo de la luz* → Estadística, porque el número de personas en cada casa no nos permite saber exactamente el coste del recibo de la luz. Hay correlación positiva. Normalmente, cuantas más personas hay en una casa, más luz se consume.

e) *Posición al finalizar la liga - Número de partidos perdidos* → Estadística, porque la posición al finalizar la liga no nos permite saber exactamente el número de partidos perdidos. Hay correlación negativa. Normalmente, cuanto más alta es la posición en la liga, menos partidos se han perdido.

*Posición al finalizar la liga - Número de partidos ganados* → Estadística, porque la posición al finalizar la liga no nos permite saber exactamente el número de partidos ganados. Hay correlación positiva. Normalmente, cuanto más alta es la posición en la liga, más partidos se han ganado.

### Ejemplo de relación estadística

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




- Identifica a Guillermo y Gabriel, hermanos de buena estatura, cuyo padre es bajito.
- Identifica a Sergio, de estatura normalita, cuyo padre es muy alto.
- ¿Podemos decir que hay una cierta relación entre las estaturas de estos 15 chicos y las de sus padres?
  - Guillermo y Gabriel están representados mediante los puntos  $(160, 175)$  y  $(160; 177,5)$ .
  - Sergio está representado con el punto  $(192,5; 172,5)$ .
  - Sí; en general, cuanto más alto sea el padre, más altos son los hijos.

# 1 ► DISTRIBUCIONES BIDIMENSIONALES. NUBES DE PUNTOS

C.E.: CE 1.3. (EA 1.3.1.-EA 1.3.2.) CE 5.1. (EA 5.1.1.-EA 5.1.2.) CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.)

Página 357

---

- 1  [La lectura de los enunciados permite trabajar la destreza expresión escrita de esta clave].
- ¿Verdadero o falso?
- a) En una distribución bidimensional, para cada valor de  $x$  solo puede haber un valor de  $y$ .
  - b) Cuantos más puntos tenga una distribución bidimensional, más fuerte es su correlación.
  - c) Las series temporales son distribuciones estadísticas en las que una de las variables es el tiempo. Aunque no sean distribuciones bidimensionales propiamente dichas, pueden tratarse del mismo modo que estas.
- a) Falso, se pueden mirar las nubes de puntos de esta misma página.
  - b) Falso, la correlación depende de la relación entre las características que se estudian en una población, no del número de elementos de la población.
  - c) Verdadero.

## 2 ▶ CORRELACIÓN LINEAL

C.E.: CE 1.3. (EA 1.3.1.-EA 1.3.2.) CE 5.1. (EA 5.1.1.-EA 5.1.2.) CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.)

Página 359

### 1 ¿Verdadero o falso?

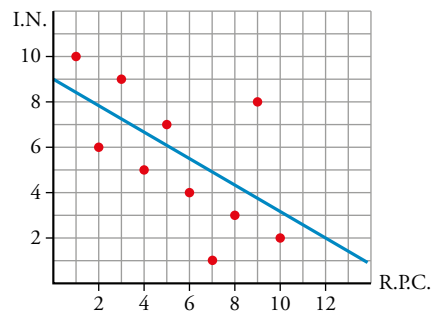
- Cuanto más próximos estén a una recta los puntos de una distribución bidimensional, más fuerte es su correlación lineal.
  - Si la recta de regresión tiene pendiente negativa, la correlación lineal es negativa.
  - Si los puntos de la nube no se aproximan a ninguna recta, entonces las variables están incorreladas.
- Verdadero. Porque la correlación estudia las distancias de los puntos a la recta de regresión. Cuanto más pequeña es la distancia a la recta, mayor es la correlación.
  - Verdadero. Una recta de pendiente negativa indica, como el signo del coeficiente de correlación, que al aumentar una variable, la otra disminuye.
  - Verdadero.

### 2 [La interpretación de los datos de la tabla requiere poner en práctica la iniciativa (dimensión productiva de esta clave)].

La siguiente tabla 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 ( $-0,62$ ).



### 3 ▶ PARÁMETROS ASOCIADOS A UNA DISTRIBUCIÓN BIDIMENSIONAL

C.E.: CE 1.3. (EA 1.3.1.-EA 1.3.2.) CE 1.13. (EA 1.13.1.) CE 5.1. (EA 5.1.1.-EA 5.1.4.) CE 5.3. (EA 5.3.1.)

Página 361

#### 1 ¿Verdadero o falso?

- El signo de la correlación ( $r$ ) coincide con el de la covarianza ( $\sigma_{xy}$ ).
- Si cambiamos las unidades en que se expresa la variable  $x$ , entonces se modifican los valores de  $\bar{x}$ ,  $\sigma_x$  y  $\sigma_{xy}$ .
- Aunque cambiemos las unidades en que se da la variable  $x$  (o  $y$ , o ambas) el valor de la correlación,  $r$ , no cambia.

a) Verdadero,  $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ ; como  $\sigma_x$  y  $\sigma_y$  son positivas, el signo de  $r$  es el de  $\sigma_{xy}$ .

b) Falso. Varían todos los parámetros menos  $r$ , porque  $r$  es el único que no tiene dimensiones.

c) Verdadero.

#### 2 Obtén mediante cálculos manuales los coeficientes de correlación de las distribuciones del epígrafe anterior:

*Salto de altura-Salto con pértiga*

*Salto de altura-1500 m lisos*

*Salto de altura-Lanzamiento de peso*

Comprueba tus resultados con la calculadora.

$x$ : salto de altura

$y$ : salto con pértiga

Elaboramos la tabla como en el ejercicio resuelto:

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i \cdot y_i$
1	1	1	1	1
2	4	4	16	8
3	2	9	4	6
4	3	16	9	12
5	5	25	25	25
6	7	36	49	42
7	6	49	36	42
8	8	64	64	64
36	36	204	204	200

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

$$\sigma_x = \sqrt{\frac{200}{8} - 4,5^2} = 2,2913$$

$$\sigma_y = \sqrt{\frac{200}{8} - 4,5^2} = 2,2913$$

$$\sigma_{xy} = \frac{200}{8} - 4,5 \cdot 4,5 = 4,75$$

$$r = \frac{4,75}{2,2913 \cdot 2,2913} = 0,90475$$

$x$ : salto de altura

$y$ : 1 500 m lisos

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i \cdot y_i$
1	3	1	9	3
2	2	4	4	4
3	5	9	25	15
4	1	16	1	4
5	7	25	49	35
6	6	36	36	36
7	4	49	16	28
8	8	64	64	64
36	36	204	204	189

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

$$\sigma_x = \sqrt{\frac{204}{8} - 4,5^2} = 2,2913$$

$$\sigma_y = \sqrt{\frac{204}{8} - 4,5^2} = 2,2913$$

$$\sigma_{xy} = \frac{189}{8} - 4,5 \cdot 4,5 = 3,375$$

$$r = \frac{3,375}{2,2913 \cdot 2,2913} = 0,64285$$

$x$ : salto de altura

$y$ : lanzamiento de peso

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i \cdot y_i$
1	7	1	49	7
2	5	4	25	10
3	8	9	64	24
4	6	16	36	24
5	4	25	16	20
6	1	36	1	6
7	3	49	9	21
8	2	64	4	16
36	36	204	204	128

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

$$\sigma_x = \sqrt{\frac{204}{8} - 4,5^2} = 2,2913$$

$$\sigma_y = \sqrt{\frac{204}{8} - 4,5^2} = 2,2913$$

$$\sigma_{xy} = \frac{128}{8} - 4,5 \cdot 4,5 = -4,25$$

$$r = \frac{-4,25}{2,2913 \cdot 2,2913} = -0,80952$$

## 4 ► RECTA DE REGRESIÓN

C.E.: CE 1.3. (EA 1.3.1.-EA 1.3.2.) CE 1.13. (EA 1.13.1.) CE 5.1. (EA 5.1.5.) CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.)

Página 363

---

### 1 ¿Verdadero o falso?

- a) **Cuanto más fuerte sea la correlación, más puntos habrá de la nube que se encuentren exactamente sobre la recta de regresión.**
  - b) **Cuanto más fuerte sea la correlación, más cerca de la recta de regresión estarán los puntos de la nube.**
  - c) **Cuanto más fuerte sea la correlación, más fiables serán las estimaciones hechas a partir de la recta de regresión.**
- a) Falso. Aunque la correlación sea muy grande, es posible que ningún punto de la nube de puntos esté sobre la recta.
  - b) Falso. Habrá muchos puntos cerca de la recta, pero puede haber puntos aislados lejos de la recta.
  - c) Verdadero. Los valores de una de las variables son más predecibles, puesto que están muy próximos a la recta de regresión.

## 5 ► HAY DOS RECTAS DE REGRESIÓN

C.E.: CE 1.3. (EA 1.3.1.-EA 1.3.2.) CE 5.1. (EA 5.1.5.) CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.) CE 5.3. (EA 5.3.1.)

Página 364

### 1 ¿Verdadero o falso?

- a) En una distribución bidimensional en la que se estudien conjuntamente las estaturas ( $x$ ) y los pesos ( $y$ ) de un grupo de jóvenes en la cual  $\bar{x} = 170$  cm e  $\bar{y} = 65$  kg, es imposible que las rectas de regresión sean  $y = 0,8x - 67$  e  $y = 1,1x - 121$ .
- b) Si en una distribución bidimensional es  $\bar{x} = 3$  e  $\bar{y} = 5$ , entonces es posible que las rectas de regresión sean  $y = 2x - 1$  e  $y = -x + 8$ , pues ambas se cortan en  $(3, 5)$ .
- c) Si las rectas de regresión son  $y = \frac{1}{5}x + 10$  e  $y = 11x - 2$ , entonces la correlación es débil porque las rectas forman un ángulo próximo a  $90^\circ$ .

a) 
$$\left\{ \begin{array}{l} y = 0,8x - 67 \\ y = 1,1x - 121 \end{array} \right\} \quad x = 180,0; \quad y = 77,0 \rightarrow \text{Se cortan en } (180, 77).$$

El punto de corte de las rectas de regresión debe ser  $(\bar{x}, \bar{y}) = (170, 65)$ , luego es verdadera la afirmación.

- b) Falso. El signo de la pendiente de las dos rectas de regresión debe ser igual.
- c) Verdadero. Se puede observar en las gráficas de esta página.



## 6 ▶ TABLAS DE CONTINGENCIA

C.E.: CE 1.13. (EA 1.13.1.) CE 5.1. (EA 5.1.1.-EA 5.1.2.-EA 5.1.3.-EA 5.1.4.-EA 5.1.5.) CE 5.3. (EA 5.3.1.)


Página 365

- 1 Calcula la media y la desviación típica de la distribución marginal de la  $x$ . Para ello, asigna a cada intervalo de edades su marca de clase (punto medio) y al último intervalo asígnale el valor 75.

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2$	$x_i^2 \cdot f_i$
21,5	50	1 075	462,25	23 112,5
30,5	85	2 592,5	930,25	79 071,25
43	140	6 020	1 849	258 860
58	100	5 800	3 364	336 400
75	125	9 375	5 625	703 125
	500	24 862,5		1 400 568,75

$$\bar{x} = \frac{24\,862,5}{500} = 49,725$$

$$\sigma_x = \sqrt{\frac{1\,400\,568,75}{500} - 49,725^2} = 18,126$$

- 2  ¿Qué te hace decir eso? [Esta estrategia de pensamiento se puede trabajar en esta actividad].

La distribución marginal de la  $y$  corresponde a una variable cualitativa. Por tanto, no tiene media ni desviación típica. El único parámetro que podemos asignarle es la moda. ¿Cuál es?

Moda = Deportes.

Página 366

- 3 Comprueba que la siguiente tabla corresponde a la distribución de  $x$  condicionada a  $y \in \{\text{INF., DOC.}\}$ .

$x$	18-25	26-35	36-50	51-65	más de 65
$f$	9	21	36	26	46

Halla su media y su desviación típica.

$x_i$	21,5	30,5	43	58	75	
INF	4	6	15	11	25	61
DOC	5	15	21	15	21	77
INF-DOC	9	21	36	26	46	138

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2$	$x_i^2 \cdot f_i$
21,5	9	193,5	462,25	4 160,25
30,5	21	640,5	930,25	19 535,25
43	36	1 548	1 849	66 564
58	26	1 508	3 364	87 464
75	46	3 450	5 625	258 750
	138	7 340		436 473,5

$$\bar{x} = \frac{7\,340}{138} = 53,188$$

$$\sigma_x = \sqrt{\frac{436\,473,5}{138} - 53,188^2} = 18,273$$

4 Haz la distribución de  $y$  condicionada a  $x < 36$ .

$y_i$	$f_i$
INF	10
DOC	20
ENT	20
DEP	54
PEL	26
OTR	5

5 Comprueba, calculando las frecuencias relativas, que el suceso PEL. no es independiente de la edad.

$x_j$	21,5	30,5	43	58	75	
PEL	11	15	20	16	11	73
	0,15068493	0,20547945	0,2739726	0,21917808	0,15068493	

Se observa que las frecuencias relativas varían según la edad.

6 Haz la distribución de  $x$  condicionada a NO DEPORTE y compara sus frecuencias relativas con las de la distribución marginal de la  $x$ .

$x_i$	21,5	30,5	43	58	75	
NO DEP	61	105	166	119	138	589

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2$	$x_i^2 \cdot f_i$
21,5	61	1 311,5	462,25	28 197,25
30,5	105	3 202,5	930,25	97 676,25
43	166	7 138	1 849	306 934
58	119	6 902	3 364	400 316
75	138	10 350	5 625	776 250
	589	28 904		1 609 373,5

$$\bar{x} = \frac{28\,904}{589} = 49,073$$

$$\sigma_x = \sqrt{\frac{1\,609\,373,5}{589} - 49,073^2} = 33,56$$

La media es similar; sin embargo, la desviación típica es mayor si consideramos los datos de las personas que no ven deportes.

Página 368

7 Otro grupo de 154 personas han realizado los mismos test, con los resultados que se dan en la tabla de la derecha. Halla el coeficiente de correlación.

De los datos obtenemos las siguientes tablas:

$x_i \backslash y_i$	0	1	2	3	4	
0	17	22	6	4	1	50
1	15	14	8	2	0	39
2	13	6	10	5	1	35
3	5	4	2	6	2	19
4	3	1	0	3	4	11
	53	47	26	20	8	154

$x_i \backslash y_i$	0	1	2	3	4
0	17	22	6	4	1
1	15	14	8	2	0
2	13	6	10	5	1
3	5	4	2	6	2
4	3	1	0	3	4

Distribución marginal de la  $x$ :

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2$	$x_i^2 \cdot f_i$
0	53	0	0	0
1	47	47	1	47
2	26	52	4	104
3	20	60	9	180
4	8	32	16	128
	154	191		459

$$\bar{x} = \frac{191}{154} = 1,24$$

$$\sigma_x = \sqrt{\frac{459}{154} - 1,24^2} = 1,20$$

Distribución marginal de la  $y$ :

$y_i$	$f_i$	$y_i \cdot f_i$	$y_i^2$	$y_i^2 \cdot f_i$
0	50	0	0	0
1	39	39	1	39
2	35	70	4	140
3	19	57	9	171
4	11	44	16	176
	154	210		526

$$\bar{y} = \frac{210}{154} = \frac{15}{11} = 1,36$$

$$\sigma_y = \sqrt{\frac{526}{154} - 1,36^2} = 1,25$$

$$\sigma_{xy} = \frac{332}{154} - 1,36 \cdot 1,20 = 0,52$$

$$r = \frac{0,52}{1,25 \cdot 1,20} = 0,35$$

## EJERCICIOS Y PROBLEMAS RESUELTOS

C.E.: CE 1.6. (EA 1.6.1.-EA 1.6.2.) CE 1.13. (EA 1.13.1.) CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.)

Página 370

### 3. Obtención de la correlación a partir de las dos rectas de regresión

Hazlo tú

- **Calcula las medias de las distribuciones y el coeficiente de correlación a partir de estas dos rectas de regresión:**

$$y = 0,77x + 4,64$$

$$x = 1,2y - 4,73$$

- a)  $(\bar{x}, \bar{y})$  son las coordenadas del punto de intersección de las dos rectas de regresión.

$$\begin{cases} y = 0,77x + 4,64 \\ x = 1,2y - 4,73 \end{cases}$$

$$\rightarrow y = 0,77(1,2y - 4,73) + 4,64 \rightarrow y = \frac{9979}{760} = 13,13 \rightarrow x = 11,02$$

Por tanto,  $(\bar{x}, \bar{y}) = (11,02; 13,13)$ .

- b) Calculamos el coeficiente de correlación a partir de las pendientes de las rectas de regresión:

$$m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = 0,77$$

$$m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} = 1,2$$

$$r = \sqrt{m_{yx} \cdot m_{xy}} = \sqrt{0,77 \cdot 1,2} = 0,96$$

La correlación es positiva y fuerte.

## EJERCICIOS Y PROBLEMAS GUIADOS

C.E.: CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.)

Página 371

### 1. Dos rectas de regresión. Estimaciones

- La siguiente tabla relaciona las variables  
 $x$ : gastos en publicidad (miles de euros)  
 $y$ : ventas (miles de euros)

$x$	1	2	3	4	5	6
$y$	10	17	30	28	39	47

durante los 6 primeros meses de promoción de un cierto producto:

- Hallar las dos rectas de regresión.
- Efectuar la estimación  $\hat{y}(5,5)$  y explicar su significado.
- Para obtener unas ventas de 20 000 €, ¿cuántos miles de euros se estima que hay que gastar en publicidad?

¿Serán fiables estas estimaciones?

a)

$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i \cdot y_i$
1	10	1	100	10
2	17	4	289	34
3	30	9	900	90
4	28	16	784	112
5	39	25	1 521	195
6	47	36	2 209	282
21	171	91	5 803	723

$$\bar{x} = \frac{21}{6} = 3,5$$

$$\bar{y} = \frac{171}{6} = 28,5$$

$$\sigma_x = \sqrt{\frac{91}{6} - 3,5^2} = 1,71$$

$$\sigma_y = \sqrt{\frac{5\,803}{6} - 28,5^2} = 12,45$$

$$\sigma_{xy} = \frac{723}{6} - 3,5 \cdot 28,5 = 20,75$$

Pendiente de la recta de regresión de  $Y$  sobre  $X$ :

$$m_{yx} = \frac{20,75}{1,71^2} = 7,1$$

$$y - 28,5 = 7,1(x - 3,5)$$

Pendiente de la recta de regresión de  $X$  sobre  $Y$ :

$$m_{xy} = \frac{12,45^2}{20,75} = 7,47$$

$$y - 28,5 = 7,47(x - 3,5)$$

b)  $\hat{y}(5,5) = 7,1(5,5 - 3,5) + 28,5 = 42,7$

c)  $\hat{x}(20) \rightarrow 20 - 28,5 = 7,47(x - 3,5) \rightarrow y = 2,36$

$$r = \frac{20,75}{1,71 \cdot 12,45} = 0,97$$

## 2. Tabla de doble entrada

- Una compañía discográfica ha recopilado en la tabla de la derecha la siguiente información sobre el número de conciertos dados por 15 grupos musicales durante un verano, y las ventas de discos de estos grupos (en miles).

CONC. (y) \ DISCOS (x)	10-30	30-40	40-80
1-5	3	0	0
5-10	1	4	1
10-20	0	1	5

- Calcular el número medio de discos vendidos.
- ¿Cuál es el coeficiente de correlación?
- Obtener la recta de regresión de  $Y$  sobre  $X$ .
- Si un grupo musical vende 18 000 discos, ¿qué número de conciertos se prevé para él?

a)

CONC. ( $y_i$ ) \ DISCOS ( $x_j$ )	20	35	60	
3	3	0	0	3
7,5	1	4	1	6
15	0	1	5	6
	4	5	6	

$$\bar{x} = \frac{3 \cdot 3 + 7,5 \cdot 6 + 15 \cdot 6}{15} = 9,6$$

b)

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2$	$x_i^2 \cdot f_i$
3	3	9	9	27
7,5	6	45	56,25	337,5
15	6	90	225	1350
	15	144		1714,5

$$\sigma_x = \sqrt{\frac{1714,5}{15} - 9,6^2} = 4,71$$

$y_i$	$f_i$	$y_i \cdot f_i$	$y_i^2$	$y_i^2 \cdot f_i$
20	4	80	400	1600
35	5	175	1225	6125
60	6	360	3600	21600
	15	615		29325

$$\bar{y} = \frac{20 \cdot 4 + 35 \cdot 5 + 60 \cdot 6}{15} = 41$$

$$\sigma_y = \sqrt{\frac{29325}{15} - 41^2} = 16,55$$

$$\Sigma x \cdot y \cdot f = 6855$$

$$\sigma_{xy} = \frac{6855}{15} - 9,6 \cdot 41 = 63,4$$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{63,4}{4,71 \cdot 16,55} = 0,81$$

c)  $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{63,4}{4,71^2} = 2,86$

$$y - 41 = 2,86(x - 9,6) \rightarrow y = 2,86x + 13,51$$

- d) La previsión de conciertos será:

$$\hat{y}(18) = 2,86 \cdot 18 + 13,51 = 65$$


**EJERCICIOS Y PROBLEMAS PROPUESTOS**

C.E.: CE todos los tratados en la unidad (EA todos los tratados en la unidad)

Página 372

**Para practicar**

## Sin fórmulas


1  [Las dudas que surjan sobre las relaciones entre las variables planteadas pueden ser tratadas según esta técnica].

Para cada uno de los siguientes casos, indica:

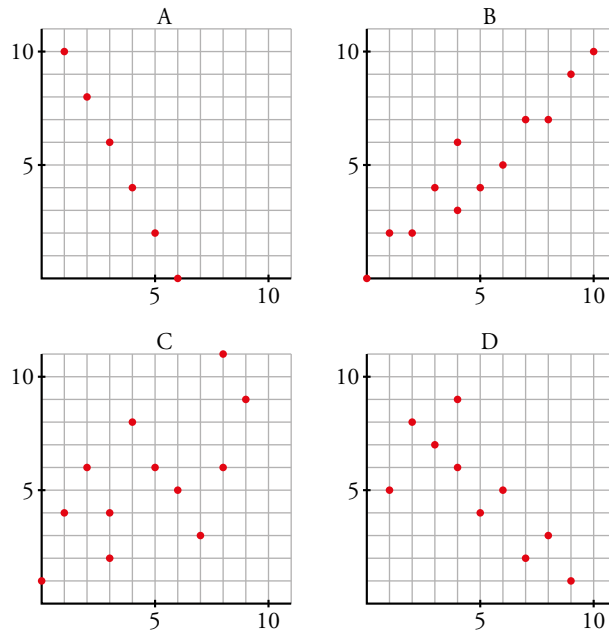
- Cuáles son las variables que se relacionan.
  - Si se trata de una relación funcional o de una relación estadística y, en este último caso, determina el signo de la correlación.
- a) *Renta mensual de una familia-Gasto mensual en electricidad*
  - b) *Radio de una esfera-Volumen de esta*
  - c) *Litros de lluvia recogidos en una ciudad*  
*Tiempo dedicado a ver la televisión por sus habitantes*
  - d) *Longitud del trayecto recorrido en una línea de cercanías*  
*Precio del billete*
  - e) *Peso de los alumnos de 1.º de Bachillerato*  
*Número de calzado que usan*
  - f) *Toneladas de tomate recogidas en una cosecha*  
*Precio del kilo de tomate en el mercado*
  - g) *Superficie de una vivienda-Valor de la misma*

Indica en cada caso cómo crees que será la correlación: fuerte, intermedia o débil.

- 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 estarán los tomates).
- g) Correlación positiva.

2  [Las diferencias entre las gráficas que dibujarán los alumnos pueden ser analizadas según esta técnica].

a) Copia en tu cuaderno y traza a ojo una recta de regresión para cada una de estas distribuciones bidimensionales:

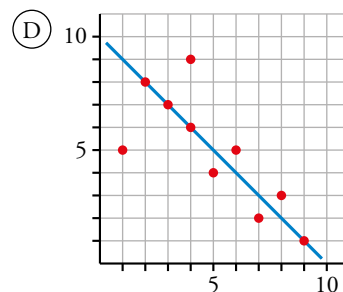
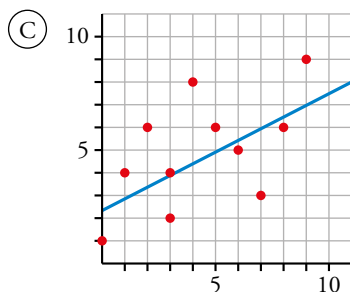
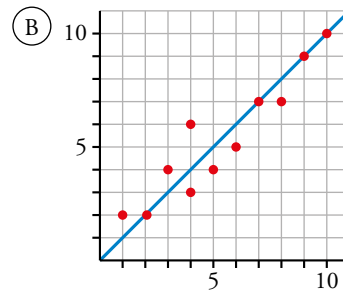
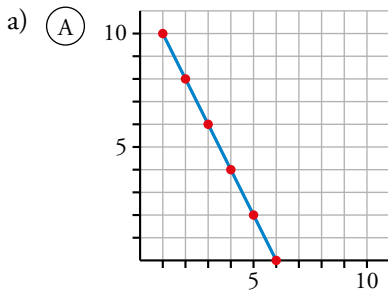


b) ¿Cuáles de ellas tienen correlación positiva y cuáles tienen correlación negativa?

c) Sin hacer cálculos, elige, de entre los siguientes valores, la correlación de cada una de las distribuciones:

0      0,64      1      -0,98      0,95      -1      -0,76

d) Una de ellas presenta relación funcional; ¿cuál? Da la expresión analítica de la función que relaciona las dos variables.



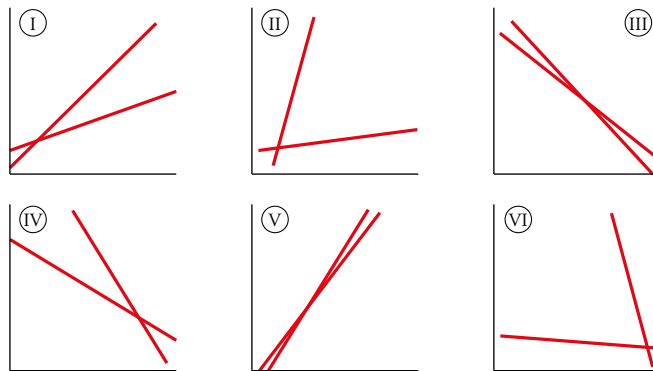
b) B y C tienen correlación positiva; A y D, negativa.

c) A  $\rightarrow -1$ ; B  $\rightarrow 0,95$ ; C  $\rightarrow 0,64$ ; D  $\rightarrow -0,76$

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



3 Cada una de estas seis distribuciones bidimensionales está representada por sus dos rectas de regresión:



Sus coeficientes de correlación son, no respectivamente:

-0,9      0,99      0,6      -0,2      -0,5      0,1

Asigna, razonadamente, a cada una su valor.

I → 0,6

II → 0,1

III → -0,9

IV → -0,5

V → 0,99

VI → -0,2

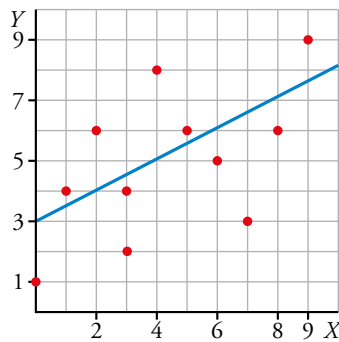
4 Representa la nube de puntos de esta distribución y estima cuál de estos tres puede ser el coeficiente de correlación:

a)  $r = 0,98$

b)  $r = -0,87$

c)  $r = 0,58$

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

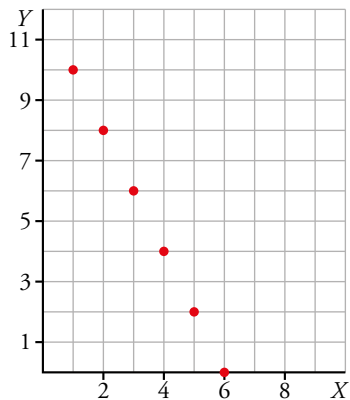


El coeficiente de correlación es  $r = 0,58$ .


5 Representa sobre papel cuadrulado la nube de puntos correspondiente a esta distribución:

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

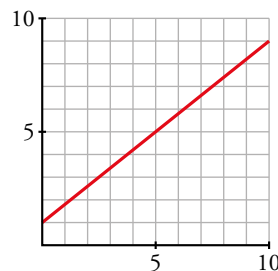
¿Cuál crees que es el coeficiente de correlación?



$r = -1$  porque están alineados.

6  **Sumamos.** [La creación de un conjunto de puntos que cumpla las condiciones indicadas permite trabajar la innovación (dimensión productiva)].

- a) En tu cuaderno, en una cuadrícula como esta, sitúa diez puntos de modo que estimes que su correlación sea 0,9 y una de sus rectas de regresión sea la que ves.
- b) Repite la experiencia para conseguir un coeficiente de correlación de 0,6.
- c) Haz lo mismo para un coeficiente de 0,3.

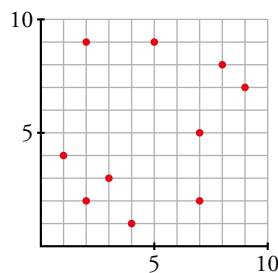
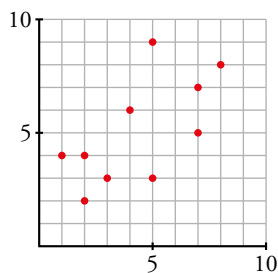
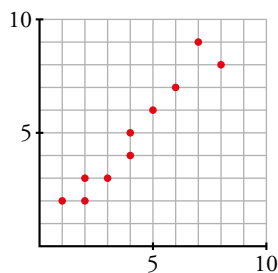


\* Atención: se pide estimar, pero no calcular.

a)  $r = 0,9$

b)  $r = 0,6$

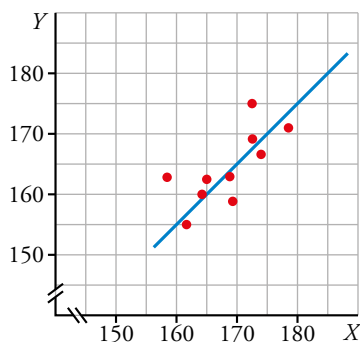
c)  $r = 0,3$



7 Las estaturas de 10 chicas,  $x$ , y las de sus madres,  $y$ , 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

- a) Representa estos valores mediante una nube de puntos.  
 b) Traza a ojo una recta de regresión y di si la correlación es positiva o negativa y más o menos fuerte de lo que esperabas.



La correlación es positiva y fuerte.

## Página 373

### Con fórmulas

8 Esta es la distribución bidimensional dada por la nube de puntos B del ejercicio 2:

$x$	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 mediante cálculos manuales:

- a)  $\bar{x}$ ,  $\bar{y}$ ,  $\sigma_x$ ,  $\sigma_y$ ,  $\sigma_{xy}$ .  
 b) El coeficiente de correlación,  $r$ . Interpretalo.  
 c) Las ecuaciones de las dos rectas de regresión.  
 d) Comprueba los resultados con la calculadora.

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

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

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  $Y$ :

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

9 a) Representa la nube de puntos correspondiente a la siguiente distribución bidimensional:

<b>x</b>	0	1	2	3	3	4	5	6	7	8	9
<b>y</b>	1	4	6	2	4	8	6	5	3	6	9

b) Comprueba con la calculadora que sus parámetros son:

$$\bar{x} = 4,4$$

$$\bar{y} = 4,9$$

$$\sigma_{xy} = 3,67$$

$$\sigma_x = 2,77$$

$$\sigma_y = 2,31$$

$$r = 0,57$$

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 4.

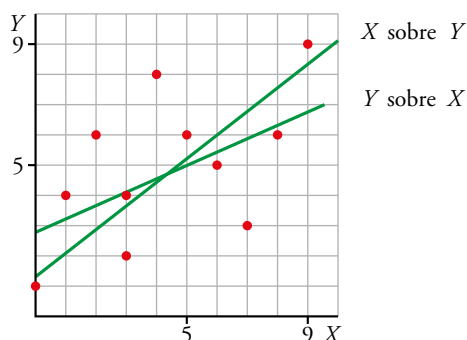
b) Se comprueba.

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  $Y$ :

$$m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} = \frac{3,67}{2,31^2} = 0,69 \rightarrow \frac{1}{m_{xy}} = 1,45 \rightarrow y = 4,9 + 1,45(x - 4,4) \rightarrow y = 1,45x - 1,48$$



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  $y = 10,5 + 3,2x$ .

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

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

c) 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».

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

b)  $\hat{y}(13)$  e  $\hat{y}(20)$  son estimaciones fiables,  $\hat{y}(30)$  es poco fiable e  $\hat{y}(100)$  es una estimación nada fiable.

c) 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].

**11** 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.

- d) Calcula  $\hat{y}(4,5)$ ,  $\hat{y}(11)$ ,  $\hat{y}(20)$  dilucidando cuánto de fiables son dichas estimaciones.

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

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

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

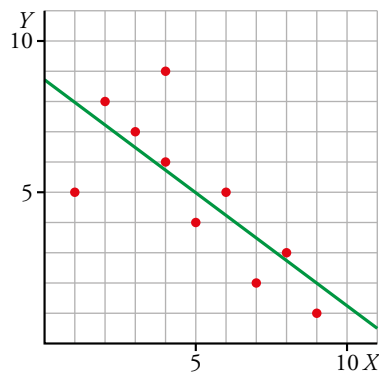
$\Sigma y^2 = 310$        $\sigma_y = \sqrt{\frac{301}{10} - 5^2} = 2,45$

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

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

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

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



- d)  $\hat{y}(4,5) = 5,56$
- $\hat{y}(11) = -3,04$
- $\hat{y}(20) = -14,95$

Como  $r = 0,76$ , la estimación para 4,5 la podemos considerar fiable, pero las de 11 y 20, que no están en el intervalo de datos, no se pueden considerar muy fiables.



- 15 ODS** **Meta 11.7.** [Tras el visionado del vídeo el docente puede plantear un debate sobre cuáles son las mejores alternativas de movilidad a los vehículos contaminantes].

El equipo de gobierno de una gran ciudad ha introducido una tasa para disminuir el tráfico en el centro. La tasa,  $x$ , se fijó en 4 €/día el primer año y ha subido 2 €/día cada año. La siguiente tabla muestra la media diaria de vehículos,  $y$ , en millones, que entran cada día a la ciudad, durante los ocho primeros años.

<b>x: TASA</b>	4	6	8	10	12	14	16	18
<b>y: N.º DE VEHÍCULOS</b>	2,4	2,5	2,2	2,3	2,0	1,8	1,7	1,5

- a) Calcula el coeficiente de correlación y la recta de regresión de  $Y$  sobre  $X$ .  
b) Si el gobierno quiere llegar a reducir el número medio de vehículos diarios a 1 millón, ¿qué tasa se estima que debe imponer? ¿Es fiable esta estimación?

a)  $\bar{x} = 11$

$$\sigma_x = \sqrt{\frac{1136}{8} - 11^2} = 4,58$$

$$\bar{y} = 2,05$$

$$\sigma_y = \sqrt{\frac{34,52}{8} - 2,05^2} = 0,335$$

$$\sigma_{xy} = \frac{168,6}{8} - 11 \cdot 2,05 = -1,475$$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-1,475}{4,58 \cdot 0,335} = -0,96$$

Busquemos la recta de regresión de  $Y$  sobre  $X$ :

$$y = 2,05 + \frac{-1,475}{4,58^2}(x - 11) = 2,05 - 0,07(x - 11)$$

- b) Buscamos la recta de regresión de  $X$  sobre  $Y$ :

$$x = \bar{x} + \frac{\sigma_{xy}}{\sigma_y^2}(y - \bar{y}) = 11 - \frac{1,475}{0,335^2}(y - 2,05) = 11 - 13,14(y - 2,05)$$

$$\hat{x}(1) = 11 - 13,14(1 - 2,05) = 24,8$$

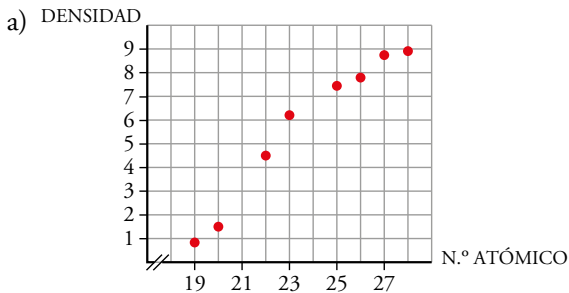
La tasa debe ser de 24,8 euros diarios.

Aunque  $|r| = 0,96$  es muy próximo a 1, el valor estudiado no está en el intervalo observado, así que la estimación hay que tomarla con reservas.

**16** La siguiente tabla relaciona el número atómico de varios metales,  $x$ , con su densidad,  $y$ , en  $\text{g/cm}^3$ :

ELEMENTO	K	Ca	Ti	V	Mn	Fe	Co	Ni
N.º ATÓMICO	19	20	22	23	25	26	27	28
DENSIDAD	0,86	1,54	4,50	6,11	7,44	7,88	8,86	8,91

- a) Representa los puntos, halla el coeficiente de correlación y calcula la ecuación de la recta de regresión de  $Y$  sobre  $X$ .
- b) Estima la densidad del cromo sabiendo que su número atómico es 24  $\rightarrow$  Cr (24).



$$r = 0,98$$

$$y = -16,69 + 0,95x$$

b)  $\hat{y}(24) = 4,9$

La densidad del cromo se estima en, aproximadamente, 6,11. Su valor real es 7,1.

## Página 374

**17** Esta tabla recoge tres variables socio-métricas de doce países:

- a) Halla manualmente el coeficiente de correlación entre las variables  $x$ - $y$  y entre las variables  $x$ - $z$ .
- b) ¿Qué conclusiones sacas de los resultados obtenidos?
- c) Comprueba los resultados con la calculadora.

PAÍS	x: RENTA PER CÁPITA (\$)	y: ÍNDICE DE NATALIDAD (‰)	z: EXPECTATIVA DE VIDA AL NACER (AÑOS)
A	873	50	49
B	402	48	50
C	536	47	54
D	869	44	57
E	1 171	41	61
F	636	36	64
G	1 417	35	59
H	2 214	31	63
I	1 334	28	63
J	769	26	61
K	1 720	25	64
L	2 560	24	70

a)  $x$ : renta per cápita (\$).

$y$ : índice de natalidad (‰).

$$r = -0,68$$

La correlación es negativa; es decir, si aumenta la renta per cápita, disminuye el índice de natalidad.

$x$ : renta per cápita (\$).

$z$ : expectativa de vida al nacer (años).

$$r = 0,82$$

La correlación es positiva; es decir, si aumenta la renta per cápita, aumenta la expectativa de vida al nacer.

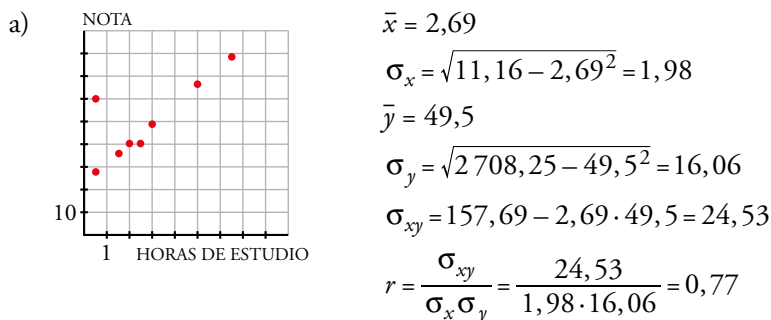
- b) La correlación es mayor en valor absoluto en el segundo caso, luego la renta per cápita es más determinante de la expectativa de vida al nacer que del índice de natalidad.



**18** La siguiente tabla muestra el tiempo,  $x$ , diario de estudio de matemáticas y la nota,  $y$ , en el último examen correspondiente a 8 estudiantes (100 es la nota máxima).

$x$	1,5	0,5	2,5	3	0,5	2	5	6,5
$y$	36	27	40	49	60	40	66	78

- Dibuja la correspondiente nube de puntos y calcula el coeficiente de correlación.
- Identifica en la nube un punto que se sale de la tendencia de los demás en el contexto del problema y no lo tengas en cuenta para calcular el coeficiente de correlación y la recta de regresión de  $Y$  sobre  $X$ .
- Si Ana estudió unas 8 horas, ¿qué nota estimas que le corresponderá? Ten en cuenta la recta de regresión que hallaste en el apartado b).
- Estudia la fiabilidad del resultado del apartado c).



b) Repetimos los cálculos sin tener en cuenta el punto (0,5; 60):

$$\bar{x} = 3$$

$$\sigma_x = 1,93$$

$$\bar{y} = 48$$

$$\sigma_y = 16,64$$

$$\sigma_{xy} = 31,93$$

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{31,93}{1,93 \cdot 16,64} = 0,994$$

El coeficiente de correlación se acerca mucho a 1 por lo que la correlación una vez eliminado el punto más alejado ha pasado a ser fuerte. Se puede decir que tenemos casi una relación funcional.

La recta de regresión será:

$$y = 48 + \frac{31,93}{1,93^2}(x - 3) \rightarrow y = 48 + 8,57(x - 3)$$

$$\hat{y}(8) = 48 + 8,57(8 - 3) = 90,85$$

Le corresponderán 90,85 puntos.

c) Aunque  $r$  es muy cercano a 1, el resultado no se puede considerar muy preciso porque el valor buscado no se encuentra dentro del rango donde teníamos datos.

**19** Elegimos seis automóviles al azar. Su antigüedad, en años, y el número de kilómetros que han rodado, en miles de kilómetros, están relacionados por la siguiente tabla:

ANTIGÜEDAD	1	2	4	4	5	6	7
KILÓMETROS RECORRIDOS	15	45	32	61	60	132	93

- Calcula la media y la desviación típica.
- Calcula el coeficiente de correlación e interprétalo.
- Si un automóvil tiene tres años, ¿cuántos kilómetros estimas que ha rodado?
- ¿Y si tiene cinco años? ¿Y diez? Justifica tus respuestas.

$x$ : antigüedad

$y$ : kilómetros recorridos

a)  $\bar{x} = 4,14$

$$\sigma_x = 1,96$$

$$\bar{y} = 62,57$$

$$\sigma_y = 36,37$$

b)  $r = 0,81$

Es positiva; es decir, si aumenta la antigüedad, aumentan los kilómetros recorridos. La correlación es fuerte porque  $r$  está próximo a 1.

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

$$y = 15,1x$$


$$\hat{y}(3) = 15,1 \cdot 3 = 45,3 \rightarrow \text{Se estima que recorre 45 300 km en 3 años.}$$

$$\hat{y}(5) = 15,1 \cdot 5 = 75,5 \rightarrow \text{Se estima que recorre 75 500 km en 5 años.}$$

$$\hat{y}(10) = 15,1 \cdot 10 = 151 \rightarrow \text{Se estima que recorre 151 000 km en 10 años.}$$


Esta última estimación es menos precisa que las anteriores, pues 10 no está en el intervalo  $[0, 7]$  del que se tienen los datos.

### Cuestiones teóricas

**20**  **Piensa y comparte en pareja.** [El alumnado puede plantear las razones por las que ha llegado a una determinada conclusión tal y como se explica en esta estrategia].

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 la nueva distribución?

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

**21**  [La justificación de la respuesta permite trabajar al alumnado la destreza expresión oral].

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.

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Como  $\sigma_x$  y  $\sigma_y$  son positivas, el signo de  $r$  es el mismo que el de  $\sigma_{xy}$ , luego si la covarianza es negativa,  $r$  también lo es.

$$m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}, \text{ cuyo signo es el mismo que el signo de } \sigma_{xy}.$$

$$m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} \text{ cuyo signo es el mismo que el signo de } \sigma_{xy}.$$

Luego si la covarianza es negativa,  $m_{yx}$  y  $m_{xy}$  son negativas.

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

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

**23** ¿Qué condición debe cumplir  $r$  para que las estimaciones hechas con la regresión sean fiables?

$|r|$  debe estar próximo a 1.

**24** Prueba que el producto de  $m_{yx}$  y  $\frac{1}{m_{xy}}$  es igual al coeficiente de determinación,  $r^2$ .

Sabemos que  $m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}$  y  $m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2}$ :

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

**25** Sabiendo que  $m_1$  y  $m_2$  son las pendientes de las dos rectas de regresión, expresa en función de ellas el coeficiente de correlación lineal.

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$m_{yx} \cdot \frac{1}{m_{xy}} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot \frac{\sigma_{xy}}{\sigma_y^2} = \frac{\sigma_{xy}^2}{\sigma_x^2 \sigma_y^2} = r^2$$

Luego  $r = \sqrt{m_{yx} \cdot \frac{1}{m_{xy}}}$

**26** La estatura media de 100 escolares es de 155 cm con una desviación típica de 15,5 cm.

La recta de regresión de la estatura respecto al peso es  $y = 80 + 1,5x$  ( $x$ : en kg;  $y$ : estatura en cm).

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 pasa por  $(\bar{x}, \bar{y})$ , luego el peso medio será la solución de la ecuación:

$$\bar{y} = 80 + 1,5\bar{x} \rightarrow 155 = 80 + 1,5\bar{x} \rightarrow \bar{x} = 50 \text{ kg}$$

b) El signo del coeficiente de correlación entre peso y estatura es el mismo que el de la pendiente de la recta de regresión, luego es positivo.

**27** ¿Verdadero o falso?

a) Si la pendiente de la recta de regresión de  $Y$  sobre  $X$  es 1, la correlación entre las dos variables es muy fuerte.

b) Si la pendiente de la recta de regresión de  $Y$  sobre  $X$  es negativa, la pendiente de la recta de regresión de  $X$  sobre  $Y$  también es negativa.

c) En una relación funcional lineal las dos rectas de regresión coinciden.

d) Cuanto más fuerte sea la correlación entre dos variables, mayor es su coeficiente de determinación.

e) En una distribución bidimensional de dos puntos distintos el coeficiente de correlación es 1.

f) Imagina dos nubes de puntos, A y B, con el mismo coeficiente de correlación, 0,98. La distribución A tiene 8 puntos y la B, 100. Si añadimos en cada una un nuevo punto que se separa «mucho» de la recta de regresión, el coeficiente de correlación de A disminuirá mucho más que el de B.

a) Falso. Si la pendiente de la recta de regresión de  $Y$  sobre  $X$  es 1, sabemos que la covarianza es igual a la varianza de  $x$ , pero no que  $r$  esté próximo a 1.

b) Verdadero, porque  $m_{yx} \cdot \frac{1}{m_{xy}} = r^2 > 0$

El producto es un número positivo, luego las dos pendientes tienen que tener el mismo signo.

c) Verdadero. En una relación funcional,  $r = 1$ .

$$r = \sqrt{m_{yx} \cdot \frac{1}{m_{xy}}} \rightarrow 1 = m_{yx} \cdot \frac{1}{m_{xy}} \rightarrow m_{xy} = m_{yx}$$

Como las dos rectas pasan por  $(\bar{x}, \bar{y})$  y tienen la misma pendiente, coinciden.

d) Verdadero, porque  $0 \leq r^2 \leq 1$ .

Si la correlación es muy fuerte,  $|r|$  está próximo a 1, luego  $r^2$  se aproxima a 1.

e) Verdadero, porque la recta pasará por los dos puntos que tenemos. Por dos puntos pasa una única recta, y  $r$  será exactamente 1. Se dice que la correlación es perfecta.

f) Verdadero. Añadiendo un punto a una nube de 100 puntos la importancia de este punto queda más disimulada que entre 8 puntos. Cuantos menos puntos tiene una nube, más notoriedad tiene cada punto en ella.

## Página 375

### Para profundizar

**28** En una autoescuela, cada alumno realiza un total de 80 tests repartidos en 4 tandas de 20. La siguiente tabla relaciona las variables número de la tanda ( $x$ ) y número de fallos ( $y$ ):

Por ejemplo: En la tercera tanda, en 12 de los tests se encontraron de 0 a 3 fallos; en 7, de 4 a 7 fallos...

$x \backslash y$	0-3	4-7	8-11	1-15
1	0	4	11	5
2	1	10	7	2
3	12	7	1	0
4	16	4	0	0

a) Calcula manualmente el coeficiente de correlación y halla la ecuación de la recta de regresión de  $Y$  sobre  $X$ .

b) ¿Cuántos fallos se estima que tendrá un alumno en la primera tanda? ¿Y en la segunda? ¿Y en la última?

c) Comprueba los resultados con la calculadora.

a)

TANDA = $x_i$ \ FALLOS = $y_j$	1,5	5,5	9,5	13,5	
1	0	4	11	5	20
2	1	10	7	2	20
3	12	7	1	0	20
4	16	4	0	0	20
	29	25	19	7	80

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2$	$x_i^2 \cdot f_i$
1	20	20	1	20
2	20	40	4	80
3	20	60	9	180
4	20	80	16	320
	80	200		600

$$\bar{x} = \frac{200}{80} = \frac{5}{2} = 2,5$$

$$\sigma_x = \sqrt{\frac{600}{80} - 2,5^2} = 1,12$$

$y_i$	$f_i$	$y_i \cdot f_i$	$y_i^2$	$y_i^2 \cdot f_i$
1,5	29	43,5	2,25	65,25
5,5	25	137,5	30,25	756,25
9,5	19	180,5	90,25	1714,75
13,5	7	94,5	182,25	1275,75
	80	456		3812

$$\bar{y} = \frac{456}{80} = \frac{57}{10} = 5,7$$

$$\sigma_y = \sqrt{\frac{3812}{80} - 5,7^2} = 3,89$$

$$\Sigma x \cdot y \cdot f = 876$$

$$\sigma_{xy} = \frac{876}{80} - 2,5 \cdot 5,7 = -3,3$$

$$r = \frac{-3,3}{1,12 \cdot 3,89} = -0,76$$

$$m_{yx} = \frac{-3,3}{1,12^2} = -2,63$$

Recta de regresión de  $Y$  sobre  $X$ :  $y - 5,7 = -2,63(x - 2,5)$

b)  $\hat{y}(1) = -2,63(1 - 2,5) + 5,7 = 9,645$

Se estima que tendrá entre 9 y 10 fallos en la primera tanda.

$$\hat{y}(2) = -2,63(2 - 2,5) + 5,7 = 7,015$$

Se estima que tendrá 7 fallos en la segunda tanda.

$$\hat{y}(4) = -2,63(4 - 2,5) + 5,7 = 1,755$$

Se estima que tendrá entre 1 y 2 fallos en la cuarta tanda, más veces 2 fallos que 1.

**29** En un estudio realizado a los trabajadores de una cadena de fabricación de piezas de coches sobre su productividad quincenal, se relacionan las horas trabajadas ( $x$ ) con las unidades producidas ( $y$ ).

Sabemos que:

- La recta de regresión de  $Y$  sobre  $X$  es:

$$y = 3,47x + 32,01$$

- La recta de regresión de  $X$  sobre  $Y$  es:

$$y = 3,81x + 5,36$$

- El intervalo de horas empleadas por los trabajadores es  $[60, 85]$ .

a) Halla  $\bar{x}$ ,  $\bar{y}$  y el coeficiente de correlación.

b) Si un operario trabaja 70 horas en una quincena, ¿cuántas unidades se estima que produzca? ¿Cómo de fiable es esta estimación? ¿Y si trabaja en total 40 horas? ¿Y si fueran 120 horas?

c) Si un empleado esta quincena ha llegado a producir 300 piezas, ¿cuántas horas se estima que ha trabajado?

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

$$\left\{ \begin{array}{l} y = 3,47x + 32,01 \\ y = 3,81x + 5,36 \end{array} \right\} \rightarrow \bar{x} = 78,38; \bar{y} = 304$$

$$r^2 = \frac{m_{yx}}{m_{xy}} = \frac{3,47}{3,81} = 0,91 \rightarrow r = \sqrt{0,91} = 0,95394$$

b)  $\hat{y}(70) = 3,47 \cdot 70 + 32,01 = 274,91$

Se estima que el operario produzca unas 275 unidades trabajando 70 horas.

Como  $r$  es muy próximo a 1 y, además, 70 está en el intervalo de horas empleadas, la estimación es muy fiable.

$\hat{y}(40) = 3,47 \cdot 40 + 32,01 = 170,81$

Se estima que el operario produzca casi 171 unidades trabajando 40 horas. Esta estimación no es tan fiable como la anterior porque  $40 \notin [60, 85]$ .

$\hat{y}(120) = 3,47 \cdot 120 + 32,01 = 448,41$

Se estima que el operario produzca alrededor de 448 unidades trabajando 120 horas. Esta estimación no es muy fiable porque  $120 \notin [60, 85]$ .

c)  $300 = 3,81x + 5,36 \rightarrow x = 77,33$

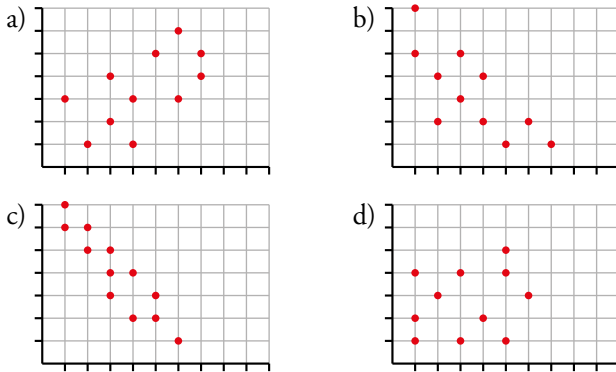
Se estima que ha trabajado entre 77 y 78 horas.

## AUTOEVALUACIÓN

C.E.: CE 5.1. (EA 5.1.1.-EA 5.1.2.) CE 5.2. (EA 5.2.1.-EA 5.2.2.-EA 5.2.3.-EA 5.2.4.) CE 5.3. (EA 5.3.1.)

Página 375

### 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

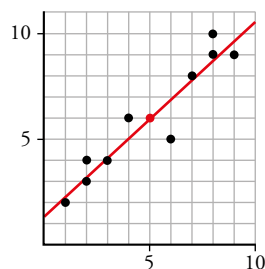
La correlación de a) es positiva, y las de b) y c), negativas. En d) no se aprecia correlación. La correlación de c) es más fuerte que la de b). Por tanto:

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

### 2 Representa esta distribución bidimensional:

<b>x</b>	1	2	2	3	4	6	7	8	8	9
<b>y</b>	2	4	3	4	6	5	8	9	10	9

- Calcula los parámetros  $\bar{x}$ ,  $\bar{y}$ ,  $\sigma_x$ ,  $\sigma_y$  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 distribución bidimensional es  $y = 1,6x - 3$ . Sabemos que  $\bar{x} = 10$  y  $r = 0,8$ .

- a) Calcula  $\bar{y}$ .  
 b) 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  $Y$ .

a) 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  $\bar{x}$ .

c) 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 mensual 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

- a) Calcula la recta de regresión de  $Y$  sobre  $X$ .  
 b) Halla el coeficiente de correlación entre el consumo y la renta.  
 c) ¿Qué predicción podemos hacer sobre el consumo de energía per cápita de un país cuya renta per cápita es de 4 400 €? (Recuerda que en la tabla se da la renta en miles de euros.)  
 d) Estima la renta per cápita que tendrá un país en el cual el consumo de energía per cápita ha sido de 9 000 kWh.  
 e) ¿Cómo de fiables son estas estimaciones?

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

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

a) Recta de regresión de  $Y$  sobre  $X$ :  $y = 4,37 + \frac{2,51}{2,46^2}(x - 8,63) \rightarrow y = 0,80 + 0,41x$

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 kWh por habitante.

d)  $9 = 0,80 + 0,41\hat{x}(9) \rightarrow \hat{x}(9) = 20 \rightarrow$  Se estima una renta per cápita de 20 000 €.

e) En la primera estimación (apartado c), el valor  $x = 4,4$  es próximo a los valores de la tabla. Como el coeficiente de correlación es alto (0,93), la estimación es razonablemente fiable. En la segunda estimación (apartado d), el valor  $y = 9$  es lejano a los de la tabla. Por tanto, la estimación es poco fiable.