

Resol

Pàgina 215

Relació funcional i relació estadística

En cadascun dels casos següents has de dir si, entre les dues variables que s'esmenten, hi ha relació funcional o relació estadística (correlació) i, en aquest darrer cas, indicar si és positiva o negativa:

a) En un conjunt de famílies:

Estatuta mitjana dels pares - Estatuta mitjana dels fills

b) Entre els països del món respecte a Catalunya:

Volum d'exportació - Volum d'importació

c) En els països del món:

Taxa de mortalitat infantil - Metges per cada 1 000 habitants

d) A les cases d'una ciutat:

kWh consumits durant el gener - Cost del rebut de la llum

Nombre de persones en cada casa - Cost del rebut de la llum

e) En els equips de futbol:

Posició en finalitzar la Lliga - Nombre de partits perduts

Posició al finalitzar la Lliga - Nombre de partits guanyats

a) Estadística, perquè l'estatura mitjana dels pares no ens permet saber exactament l'estatura mitjana dels fills. Hi ha correlació positiva. Normalment, els fills de pares alts són alts.

b) Estadística, perquè el volum d'exportació no ens permet saber exactament el volum d'importació. Hi ha correlació negativa. Normalment, els països que exporten molt, importen poc.

c) Estadística, perquè la taxa de mortalitat infantil no ens permet saber exactament el nombre de metges per cada 1 000 habitants. Hi ha correlació negativa. Normalment, els països que tenen una taxa de mortalitat infantil gran tenen pocs metges per cada 1 000 habitants.

d) *kWh consumits durant gener - Cost del rebut de la llum* → Funcional; si coneixem els kWh consumits durant gener, podem calcular el cost del rebut de la llum.

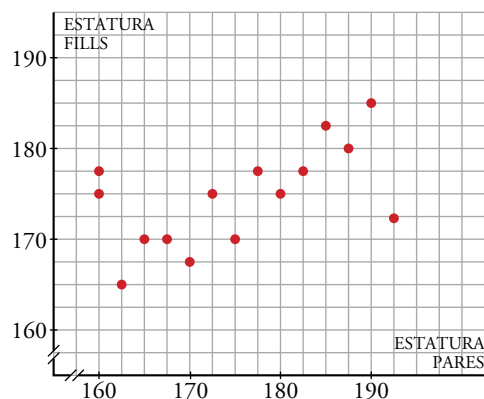
Nombre de persones en cada casa - Cost del rebut de la llum → Estadística, perquè el nombre de persones en cada casa no ens permet saber exactament el cost del rebut de la llum. Hi ha correlació positiva. Normalment, com més persones hi ha en una casa, més llum es consumeix.

e) *Posició en finalitzar la Lliga - Nombre de partits perduts* → Estadística, perquè la posició en finalitzar la Lliga no ens permet saber exactament el nombre de partits perduts. Hi ha correlació negativa. Normalment, com més alta és la posició en la Lliga, menys partits s'han perdut.

Posició en finalitzar la Lliga - Nombre de partits guanyats → Estadística, perquè la posició en finalitzar la Lliga no ens permet saber exactament el nombre de partits guanyats. Hi ha correlació positiva. Normalment, com més alta és la posició en la Lliga, més partits s'han guanyat.

Exemple de relació estadística

En el gràfic següent, cada punt correspon a un nen. L'abscissa és l'estatura del seu pare, i l'ordenada la del nen:



- Identifica en Guillem i en Biel, germans de bona estatura, sabent que el seu pare és baix.
- Identifica en Sergi, d'estatura normal, si saps que el seu pare és molt alt.
- Podem dir que hi ha una certa relació entre les estatures d'aquests 15 nens i les dels pares?
 - En Guillem i en Biel estan representats mitjançant els punts (160, 175) i (160; 177,5).
 - En Sergi està representat amb el punt (192,5; 172,5).
 - Sí; en general, com més alt sigui el pare, més alts són els fills.

1 Distribucions bidimensionals. Núvols de punts

Pàgina 217

1 Cert o fals?

- a) En una distribució bidimensional, per a cada valor de x només hi pot haver un valor de y .
 - b) Com més punts tingui una distribució bidimensional, més forta n'és la correlació.
 - c) Les sèries temporals són distribucions estadístiques en què una de les variables és el temps.
Encara que no siguin distribucions bidimensionals pròpiament dites, es poden tractar de la mateixa manera que aquestes.
-
- a) Fals; es poden mirar els núvols de punts d'aquesta mateixa pàgina.
 - b) Fals; la correlació depèn de la relació entre les característiques que s'estudien en una població; no del nombre d'elements de la població.
 - c) Cert.

2 Correlació lineal

Pàgina 219

2 Cert o fals?

a) Com més pròxims estiguin a una recta els punts d'una distribució bidimensional, més forta n'és la correlació lineal.

b) Si la recta de regressió té pendent negatiu, la correlació lineal és negativa.

c) Si els punts del núvol no s'aproximen a cap recta, llavors les variables estan incorrelades.

a) Cert. Perquè la correlació estudia les distàncies dels punts a la recta de regressió. Com més petita és la distància a la recta, més gran és la correlació.

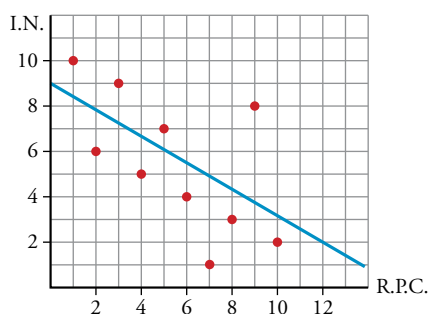
b) Cert. Una recta de pendent negatiu indica, com el signe del coeficient de correlació, que en augmentar una variable, l'altra disminueix.

c) Cert.

3 La taula següent mostra com s'ordenen entre si deu països, A, B, C..., segons dues variables, R.P.C. (*renta per càpita*) i I.N. (*índex de natalitat*). Representa els resultats en un núvol de punts, traça la recta de regressió i digues què et sembla la correlació.

PAÏSOS	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 correlació és negativa i moderadament alta ($-0,62$).



3 Paràmetres associats a una distribució bidimensional

Pàgina 221

4 Cert o fals?

- a) El signe de la correlació (r) coincideix amb el de la covariància (σ_{xy}).
- b) Si canviem les unitats en què s'expressa la variable x , llavors es modifiquen els valors de $\bar{x}$, σ_x , σ_{xy} i r .
- c) Encara que canviem les unitats de la variable x (o de la y , o de totes dues), el valor de la correlació, r , no es modifica.
- a) Cert, $r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$; com que σ_x i σ_y són positives, el signe de r és el de σ_{xy} .
- b) Fals. Varien tots els paràmetres menys r , perquè r és l'únic que no té dimensions.
- c) Cert.

5 Obtén mitjançant càlculs manuals els coeficients de correlació de les distribucions amb què hem treballat en l'epígraf anterior:

Salt d'altura - Salt amb perxa

Salt d'altura - 1 500 m llisos

Salt d'altura - Llançament de pes

Fes-ho també amb una calculadora amb mode LR.

- x : salt d'altura
- y : salt amb perxa

Elaborem la taula com en l'exercici resolt:

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

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

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

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

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

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

- x : salt d'altura
- y : 1 500 m llisos

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

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

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

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

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

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

Unitat 8. Distribucions bidimensionals

- x : salt d'altura

y : llançament de pes

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 regressió

Pàgina 223

6 Cert o fals?

- a) **Com més forta sigui la correlació, més punts del núvol hi haurà exactament sobre la recta de regressió.**
 - b) **Com més forta sigui la correlació, més a prop de la recta de regressió estaran els punts del núvol.**
 - c) **Com més forta sigui la correlació, més fiables seran les estimacions fetes a partir de la recta de regressió.**
- a) Cert. Com que r és molt gran, la distància dels punts a la recta és molt petita o nul·la.
- b) Fals. Hi haurà molts punts a prop de la recta, però hi pot haver punts aïllats lluny de la recta.
- c) Cert. Els valors d'una de les variables són més predictibles, ja que són molt pròxims a la recta de regressió.

5 Hi ha dues rectes de regressió

Pàgina 224

7 Cert o fals?

- a) En una distribució bidimensional en què s'estudien conjuntament les estatures (x) i els pesos (y) d'un grup de joves en la qual $\bar{x} = 170$ cm i $\bar{y} = 65$ kg, és impossible que les rectes de regressió siguin $y = 0,8x - 67$ i $y = 1,1x - 121$.
- b) Si en una distribució bidimensional és $\bar{x} = 3$ i $\bar{y} = 5$, llavors és possible que les rectes de regressió siguin $y = 2x - 1$ i $y = -x + 8$, perquè ambdues es tallen en $(3, 5)$.
- c) Si les rectes de regressió són $y = \frac{1}{5}x + 10$ i $y = 11x - 2$, llavors la correlació és dèbil perquè les rectes formen un angle pròxim a 90° .

a) $\begin{cases} y = 0,8x - 67 \\ y = 1,1x - 121 \end{cases} \quad x = 180,0; \quad y = 77,0 \rightarrow \text{Es tallen en } (180, 77).$

El punt de tall de les rectes de regressió ha de ser $(\bar{x}, \bar{y}) = (170, 65)$; aleshores és certa l'afirmació.

- b) Cert, pel raonament anterior.
- c) Cert. Es pot observar en les gràfiques d'aquesta pàgina.

6 Taules de contingència

Pàgina 225

- 8 Calcula la mitjana i la desviació típica de la distribució marginal de la x . Per a això, assigna a cada interval d'edats la seva marca de classe (punt mitjà) i al darrer interval assigna-li 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

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

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

- 9 La distribució marginal de la y correspon a una variable qualitativa. Per tant, no té mitjana ni desviació típica. L'únic paràmetre que podem assignar-li és la moda. Quina és?

Moda = Esports.

Pàgina 226

- 10 Comprova que la taula següent correspon a la distribució de x condicionada a $y \in \{\text{INF}, \text{DOC}\}$.

Troba'n la mitjana i la desviació típica.

x	18-25	26-35	36-50	51-65	MÁS DE 65
f	9	21	36	26	46

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

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

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

Unitat 8. Distribucions bidimensionals

11 Fes la distribució de y condicionada a $x < 36$.

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

12 Comprova, calculant les freqüències relatives, que l'esdeveniment PEL no és independent de l'edat.

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	

S'observa que les freqüències relatives varien segons l'edat.

13 Fes la distribució de x condicionada a NO ESPORTS i compara'n les freqüències relatives amb les de la distribució marginal de la x .

x_i	21,5	30,5	43	58	75	
NO ESP.	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

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

$$\sigma_x = \sqrt{\frac{1\,609\,373,5}{589} - 49,073^2} = 28,471$$

La mitjana és similar; tanmateix, la desviació típica és més gran si considerem les dades de les persones que no es miren els esports.

Pàgina 228

- 14** Un altre grup de 154 persones han realitzat els mateixos tests, amb els resultats que es donen en la taula de la dreta. Troba'n el coeficient de correlació.

De les dades obtenim les taules següents:

$x_i \backslash y_i$	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

Distribució 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

$$\bar{x} = \frac{191}{154} = 1,26$$

$$\sigma_x = \sqrt{\frac{459}{154} - 1,26^2} = 1,18$$

Distribució 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

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

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

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

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

Exercicis i problemes resolts

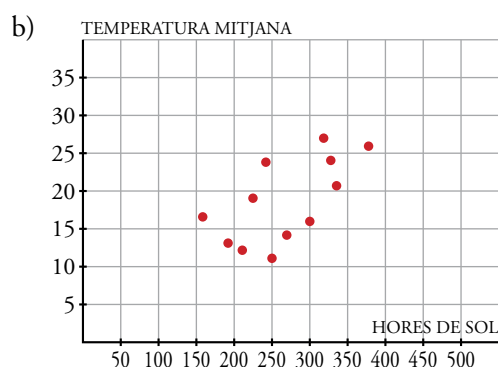
Pàgina 229

1. Relació funcional i relació estadística

Fes-ho tu. Fes el mateix amb les variables *Hores de sol* - *Temperatura mitjana a Almeria*.

	g	F	Mç	Ab	Mg	Jn	Jl	Ag	S	O	N	D
HORES DE SOL	216	251	271	300	335	329	377	321	244	225	159	194
TEMPERATURA MITJANA	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) És una distribució bidimensional en què es relacionen les variables x : hores de sol i y : temperatura mitjana a Almeria, corresponents a un any.



c) És una relació estadística; el nombre d'hores de sol no determina la temperatura mitjana.

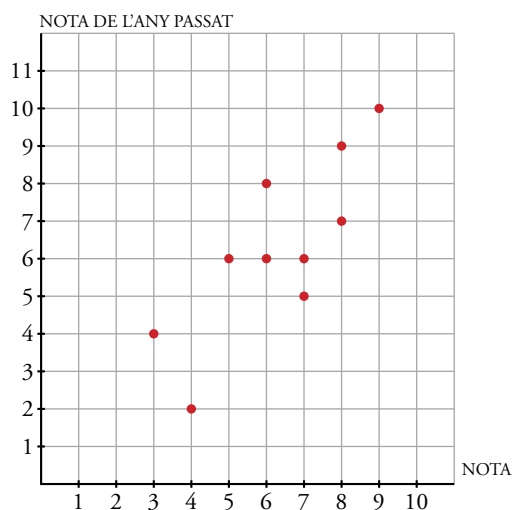
2. Anàlisi gràfica d'una distribució bidimensional

Fes-ho tu. Estudia la correlació entre la NOTA d'un examen de Matemàtiques i les altres variables que apareixen en la taula següent:

NOTA	3	4	5	6	6	7	7	8	8	9
NOTA DE L'ANY PASSAT	4	2	6	6	8	5	6	9	7	10
QUOCIENT INTEL·LECTUAL	120	102	118	105	95	110	103	107	115	125

• x : Nota

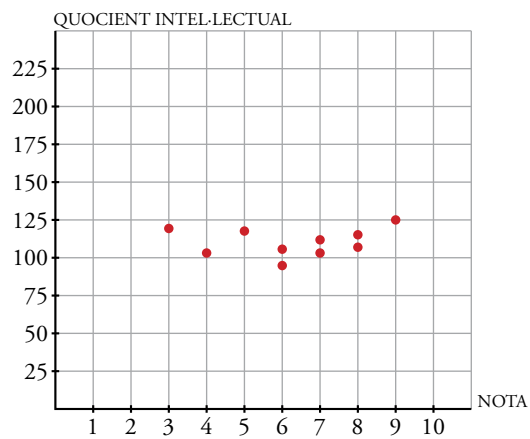
y : Nota de l'any passat



Hi ha una correlació positiva bastant forta.

- x : Nota

y : Quocient intel·lectual



Hi ha una correlació positiva bastant forta.

Exercicis i problemes guiats

Pàgina 231

1. Dues rectes de regressió. Estimacions

La taula següent relaciona les variables

x : despeses en publicitat (milers d'euros)

y : vendes (milers d'euros)

durant els 6 primers mesos de promoció d'un cert producte:

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

a) Trobar les dues rectes de regressió.

b) Efectuar l'estimació $\hat{y}(5,5)$ i explicar el seu significat.

c) Per obtenir unes vendes de 20 000 €, quants milers d'euros s'estima que han de gastar en publicitat?

Seran fiables aquestes estimacions?

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	1521	195
6	47	36	2209	282
21	171	91	5803	723

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

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

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

$$\sigma_y = \sqrt{\frac{5803}{6} - 28,5^2} = 12,45$$

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

Pendent de la recta de regressió de Y sobre X :

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

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

Pendent de la recta de regressió 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) $\hat{x}(20) \rightarrow 20 - 28,5 = 7,47(x - 3,5) \rightarrow y = 2,36$

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

Unitat 8. Distribucions bidimensionals

2. Taula de doble entrada

Una companyia discogràfica ha recopilat en la taula de la dreta la següent informació sobre el nombre de concerts efectuats per 15 grups musicals durant un estiu, i les vendes de discos d'aquests grups (en milers).

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

a) Calcular el nombre mitjà de discos venuts.

b) Quin és el coeficient de correlació?

c) Obtén la recta de regressió de Y sobre X .

d) Si un grup musical ven 18 000 discos, quin nombre de concerts es preveuen?

a) y b)

DISCOS (x_i) \ CONC. (y_j)	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

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

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

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

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

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

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

Recta de regressió de Y sobre X :

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

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

Exercicis i problemes proposats

Pàgina 232

Per practicar

■ Sense fórmules

1 Per a cadascun dels casos següents, indica:

- Quines són les variables que s'hi relacionen.
- Si es tracta d'una relació funcional o d'una relació estadística i, en aquest darrer cas, determina el signe de la correlació.

a) *Renda mensual d'una família - Despesa mensual en electricitat*

b) *Radi d'una esfera - Volum d'aquesta esfera*

c) *Litres de pluja recollits en una ciutat - Temps dedicat a mirar la televisió pels seus habitants*

d) *Longitud del trajecte recorregut en una línia de rodalia - Preu del bitllet*

e) *Pes dels alumnes de 1r de Batxillerat - Número de calçat que duen*

f) *Tones de tomàquets en una collita - Preu del quilo de tomàquets al mercat*

g) *Superfície d'una casa - Valor de la casa*

a) Renda (€), despesa (€).

Correlació positiva.

b) Relació funcional.

c) Relació estadística. Segurament molt dèbil. Positiva (¿hem de pensar que, com més ploqui, més temps passaran a casa i, per tant, més es miraran la televisió?).

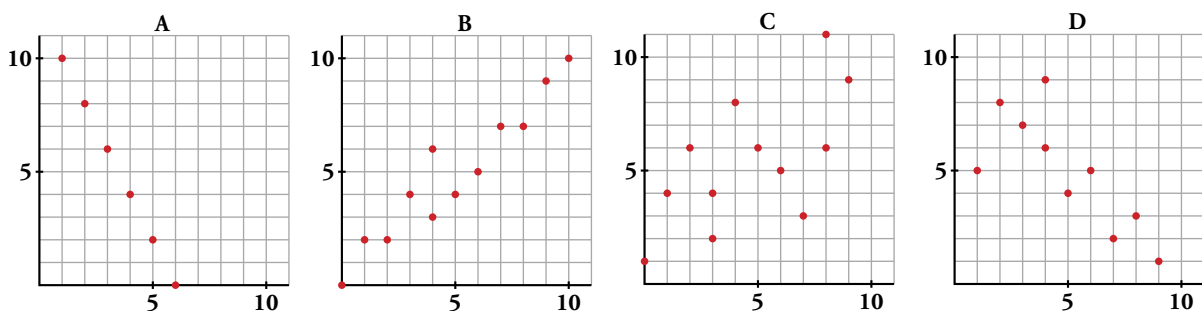
d) Encara que ho sembli *a priori*, segurament la relació no és funcional. És una correlació positiva forta.

e) Correlació positiva.

f) Correlació negativa (com més gran sigui la collita, més barats seran els tomàquets).

g) Correlació positiva.

2 a) Copia en el teu quadern i traça a ull una recta de regressió per a cadascuna d'aquestes distribucions bidimensionals:

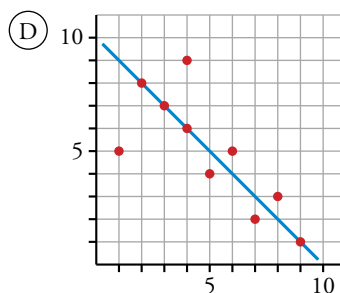
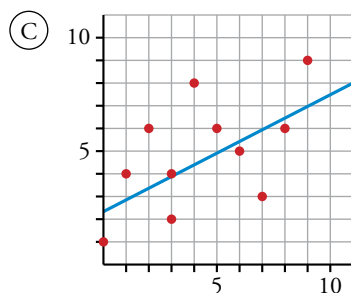
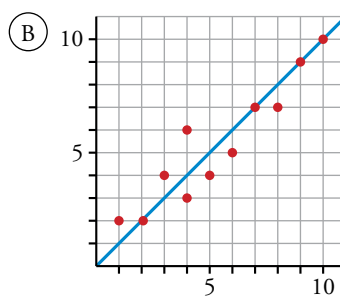
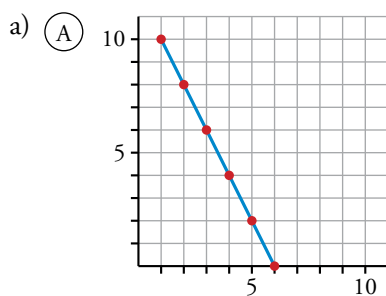


b) Quines tenen correlació positiva i quines tenen correlació negativa?

c) Sense fer càlculs, tria, d'entre els valors següents, la correlació de cadascuna de les distribucions:

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

d) Una d'aquestes distribucions presenta una relació funcional; quina? Dóna l'expressió analítica de la funció que relaciona les dues variables.

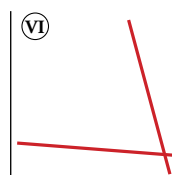
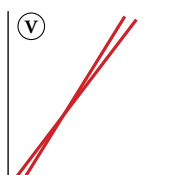
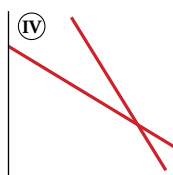
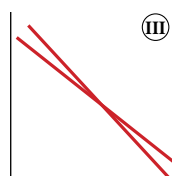
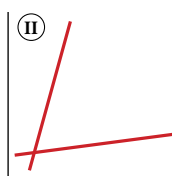
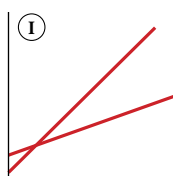


b) B i C tenen correlació positiva; A i D, negativa.

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

d) La A és relació funcional: $y = 12 - 2x$.

3 Cadascuna d'aquestes sis distribucions bidimensionals està representada per les seves dues rectes de regressió:



Els seus coeficients de correlació són, desordenats:

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

Assigna'n, raonadament, el valor a cada distribució.

I $\rightarrow 0,6$

II $\rightarrow 0,1$

III $\rightarrow -0,9$

IV $\rightarrow -0,5$

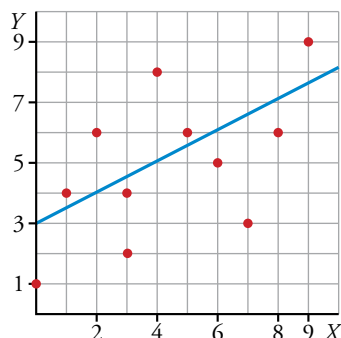
V $\rightarrow 0,99$

VI $\rightarrow -0,2$

- 4 Representa el núvol de punts d'aquesta distribució i estima quin d'aquests tres pot ser el coeficient de correlació:

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

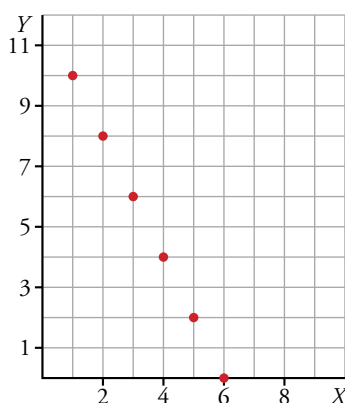


c) $r = 0,58$

- 5 Representa sobre paper quadriculat el núvol de punts corresponent a aquesta distribució:

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

Quin creus que és el coeficient de correlació?



$r = -1$, perquè estan alineats.

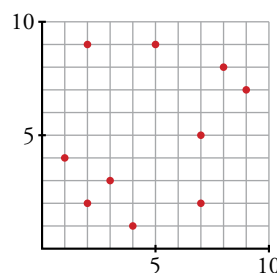
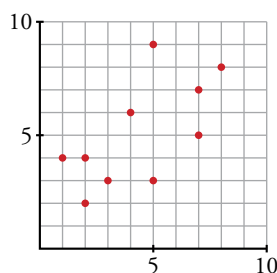
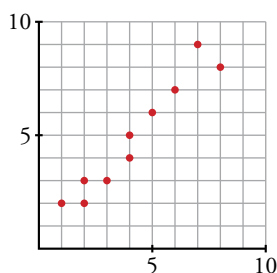
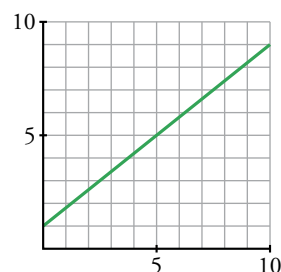
- 6 a) Al teu quadern, en una quadrícula com aquesta, situa deu punts de manera que estimis que la seva correlació sigui 0,9 i una de les rectes de regressió sigui la que veus.
b) Repeteix l'experiència per aconseguir un coeficient de correlació de 0,6.
c) Fes el mateix per a un coeficient de 0,3.

* Atenció: es demana estimar, però no calcular.

a) $r = 0,9$

b) $r = 0,6$

c) $r = 0,3$



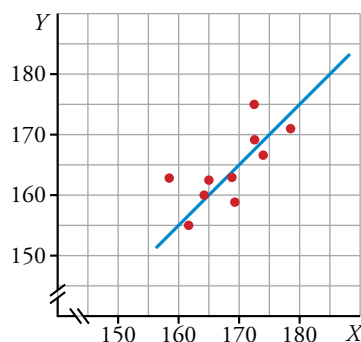
Unitat 8. Distribucions bidimensionals

7 Les estatures de 10 nenes i les de les seves mares són:

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 aquests valors mitjançant un núvol de punts.

b) Traça a ull una recta de regressió i digues si la correlació és positiva o negativa i més o menys forta del que esperaves.



La correlació és positiva i forta.

Pàgina 233

Amb fórmules

8 Aquesta és la distribució bidimensional donada pel núvol de punts B de l'exercici 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

Troba:

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

b) El coeficient de correlació, r . Interpreta'l.

c) Les equacions de les dues rectes de regressió.

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

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

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

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

$$b) r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = 0,95. \text{ Es tracta d'una correlació forta i positiva.}$$

c) Recta de regressió de Y sobre X :

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

Recta de regressió de X sobre Y :

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

Unitat 8. Distribucions bidimensionals

- 9 a) Representa el núvol de punts corresponent a la distribució bidimensional següent:

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) Comprova amb la calculadora que els paràmetres són:

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

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

- c) Troba les equacions de les dues rectes de regressió, X sobre Y i Y sobre X , i representa-les juntament amb el núvol de punts.

a) Representada en l'exercici 4.

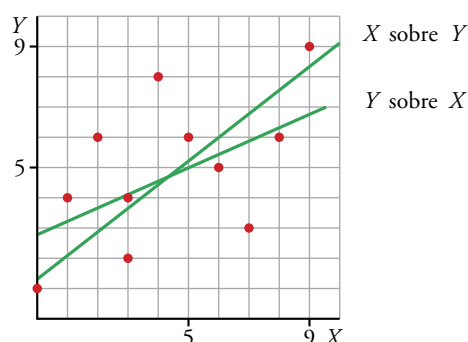
b) Es comprova.

- c) • Recta de regressió 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 regressió 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 distribució bidimensional en què els valors de x són 12, 15, 17, 21, 22 i 25, té una correlació $r = 0,99$ i la recta de regressió és $y = 10,5 + 3,2x$.

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

- b) Quines de les estimacions anteriors són fiables, quina poc fiable i quina no s'ha de fer?

- c) Expressa els resultats en termes adequats.

Per exemple:

$\hat{y}(13) = 52,1$. "Per a $x = 13$ és molt probable que el valor corresponent de y sigui pròxim a 52".

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

- b) Són fiables $\hat{y}(13)$ i $\hat{y}(20)$, perquè 13 i 20 estan en l'interval de valors utilitzats per obtenir la recta de regressió.

$\hat{y}(30)$ és menys fiable, ja que 30 es troba fora de l'interval, encara que a prop d'ell.

$\hat{y}(100)$ és una estimació gens fiable, ja que 100 és molt lluny de l'interval [12, 25].

11 Observa la distribució D de l'exercici 2.

- a) Descriu-la mitjançant una taula de valors.
- b) Realitza els càlculs per obtenir-ne el coeficient de correlació.
- c) Representa els punts al teu quadern.

Troba l'equació de la recta de regressió de Y sobre X i representa-la.

- d) Calcula $\hat{y}(4,5)$, $\hat{y}(11)$, $\hat{y}(20)$ dilucidant com de fiables són aquestes estimacions.

a)

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

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

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

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

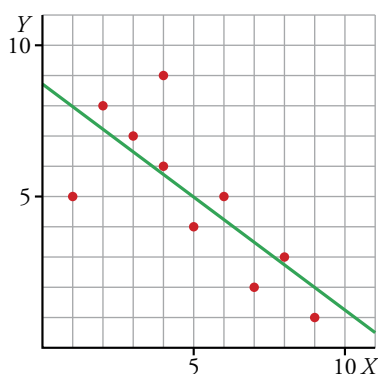
$\Sigma y^2 = 310$ $\sigma_y = \sqrt{\frac{310}{10} - 5^2} = 2,45$

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

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

- c) Recta de regressió de Y sobre X :

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



d) $\hat{y}(4,5) = 5,56$

$\hat{y}(11) = -3,04$

$\hat{y}(20) = -14,95$

Com que $r = 0,76$, l'estimació per a 4,5 la podem considerar fiable, però les de 11 i 20, que no estan en l'interval de dades, no es poden considerar gaire fiables.

12 Calcula les correlacions corresponents als núvols de punts que has inventat en l'exercici 6. Comprova si les correlacions obtingudes s'assemblen a les que pretenies aconseguir.

- a) $r = 0,97$
- b) $r = 0,64$
- c) $r = 0,25$

Per resoldre

13 La taula següent recull les dades econòmiques d'algunes de les pel·lícules més rendibles d'un any (les quantitats estan donades en milions d'euros):

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

- a) Troba'n el coeficient de correlació.
- b) Obtén la recta de regressió de Y sobre X i estima quina recaptació han d'esperar si inverteixen 30 milions d'euros en una pel·lícula.

- a) $r = 0,6$
- b) $y = 3,05x + 13,05$
 $\hat{y}(30) = 104,55$

Han d'esperar que es recaptin 104,55 milions d'euros.

14 Un excursionista, en deu marxes diferents, pren les mesures següents:

x : altura del lloc (en m)

y : pressió atmosfèrica (en mmHg)

z : nombre de pulsacions en repòs

x	0	184	231	481	730	911	1343	1550	1820	2184
y	760	745	740	720	700	685	650	630	610	580
z	73	78	75	78	83	80	89	80	85	92

Troba el coeficient de correlació i la recta de regressió per a la distribució x - y i per a la distribució x - z i analitza'n els resultats.

- a) x : altura del lloc (en m)
- y : pressió atmosfèrica (en mmHg)
- $r = -0,99$

Recta de regressió: $y = -0,08x + 759$

Hi ha gairebé una relació funcional entre l'altura d'un lloc i la seva pressió atmosfèrica. A més, quan augmenta l'altura, disminueix la pressió.

- b) x : altura del lloc (en m)
- z : nombre de pulsacions en repòs
- $r = 0,85$

Recta de regressió: $y = 6,87x + 74,8$

Hi ha una correlació forta entre l'altura d'un lloc i el nombre de pulsacions, en repòs, d'una persona. A més, quan augmenta l'altura, augmenten les pulsacions en repòs.

Unitat 8. Distribucions bidimensionals

- 15** En la taula següent es consignen els gols a favor (x) i els gols en contra (y) al final del campionat de Lliga de 1a divisió de futbol de 10 equips:

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

Troba el coeficient de correlació entre les dues variables i analitza el que has obtingut.

x : gols a favor

y : gols en contra

$$r = -0,61$$

r és negatiu; per tant, com més gols a favor té un equip, menys gols en contra té. La correlació no és gaire forta, per la qual cosa no és massa fiable estimar els gols en contra sabent els gols a favor.

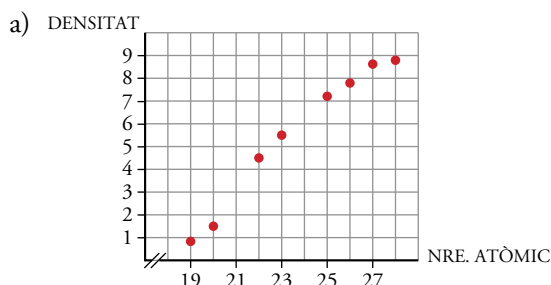
- 16** La taula següent relaciona el nombre atòmic de diversos metalls, x , amb la seva densitat, y :

ELEMENT	K	Ca	Ti	V	Mn	Fe	Co	Ni
NRE. ATÒMIC	19	20	22	23	25	26	27	28
DENSITAT	0,86	1,54	4,50	5,60	7,11	7,88	8,70	8,80

- a) Representa els punts, troba el coeficient de correlació i calcula l'equació de la recta de regressió de Y sobre X .

- b) Estima la densitat del crom sabent que el seu nombre atòmic és 24 $\rightarrow$ Cr (24).

- c) Estima la densitat de l'escandi $\rightarrow$ Sc (21).



$$r = 0,98$$

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

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

La densitat del crom s'estima en, aproximadament, 5,86. El seu valor real és 7,1.

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

La densitat de l'escandi s'estima en, aproximadament, 3,01. El seu valor real és 2,9.

Pàgina 234

17 La taula següent relaciona tres variables sociomètriques relatives a dotze països:

a) Troba el coeficient de correlació entre les variables x - y i entre les variables x - z .

b) Quines conclusions extreus dels resultats obtinguts?

a) x : renda per càpita (\$)

y : índex de natalitat (‰)

$$r = -0,68$$

La correlació és negativa; és a dir, si augmenta la renda per càpita, disminueix l'índex de natalitat.

x : renda per càpita (\$)

z : expectativa de vida en néixer (anys)

$$r = 0,82$$

La correlació és positiva; és a dir, si augmenta la renda per càpita, augmenta l'expectativa de vida en néixer.

b) La correlació és més gran en valor absolut en el segon cas; per tant, la renda per càpita és més determinant de l'expectativa de vida en néixer que de l'índex de natalitat.

PAÍS	x: RENDA PER CÀPITA (\$)	y: ÍNDEX DE NATALITAT (‰)	z: EXPECTATIVA DE VIDA EN NÉIXER (anys)
A	873	50	49
B	402	48	50
C	536	47	54
D	869	44	57
E	1171	41	61
F	636	36	64
G	1417	35	59
H	2214	31	63
I	1334	28	63
J	769	26	61
K	1720	25	64
L	2560	24	70

18 La taula següent expressa les dades de vuit alumnes de secundària d'un centre escolar, relatives al nombre d'hores per dia que miren la televisió i la nota mitjana obtinguda en la darrera avaluació:

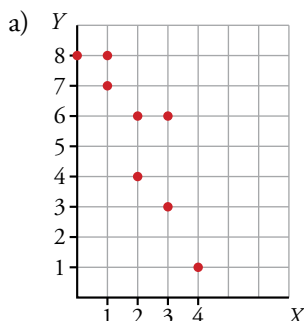
x: NRE. D'HORES DE TV	0	2	1	3	4	2	1	3
y: NOTA MITJANA	8	6	8	6	1	4	7	3

a) Representa'n gràficament les dades. Permet l'anàlisi gràfica deduir el signe del coeficient de correlació de Y sobre X ? I el de X sobre Y ?

b) Troba el coeficient de correlació de les dues variables.

c) Calcula l'equació de la recta de regressió de Y sobre X .

d) Quina nota mitjana pot esperar qui mira la televisió tres hores i mitja diàries? I qui la mira cinc hores?



Els dos signes són negatius; és a dir, si augmenten les hores per dia que miren la televisió, disminueix la nota mitjana obtinguda en l'última avaluació.

b) $r = -0,87 \rightarrow$ correlació forta

c) Recta de regressió 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 mira la televisió tres hores i mitja diàries, hem d'esperar que tregui un 2,86.

Si mira la televisió cinc hores, hem d'esperar que tregui un 0,35.

- 19** Triem sis automòbils a l'atzar. L'antiguitat, en anys, i el nombre de quilòmetres que han rodat, en milers de quilòmetres, estan relacionats per la taula següent:

ANTIGUITAT	1	2	4	4	5	6	7
QUILÒMETRES RECORREGUTS	15	45	32	61	60	132	93

- a) Calcula la mitjana i la desviació típica de les dues variables que hi intervenen.
b) Calcula'n el coeficient de correlació i interpreta'l.
c) Si un automòbil té tres anys, quants quilòmetres estimes que ha recorregut?
d) I si té cinc anys? I deu? Justifica'n les respostes.

x : antiguitat

y : quilòmetres recorreguts

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

$$\sigma_x = 1,96$$

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

$$\sigma_y = 36,37$$

b) $r = 0,81$

És positiva; és a dir, si augmenta l'antiguitat, augmenten els quilòmetres recorreguts. La correlació és forta perquè r és pròxim a 1.

- c) i d) Recta de regressió de Y sobre X :

$$y = 15,1x$$

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

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

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

Aquesta última estimació és menys precisa que les anteriors, ja que 10 no és en l'interval $[0, 7]$ del qual tenim les dades.

Qüestions teòriques

- 20** El coeficient de correlació d'una distribució bidimensional és 0,87.

Si els valors de les variables es multipliquen per 10, quin serà el coeficient de correlació d'aquesta nova distribució?

El mateix, ja que r no depèn de les unitats; és adimensional.

- 21** Hem calculat la covariància d'una certa distribució i ha resultat negativa.

Justifica per què podem afirmar que tant el coeficient de correlació com els pendents de les dues rectes de regressió són nombres negatius.

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

Com que σ_x i σ_y són positives, el signe de r és el mateix que el de σ_{xy} ; per tant, si la covariància és negativa, r també ho és.

$$m_{yx} = \frac{\sigma_{xy}}{\sigma_x^2}, \text{ el signe del qual és el mateix que el signe de } \sigma_{xy}.$$

$$m_{xy} = \frac{\sigma_{xy}}{\sigma_y^2} \text{ el signe del qual és el mateix que el signe de } \sigma_{xy}.$$

Aleshores, si la covariància és negativa, m_{yx} i m_{xy} són negatives.

22 Quin punt tenen en comú les dues rectes de regressió?

El centre de gravetat de la distribució, $(\bar{x}, \bar{y})$.

23 Quina condició ha de complir r perquè les estimacions fetes amb la recta de regressió siguin fiables?

$|r|$ ha de ser pròxim a 1.

24 Prova que el producte dels coeficients de regressió m_{yx} i m_{xy} és igual al quadrat del coeficient de correlació.

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

25 Explica com es calculen $\bar{x}$ i $\bar{y}$ a partir de les dues rectes de regressió d'una distribució bidimensional. Aplica-ho a aquest cas:

La recta de regressió de Y sobre X és: $y = 8,7 - 0,76x$

La recta de regressió de X sobre Y és: $y = 11,36 - 1,3x$

$(\bar{x}, \bar{y})$ són les coordenades del punt d'intersecció de les dues rectes de regressió.

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

26 Explica com es troba el coeficient de correlació a partir de les dues rectes de regressió d'una distribució bidimensional. Aplica-ho al cas de l'exercici anterior.

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

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

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

En l'exercici anterior:

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

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

27 L'estatura mitjana de 100 escolars és de 155 cm amb una desviació típica de 15,5 cm.

La recta de regressió de l'estatura respecte al pes és $y = 80 + 1,5x$ (x : pes; y : estatura)

a) Quin és el pes mitjà d'aquests escolars?

b) Quin és el signe del coeficient de correlació entre pes i estatura?

a) La recta de regressió passa per $(\bar{x}, \bar{y})$; aleshores el pes mitjà serà la solució de l'equació:

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

b) El signe del coeficient de correlació entre pes i estatura és el mateix que el del pendent de la recta de regressió; per tant, és positiu.

28 Cert o fals?

- a) Si el pendent de la recta de regressió de Y sobre X és 1, la correlació entre les dues variables és molt forta.
- b) Si el pendent de la recta de regressió de Y sobre X és negatiu, el pendent de la recta de regressió de X sobre Y també és negatiu.
- c) En una relació funcional lineal les dues rectes de regressió coincideixen.
- d) Com més forta sigui la correlació entre les variables x i y , més gran és r^2 .
- a) Fals. Si el pendent de la recta de regressió de Y sobre X és 1, sabem que la covariància és igual a la variància de x , però no que r sigui pròxim a 1.
- b) Cert, perquè $m_{yx} \cdot \frac{1}{m_{xy}} = r^2 > 0$

El producte és un nombre positiu; aleshores els dos pendents han de tenir el mateix signe.

- c) Cert. En una relació funcional, $r = 1$.

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

Com que les dues rectes passen per $(\bar{x}, \bar{y})$ i tenen el mateix pendent, coincideixen.

- d) Cert, perquè $0 \leq r^2 \leq 1$.

Si la correlació és molt forta, $|r|$ és pròxim a 1; aleshores r^2 s'aproxima a 1.

Pàgina 235**Per aprofundir**

- 29** En una autoescola, cada alumne realitza un total de 80 tests repartits en 4 tandes de 20. La taula següent relaciona les variables número de tanda (x) i nombre d'errors (y):

$y \backslash x$	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

Per exemple: En la tercera tanda, en 12 dels tests es van trobar de 0 a 3 errors; en 7, de 4 a 7 errors...

- a) Calcula'n el coeficient de correlació i troba l'equació de la recta de regressió de Y sobre X .
- b) Quants errors s'estima que tindrà un alumne en la primera tanda? I en la segona? I en la darrera?

a)

$\text{ERRORS} = y_i \backslash \text{TANDA} = x_i$	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

Unitat 8. Distribucions bidimensionals

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

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

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

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

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

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

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

$$\text{Recta de regressió de } Y \text{ sobre } X: y - 5,7 = -2,63(x - 2,5)$$

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

S'estima que tindrà entre 9 i 10 errors en la primera tanda.

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

S'estima que tindrà 7 errors en la segona tanda.

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

S'estima que tindrà entre 1 i 2 errors en la quarta tanda, més vegades 2 errors que 1.

30 En un estudi realitzat als treballadors d'una cadena de fabricació de peces de cotxes sobre la seva productivitat quinzenal, es relacionen les hores treballades (x) amb les unitats produïdes (y).

Sabem que:

- La recta de regressió de Y sobre X és:

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

- La recta de regressió de X sobre Y és:

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

- L'interval d'hores emprades pels treballadors és $[60, 85]$.

a) Troba $\bar{x}$, $\bar{y}$ i el coeficient de correlació.

b) Si un operari treballa 70 hores en una quinzena, quantes unitats s'estima que produirà? Com és de fiable aquesta estimació? I si treballa en total 40 hores? I si fossin 120 hores?

c) Si un empleat ha arribat a produir 300 peces aquesta quinzena, quantes hores s'estima que ha treballat?

a) $(\bar{x}, \bar{y})$ es el punt de tall de les dues rectes de regressió:

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

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

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

S'estima que l'operari produeix unes 275 unitats treballant 70 hores.

Com que r és molt pròxim a 1 i, a més, 70 és en l'interval d'hores emprades, l'estimació és molt fiable.

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

S'estima que l'operari produeix gairebé 171 unitats treballant 40 hores. Aquesta estimació no és tan fiable com l'anterior perquè $40 \notin [60, 85]$

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

S'estima que l'operari produeixi al voltant de 448 unitats treballant 120 hores. Aquesta estimació no és gaire fiable perquè $120 \notin [60, 85]$.

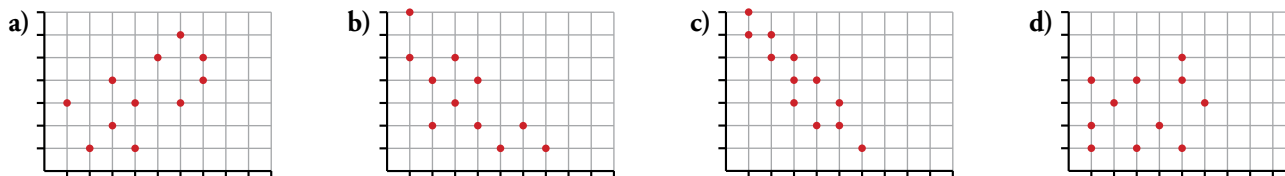
c) $300 = 3,81x + 5,36 \rightarrow x = 77,33$

S'estima que ha treballat entre 77 i 78 hores.

Autoavaluació

Pàgina 235

1 Observa aquestes distribucions bidimensionals:



Assigna raonadament un dels coeficients de correlació següents a cada gràfic:

0,2 -0,9 -0,7 0,6

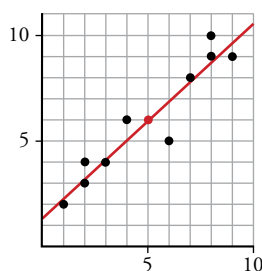
La correlació de a) és positiva, i les de b) i c), negatives. En d) no s'aprecia correlació. La correlació de c) és més forta que la de b). Per tant:

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

2 Representa aquesta distribució 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 els paràmetres $\bar{x}$, $\bar{y}$, σ_x , σ_y i σ_{xy} .
- b) Troba'n el coeficient de correlació.
- c) Troba la recta de regressió de Y sobre X .
- d) Estima el valor de y per a $x = 5$ i per a $x = 10$. Són bones aquestes estimacions?



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

Les estimacions són molt fiables perquè $r = 0,95$ és un valor molt alt. Si es tractés de "notes" (de 0 a 10), la segona estimació caldria "fer-la real" i donar-li el valor 10.

3 La recta de regressió de Y sobre X d'una certa distribució bidimensional és $y = 1,6x - 3$. Sabem que $\bar{x} = 10$ i $r = 0,8$.

a) Calcula $\bar{y}$.

b) Estima el valor de y per a $x = 12$ i per a $x = 50$. Quina estimació et sembla més fiable?

c) Troba la recta de regressió de X sobre Y .

a) Com que la recta passa per $(\bar{x}, \bar{y})$:

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

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

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

La primera estimació és acceptable pel fet que 12 és pròxim a $\bar{x} = 10$ (no tenim informació sobre els valors que pren x). La segona estimació és molt poc significativa, ja que 50 se separa massa de $\bar{x}$.

c) Coneixent $r = 0,8$ i el coeficient de regressió de Y sobre X (pendent de la recta), 1,6:

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

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

$$\text{Per tant, el pendent de la recta de regressió de } X \text{ sobre } Y \text{ és } m_{xy} = \frac{1}{0,4} = 2,5.$$

$$\text{Equació de la recta de regressió de } X \text{ sobre } Y: y = 6 + 2,5(x - 5)$$

4 El consum mensual d'energia per càpita, y , en milers de kWh, i la renda per càpita, x , en milers d'euros, de sis països són:

	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 regressió de Y sobre X .

b) Troba el coeficient de correlació entre el consum i la renda.

c) Quina predicció podem fer sobre el consum d'energia per càpita d'un país la renda per càpita del qual és de 4 400 €? (Recorda que en la taula es dona la renda en milers d'euros.)

d) Estima la renda per càpita que tindrà un país en el qual el consum d'energia per càpita ha estat de 9 000 kWh.

e) Com són de fiables aquestes estimacions?

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

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

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

$$\text{b) Coeficient de correlació: } r = \frac{2,51}{1,09 \cdot 2,46} = 0,93$$

$$\text{c) Per a } x = 4,4 \text{ estimem el valor de } y: \hat{y}(4,4) = 0,79 + 0,41 \cdot 4,4 = 2,59$$

Hom estima un consum d'energia de 2,59 milers de kWh per habitant.

$$\text{d) } 9 = 0,80 + 0,41\hat{x}(9) \rightarrow \hat{x}(9) = 20 \rightarrow \text{S'estima una renda per càpita de 20 000 €}.$$

e) En la primera estimació (apartat c), el valor $x = 4,4$ és pròxim als valors de la taula. Com que el coeficient de correlació és alt (0,93), l'estimació és raonablement fiable. En la segona estimació (apartat d), el valor $y = 9$ és lluny dels de la taula. Per tant, l'estimació és poc fiable.