Resuelve

Página 215

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:

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.

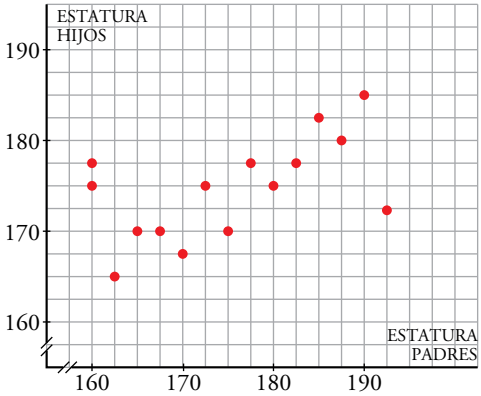
Número de personas en cada casa - Coste del recibo de la luz → Estadística, porque el número de personas en cada casa no nos permite saber exactamente el coste del recibo de la luz. Hay correlación positiva. Normalmente, cuantas más personas hay en una casa, más luz se consume.

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 corresponde a un chico. La abscisa es la estatura de su padre, y la ordenada, su propia altura:



a) Identifica a Guillermo y Gabriel, hermanos de buena estatura, cuyo padre es bajito.

b) Identifica a Sergio, de estatura normalita, cuyo padre es muy alto.

c) ¿Podemos decir que hay una cierta relación entre las estaturas de estos 15 chicos y las de sus padres?

a) Guillermo y Gabriel están representados mediante los puntos (160, 175) y (160; 177,5).

b) Sergio está representado con el punto (192,5; 172,5).

c) Sí; en general, cuanto más alto sea el padre, más altos son los hijos.

1 Distribuciones bidimensionales. Nubes de puntos

Página 217

1 ¿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

Página 219

1 ¿Verdadero o falso?

a) Cuanto más próximos estén a una recta los puntos de una distribución bidimensional, más fuerte es su correlación lineal.

b) Si la recta de regresión tiene pendiente negativa, la correlación lineal es negativa.

c) Si los puntos de la nube no se aproximan a ninguna recta, entonces las variables están incorreladas.

a) 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.

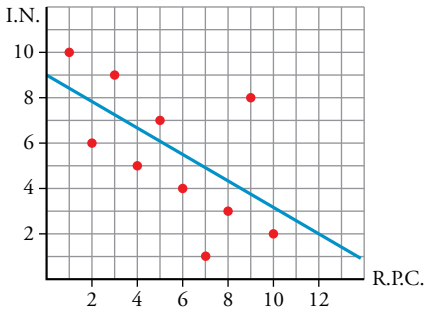
b) Verdadero. Una recta de pendiente negativa indica, como el signo del coeficiente de correlación, que al aumentar una variable, la otra disminuye.

c) Verdadero.

2 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

Página 221

1 ¿Verdadero o falso?

a) El signo de la correlación (r) coincide con el de la covarianza (σ_{xy}).

b) Si cambiamos las unidades en que se expresa la variable x, entonces se modifican los valores de $\bar{x}$, σ_x, σ_{xy} y r.

c) Aunque cambiemos las unidades en que se da la variable x (o la y, o ambas) el valor de la correlación, r, no se modifica.

a) Verdadero, $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$; como σ_x y σ_y son positivas, el signo de r es el de σ_{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 con las que hemos trabajado en el epígrafe anterior:

Salto de altura - Salto con pértiga

Salto de altura - 1 500 m lisos

Salto de altura - Lanzamiento de peso

Hazlo también con una calculadora con modo LR.

- 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

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

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

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

$$\sigma_{xy} = \frac{204}{8} - 4{,}5 \cdot 4{,}5 = 4{,}75$$

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

- 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

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

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

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

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

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

- 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

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

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

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

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

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

4 Recta de regresión

Página 223

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) Verdadero. Como r es muy grande, la distancia de los puntos a la recta es muy pequeña o nula.

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

Página 224

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°.

a) $\left.\begin{cases} y = 0{,}8x - 67 \\ y = 1{,}1x - 121 \end{cases}\right\}$ $x = 180{,}0$; $y = 77{,}0$ $\rightarrow$ 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) Verdadero, por el razonamiento anterior.

c) Verdadero. Se puede observar en las gráficas de esta página.

6 Tablas de contingencia

Página 225

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 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 226

3 Comprueba que la siguiente tabla corresponde a la distribución de x condicionada a $y \in$ {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_i	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

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

$$\sigma_x = \sqrt{\frac{1\,609\,373,5}{500} - 40,073^2} = 28,471$$

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

Página 228

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

y_i \ x_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

De los datos obtenemos las siguientes tablas:

y_i \ x_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

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

$$\bar{x} = \frac{194}{154} = 1,26$$

$$\sigma_x = \sqrt{\frac{459}{154} - 126^2} = 1,18$$

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

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

$$\sigma_y = \sqrt{\frac{526}{154} - 136^2} = 1,25$$

$$\sigma_{xy} = \frac{400}{154} - 1,36 \cdot 1,18 = 0,99$$

$$r = \frac{0,99}{1,25 \cdot 1,18} = 0,67$$

Ejercicios y problemas resueltos

Página 229

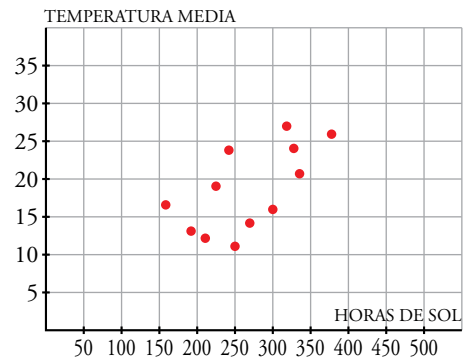
1. Relación funcional y relación estadística

Hazlo tú. Haz lo mismo con las variables *Horas de sol - Temperatura media en Almería*.

	E	F	Mr	Ab	My	Jn	Jl	Ag	S	O	N	D
HORAS DE SOL	216	251	271	300	335	329	377	321	244	225	159	194
TEMPERATURA MEDIA	12,2	10,8	14,5	15,6	20,4	24,1	25,8	26,7	23,7	19,3	16,4	13,5

a) Es una distribución bidimensional en la que se relacionan las variables x: horas de sol e y: temperatura media en Almería, correspondientes a un año.

b)



c) Es una relación estadística, el número de horas de sol no determina la temperatura media.

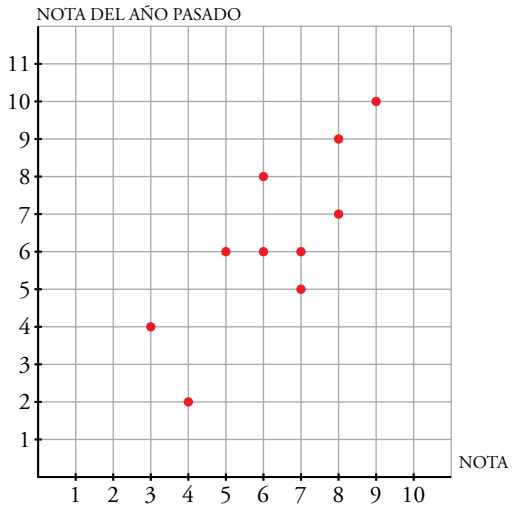
2. Análisis gráfico de una distribución bidimensional

Hazlo tú. Estudia la correlación entre la NOTA de un exámen de Matemáticas y las otras variables que aparecen en la siguiente tabla:

NOTA	3	4	5	6	6	7	7	8	8	9
NOTA DEL AÑO PASADO	4	2	6	6	8	5	6	9	7	10
COCIENTE INTELECTUAL	120	102	118	105	95	110	103	107	115	125

- x: Nota

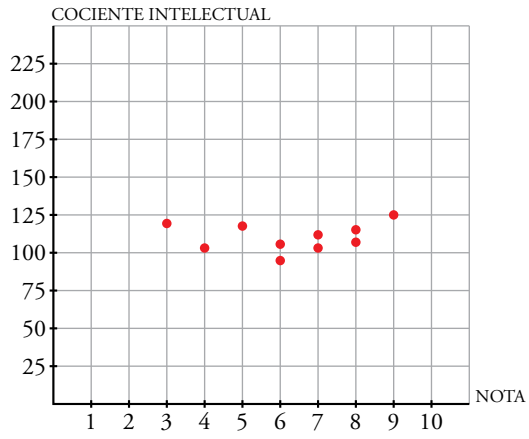
  y: Nota del año pasado



Hay una correlación positiva bastante fuerte.

- x: Nota

  y: Cociente intelectual



Hay una correlación positiva bastante fuerte.

Ejercicios y problemas guiados

Página 231

1. Dos rectas de regresión. Estimaciones

La siguiente tabla relaciona las variables

x: gastos en publicidad (miles de euros)

y: ventas (miles de euros)

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

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

a) Hallar las dos rectas de regresión.

b) Efectuar la estimación $\hat{y}(5{,}5)$ y explicar su significado.

c) 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).

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

a) Calcular el número medio de discos vendidos.

b) ¿Cuál es el coeficiente de correlación?

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

d) Si un grupo musical vende 18 000 discos, ¿qué número de conciertos se prevé para él?

a) y b)

CONC. (y_i) \ DISCOS (x_i)	20	35	60	
2,5	3	0	0	3
7,5	1	4	1	6
15	0	1	5	6
	4	5	6	

x_i	f_i	$x_i \cdot f_i$	x_i^2	$x_i^2 \cdot f_i$
2,5	3	7,5	6,25	18,75
7,5	6	45	56,25	337,5
15	6	90	225	1 350
	15	142,5		1 706,25

$$\bar{x} = \frac{142,5}{15} = 9,5$$

$$\sigma_x = \sqrt{\frac{1\,706,25}{15} - 9,5^2} = 4,85$$

y_i	f_i	$y_i \cdot f_i$	y_i^2	$y_i^2 \cdot f_i$
20	4	80	400	1 600
35	5	175	1 225	6 125
60	6	360	3 600	21 600
	15	615		29 325

$$\bar{y} = \frac{615}{15} = 41$$

$$\sigma_y = \sqrt{\frac{29\,325}{15} - 41^2} = 16,55$$

$$\Sigma x \cdot y \cdot f = 6\,825$$

$$\sigma_{xy} = \frac{6\,825}{15} - 9,5 \cdot 41 = 65,5$$

$$r = \frac{65,5}{4,85 \cdot 16,55} = 0,82$$

c) $m_{yx} = \dfrac{65,5}{4,85^2} = 2,78$

Recta de regresión de Y sobre X:

$$y - 41 = 2,78(x - 9,5)$$

d) $\hat{y}(18) = 2,78(18 - 9,5) + 41 = 64,63$

Ejercicios y problemas propuestos

Página 232

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 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***

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?).

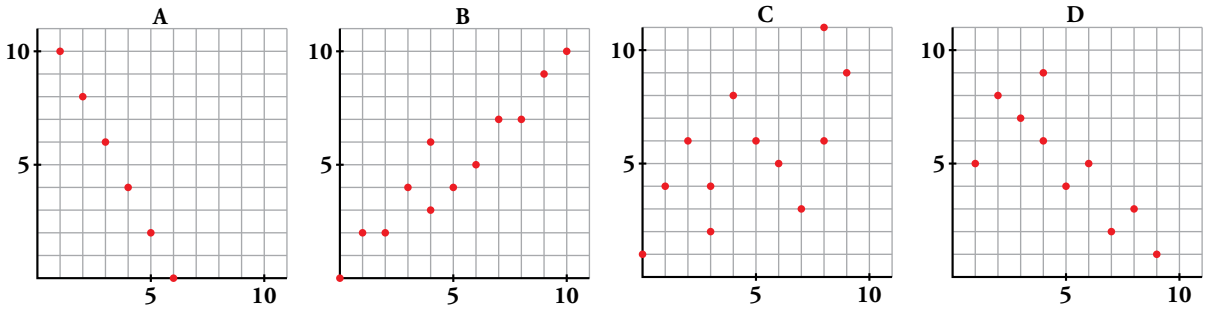
d) Aunque lo parezca *a priori*, seguramente la relación no es funcional. Es una correlación positiva fuerte.

e) Correlación positiva.

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

g) Correlación positiva.

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



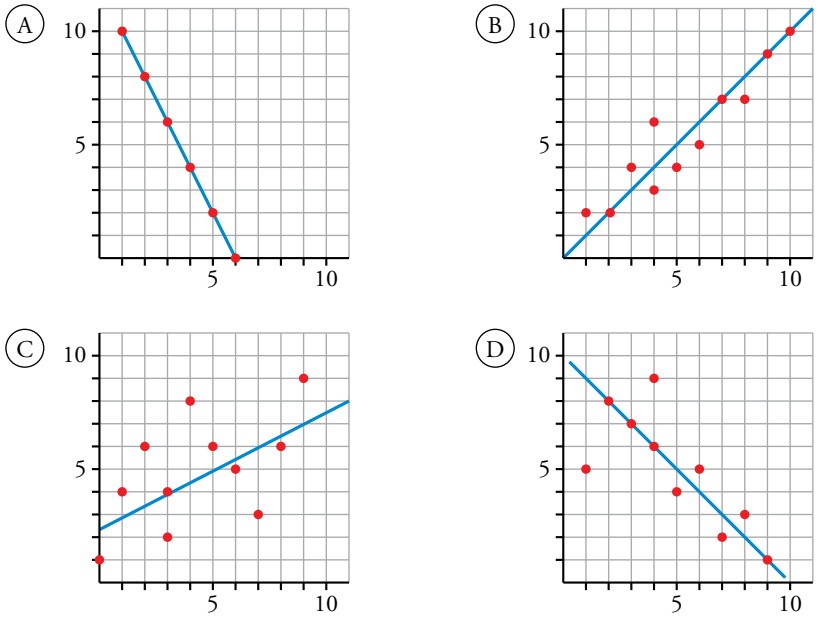
b) ¿Cuáles de ellas tienen correlación positiva y cuáles tienen correlación negativa?

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.

a)

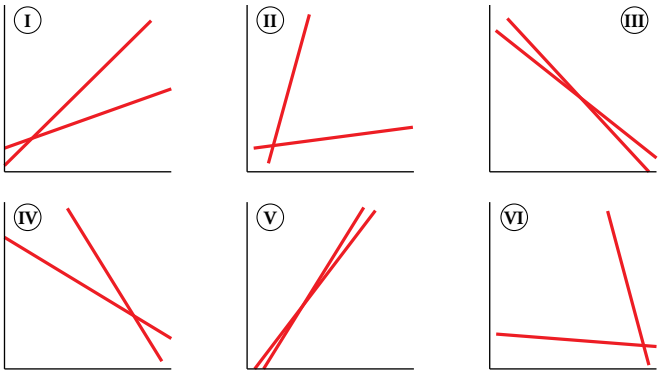


b) B y C tienen correlación positiva; A y D, negativa.

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 $\rightarrow$ 0,6

II $\rightarrow$ 0,1

III $\rightarrow$ –0,9

IV $\rightarrow$ –0,5

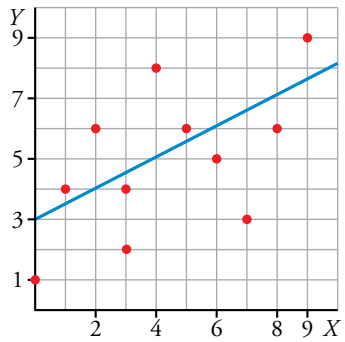
V $\rightarrow$ 0,99

VI $\rightarrow$ –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

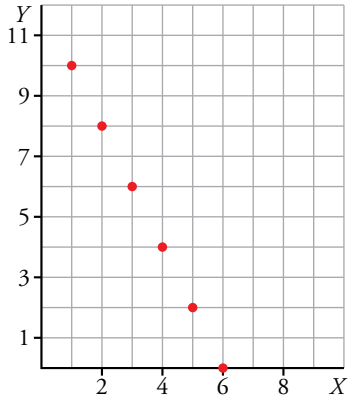


c) $r = 0{,}58$

5 **Representa sobre papel cuadriculado 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?



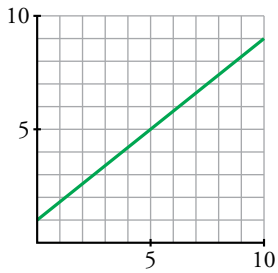
$r = -1$ porque están alineados.

6 **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.**

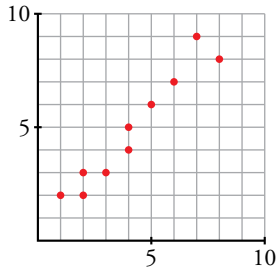
b) Repite la experiencia para conseguir un coeficiente de correlación de 0,6.

c) Haz lo mismo para un coeficiente de 0,3.

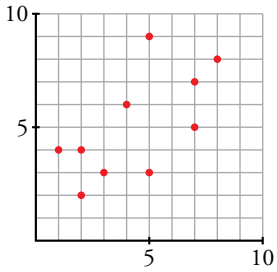
* ***Atención: se pide estimar, pero no calcular.***



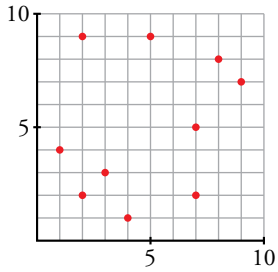
a) $r = 0{,}9$



b) $r = 0{,}6$



c) $r = 0{,}3$

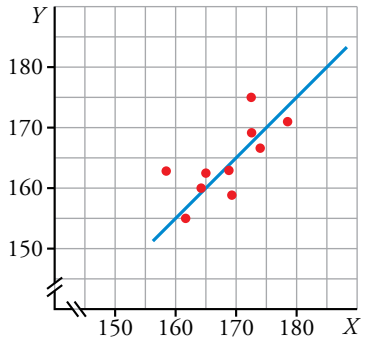


7 **Las estaturas de 10 chicas y las de sus madres son:**

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

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 233

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:

a) $\bar{x}$, $\bar{y}$, σ_x, σ_y, σ_{xy}.

b) El coeficiente de correlación, r. **Interprétalo.**

c) Las ecuaciones de las dos rectas de regresión.

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

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

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

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

b) $r = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y} = 0{,}95$. 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:**

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:

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

$\sigma_x = 2{,}77$ $\quad$ $\sigma_y = 2{,}31$ $\quad$ $r = 0{,}58$

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.

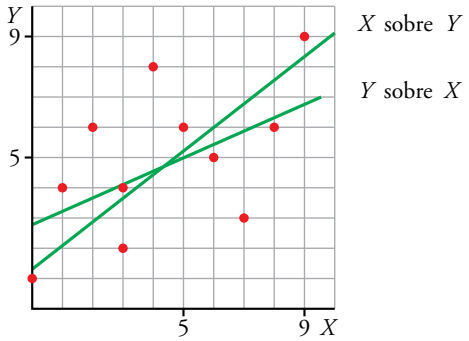
b) Se comprueba.

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

b) 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)

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

b) $n = 10$ $\qquad \Sigma x = 49 \qquad \bar{x} = \dfrac{49}{19} = 4{,}9$

$\Sigma y = 50 \qquad \bar{y} = \dfrac{50}{10} = 5$

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

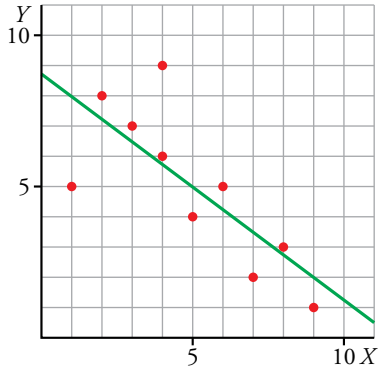
$\Sigma y^2 = 310 \qquad \sigma_y = \sqrt{\dfrac{301}{10} - 5^2} = 2{,}45$

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

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

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.

12 **Calcula las correlaciones correspondientes a las nubes de puntos que inventaste en el ejercicio 6. Comprueba si las correlaciones obtenidas se parecen a las que pretendías alcanzar.**

a) $r = 0{,}97$

b) $r = 0{,}64$

c) $r = 0{,}25$

Para resolver

13 **La siguiente tabla recoge los datos económicos de algunas de las películas más rentables de un año (las cantidades están dadas en millones de euros):**

***x*: GASTOS**	**18**	**15**	**20**	**11**	**10**	**6**	**6**	**14**	**16**	**12**
***y*: RECAUDACIÓN**	**93**	**83**	**80**	**47**	**46**	**44**	**36**	**34**	**33**	**26**

a) Halla el coeficiente de correlación.

b) Obtén la recta de regresión de Y sobre X y estima qué recaudación cabe esperar si se invierten 30 millones de euros en una película.

a) $r = 0{,}6$

b) $y = 3{,}05x + 13{,}05$

$\hat{y}(30) = 104{,}55$

Cabe esperar que se recauden 104,55 millones de euros.

14 **Un excursionista, en diez marchas distintas, toma las siguientes medidas:**

***x*: altura de lugar (en m)**

***y*: presión atmosférica (en mm Hg)**

***z*: número de pulsaciones en reposo**

x	**0**	**184**	**231**	**481**	**730**	**911**	**1 343**	**1 550**	**1 820**	**2 184**
y	**760**	**745**	**740**	**720**	**700**	**685**	**650**	**630**	**610**	**580**
z	**73**	**78**	**75**	**78**	**83**	**80**	**89**	**80**	**85**	**92**

Halla el coeficiente de correlación y la recta de regresión para la distribución x-y y para la distribución x-z y analiza los resultados.

a) x: altura del lugar (en m)

y: presión atmosférica (en mm Hg)

$r = -0{,}99$

Recta de regresión: $y = -0{,}08x + 759$

Hay casi una relación funcional entre la altura de un lugar y su presión atmosférica. Además, cuando aumenta la altura, disminuye la presión.

b) x: altura del lugar (en m)

z: número de pulsaciones en reposo

$r = 0{,}85$

Recta de regresión: $y = 6{,}87x + 74{,}8$

Hay una correlación fuerte entre la altura de un lugar y el número de pulsaciones, en reposo, de una persona. Además, cuando aumenta la altura, aumentan las pulsaciones en reposo.

15 **En la siguiente tabla se consignan los goles a favor (x) y los goles en contra (y) al final del campeonato de liga de 1.ª división de fútbol de 10 equipos:**

x	95	102	64	54	62	60	62	46	45	35	54	41
y	21	33	44	44	61	55	53	55	46	42	68	56

Halla el coeficiente de correlación entre las dos variables y analiza lo que has obtenido.

x: goles a favor

y: goles en contra

$r = -0{,}61$

r es negativo, luego cuantos más goles a favor tiene un equipo, menos goles en contra tiene. La correlación no es muy fuerte, por lo que no es demasiado fiable estimar los goles en contra sabiendo los goles a favor.

16 **La siguiente tabla relaciona el número atómico de varios metales, x, con su densidad, y:**

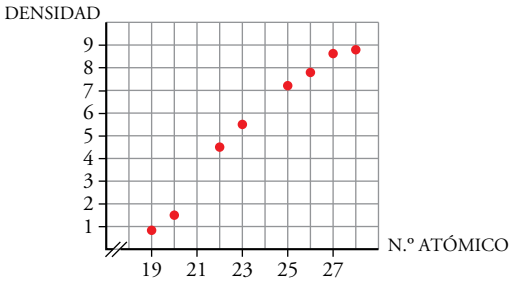
ELEMENTO	K	Ca	Ti	V	Mn	Fe	Co	Ni
N.° ATÓMICO	19	20	22	23	25	26	27	28
DENSIDAD	0,86	1,54	4,50	5,60	7,11	7,88	8,70	8,80

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 → Cr (24).

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

a)



$r = 0{,}98$

$y = -16{,}5 + 0{,}93x$

b) $\hat{y}(24) = 5{,}86$

La densidad del cromo se estima en, aproximadamente, 5,86. Su valor real es 7,1.

c) $\hat{y}(21) = 3{,}06$

La densidad del escandio se estima en, aproximadamente, 3,01. Su valor real es 2,9.

Página 234

17 **La siguiente tabla relaciona tres variables sociométricas relativas a doce países:**

PAÍS	*x*: RENTA PER CÁPITA ($)	*y*: ÍNDICE DE NATALIDAD (‰)	*z*: ESPECTATIVA 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) Halla el coeficiente de correlación entre las variables *x-y* y entre las variables *x-z*.

b) ¿Qué conclusiones sacas de los resultados obtenidos?

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 recoge datos de ocho alumnos de secundaria de un centro escolar, relativos al número de horas por día que ven la televisión y la nota media obtenida en la última evaluación:**

x: N.º DE HORAS DE TV	0	2	1	3	4	2	1	3
y: NOTA MEDIA	8	6	8	6	1	4	7	3

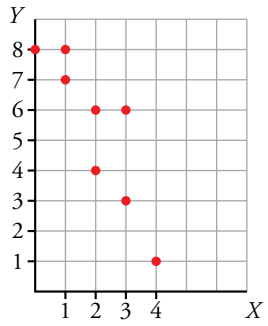
a) Representa gráficamente los datos. ¿Permite el análisis gráfico deducir el signo del coeficiente de correlación de *Y* sobre *X*? ¿Y el de *X* sobre *Y*?

b) Halla el coeficiente de correlación de las dos variables.

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

d) ¿Qué nota media puede esperar quien ve la televisión tres horas y media diarias? ¿Y quien la ve cinco horas?

a)



Los dos signos son negativos; es decir, si aumentan las horas por día que ven la televisión, disminuye la nota media obtenida en la última evaluación.

b) $r = -0{,}87 \rightarrow$ correlación fuerte

c) Recta de regresión de Y sobre X: $y = -1{,}67x + 8{,}70$

d) $\hat{y}(3{,}5) = -1{,}67 \cdot 3{,}5 + 8{,}70 = 2{,}86$

$\hat{y}(5) = -1{,}67 \cdot 5 + 8{,}70 = 0{,}35$

Si ve la televisión tres horas y media diarias, cabe esperar que saque un 2,86.

Si ve la televisión cinco horas, cabe esperar que saque un 0,35.

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

a) Calcula la media y la desviación típica de las dos variables que intervienen.

b) Calcula el coeficiente de correlación e interprétalo.

c) Si un automóvil tiene tres años, ¿cuántos kilómetros estimas que ha rodado?

d) ¿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) y d) Recta de regresión de Y sobre X:

$y = 15{,}1x$

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

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

$\hat{y}(10) = 15{,}1 \cdot 10 = 151 \rightarrow$ 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 **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.

21 **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 σ_x y σ_y son positivas, el signo de r es el mismo que el de σ_{xy}, luego si la covarianza es negativa, r también lo es.

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

$m_{xy} = \dfrac{\sigma_{xy}}{\sigma_y^2}$ cuyo signo es el mismo que el signo de σ_{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 recta de regresión sean fiables?**

$|r|$ debe estar próximo a 1.

24 **Prueba que el producto de los coeficientes de regresión es igual al cuadrado del coeficiente de correlación.**

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

25 **Explica cómo se calculan $\bar{x}$ e $\bar{y}$ a partir de las dos rectas de regresión de una distribución bidimensional. Aplícalo a este caso:**

La recta de regresión de Y sobre X es: $y = 8{,}7 - 0{,}76x$

La recta de regresión de X sobre Y es: $y = 11{,}36 - 1{,}3x$

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

$$\left.\begin{cases} y = 8{,}7 - 0{,}76x \\ y = 11{,}36 - 1{,}3x \end{cases}\right\} \rightarrow x = 4{,}9259;\ y = 4{,}9563 \begin{cases} \bar{x} = 4{,}9259 \\ \bar{y} = 4{,}9563 \end{cases}$$

26 **Explica cómo se halla el coeficiente de correlación a partir de las dos rectas de regresión de una distribución bidimensional. Aplícalo al caso del ejercicio anterior.**

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

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

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

En el ejercicio anterior:

$m_{yx} = -0{,}76;\ m_{xy} = -1{,}3$

$$r = \sqrt{-0{,}76 \cdot \frac{1}{-1{,}3}} = 0{,}7646$$

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

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

28 **¿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 las variables x e y, mayor es r^2.

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 $0 \le r^2 \le 1$.

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

Página 235

Para profundizar

29 **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):**

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

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…

a) Calcula 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?

a)

TANDA = x_i \ FALLOS = y_i	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

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

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

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	1 714,75
13,5	7	94,5	182,25	1 275,75
	80	456		3 812

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

$$\sigma = \sqrt{\frac{3\,812}{80} - 5,7^2} = 3,89$$

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

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

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

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

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.

30 **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:

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

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

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 $120 \notin [60, 85]$.

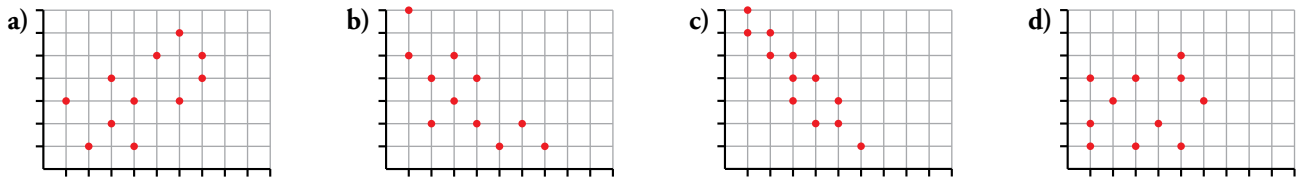
c) $300 = 3{,}81x + 5{,}36 \rightarrow x = 77{,}33$

Se estima que ha trabajado entre 77 y 78 horas.

Autoevaluación

Página 235

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) → 0,6

b) → –0,7

c) → –0,9

d) → 0,2

2 Representa esta distribución bidimensional:

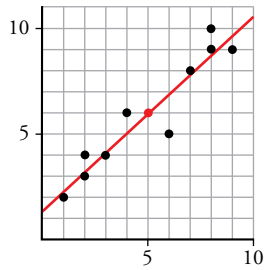
x	1	2	2	3	4	6	7	8	8	9
y	2	4	3	4	6	5	8	9	10	9

a) Calcula los parámetros $\bar{x}$, $\bar{y}$, σ_x, σ_y y σ_{xy}.

b) Halla el coeficiente de correlación.

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

d) Estima el valor de y para $x = 5$ y para $x = 10$. ¿Son "buenas" estas estimaciones?



a) $\bar{x} = 5$, $\bar{y} = 6$

$\sigma_x = 2{,}8$; $\sigma_y = 2{,}7$; $\sigma_{xy} = 7{,}1$

b) $r = 0{,}95$

c) $y = 0{,}91x + 1{,}45$

d) $\hat{y}(5) = 6$, $\hat{y}(10) = 10{,}55$

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 $r = 0{,}8$.**

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:

(Coef. Y sobre X) · (Coef. X sobre Y) = r^2

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

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

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: $y = 4{,}37 + \dfrac{2{,}51}{2{,}46^2}(x - 8{,}63) \rightarrow y = 0{,}80 + 0{,}41x$

b) Coeficiente de correlación: $r = \dfrac{2{,}51}{1{,}09 \cdot 2{,}46} = 0{,}93$

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.